NATIONAL FAIR HOUSING ALLIANCE

RFI on Use of AI by Financial Institutions: NFHA and Partners' Response

OUTLINE

Opening Remarks - Lisa Rice

Part I: Background - Stephen Hayes

Part II: Key AI Risks

- Data and Outcome Risks – Michael Akinwumi
- Model Risks – Kareem Saleh & John Merrill

Part III: Recommendations & Call to Action – Olga Akselrod & Maureen Yap

Part IV: Q&A

PART I: BACKGROUND

Overview

- AI is rapidly becoming pervasive in financial services (e.g., credit scoring/pricing; advertising; customer engagement; automated valuations; servicing/loss mitigation).
- The use of AI/ML carries serious risks, including perpetuating and amplifying discrimination.
- Agency signals to date have focused too much on innovation and not enough on ensuring safe, non-discriminatory use.

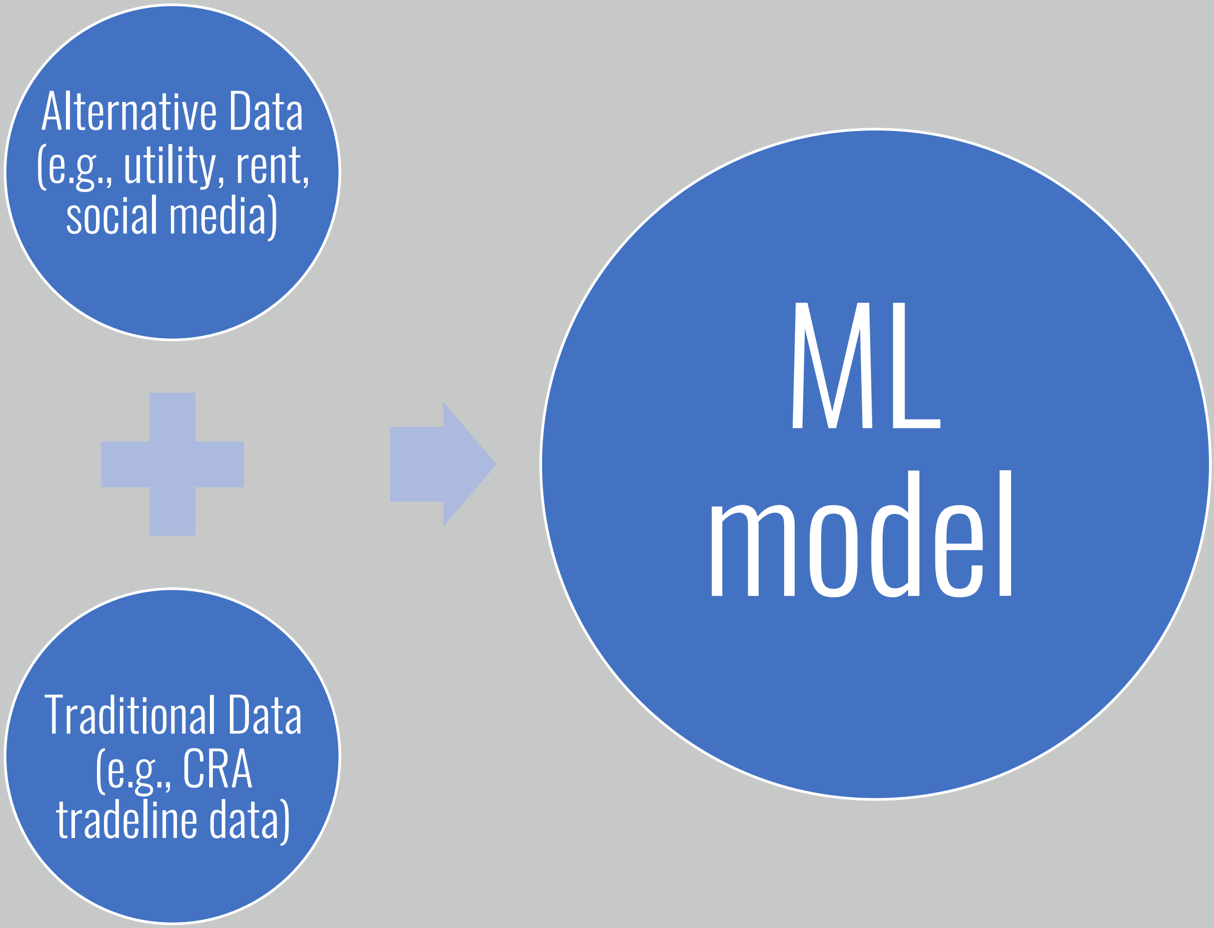
AI/ML CHALLENGES

AI/ML Models

- Hidden boxes?
- Perpetuate inequality?

Alternative Data

- Proxies for protected classes?
- Drives disparate impact?
- Inaccurate/incomplete?



AI/ML CHALLENGES

- Groups are underrepresented in data
  - Too few borrowers or properties in certain areas used in model training
- Variables reflect historic discrimination
  - Credit model penalizes victims of historic redlining
  - Model considers whether applicants were arrested
  - Model relies on past applicants to define what makes a good applicant
- Lack of diversity among developers
  - Blind spots regarding variables (e.g., gaps in employment history, last name length)

FAIRNESS GOVERNANCE & AI

- Equal Credit Opportunity Act
- Fair Housing Act
- Model Risk Management (Fed. Res. Board SR 11-7)

FAIRNESS GOVERNANCE & AI

Disparate treatment

- Including protected class (e.g., race, gender) or proxies as a variable in a model
- Using different models or segments for different protected groups

Disparate impact

- Disparate impact does <u>not</u> require proof that anyone intended to discriminate
- One need not show that protected class was "considered" at all
- In the model context, disparate impact can (and often does) exist absent inclusion of protected characteristics or proxies and it can exist even if the model works the "same" for all groups

DISPARATE IMPACT & AI

DI Step 1

Does the model result in disproportionate negative outcomes for a protected group?

