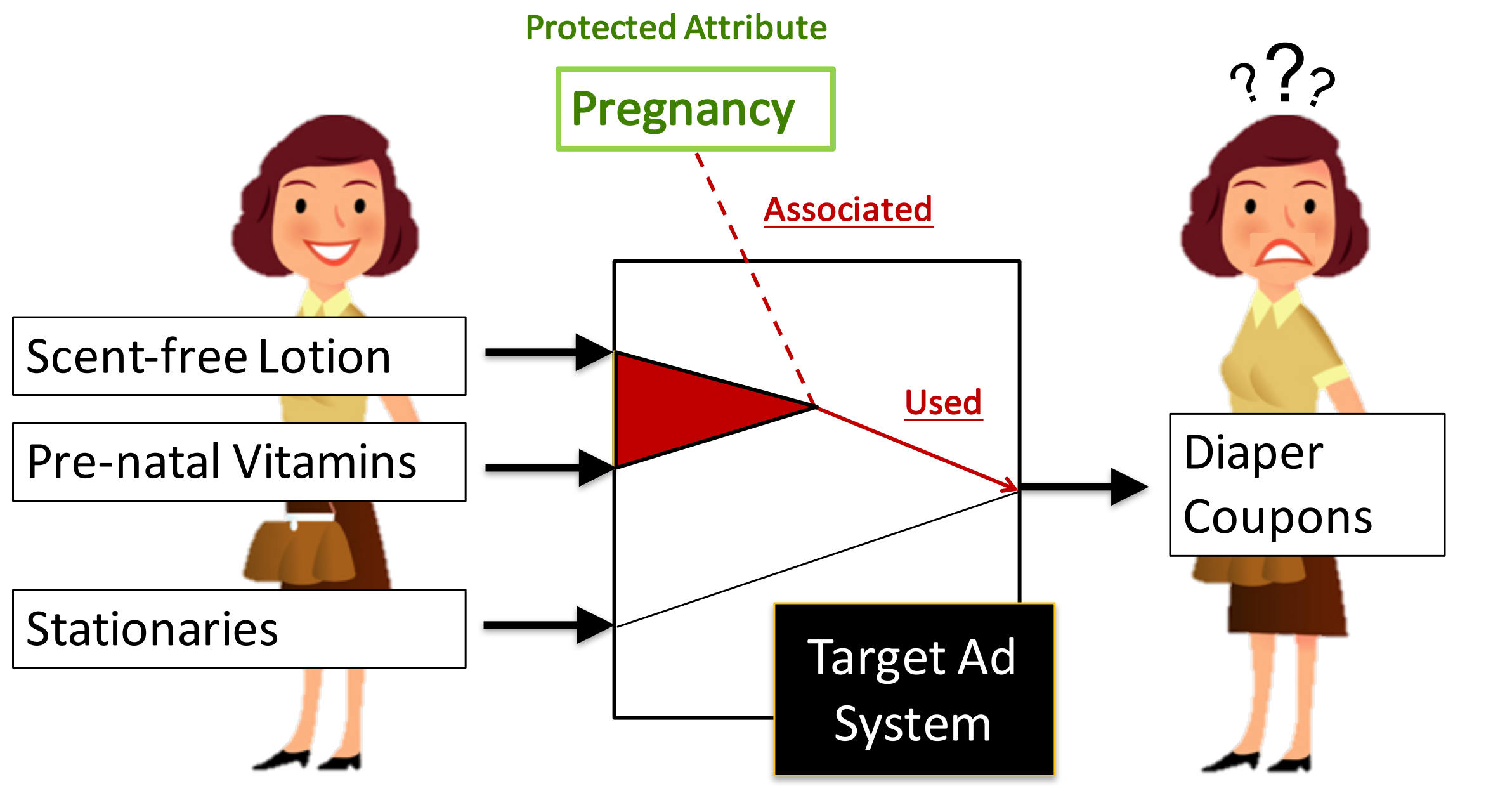
Use Privacy in Data-Driven Systems

Theory and Experiments with Machine Learnt Systems

Anupam Datta, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, Shayak Sen
Carnegie Mellon University

