Trial By Algorithm:
Auditing AI RCTs in Criminal Justice

Peter Darch
School of Information Sciences
University of Illinois at Urbana-Champaign

Valedictory Randomised Controlled Trials in the Social Sciences
University of York
May 21, 2025

Introduction

AI systems increasingly tested in field RCTs

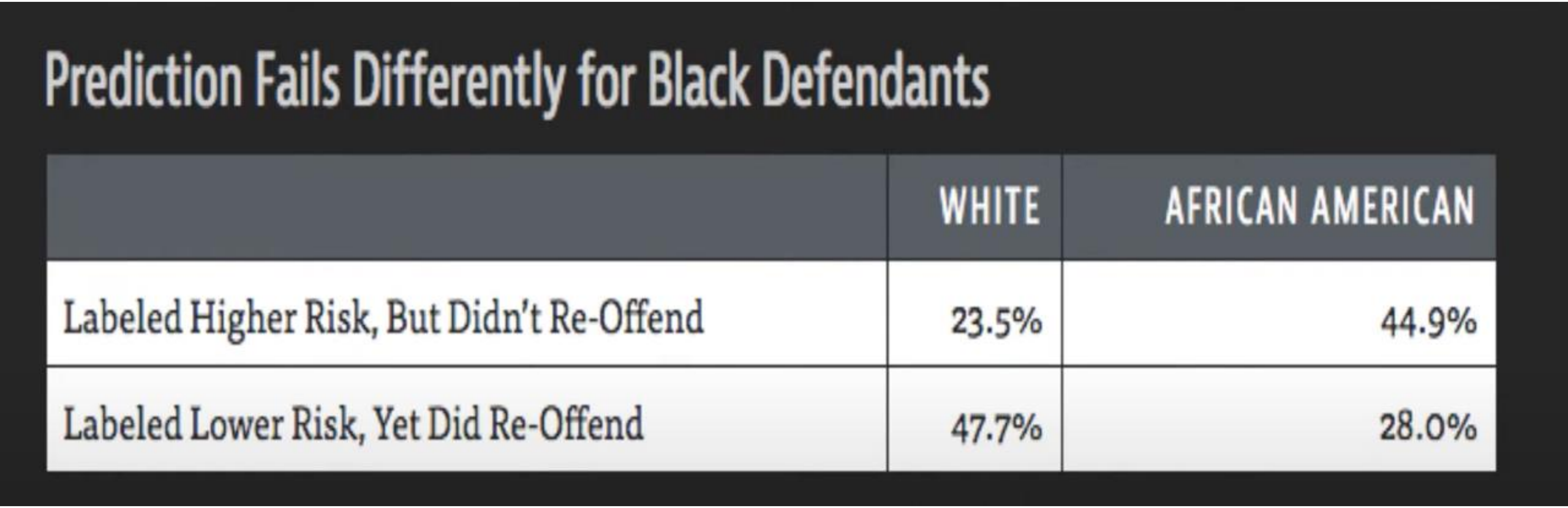
Accountability and governance are critical

Predictive policing RCTs

- IRBs not sufficient for governance
- EU: Governance via AI Act
- US & UK: no mandatory audit or transparency

Effective governance requires action by multiple stakeholders

Northpointe Algorithm



Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Northpointe Algorithm: Key Issues

Impact on civil rights

Impact on judges’ autonomy

- Lack of communication of what scores mean
- Lack of communication of data, model limitations

Due process

- Many defendants unaware of AI use
- Other defendants unaware of AI operation
- Challenges in contesting sentencing