DI Step 2

Does the model advance a legitimate business interest?

DI Step 3

Would changes to the model reduce the disparities identified in Step 1, while still serving the legitimate business interest?

DISPARATE IMPACT & AI

DI Step 1

Does the model result in disproportionate negative outcomes for a protected group?

DI Step 2

Does the model advance a legitimate business interest?

DI Step 3

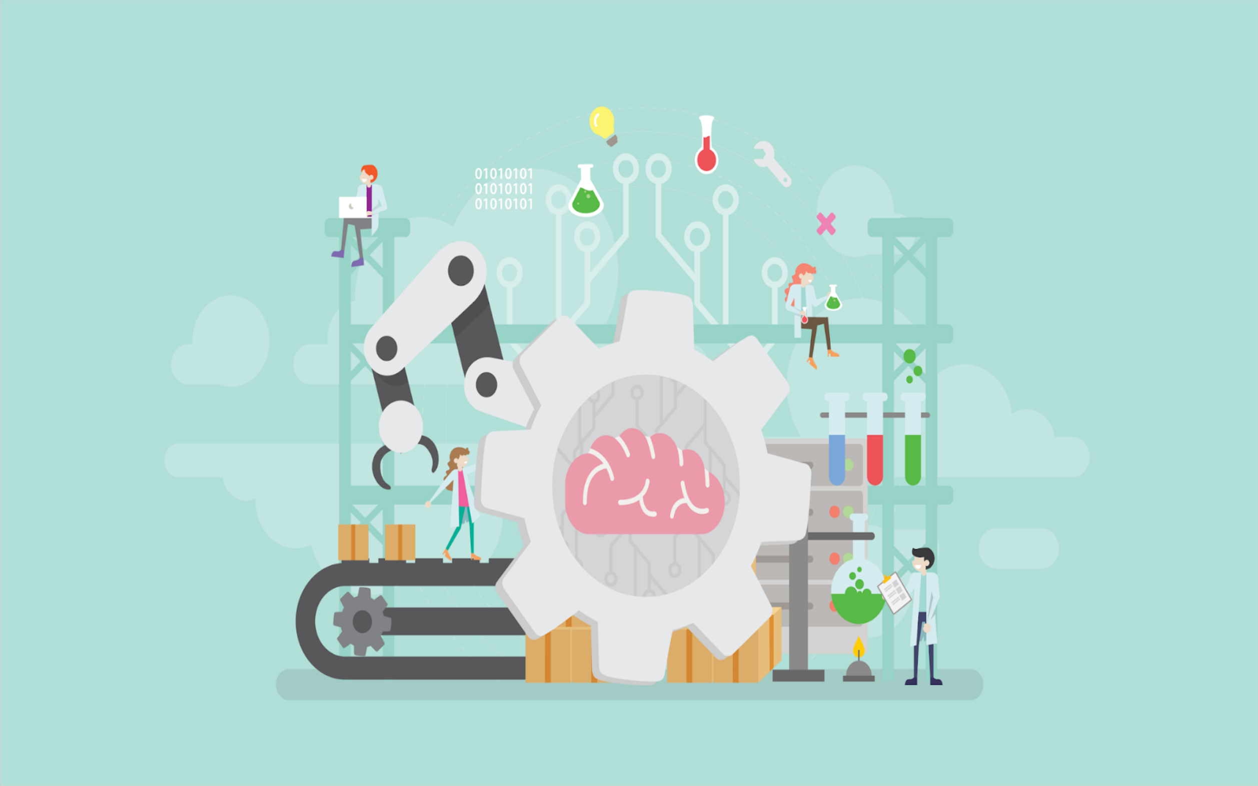
Would changes to the model reduce the disparities identified in Step 1, while still serving the legitimate business interest?

NFHA NATIONAL FAIR HOUSING ALLIANCE

PART II: KEY AI RISKS

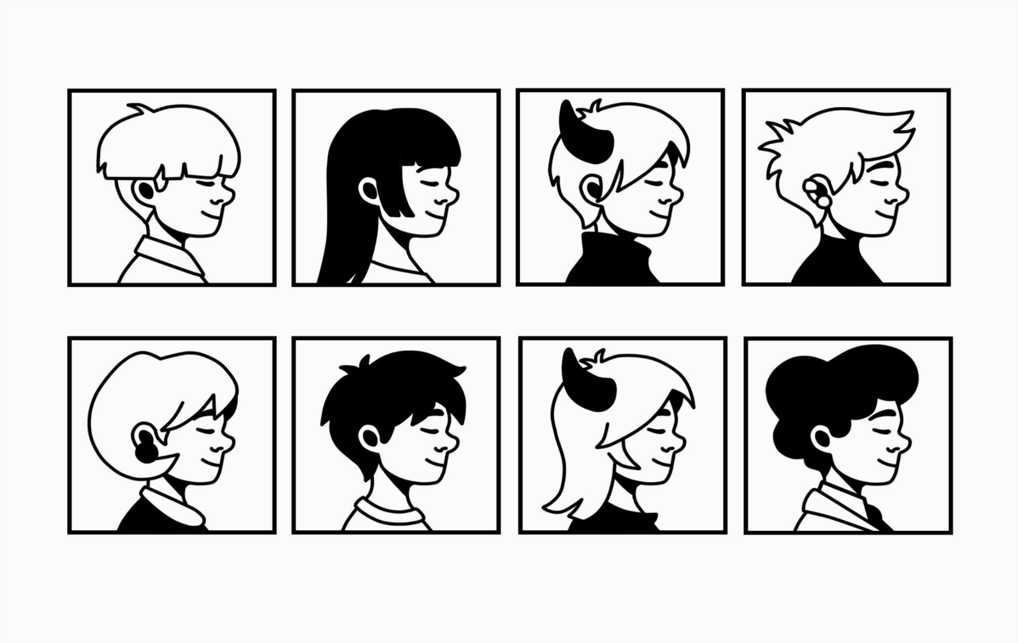
AI Risks = Data Risks + Outcome Risks + Model Risks

AI or machine learning products and services are as biased as the: holes in the data that power them; level of deterioration in their predictions; and inductive bias in their training algorithms.



