Responsible AI and Bias

Robert N. Bernard, PwC

November 2024

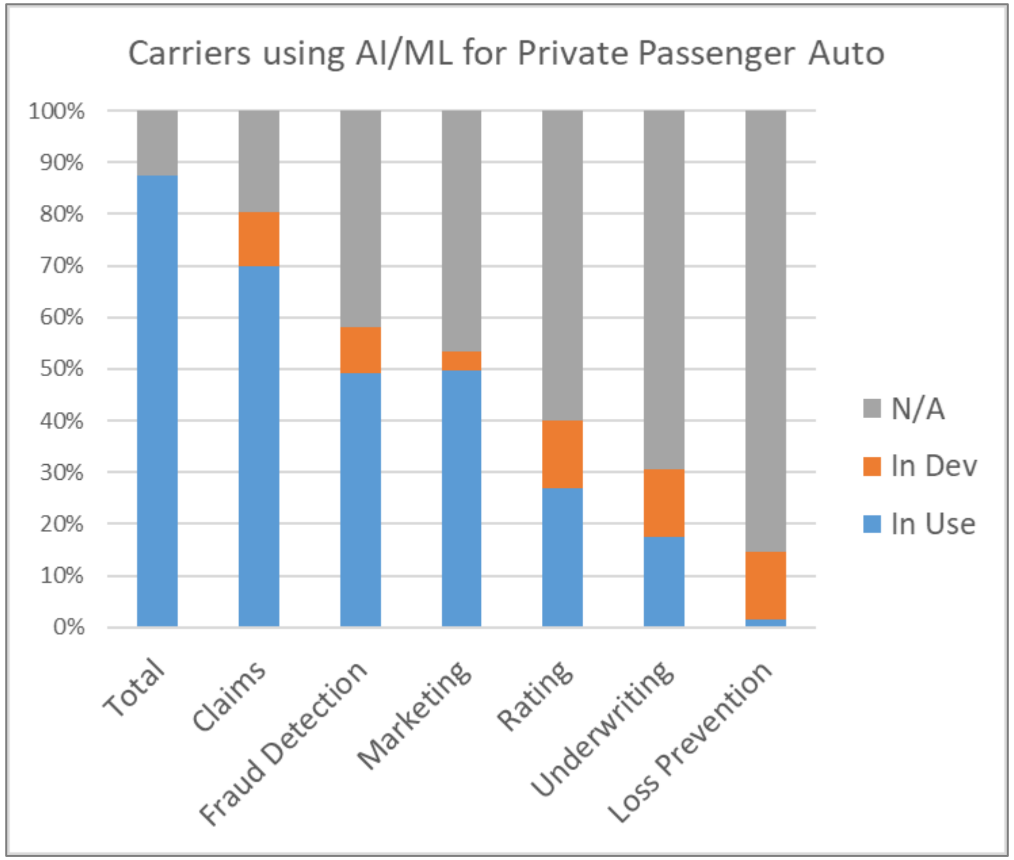
“

In Fall 2022, personal automobile insurance companies were surveyed; what percentage of them used AI/ML in their businesses?

Artificial intelligence and machine learning are now the norm

Almost 90% of personal auto insurers use AI/ML for in their business

- The NAIC recently surveyed 193 US private passenger auto insurers regarding use of Artificial Intelligence/Machine Learning (AI/ML)
- The most common use is in claims for analyzing images and fraud detection. **None use AI/ML to automatically deny a claim**
- In marketing, carriers use marketing models for targeted online advertising, targeted mail and phone advertising, provision of offers to existing customers, and direct online sales
- In rating, the most common uses are determination of rating class and relativities, with limited use for retention modeling and none for price optimization
- In underwriting, the most common uses are motor vehicle report (MVR) ordering, telematics app discount eligibility, and anomaly detection. None used AI/ML for automated eligibility denial



Source: "NAIC Big Data and Artificial Intelligence Working Group PPA AI/ML Survey December 2022"

What does the term “Responsible AI” mean to you?