Harms due to inappropriate use in data-driven systems

Credit	Web systems	Healthcare
Education	Law Enforcement	…

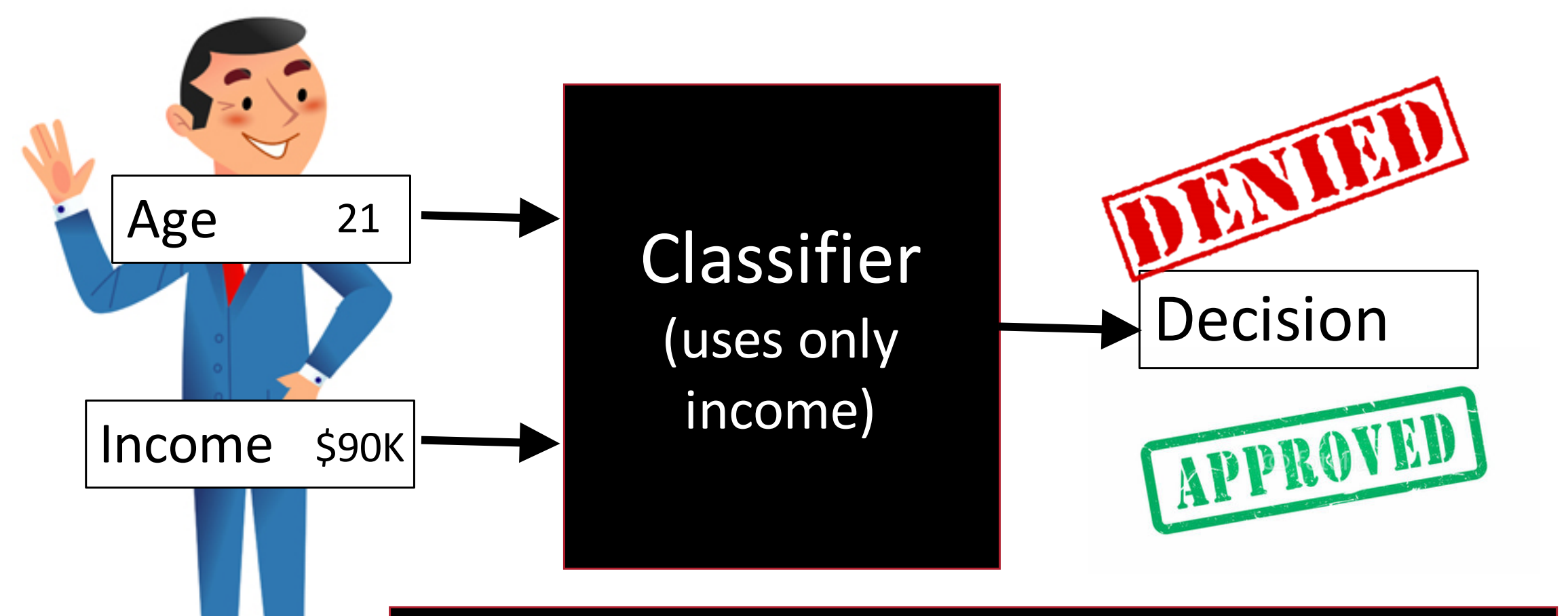


Using pregnancy status (inferred via past purchases) for marketing **[Target 2012]**

Use privacy constraints restrict the use of protected information types and some of their proxies in data-driven systems.

