

Debiasing Alternative Data for Credit Underwriting Using Causal Inference

Chris Lam

Epistamai

September 25, 2025

Limitations of Traditional Credit Scores



FEDERAL RESERVE BANK *of* KANSAS CITY

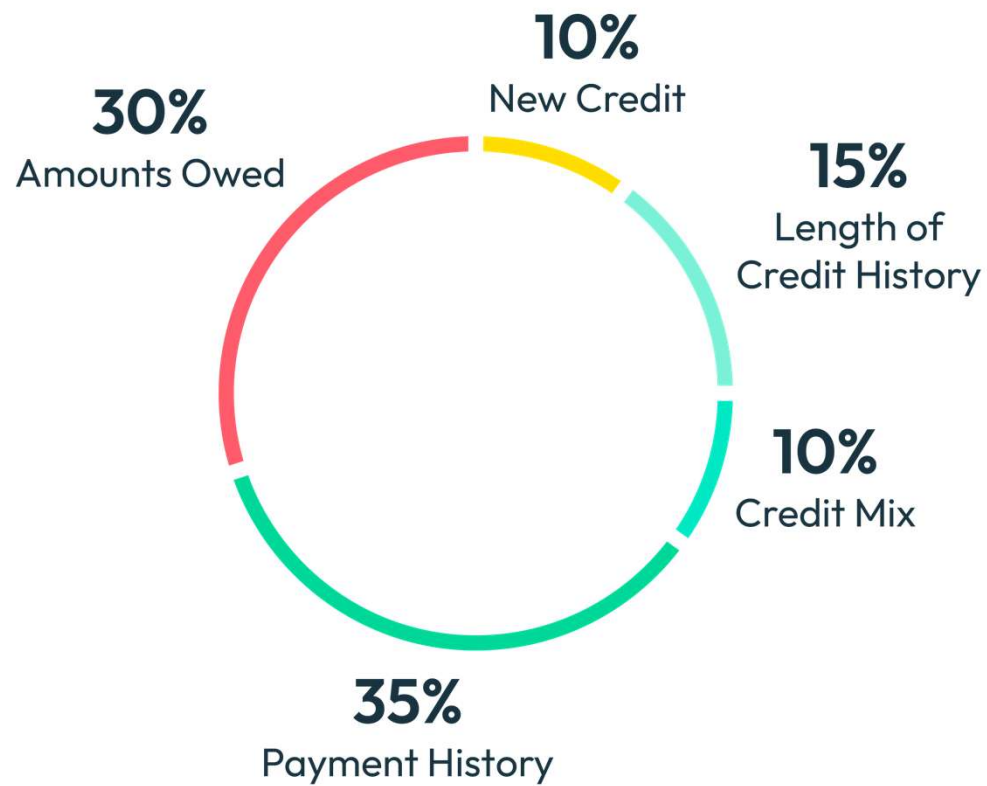


Senior Economist Ying Lei Toh

Addressing Traditional Credit Scores as a Barrier to Accessing Affordable Credit

1. Traditional credit scores do not always reflect a consumer's creditworthiness.
2. They may also disproportionately punish consumers from economically disadvantaged groups.

What Goes Into a FICO Score?