Why Responsible AI?

In a world with more AI and more demand that these AI systems be **trustworthy**, organizations need RAI to help mitigate risks.

Increasing AI usage

Increased AI adoption

across organizations and industries to deliver products and services.

The result?

- More data
- More models
- More complexity

Increasing scrutiny

Greater call for AI that is accountable and transparent:

- Emerging regulatory requirements
- More-informed consumer demands
- Evolving employee expectations

Increasing risk

Organizational & Public-Level Risks

- Economic risk
- Societal & environmental risk
- Enterprise reputational & regulatory risks

System-Level Risks

- Performance risk
- Security risk
- Control risk

Four dimensions of Responsible AI

Responsible AI at its core is simply good data science, governed by key guiding principles **from strategy to execution.**

Strategy

Data & AI Ethics
Consider the moral implication of uses of data and AI and codify them into your organization's values.

Policy & Regulation
Anticipate and understand key public policy and regulatory trends to align compliance processes.

Control

Governance
Enable oversight of systems across the three lines of defense.

Compliance
Comply with regulation, organizational policies, and industry standards.

Risk Management
Expand transitional risk detection and mitigation practices to address risks and harms unique to AI.

Responsible Practices

Interpretability & Explainability
Enable transparent model decision-making.

Sustainability
Minimize negative environmental impact.

Robustness
Enable high performing and reliable systems.

Bias & Fairness
Define and measure fairness and test systems against standards.

Security
Enhance the cybersecurity of systems.

Privacy
Develop systems that preserve data privacy.

Safety
Design and test systems to prevent physical harm.

Core Practices

Problem Formulation
Identify the concrete problem you are solving for and whether it warrants an AI / ML solution.

Standards
Follow industry standards and best practices.

Validation
Evaluate model performance and continue to iterate on design and development to improve metrics.

Monitoring
Implement continuous monitoring to identify drift and risks.

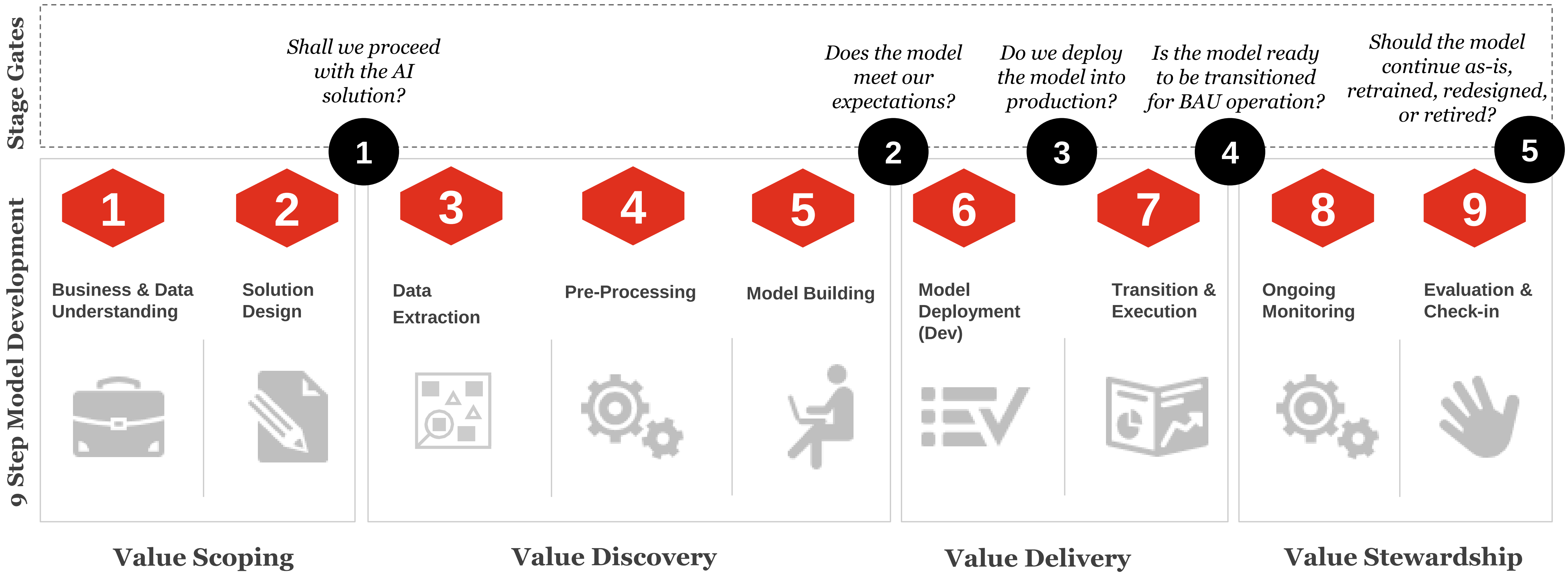
Responsible AI is not just one person's job