Underpinning these issues is accountability

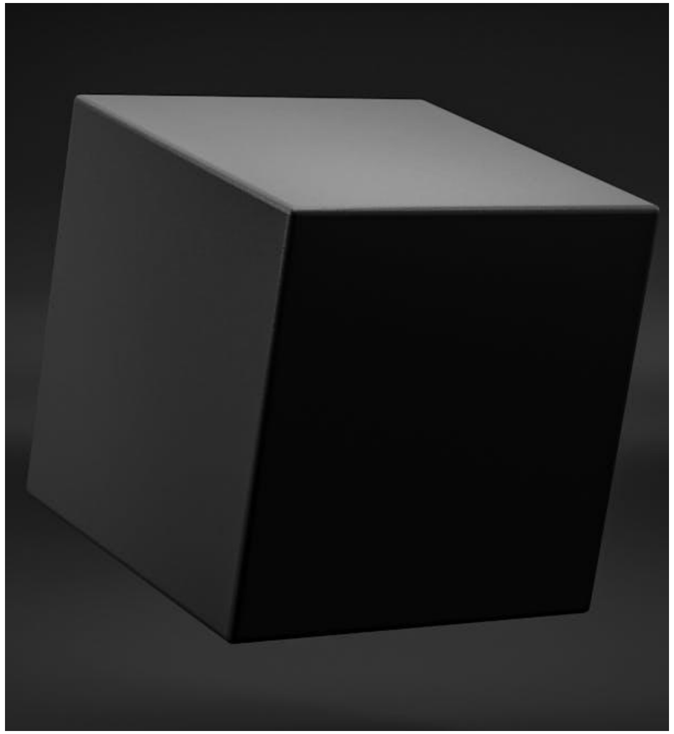
Accountability Challenges

Accountability

- Transparency, Explainability, Interpretibility
- Mechanisms to challenge

Challenges to accountability

- Proprietary or sensitive code and data
- Burden of producing documentation
- Complexity of AI system
- Inadequate access to challenge mechanisms