Potential Sources of Alternative Data



Education



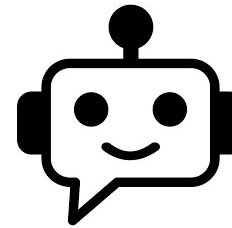
Employment



Shopping
Behavior



Psychometrics



Natural
Language

Expanding Credit Access to Underserved Borrowers



Young People



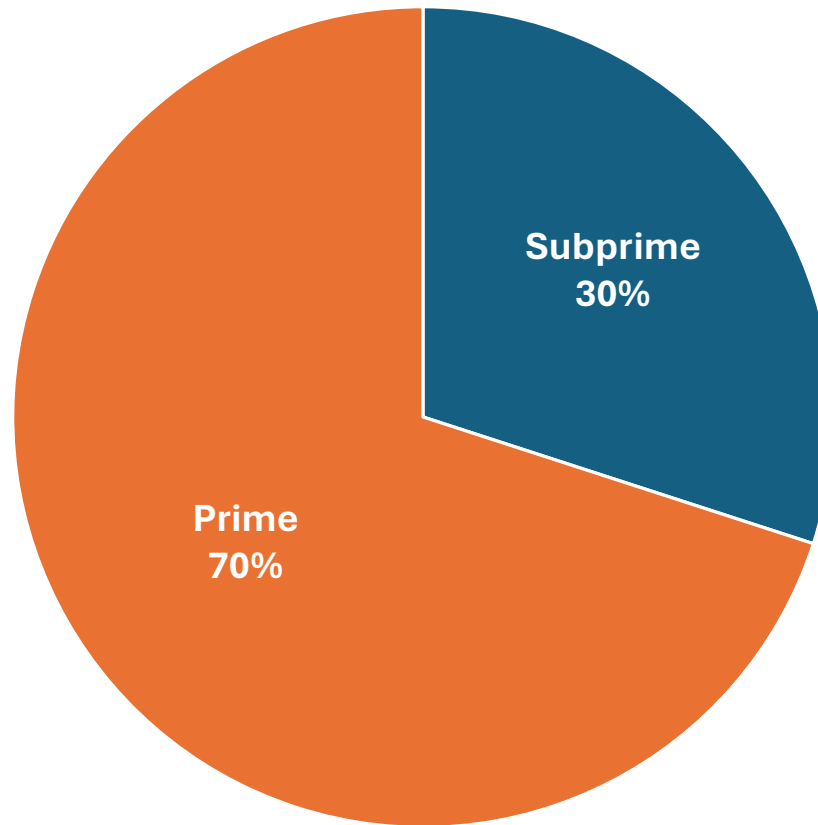
Recent Immigrants



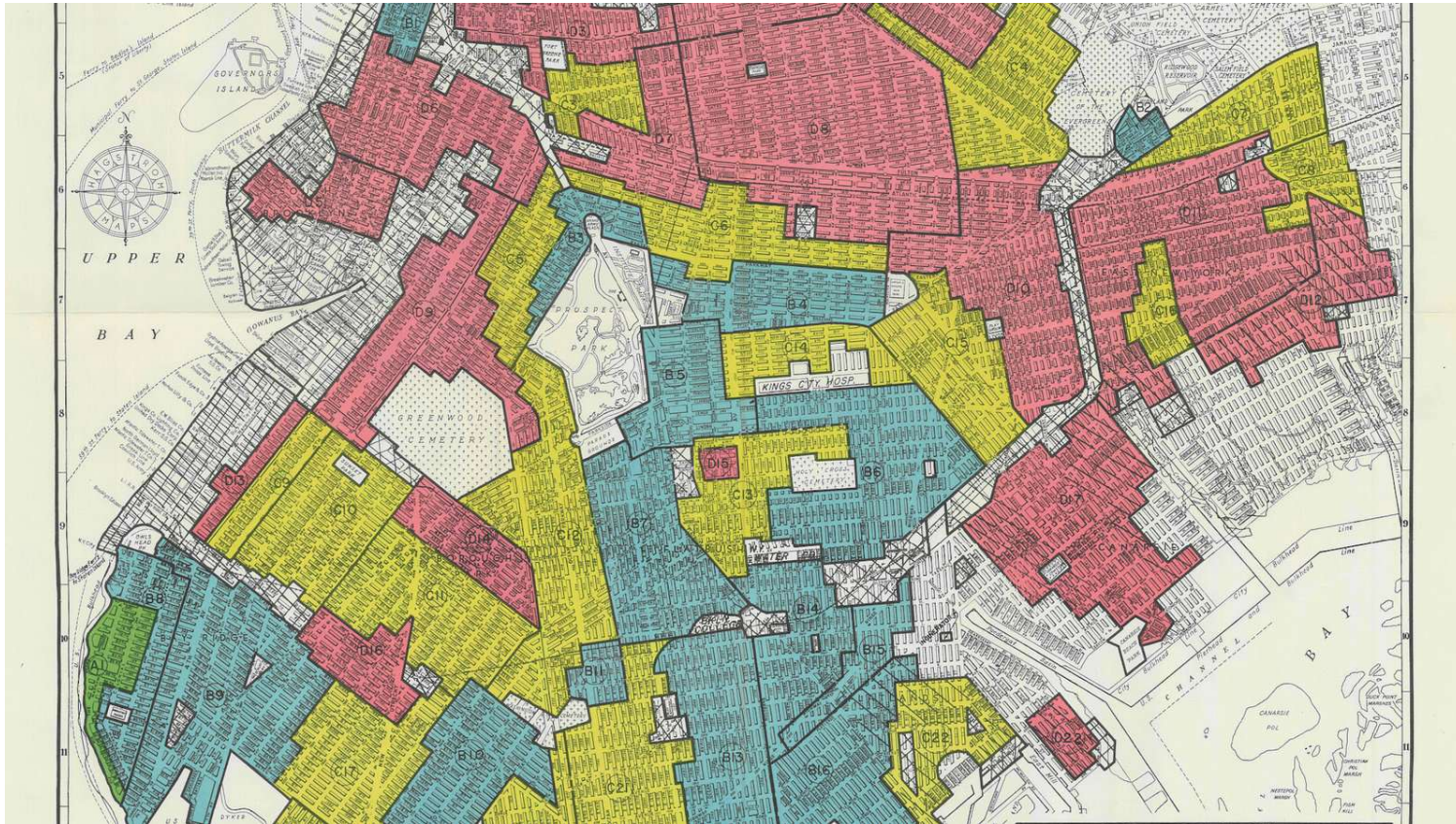
Gig Workers

Lowering Costs for Non-Prime Borrowers

US Consumers with Prime vs. Subprime Credit (Experian 2021)

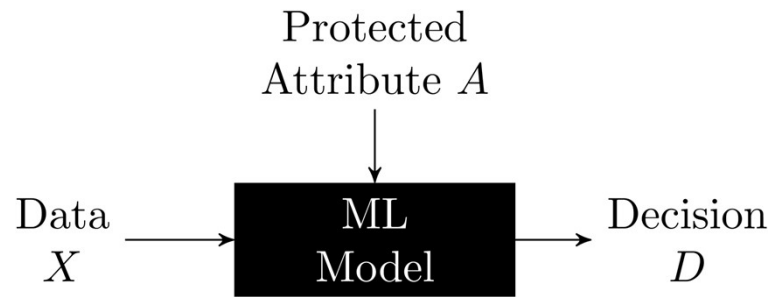


But How Do We Prevent Digital Redlining?

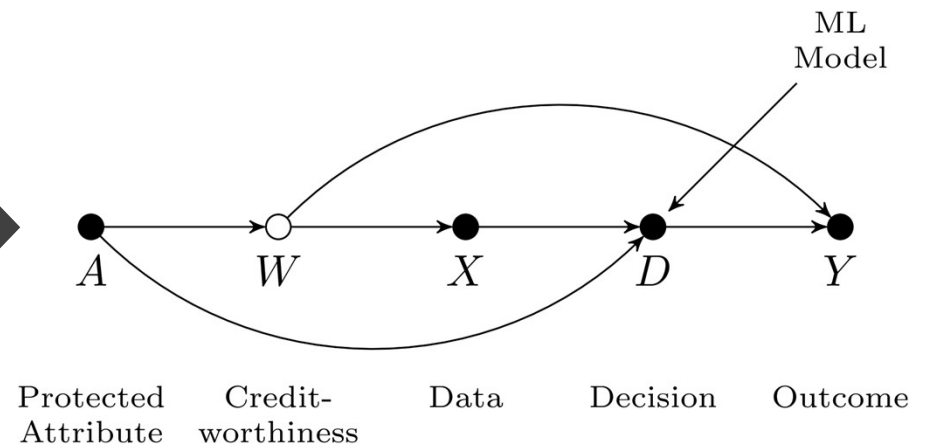


Think Outside the Black Box

Supervised Machine Learning *Block Diagram*



Causal Inference *Causal Bayesian Network*



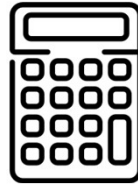
Protected Attribute A	Creditworthiness W	Data X	Decision D	Outcome Y
Race, Gender, Religion, Age	Character, Capacity, Capital, Collateral, Conditions	Income, Amounts Owed, Credit History	Approve or Deny Loan	Default or Non-Default

The 5 C's of Creditworthiness



Character

*Will the borrower
repay the loan?*



Capacity

*Can the borrower repay the
loan from their income?*



Capital

*How much is the borrower
personally invested?*



Collateral

*What can the lender claim if
the borrower fails to pay?*



Conditions

*What outside factors might
influence repayment?*

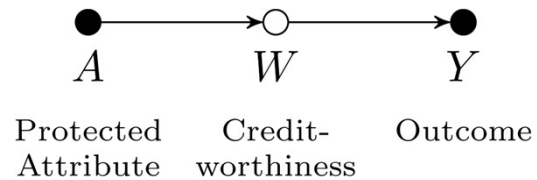
What's the Basis for Using Data in Credit Scoring?

Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks (Federal Reserve System)

“The first question to ask before using new data is the basis for considering the data. If the data are used in the credit decision-making process, what is the **nexus with creditworthiness**? Some data have an obvious link to creditworthiness and are logical extensions of current underwriting practices, while others are less obvious.”

Nexus With Creditworthiness

In causal inference, this is a **mediation analysis** problem:



Creditworthiness W may be considered biased due to structural discrimination.

Direct Effect

$$A \rightarrow W$$
$$W \rightarrow Y$$

Indirect Effect

$$A \rightarrow W \rightarrow Y$$

No Direct Effect

$$A \not\rightarrow Y$$

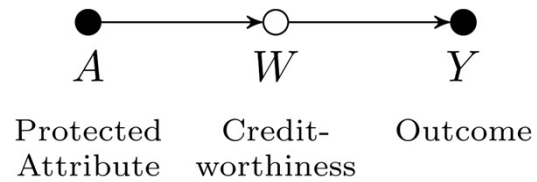
$$Y \perp\!\!\!\perp A \mid W$$

Traditional Credit Scoring (e.g. FICO)

We convert this causal model into an associational model:

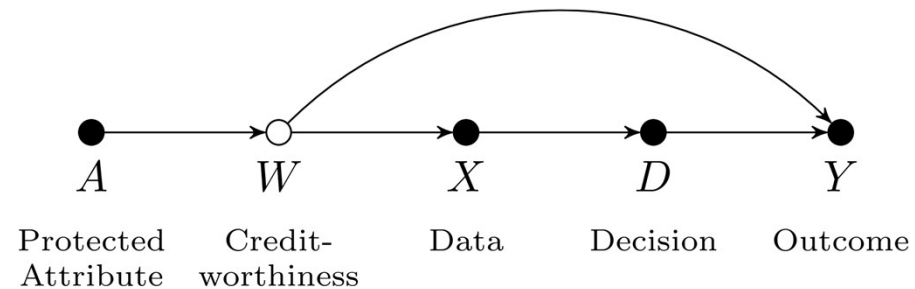
Causal Model

World Model
Mental Model



Associational Model

ML Model

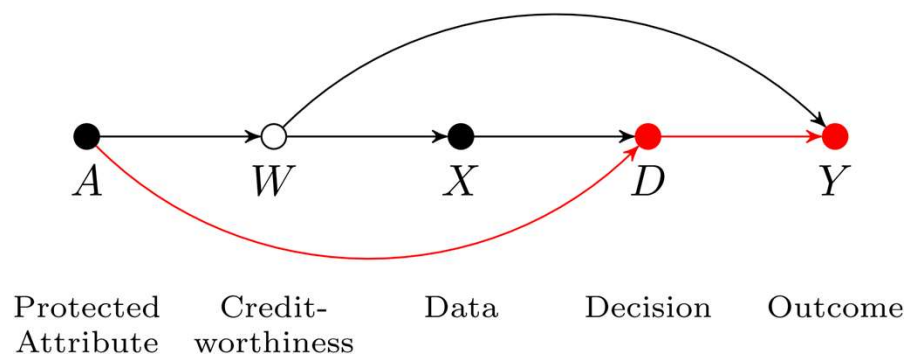


No Direct Effect

$A \nrightarrow D$

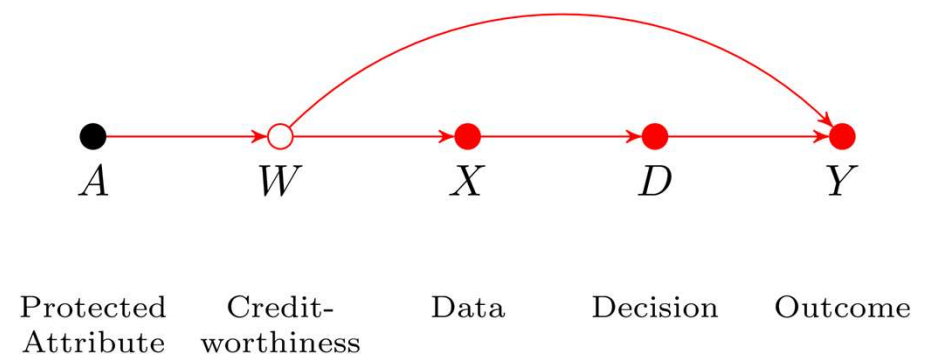
Modeling Illegal Discrimination in ML

Disparate Treatment (US)
Direct Discrimination (EU)



Negative Direct Effect
 $A \rightarrow D$

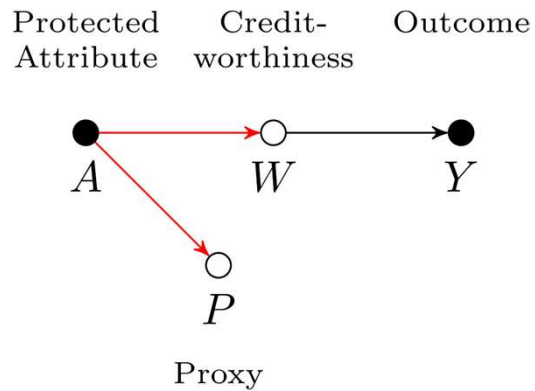
Disparate Impact (US)
Indirect Discrimination (EU)



Negative Indirect Effect
 $A \rightarrow W \rightarrow X \rightarrow D$

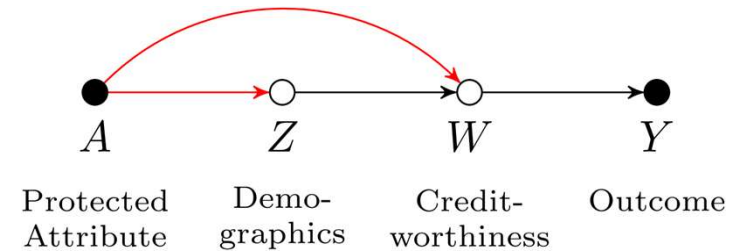
Modeling Proxies for a Protected Class

Proxy Variables P (E.g. Last Name)



Spurious Effect
 $P \leftarrow A \rightarrow W$

Demographic Variables Z (E.g. Zip Code, Education)



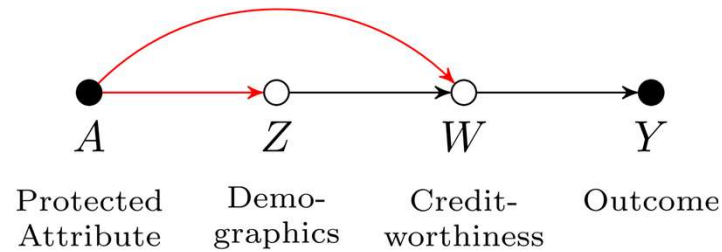
Direct Effect
 $Z \rightarrow W$

Spurious Effect
 $Z \leftarrow A \rightarrow W$

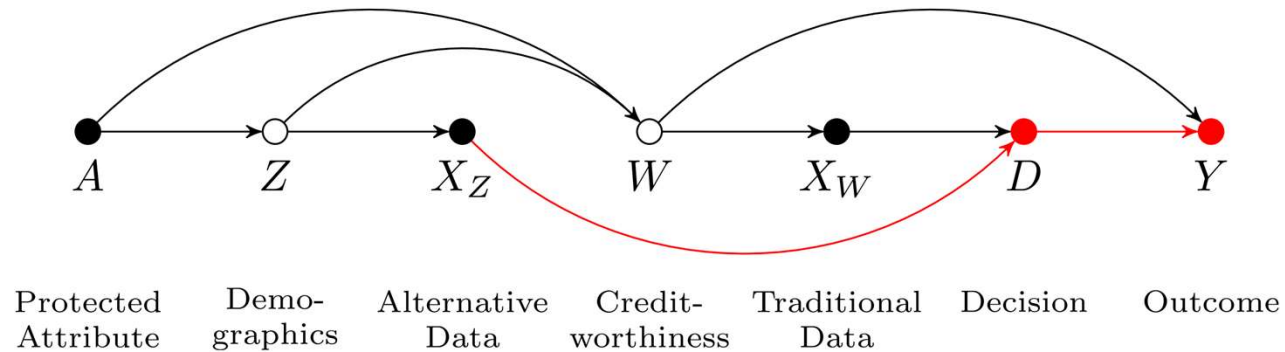
Modeling Redlining

We convert this causal model into an associational model:

Causal Model



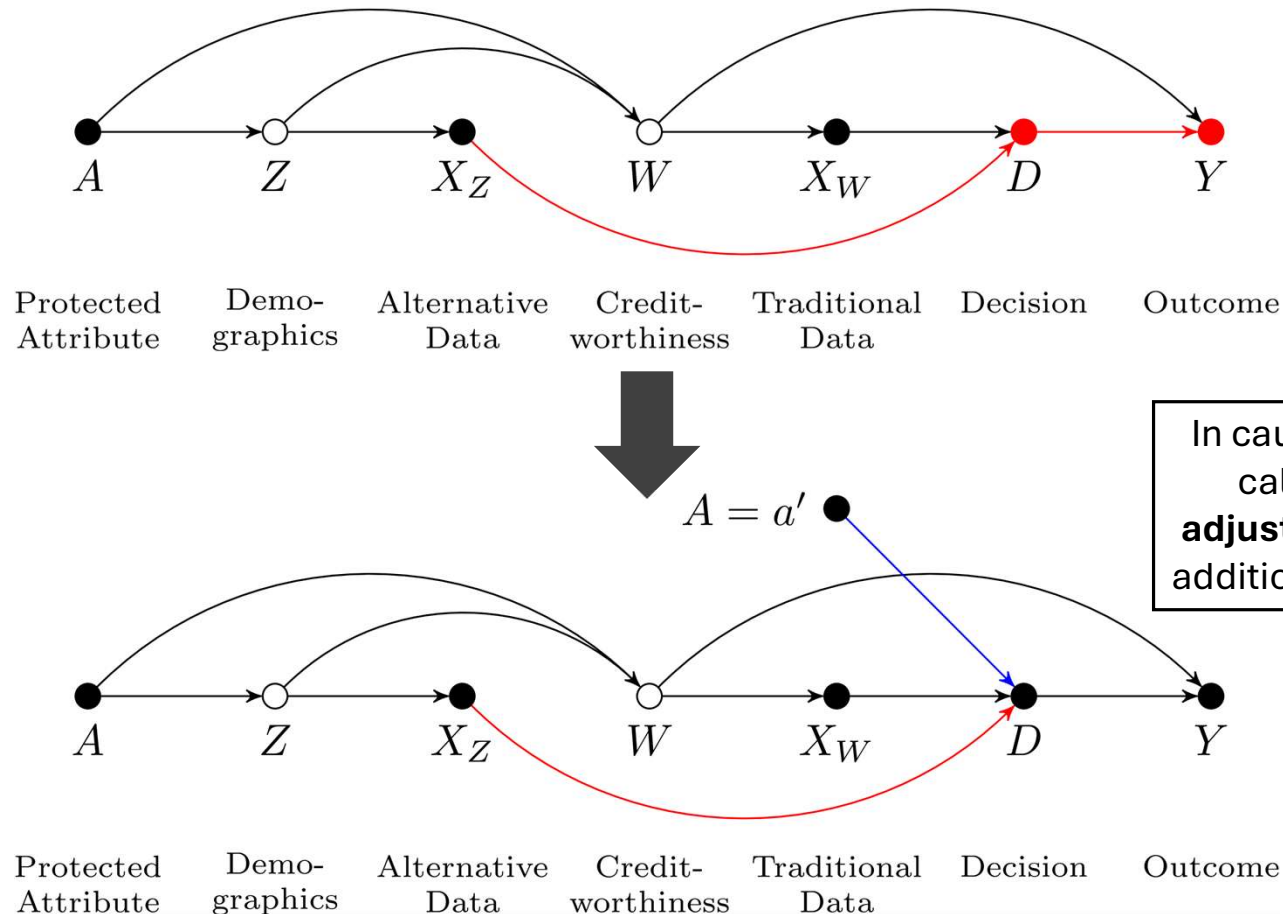
Associational Model



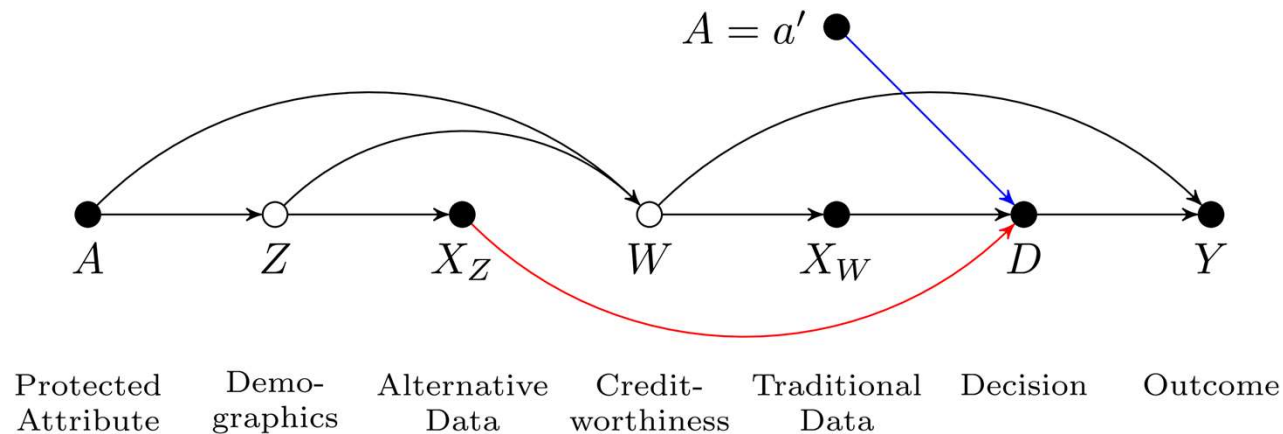
Negative Direct Effect
 $X_Z \rightarrow D$

Debiasing Alternative Data

We can perform an intervention on the protected attribute A to debias the decision D .



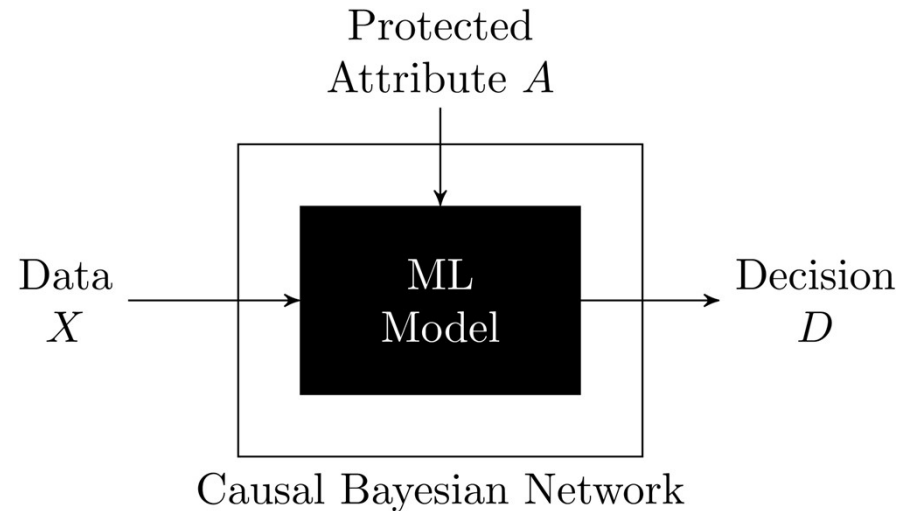
Debiasing Alternative Data



- **Training:** Perform a **backdoor adjustment** by training an ML model with the protected attribute A and alternative data X_Z .
- **Inference:** Perform an **intervention** on the protected attribute A to generate **counterfactual** predictions and take a weighted average.

Causal Data Science

A causal Bayesian network may act as a “wrapper” around a supervised machine learning model:



Causal Data Science

A causal Bayesian network may act as a “wrapper” around a supervised machine learning model:

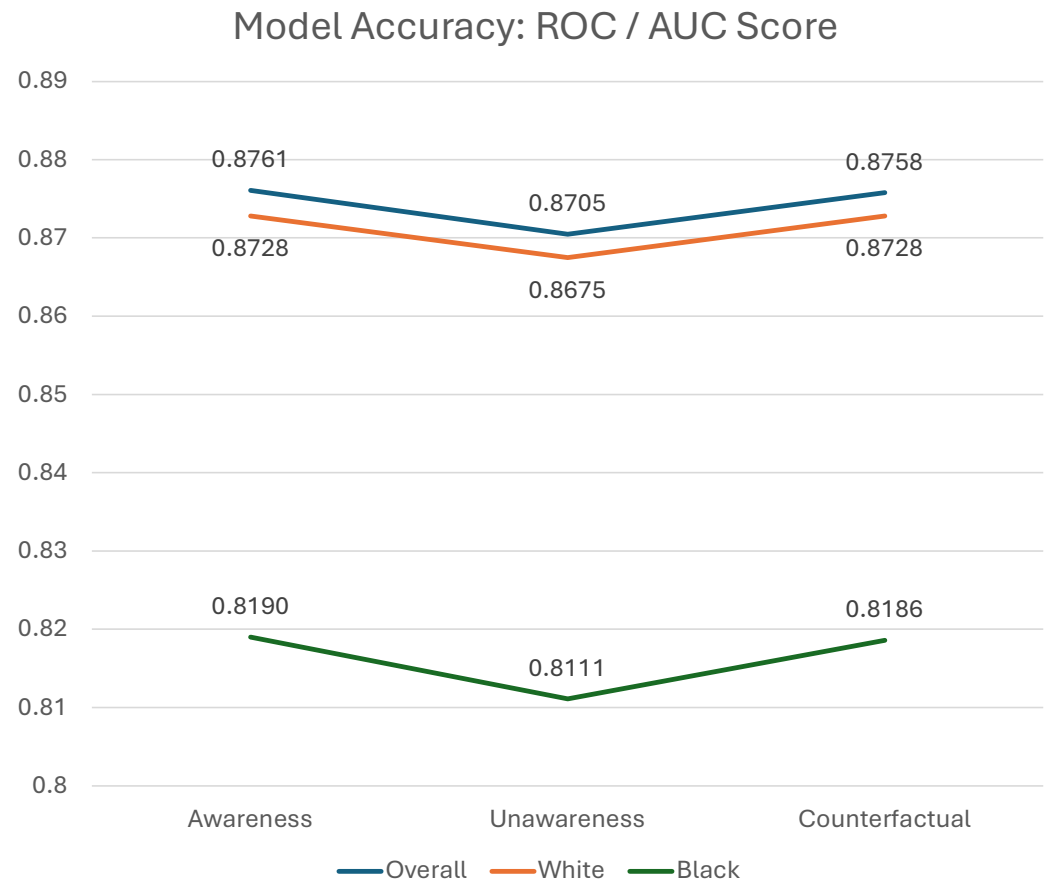
	Data Science	←	Causal Inference
1	Feature Selection	←	Mediation Analysis
2	Data Preparation	←	Overlap Testing
3	Training	←	Backdoor Adjustment
4	Inference	←	Interventions
5	Evaluation	←	Counterfactuals