Diverse stakeholders across all parts of the organization must **collaborate** to apply responsible AI practices consistently and effectively.

The 3 lines of defense

Enterprise-level	External	Regulators External auditors	External audit Regulatory requirements and legal action External standards development
Enterprise-level	3rd Line of Defense	Internal Audit Ethicists	Internal audit review Ethics review
Application-Level	2nd Line of Defense	Compliance Legal & Privacy Leadership Independent Review Group SME Model Approver	Standards and Specs Independent Review and Challenge Final Approval Legal, Data & Privacy Governance requirements
Application-Level	1st Line of Defense	Model devs Model owners Ops support	Risk Assessment Model Sourcing and/or Development Data Collection and Use Process Documentation Infrastructure Deployment Ongoing Monitoring

PwC’s 9-step process



How would you define “unfair bias”?

Purpose & Definitions

Purpose	To identify methods to avoid or mitigate unfair bias unintentionally caused or exacerbated by the use of AI models
Defining AI	Computer systems that perceive the digital or physical world, process this, & take action that may normally require human intelligence or reasoning
Defining Unfair Bias	Unexplained adverse outcomes for marginalized communities

Unfair bias in AI

Evaluating model fairness and identifying bias is predicated on the capacity for organizations to understand what drives decision making in their AI models

- In the context of AI, technologists may consider performance independent of a protected attribute (i.e., an anti-discriminatory characteristic that serves to protect individuals) one mechanism to check for fairness.
- Analyzing a dataset for bias warrants considerations such as how and when the data was sourced, how it was labeled, what attributes comprise the dataset and what populations are represented in the dataset, what language(s) are incorporated, among others.

The United Nations recognizes this as an issue…

United Nations | UN News
Global perspective Human stories

UN Women/Geno Ochieng | Girls attend a robotics bootcamp in Rwanda.

https://news.un.org/en/story/2024/10/1155446

8 October 2024 | Women

… and AI bias research is plentiful and ongoing

BMC Part of Springer Nature

BMC Medical Ethics

Home About Articles Submission Guidelines Join The Board Collections Submit manuscript

Research | Open access | Published: 17 October 2024

Exploring bias risks in artificial intelligence and targeted medicines manufacturing

Ngozi Nwebonyi & Francis McKay

External pressures on AI usage

In response to AI's impact on the insurance value chain, regulators, technologists, customers, and society as a whole are calling on the organizations developing or deploying AI systems to implement responsible practices

Regulation

- New York City Council passed a bill on automated employment decision tools requiring bias audits and communications to be sent to residents when the tool was used in a hiring or promotion decision
- Colorado insurers will be prohibited from using external data in algorithms that may unfairly discriminate, will have to provide information on external data sources, and maintain and report on a risk management framework

Customer Concerns

Customers are also expressing concerns related to AI's use and impact, according to PwC's 2021 AI predictions report:

- 27% of insurance executives consider "customer distrust of AI leading to lost business" a threat over the next five years
- 22% view "societal backlash against AI" as a threat

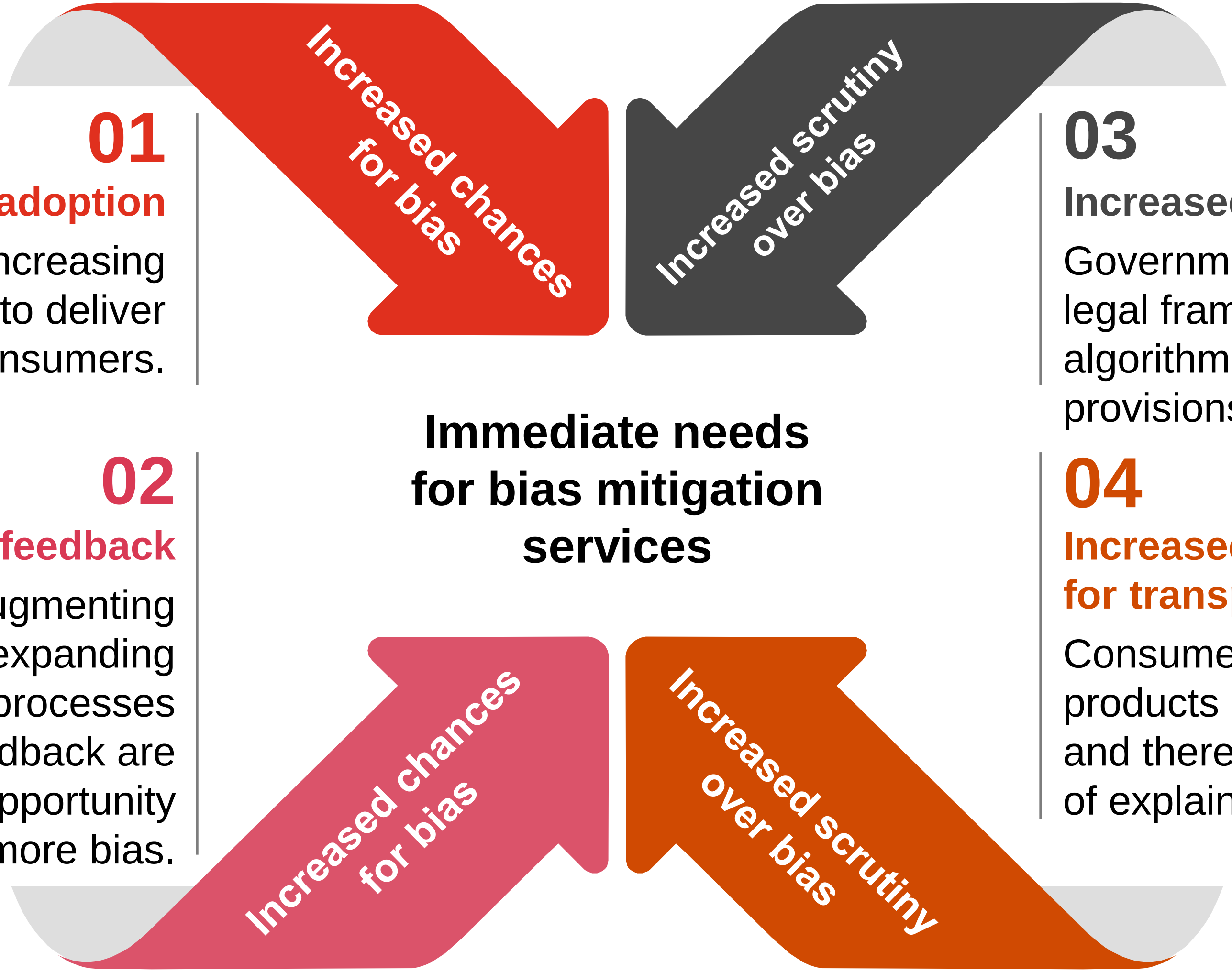
Foundational recommendations to consider: Internal policies and practices