Explicit Use

Quantitative Input Influence*

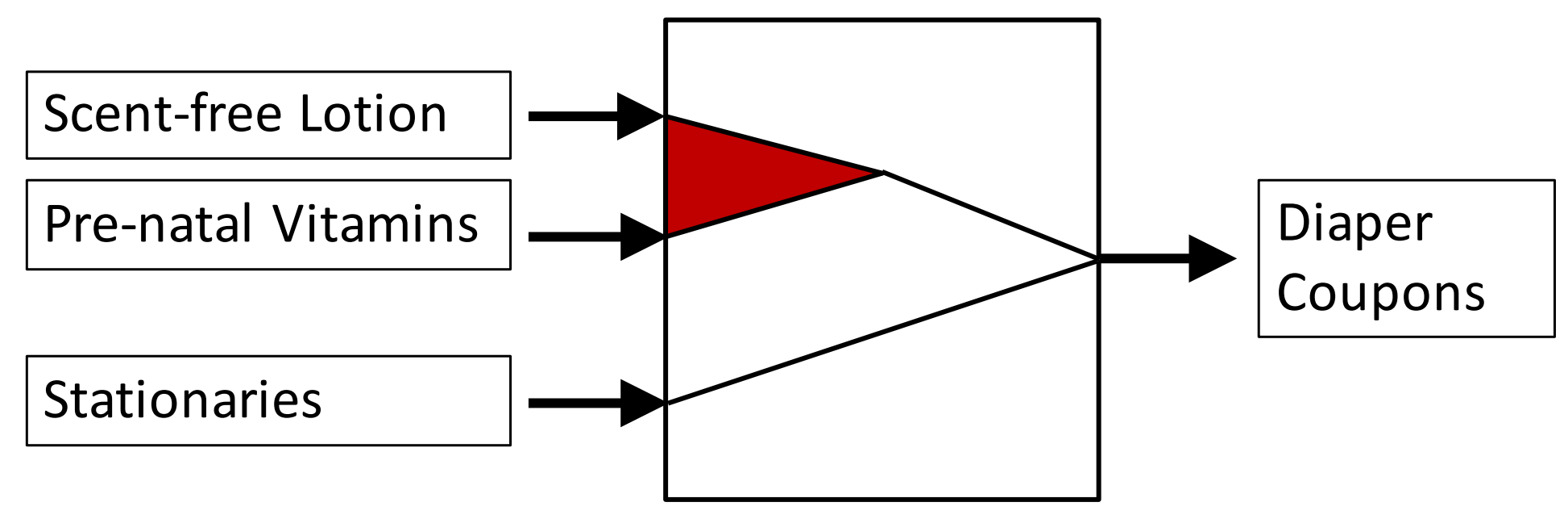


Replace feature with random values from the population, and examine distribution over outcomes.

*Anupam Datta, Shayak Sen, Yair Zick. *Algorithmic Transparency via Quantitative Input Influence.* Oakland'16

Proxy (or implicit) Use

Learning Systems as Programs



Implicate **part** (subprogram) of a system as the source of a violation

Subprogram

Program

```
OR ( IF { AND ( Shopped(Pre-natal vitamins),
                Shopped(Scent-free lotion),
                … ),
          Diaper Coupons, Nothing }
     IF { Shopped(Stationaries), Notebooks, Nothing } } )
```

Two-Phase Definition

Proxy *Is a subprogram associated to the protected attribute?*

➔ Well-studied association measures (e.g., Mutual Information)