Test Results

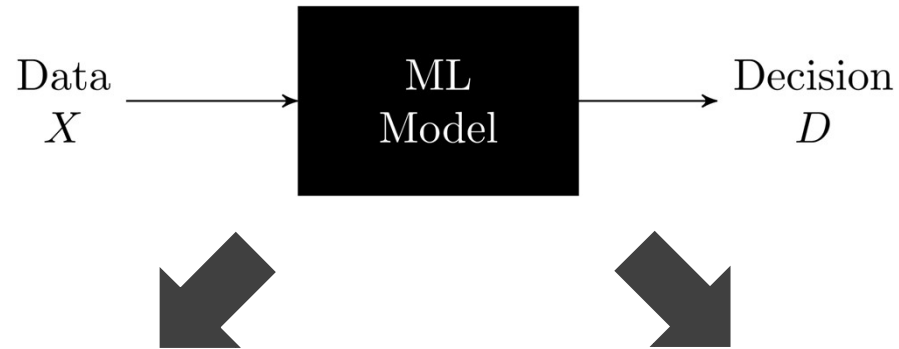
We built three models using a public dataset (National Survey of Mortgage Originations):

- **Awareness (Illegal)**
Includes all variables
- **Unawareness (Traditional)**
Dropped protected attributes and demographics
- **Counterfactual (Alternative)**
Included demographics but treated everyone equally

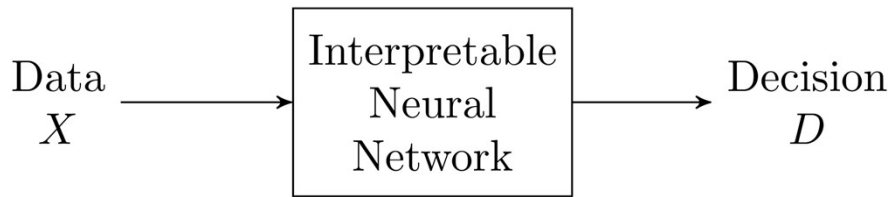
Using our debiasing algorithm, we were able to recover most of the loss (94%) by including alternative data without causing redlining.



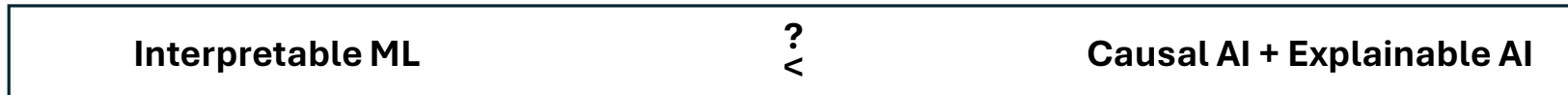
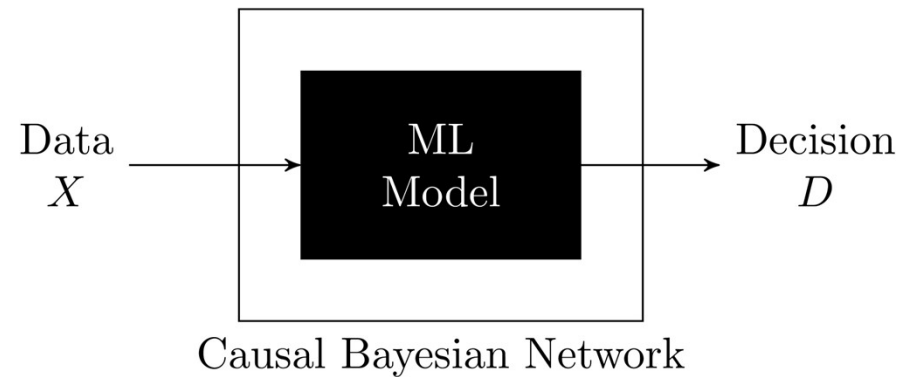
A Blind Spot in Credit Scoring?