Data Risks: Exclusion Bias

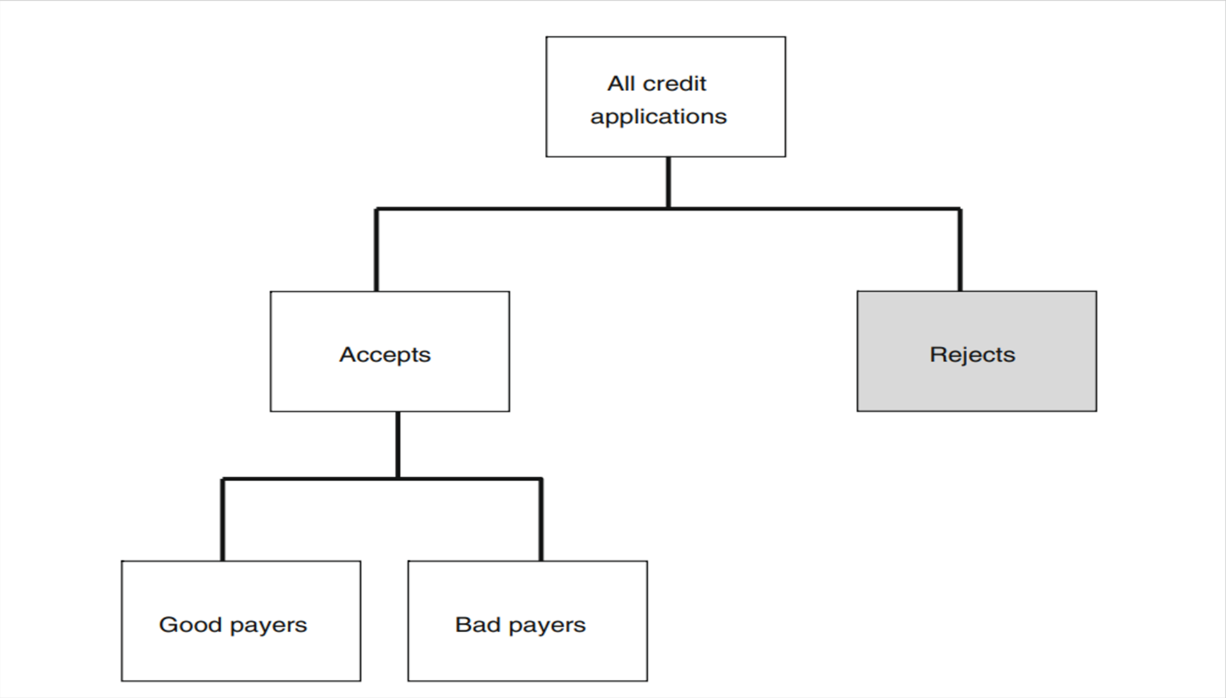
Data may be non-inclusive or may exhibit feature bias (e.g., people of color are disproportionately missing from credit data).



Data Risks: Sample Bias

Data may reflect historic discrimination or inequality
(e.g., homes in Black neighborhoods are appraised for less than properties in mostly white neighborhoods).

Consequently, in the credit scoring context, the true creditworthiness of rejected applicants **may never be known.**



Data Risks: Proxy Attributes

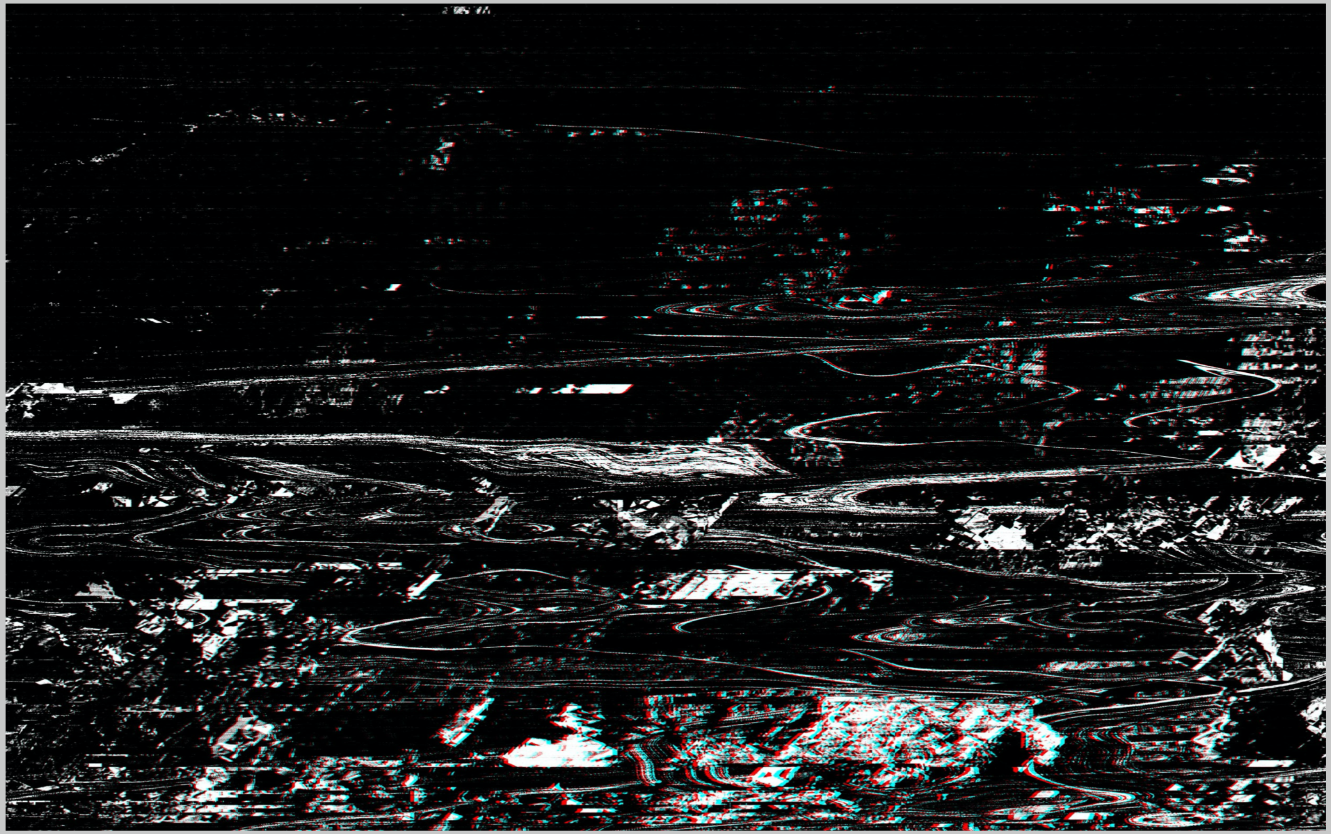
Innocuous variables that correlate with protected class variables, examples are zip code, number of divorce, or phone area code.