Use *Is a subprogram influential to the output of the program?*

➔ QII for subprograms

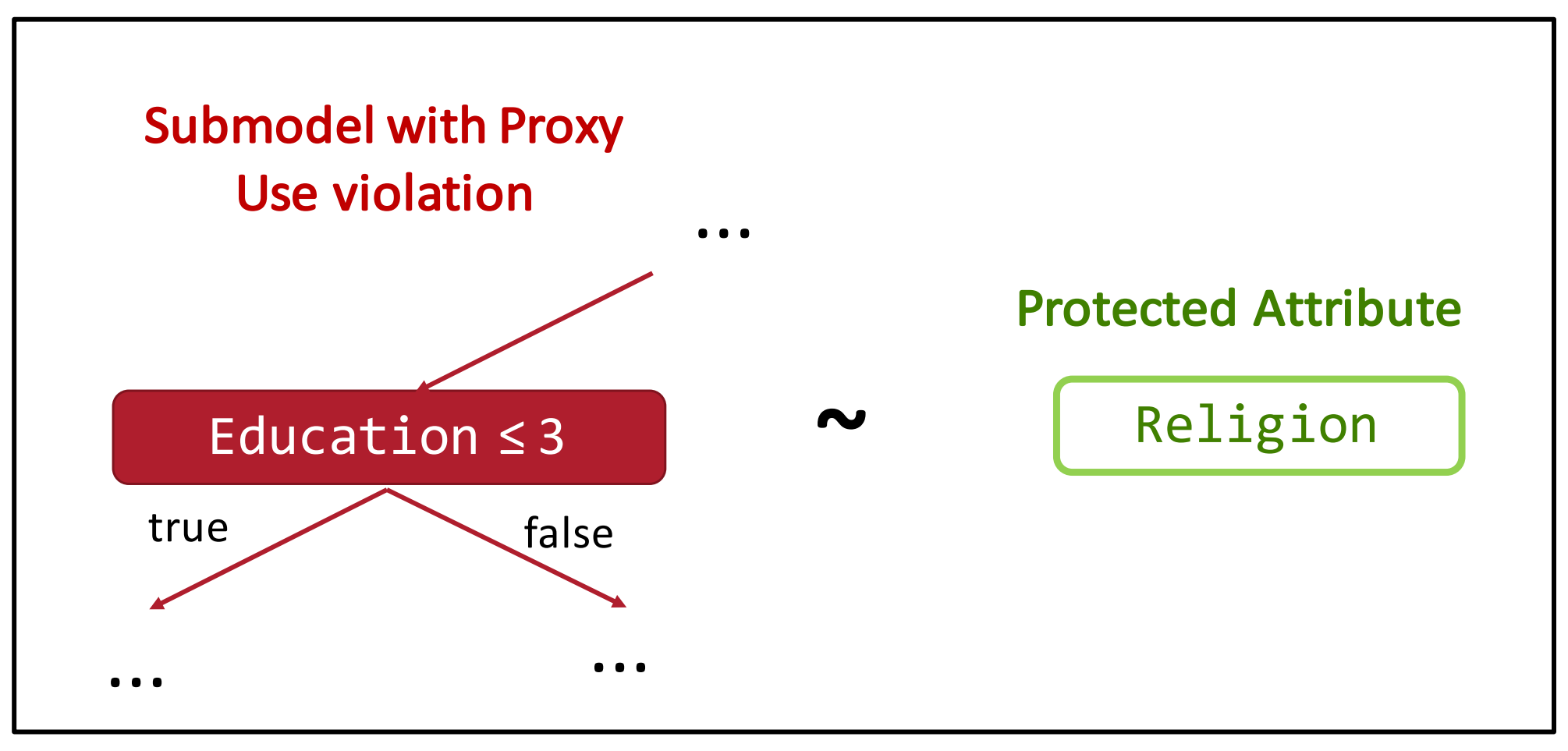
(ϵ, δ)-Proxy Use: A subprogram with association level above ϵ, and influence measure above δ exists

Experiments

contra

Advertisement targeting using the **Indonesian Contraception Dataset**

- Features: `Education, Children, Husband's Job`, etc
- Classification: `Contraception Methods`
- Protected attribute (removed in training phase): `Religion`
- ~1,500 individuals