FICO
"Replace" the black box



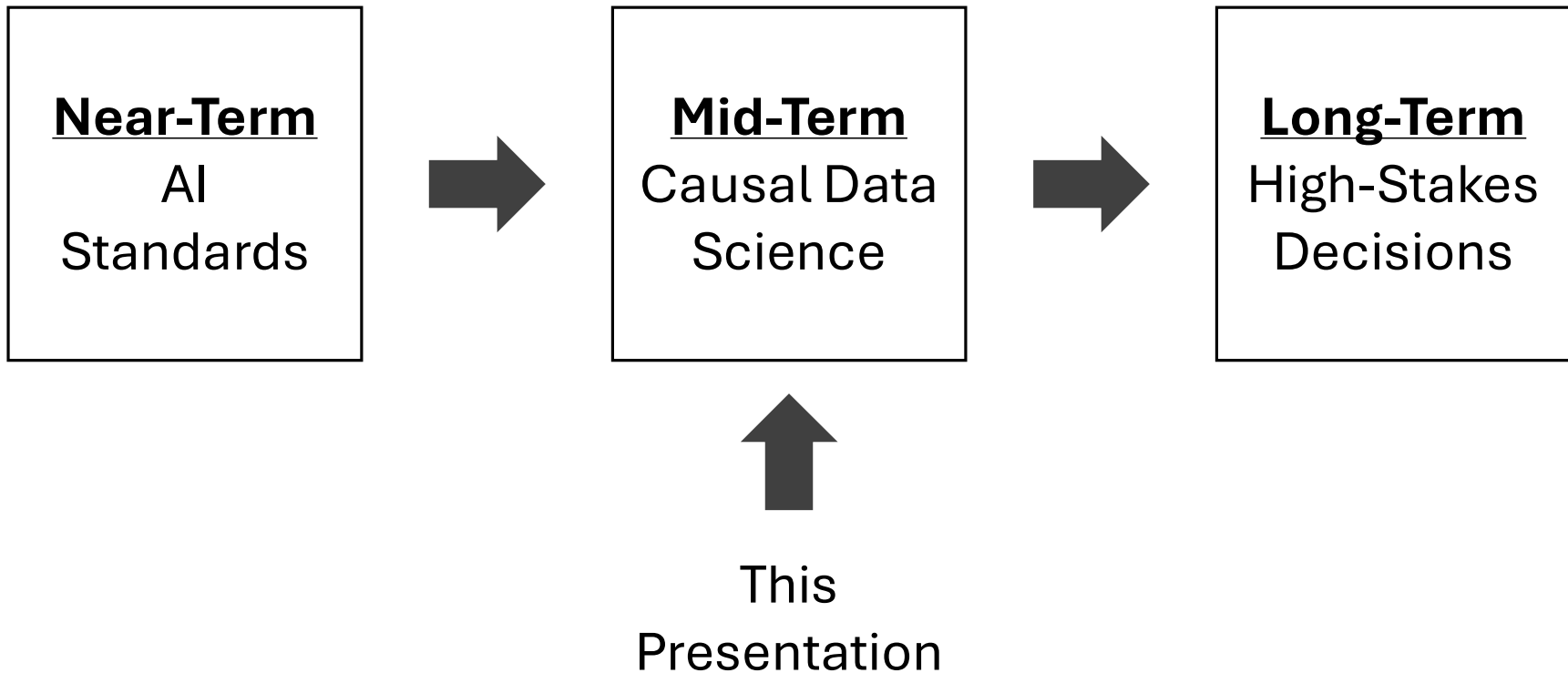
Epistamai
"Think outside" the black box



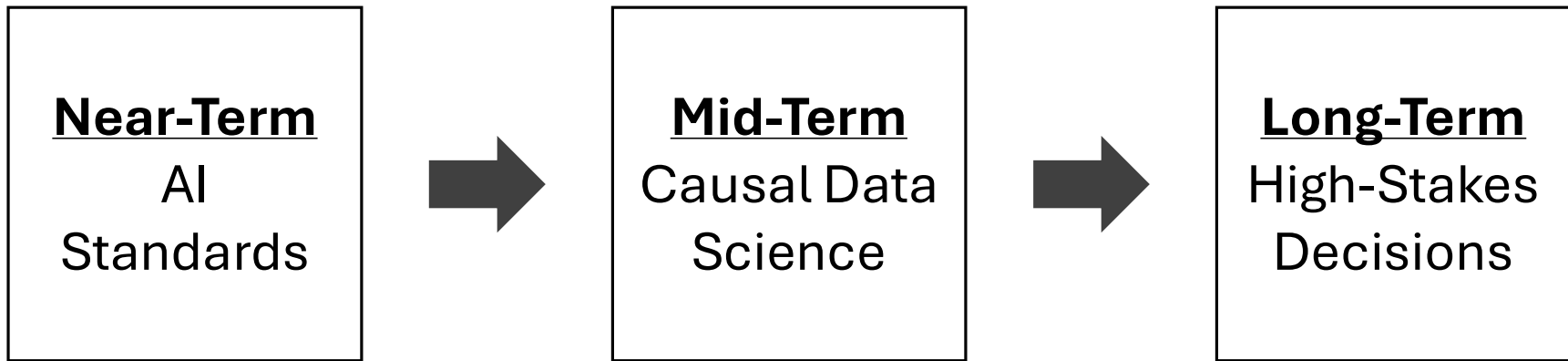
Next Generation Credit Underwriting

	1 st Generation
Analytical Technique	Logistic Regression
Traditional Data	
Income and debt level	✓
Credit and payment history	✓
Alternative Data	
Bank cash flow	
Rent and utilities	
Zip code and education	
Digital footprints	
Shopping behavior	
Personality / psychometrics	
Natural language / chatbots	