Data Risks: Label Definition

The label being used to train a model may be defined in a way that skews the label distribution in favor of Whites or Males.

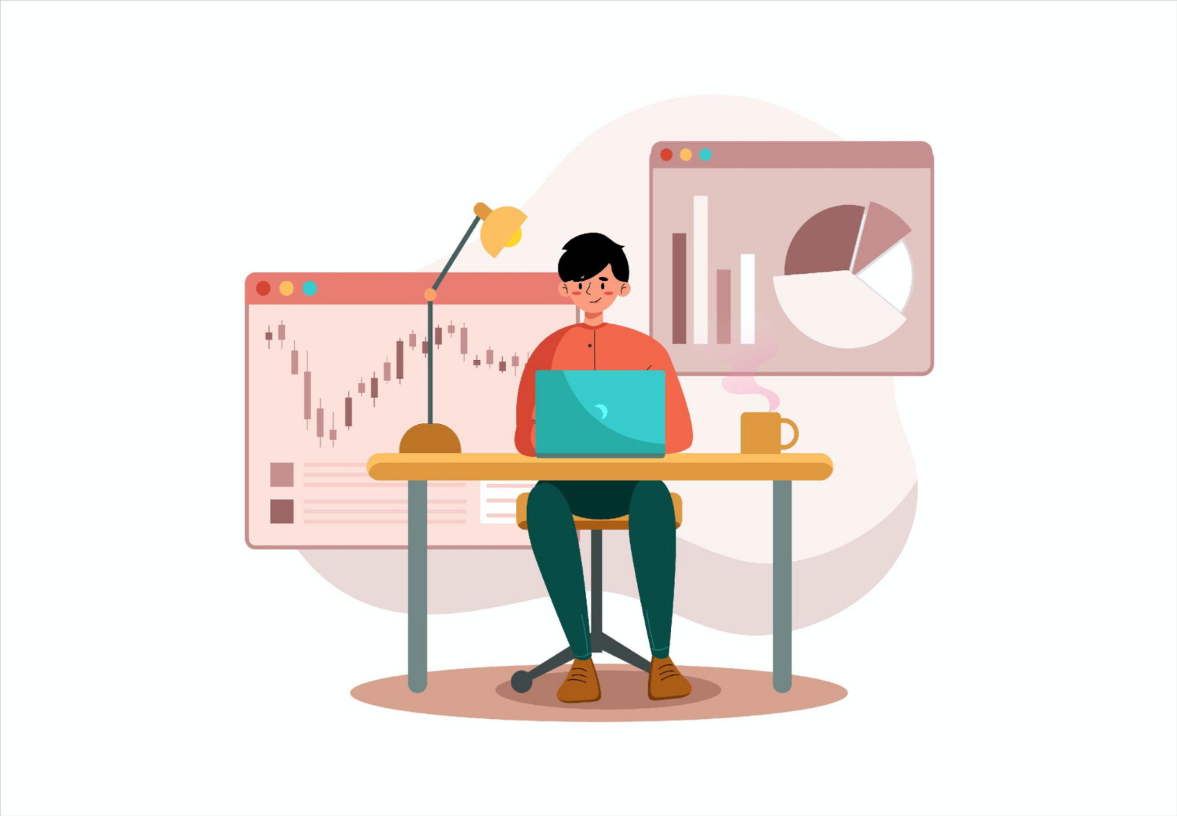
Data Risks: Irregularities in test and train data

A mismatch between patterns in the data used to train an AI model and the data used to evaluate it usually indicate high risk of generalization errors when the model is deployed.



Outcome Risks

AI or machine learning models deteriorate, and the deterioration could hurt everyone, especially People of Color, if unmonitored.



Outcome Risks: limited explainability techniques

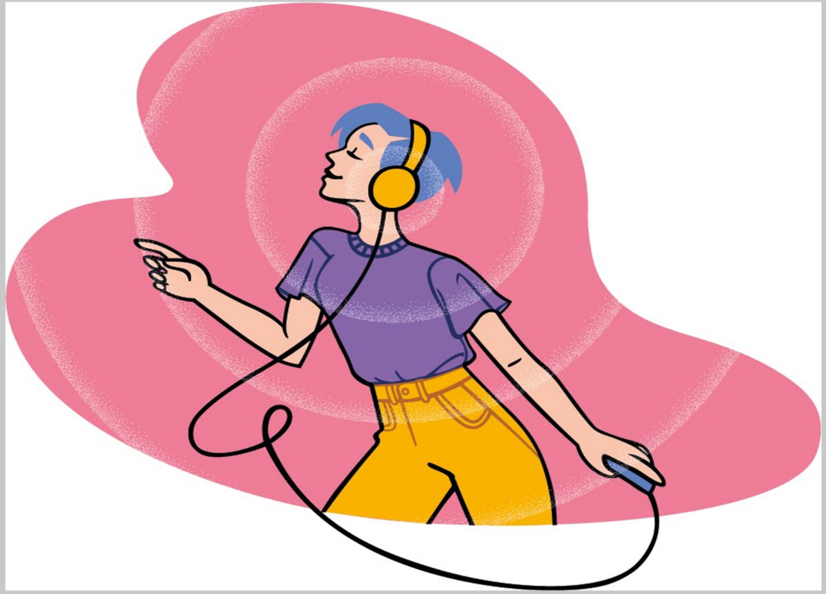
Models must be explainable in order to meet legal requirements, including ECOA's anti-discrimination provisions and requirement to provide specific and accurate adverse action notices.