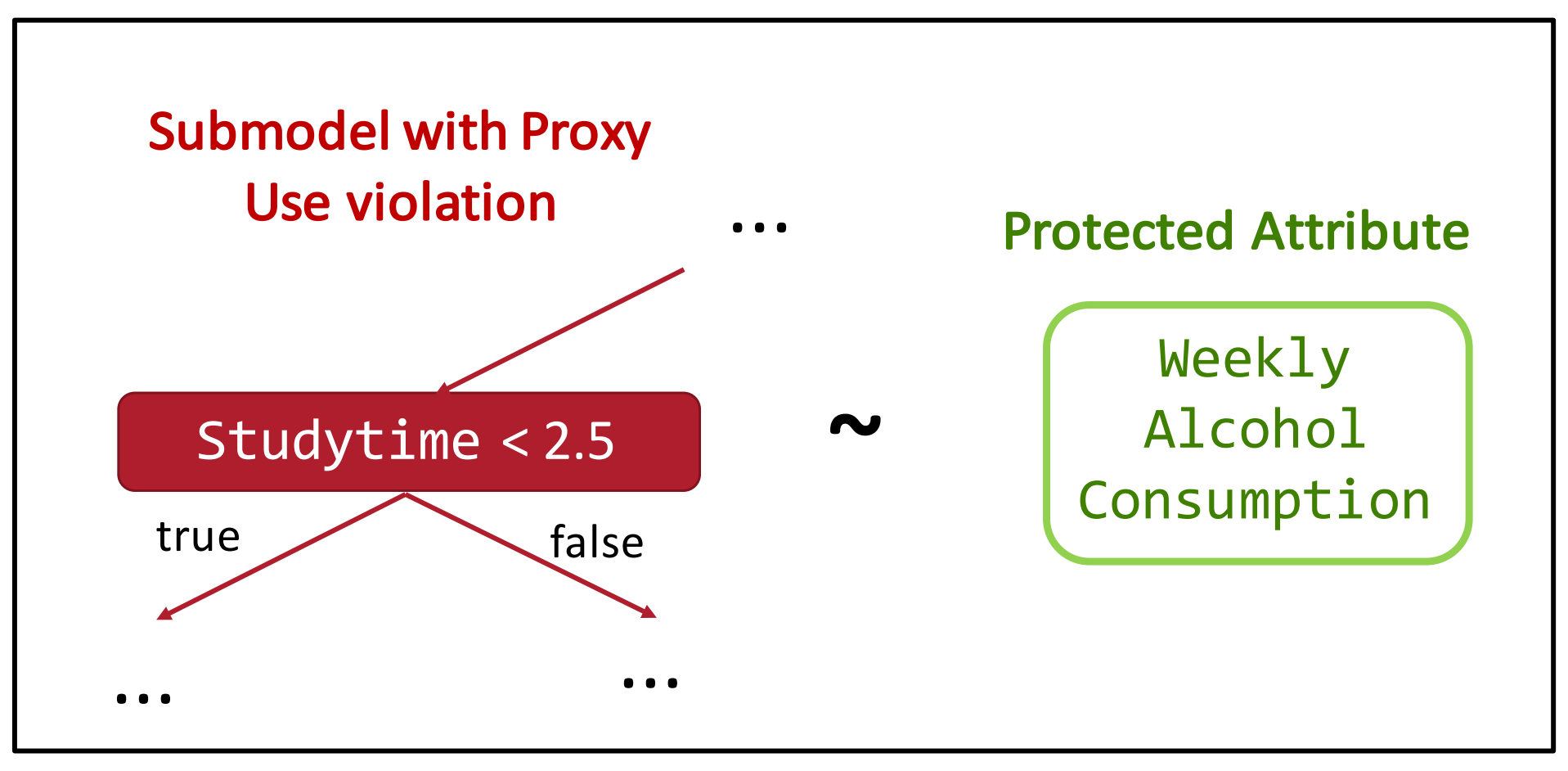
Education level is a proxy for religion

➔ **Concerning Use**

student

Academic performance prediction using **Portuguese Student Alcohol dataset**

- Features: `Failures, Studytime, Father's education level, Health status`, etc
- Classification: `Grade`
- Protected attribute : `Weekly alcohol consumption`
- ~7,000 individuals