Causal AI Roadmap

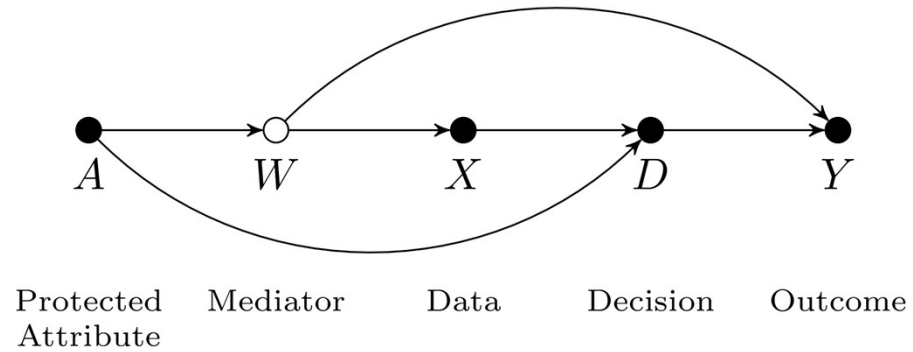


Causal AI Roadmap

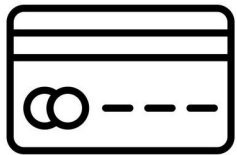


Now Let's
Look At This

Standard for Fair Decision Making Through Causal Analysis



A Standard Fairness Model for Multiple Sectors



Credit



Insurance



Housing



Employment



Education

IEEE P3591: A Rosetta Stone for AI

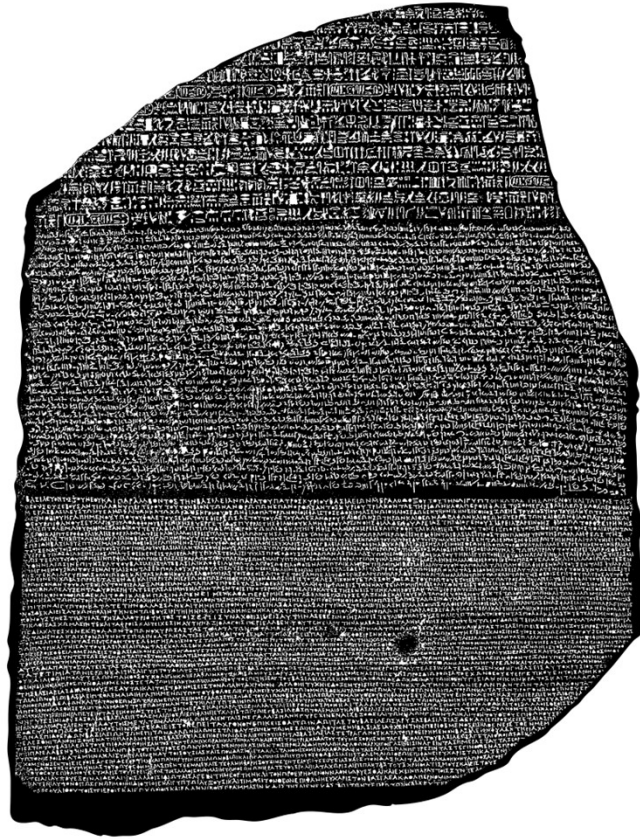
Law



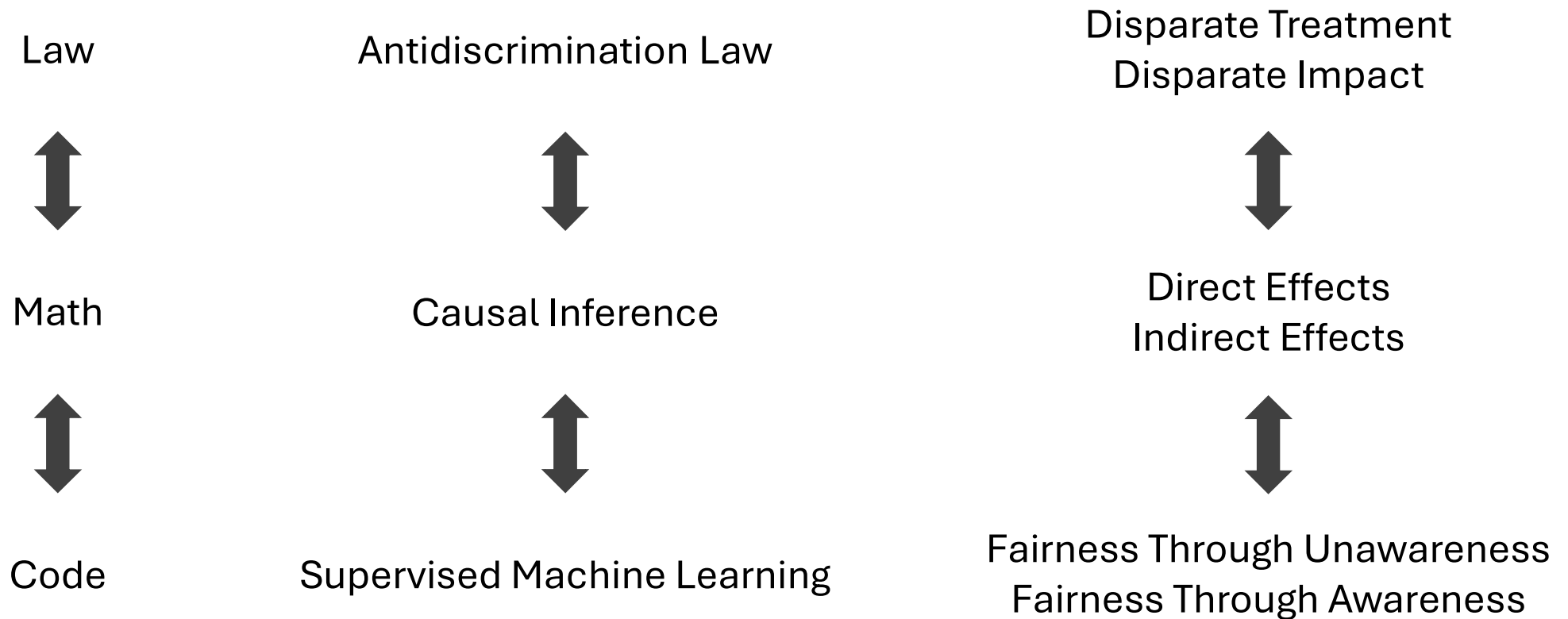
Math



Code



IEEE P3591: A Rosetta Stone for AI



IEEE P3591: Call for Participation

P3591

Standard for Fair Decision Making Through Causal Analysis

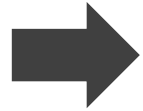
Express Interest in this Project

Active PAR

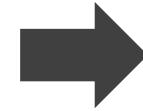
This standard describes how to perform causal fairness analysis to make fairer decisions in various high-stakes applications (e. g. credit, employment, education) that are more likely to be compliant with a country's antidiscrimination laws and regulations. It provides a standardized fairness model that encodes knowledge and assumptions about how to map the causal relationships between different variables such as a protected class (e. g. race, gender) and an outcome. The document provides the reader with a standardized language for directly translating concepts among the law, causal inference, and supervised machine learning. The standard provides criteria for selecting which variables to include in a machine learning model, how to train and deploy the model to make fairer predictions and decisions, as well as how to evaluate the model to determine the likelihood for illegal discrimination. However, specific algorithms for debiasing the data in a machine learning model to help ensure fairness are not covered in this standard. Finally, this standard is designed to be focused solely on legal compliance.

Causal AI Roadmap

Near-Term
AI
Standards



Mid-Term
Causal Data
Science

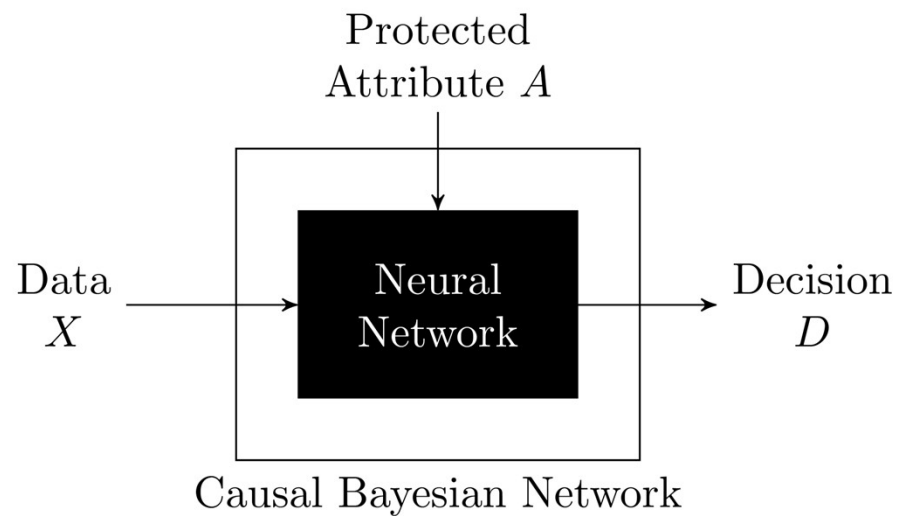


Long-Term
High-Stakes
Decisions

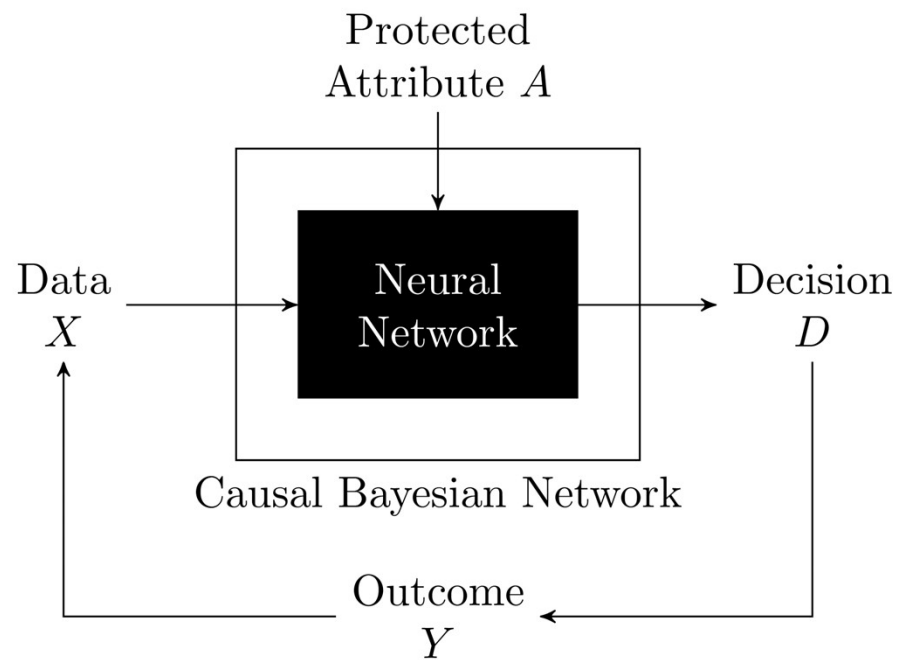


Now Let's
Look At This

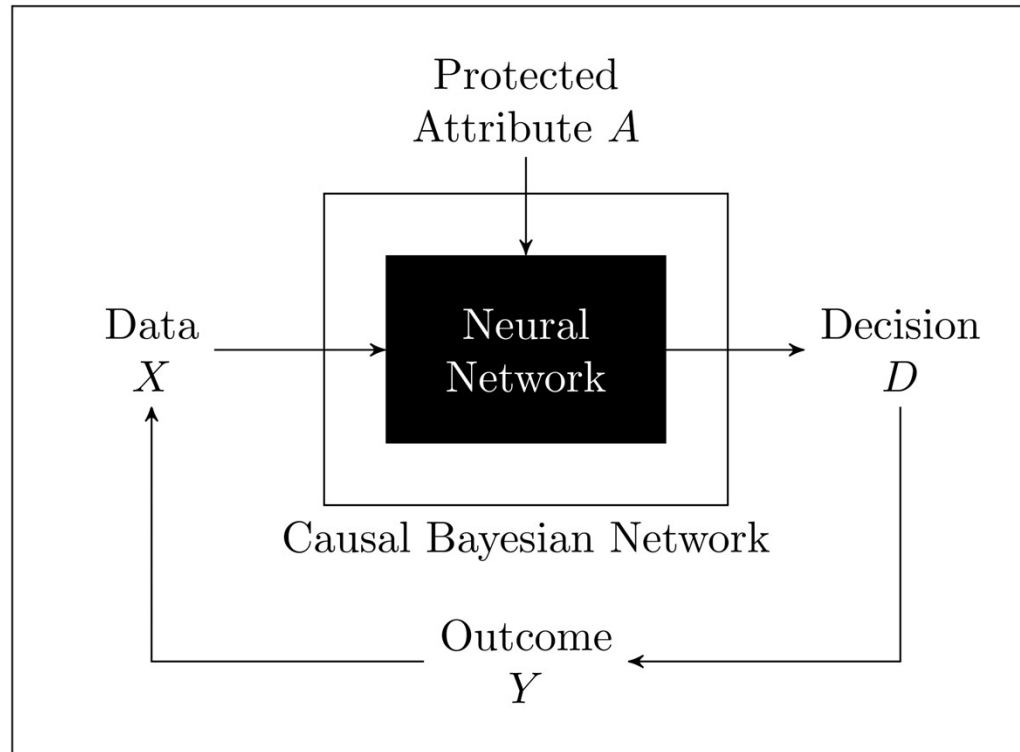
Neural Networks Have Inner and Outer Layers