If a model is not explainable, it may be difficult or impossible to:

- Assess whether variables are functioning as proxies for protected classes, and
- Remove bias from its outcomes

Outcome Risks: dynamic business environment

Business environments are dynamic, and a good AI or machine learning model is sensitive to the environment in which it is deployed.



Outcome Risks: drifts in prediction and features

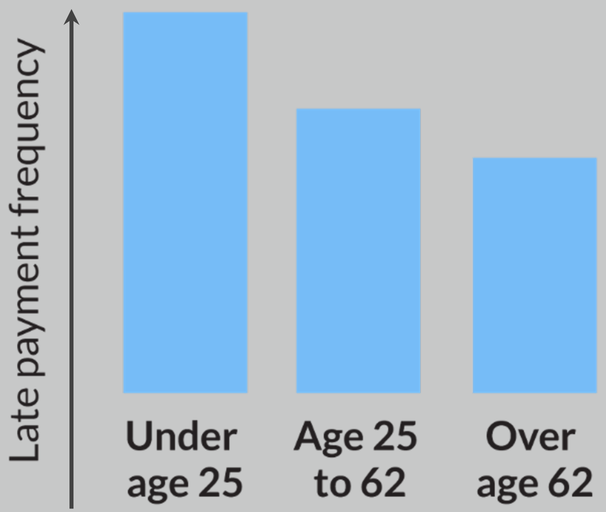
Patterns in development data may be different from those in production data, and prediction distribution in development data may be different from what is observed in production data.

PSI – population stability index;

CSI – characteristic stability index

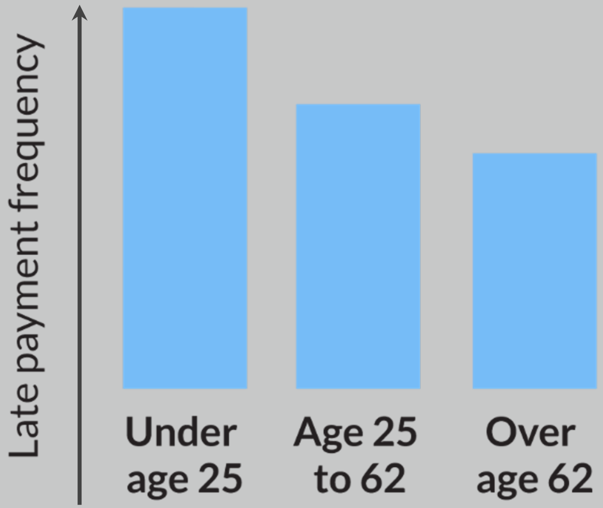
Algorithmic Bias Is Not Just a Data Problem: Model Development Decisions Can Produce Disparate Outcomes

Unbiased Source Data
Late Payment Frequency and Age

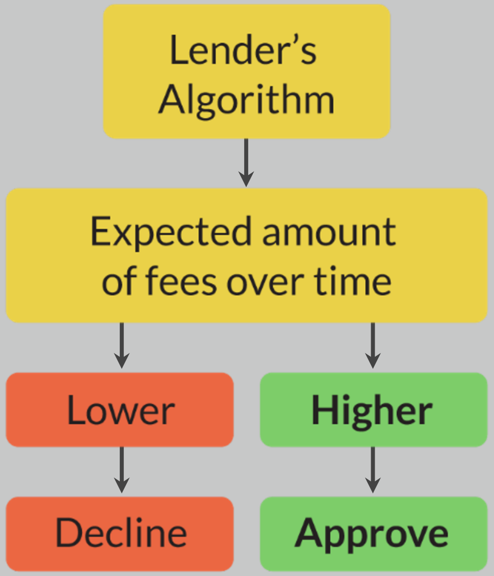


Even with unbiased data, model structure and design choices such as what to maximize can lead to bias