Study time used as a predictor for the academic performance

➔ **Acceptable Use**

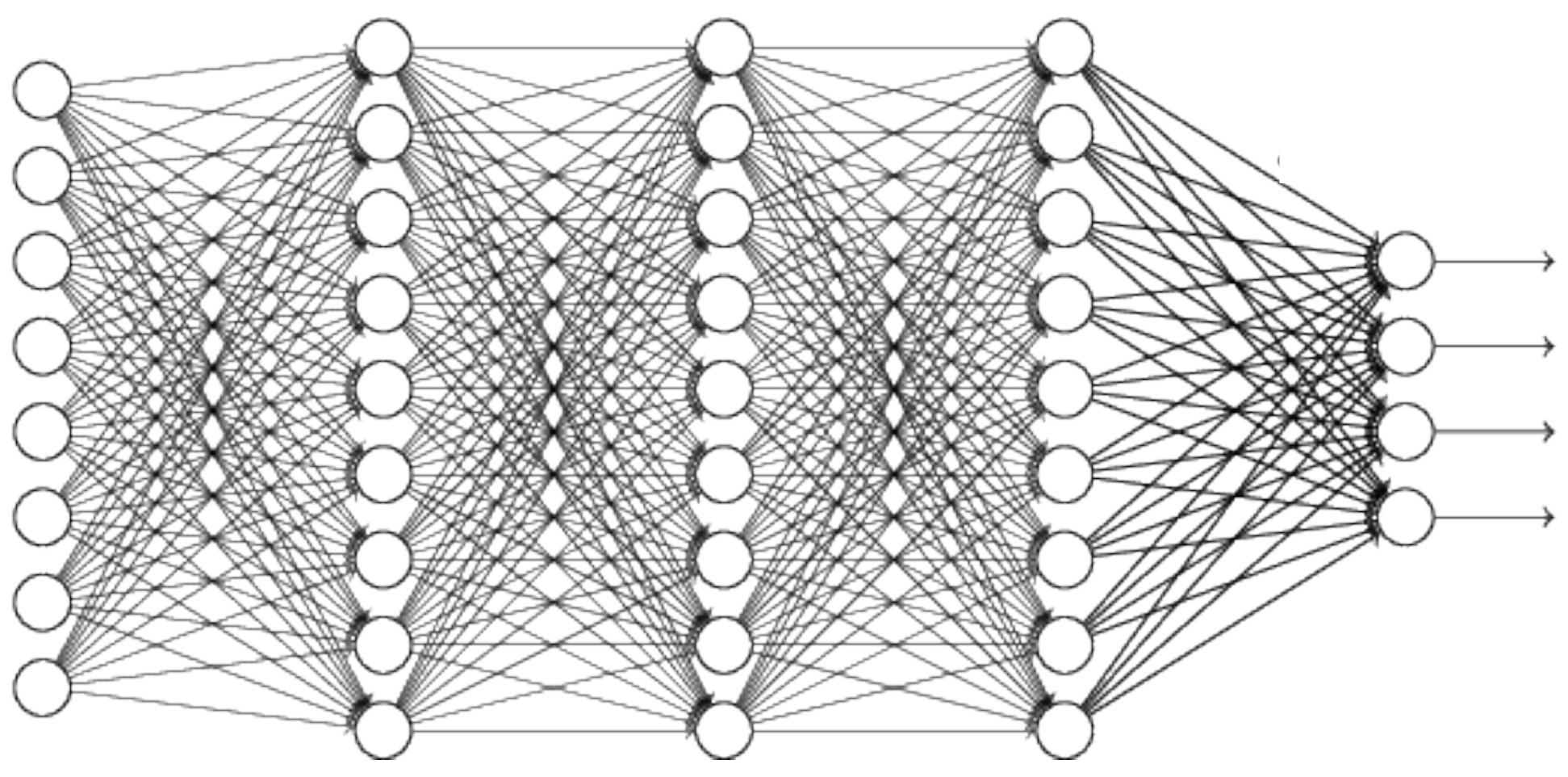
Summary

- **Use Privacy** restricts use (*explicit* or *proxy)* of protected information types
  - Axiomatically justified definition of proxy use
  - Algorithms for detection and repair
  - Implementation and evaluation for real world datasets, and commonly used ML models