Neural Networks Have Inner and Outer Layers

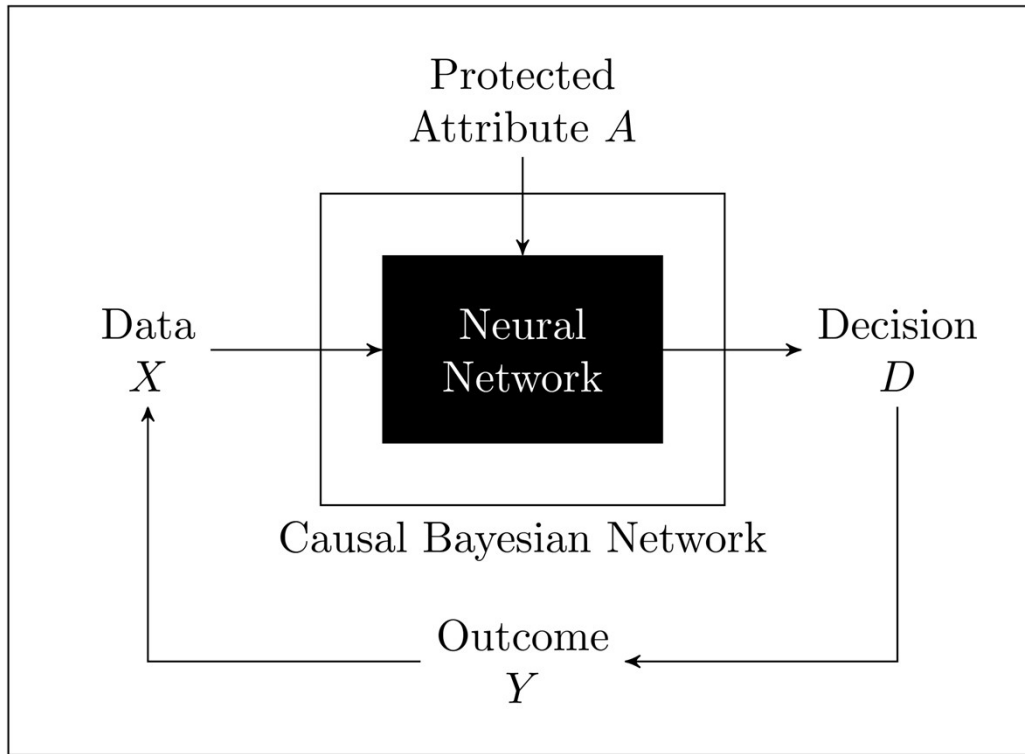


Neural Networks Have Inner and Outer Layers

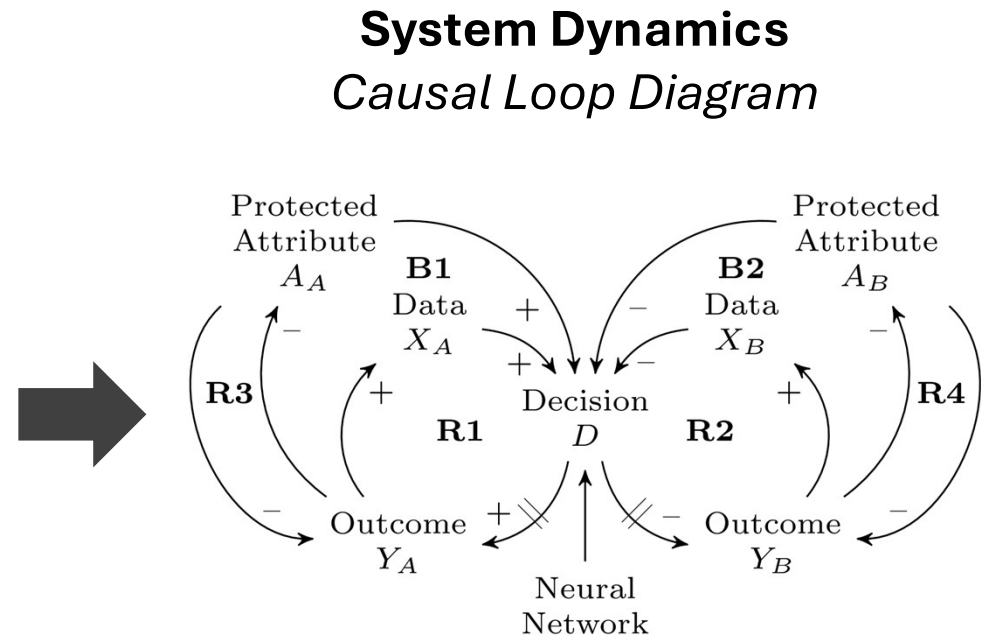


Causal Loop Diagram

Neural Networks Have Inner and Outer Layers



Causal Loop Diagram

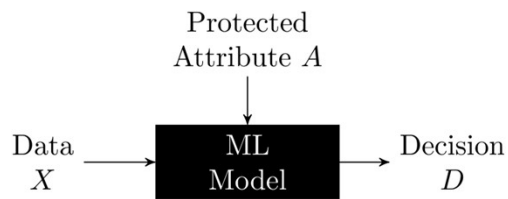


System Dynamics
Causal Loop Diagram

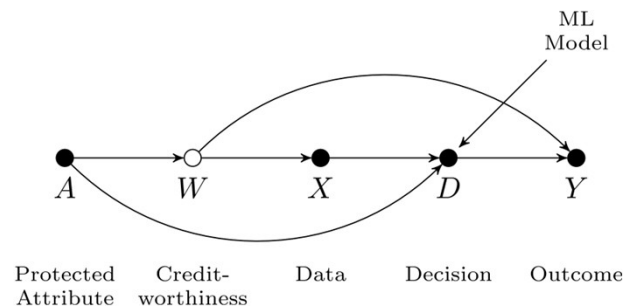
Modeling Bias in the Data Generating Process

This gives us three different representations of the same AI bias problem:

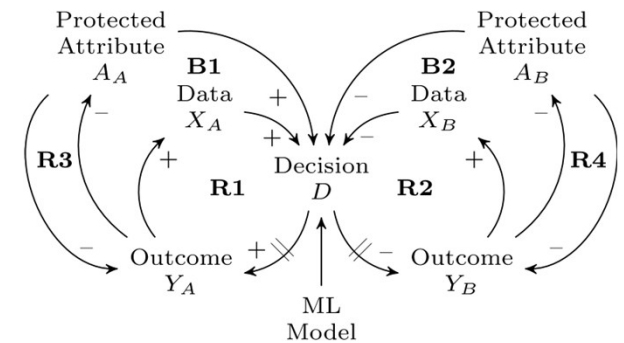
Machine Learning
Block Diagram



Causal Inference
Causal Bayesian Networks



System Dynamics
Causal Loop Diagram



A Fork in the Road for AI Research?

AI Research



Micro

*Thinking **Inside**
the Black Box*

Machine Learning
Statistics

Low-Stakes
Decisions

Macro

*Thinking **Outside**
the Black Box*

Causality
Systems

High-Stakes
Decisions