Unbiased Source Data
Late Payment Frequency and Age



Profit-Maximizing Algorithm

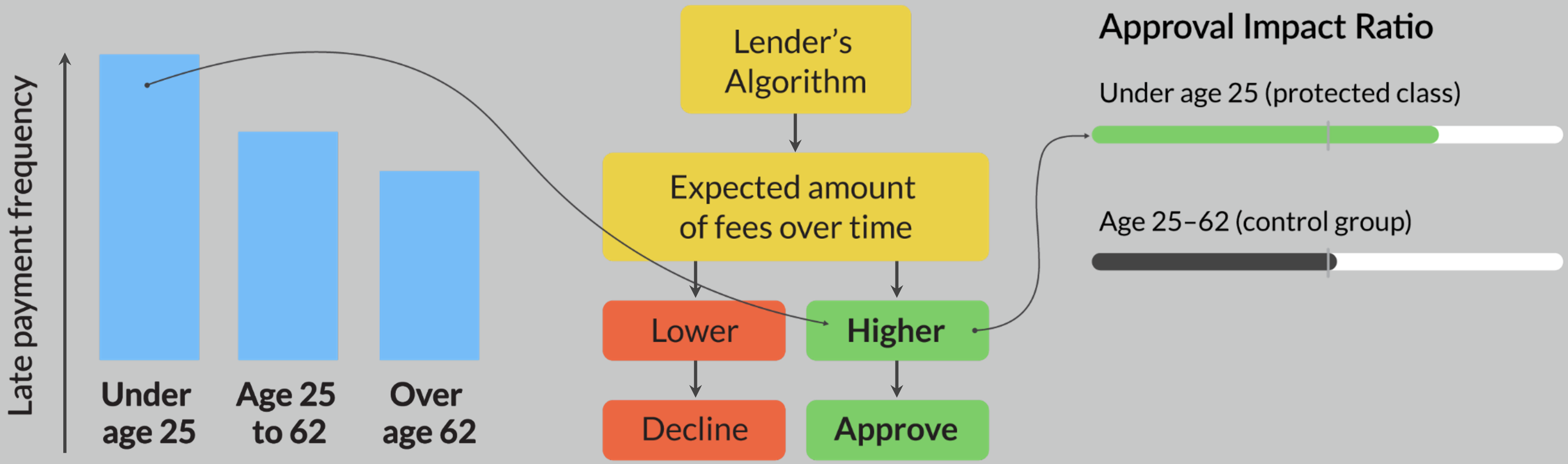


This Model Design Choice Made It More Likely For Younger Applicants to be Approved...

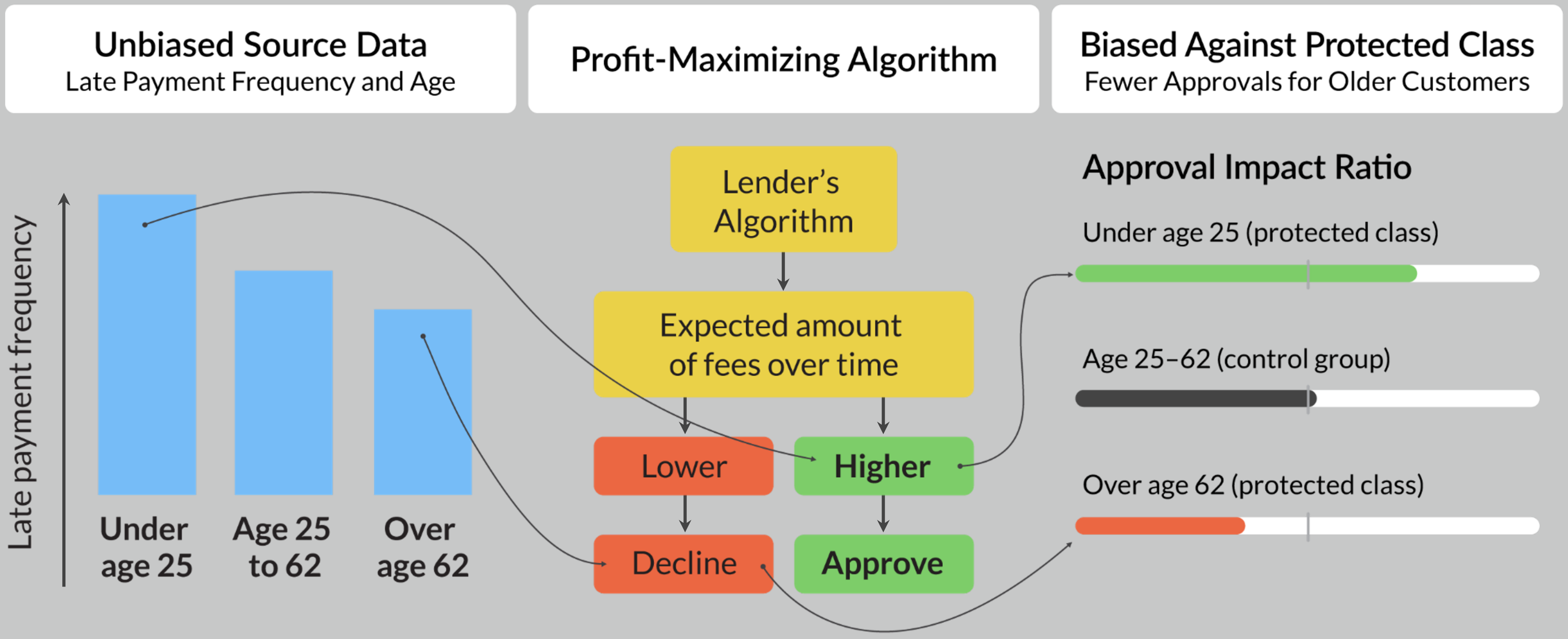
Unbiased Source Data
Late Payment Frequency and Age

