Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation

I. Elizabeth Kumar, *Brown University*
Keegan Hines, *Arthur AI*
John P. Dickerson, *Arthur AI*

EAAMO 2022

The Problem of Parallel Conversations

ML Fairness Research

- Assumes access to protected class information
- Focus on comparing outcomes or sometimes causal influence of input features
- Implicitly assumes some version of "disparate impact" theory can be directly applied to fairness statistics

U.S. Discrimination Law/Policy

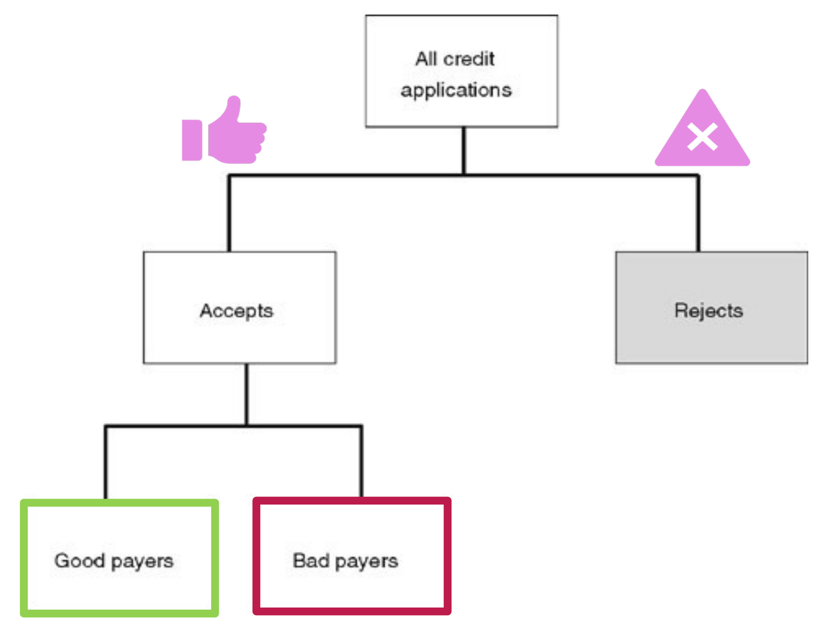
- Strong limitations on access to protected class information
- Assign meaning/appropriateness to the use of different types of input features
- "Discrimination" is based on principles of procedural justice, not defined by a statistic

In our paper, we:

- **Provide an overview of the current landscape** of credit-specific U.S. anti-discrimination law and how it pertains to algorithms for Fair ML researchers
- **Contextualize Fair ML metrics and results in terms of those metrics** to the realities of credit data to identify "discrimination risks" in the credit setting
- **Discuss regulatory opportunities** to address those risks

Discrimination Risk: Credit Invisibility

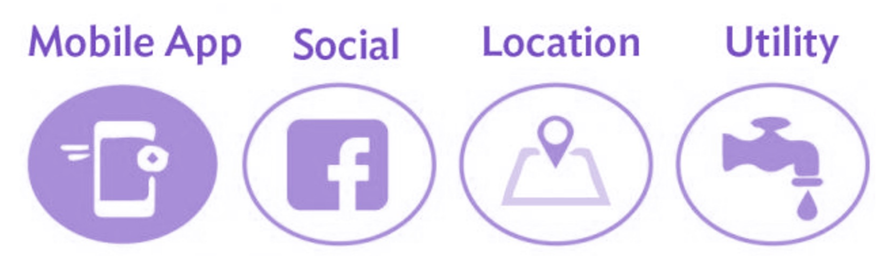
Bias due to **sampling processes** in training data



- Historical loan repayment data is less available on historically underrepresented groups, which can lead to higher error on those groups
- May result in issuing more loans that cannot be paid back