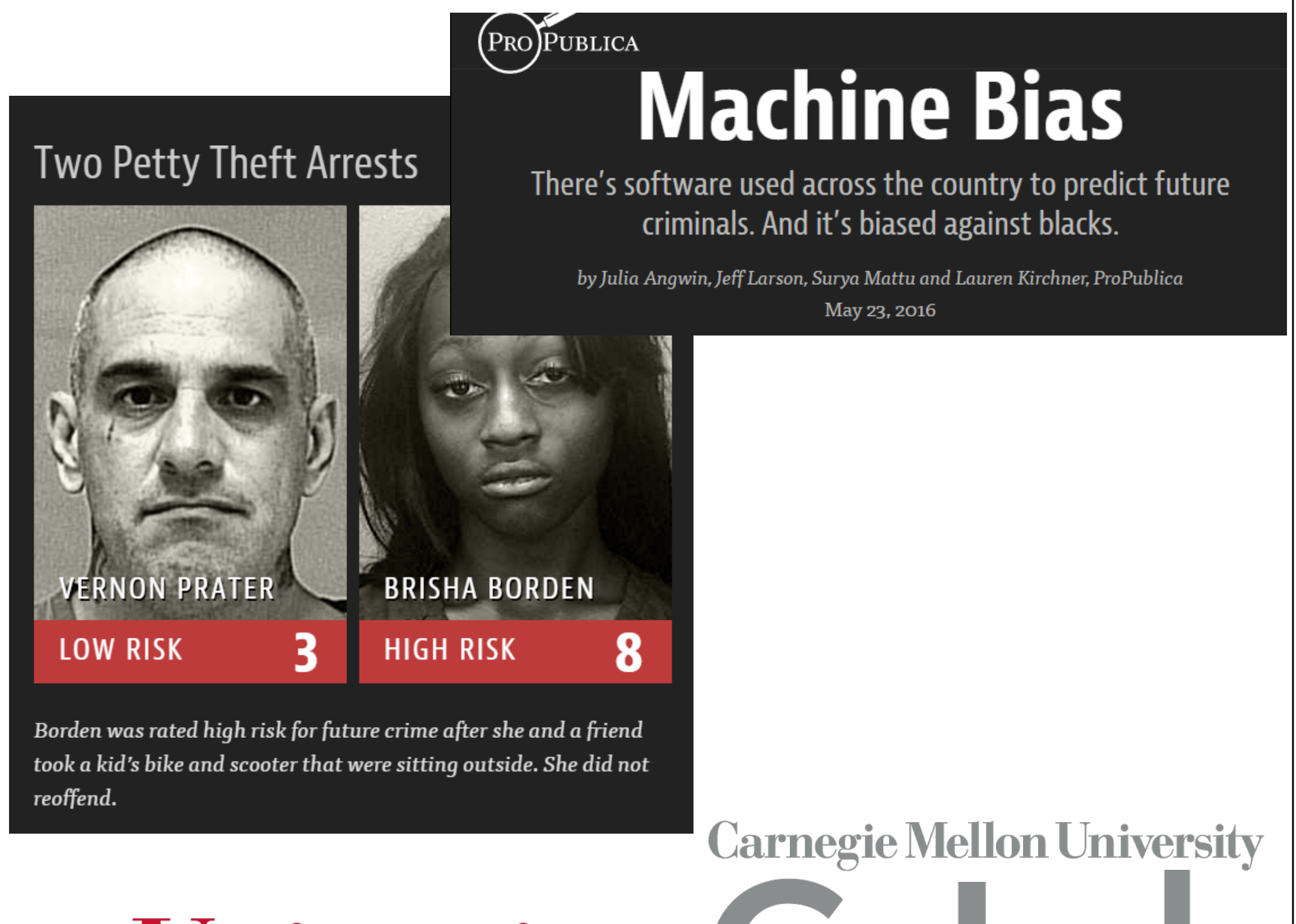
goo.gl/qfOv7G

Directions

Scale to larger systems and deep learning models



Implications for Fairness



PRO PUBLICA

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Two Petty Theft Arrests

VERNON PRATER — LOW RISK 3

BRISHA BORDEN — HIGH RISK 8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Carnegie Mellon University

Carnegie Mellon University CyLab Security and Privacy Institute