Profit-Maximizing Algorithm

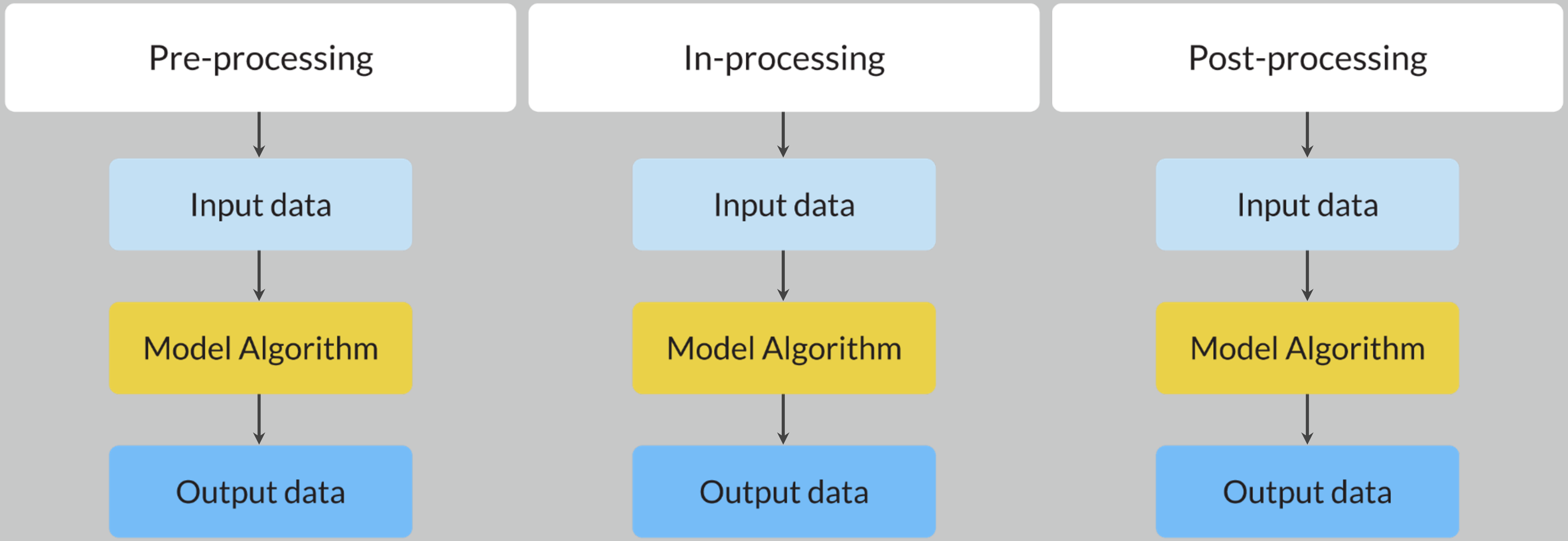
Biased Against Protected Class
Fewer Approvals for Older Customers



Final Outcome: Disparate Impact

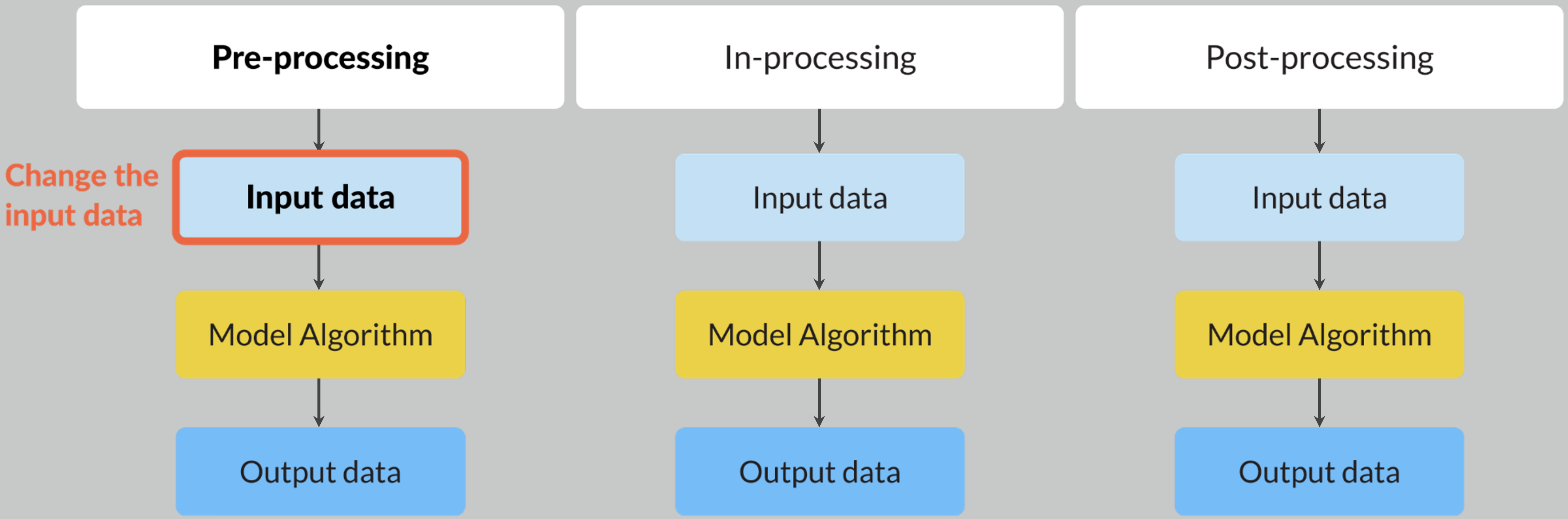


Building Fairer AI Models



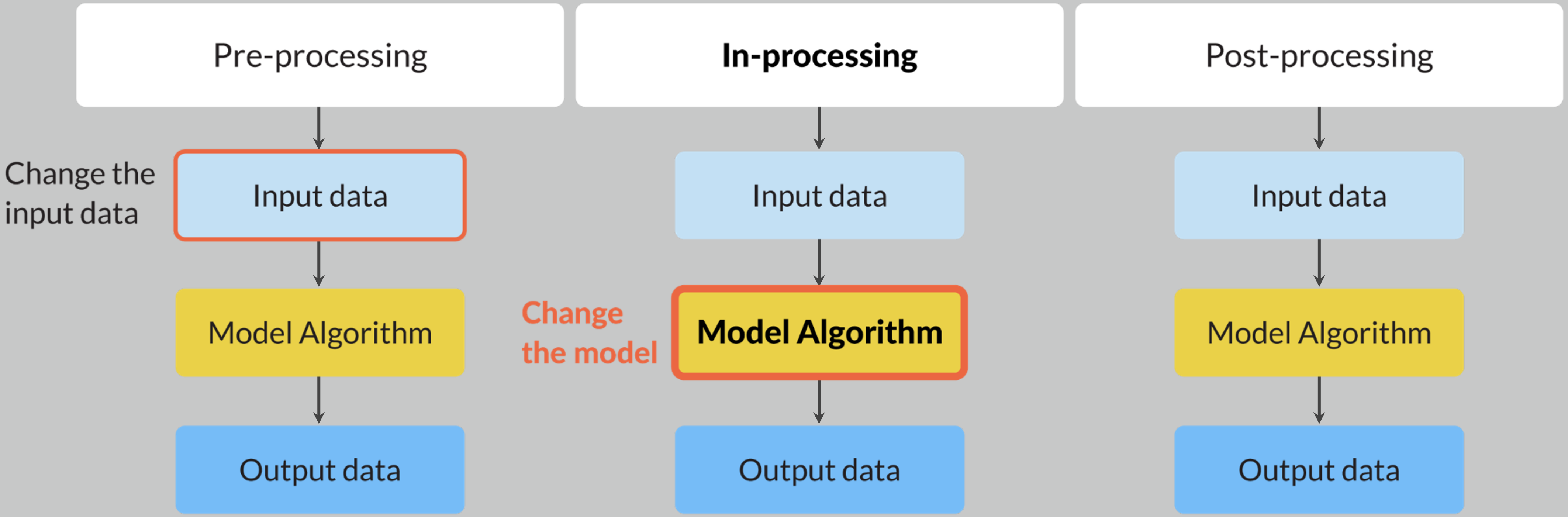
The pros and cons of using these approaches in consumer finance should be studied