Discrimination Risk: Alternative Data

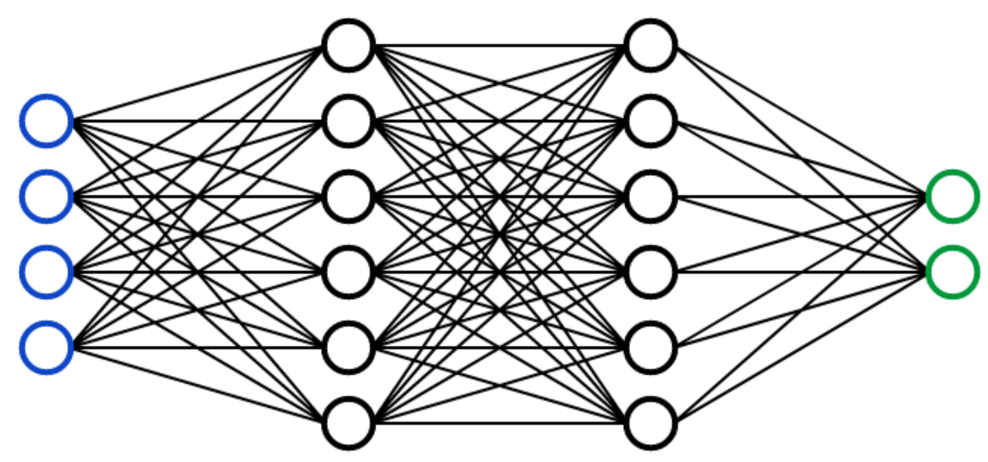
Observational bias and **measurement validity**



- Some alternative data (i.e. cash flow data) may allow more accurate underwriting of previously "credit invisible" applicants
- Other alternative data may be predictive of credit risk for different *reasons* than traditional data, and may not be related to qualities that we should accept as a reasonable basis for decision-making

Discrimination Risk: Model Complexity

Multivariate discriminatory effects are affected by **model capacity**



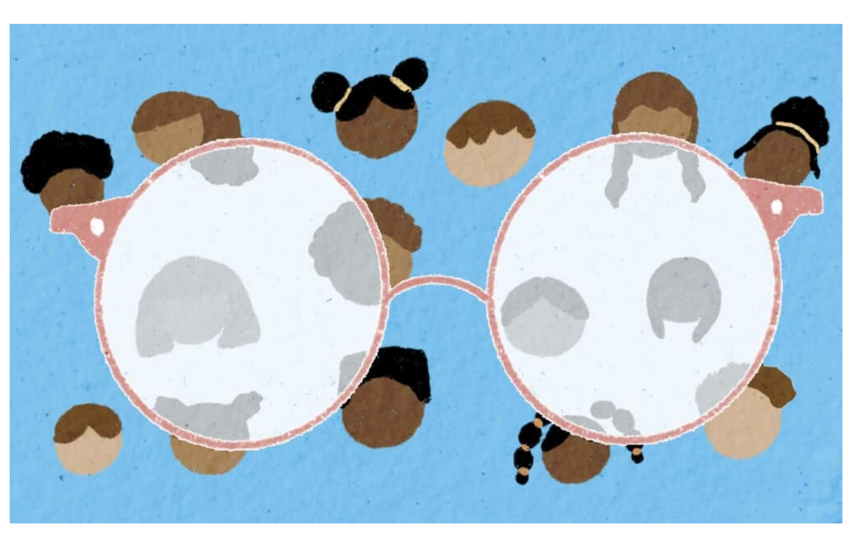
- Low-capacity models on data which is disparately predictive between classes may result in low cost-based fairness
- High-capacity models on predictive data can be have more unequal outcomes than simple models if there is bias in the labels

Equal Credit Opportunity Act

- Strict data collection rules
- Difficult to prove discrimination occurred
- Enforced by multiple agencies
- Language and history does not *neatly* imply the relevance of any **particular fairness statistics**

Regulatory Opportunities

Expanding or encouraging protected data collection



Treating discrimination risk as a form of financial model risk

Images: MIT Tech Review, Fintech Futures, The Imprint, Shutterstock