Responsibility is diffuse across people, institutions, and time

Governance for Data-Intensive Human Subjects' Research

Rise of web-scraping and A/B testing

High-profile controversies

Institutional Review Board (IRB) oversight

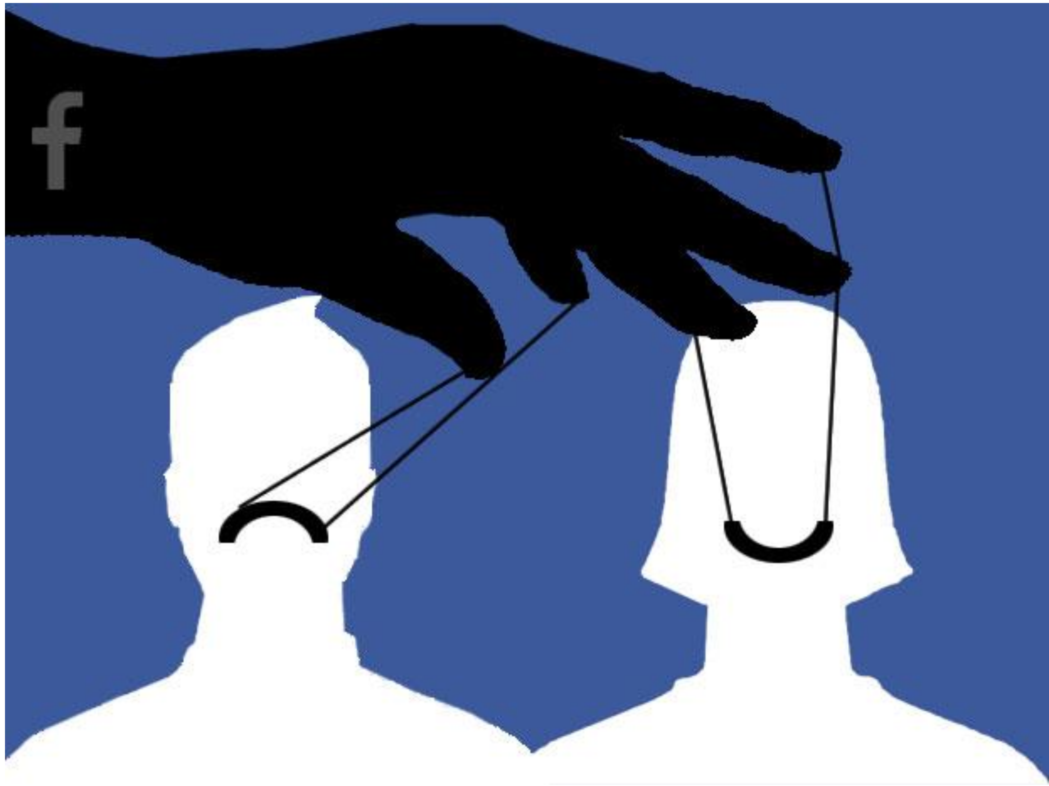
Revisions to ”the Common Rule”…

…but focused on low-risk research

Funding agency and journal policies

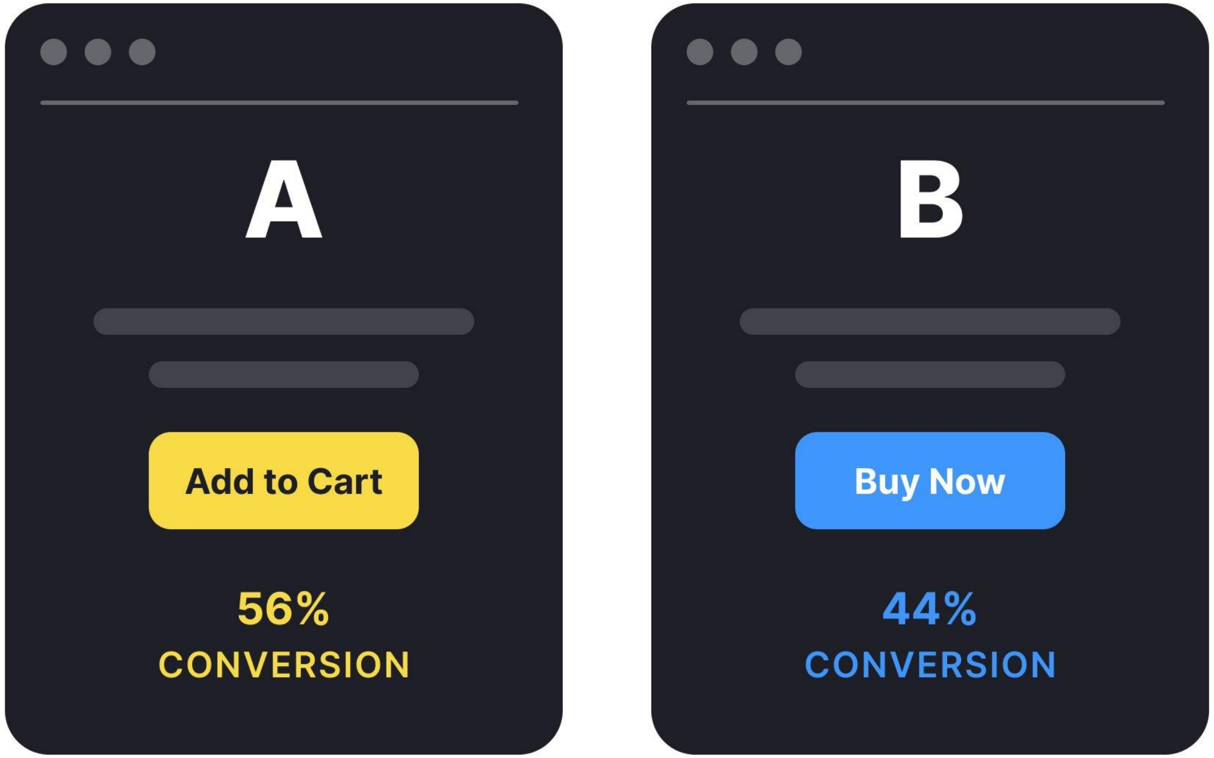
Sharing sensitive data and models?

Reproducibility not same as accountability



Field Randomized Controlled Trials of AI Systems

- Technology/Industry: A/B testing
- Tradition of Clinical RCTs
  - Healthcare
  - Education
- Criminal justice
  - Limited oversight
  - High stakes