Building Fairer AI Models



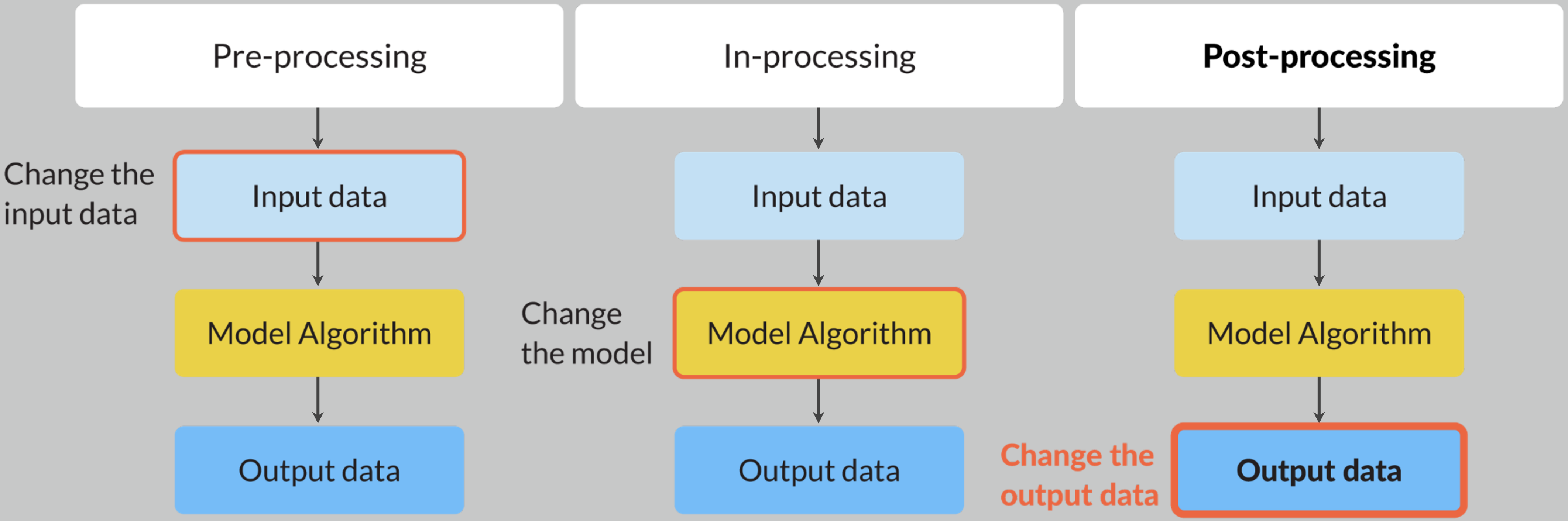
The pros and cons of using these approaches in consumer finance should be studied

Building Fairer AI Models



The pros and cons of using these approaches in consumer finance should be studied

BUILDING FAIRER AI MODELS



The pros and cons of using these approaches in consumer finance should be studied

NFHA NATIONAL FAIR HOUSING ALLIANCE

PART III: RECOMMENDATIONS & CALL TO ACTION

Recommendations

- Non-discrimination and Equity
  - Redefining model risk
- Action Plan
- Robust Supervision and Enforcement/Accountability
  - Agency review and enforcement action

Recommendations

- **Actionable Policies** - Agencies should issue policies that:
  - Redefine model risk
  - Inform on possible risks
  - Set standards for financial institution testing, documentation, and archiving and for model explainability
  - Describe Agency testing
  - Ensure transparency

Recommendations

- Public Research
  - Support public research that analyzes the efficacy and impact of AI for consumers of color and other protected classes
- Specialized Fair Lending and AI Staff
  - Hire staff that can review assessments and provide guidance
- Fair Lending Training for All AI Stakeholders
  - Ensure staff is trained: better able to recognize issues that may raise red flags

Recommendations

- Diversity, Equity, and Inclusion
  - Ensure agency and financial institution staff working on AI issues reflect diversity
- Transparency
  - Prioritize transparency for the Agencies and the financial institutions
- Engagement
  - Stay engaged with a diverse group of key stakeholders

Call to Action

- If interested, please fill out this Google Form by **Wed June 30th** to indicate that your organization would like to sign on:
  - Link to Advocate Response to RFI
  - Link to Google Form for Sign On
- For questions on the advocate response, please contact:
  - Maureen Yap: Myap@nationalfairhousing.org

NATIONAL FAIR HOUSING ALLIANCE

PART IV: QUESTIONS & ANSWERS

Contact Information

Stephen Hayes: shayes@relmanlaw.com

Michael Akinwumi: makinwumi@nationalfairhousing.org

Kareem Saleh: kareem@fair-play.ai

John Merrill: john@fair-play.ai

Maureen Yap: myap@nationalfairhousing.org

Olga Akselrod: oakselrod@aclu.org