Predictive Policing

Use data to forecast criminal activity, direct police

Location-based models

Models often supplied by private vendors

Many lawsuits already

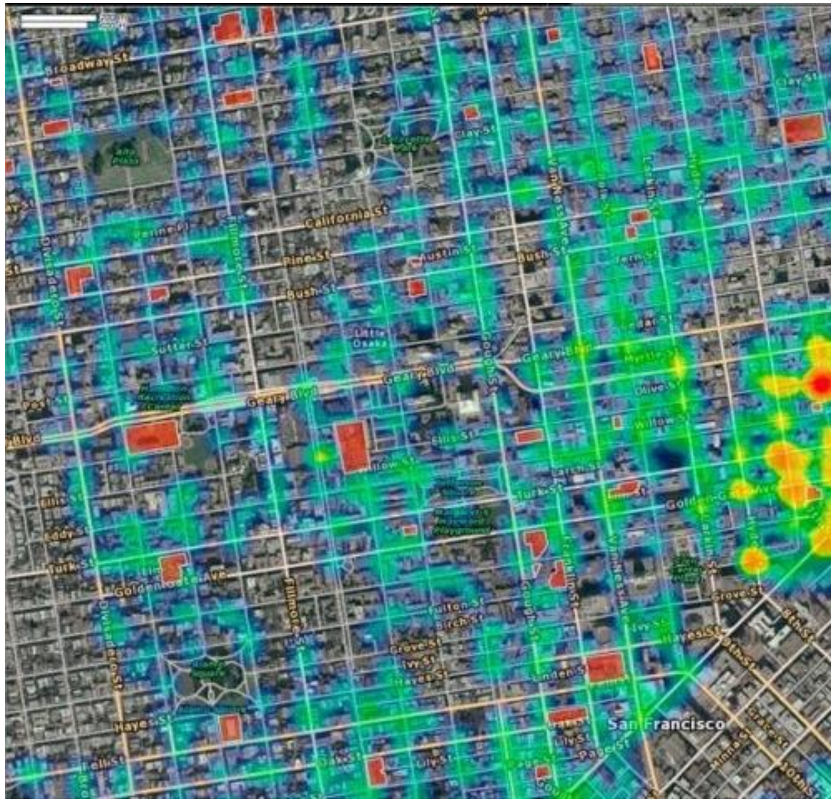
High-stakes

Significant consequences for safety

Significant consequences for liberty and civil rights

Impact on trust in law enforcement

Due process requires accountability



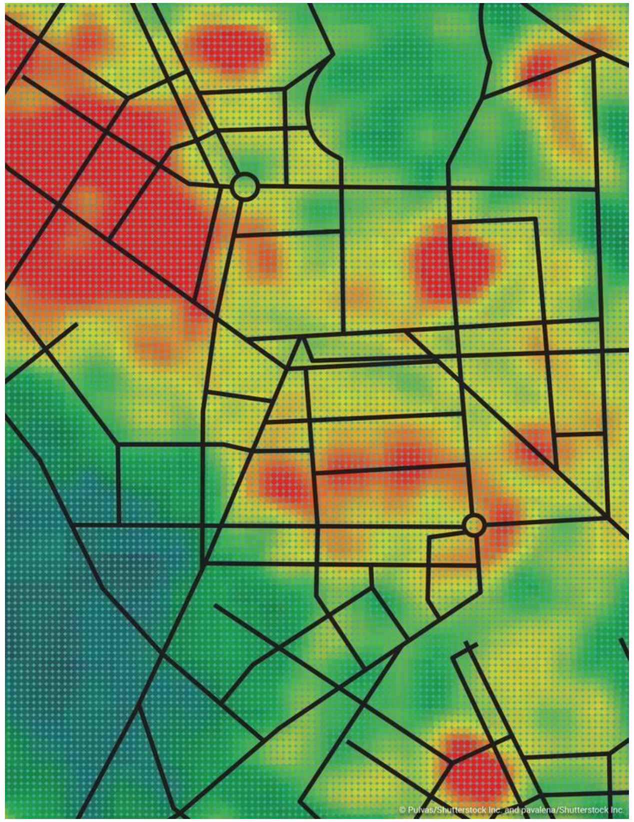
Studying Trials of Predictive Policing

Five RCTs identified

Exhaustive search for available public documentation

Qualitative content analysis

- Information to enhance transparency and accountability
- Governance issues addressed explicitly
- Governance issues addressed implicitly
- Community engagement



© Pulvas/Shutterstock Inc. and pavalena/Shutterstock Inc.

Field RCTs of Predictive Policing

Location	Year	Lead Institution	Format
Los Angeles, CA, USA	2011-12	UC Los Angeles	Journal
Shreveport, LA, USA	2012	RAND Corporation	Report
Montevideo, Uruguay	2015	University of Maryland	Journal
Philadelphia, PA, USA	2015-6	Temple University	Journal
Indianapolis, IN, USA	2019	Indiana University Indianapolis	Journal

Issues Addressed in RCT Reports

Crime reduction

Primary concern of all trials

Three of five trials showed no statistically significant impact

Impact on civil rights

Los Angeles study: no significant difference

Community perceptions and trust

Indianapolis study: “Moderate trust” in AI-driven policing

Impact on civil liberties

Two trials: no increase in use-of-force incidents or complaints

RCT Documentation Gaps

Prior assessment of likely benefits and harms

Harm mitigation strategies

Pre-registration & IRB oversight

Community consent or notification (before or after)

Oversight and accountability (local officials, IRB, police)

Transparency about models and training data

Is IRB Governance Adequate?

Existing IRB measures for data-intensive research focus on low-risk research

Notification and consent issues for affected communities

- Who speaks for the community?
- Shortcomings compared to A/B testing via a platform

External partners beyond IRB jurisdiction?

AI systems face significant accountability and transparency limitations

Public debate required about acceptability, limits, and governance

Possible Solutions (Researchers)

Open-source models or transparent vendors

Success metrics

Accessible public consultations and feedback

Greater transparency

Better record-keeping and reporting

Can researchers do all this alone?

Supporting Accountability: A “Problem of Many Hands”

New or better policies required: institutional, funding agencies, journals

Multi-stakeholder forums to establish governance strategies

New or adapted tools required for documentation

Additional digital infrastructure required

Public registries

Data and code repositories or registries

Conclusions

First generation of criminal justice RCTs: setting a precedent?

- Insufficient transparency
- Insufficient oversight

Reframe success: not just results but ensuring accountability, too

What do guardrails look like?

What trials are permissible, and under what circumstances?

Are field trials of AI in the criminal justice system ever permissible?

Thank You for Listening

Peter Darch, ptdarch@Illinois.edu