- Monitor the regulatory environment
- Engage stakeholders and establish roles and responsibilities
- Equip employees with the necessary tools and skills
- Conduct a model risk assessment
- Integrate with Model Risk Management

Increased demands for transparency and fairness should make bias mitigation a corporate necessity

01
Increased AI adoption
Organizations are rapidly increasing their use of AI algorithms to deliver products and services to consumers.

02
More data, models and feedback
The use cases for AI augmenting human activities are ever expanding and therefore datasets, processes touched and model feedback are rapidly increasing, creating opportunity for more bias.



03
Increased regulatory oversight
Governments are rapidly developing legal frameworks for holding algorithms accountable, including provisions for discrimination.

04
Increased Consumer demand for transparency
Consumer awareness of AI-enabled products and services is increasing and therefore so is their expectation of explainability.

Fair Lending programs need to be considered in the modeling process

Fine tuning Fair Lending programs

Most financial institutions have established compliance programs, which help mitigate model bias on the back end according to Fair Lending requirements. Non-Fair Lending applications still need to be assessed, and banks need to determine **how to define fairness, and for whom**

Consider Fair Lending impacts during model calibration

Assessing model bias and fair lending considerations should be an integral part of the model development/calibration/tuning process, and directly tied to organizational criteria in terms of defining, measuring, and calibrating for fairness criteria.

“

What is “fair”?

Detecting biases and bias interventions need to be part of model design and development

Fairness Definition

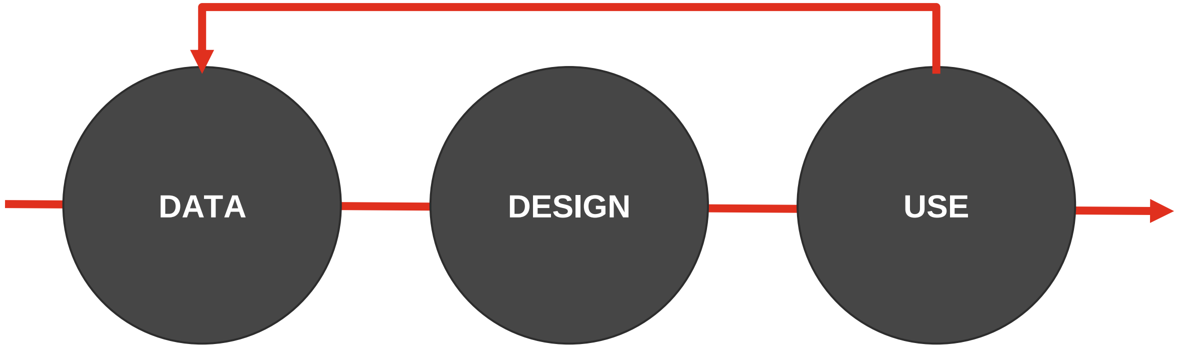
What are the different fairness definitions? Which one should we use for what purpose?

Bias Detection

How do we detect bias with respect to different decisions, protected attributes, definitions, and datasets?

Bias Intervention

How do we adjust algorithms for biases and measure tradeoff between accuracy and fairness?



Defining fairness

Fairness is not a "fuzzy" concept – it is a social construct that can be defined mathematically in multiple ways.

Statistical Measures

Predicted Outcomes

Predicted & Actual Outcomes

Predicted Probabilities

Similarity-Based Measures

Causal Definitions

Continuous (Regression)

Natural Language Processing

Bias detection

KEY CAUSES OF BIAS...

Sample Size Disparity

Selection Bias

Sampling Bias

Participation Bias

Reporting Bias

...LEAD TO ISSUES WITH

Class Imbalance

Proxies

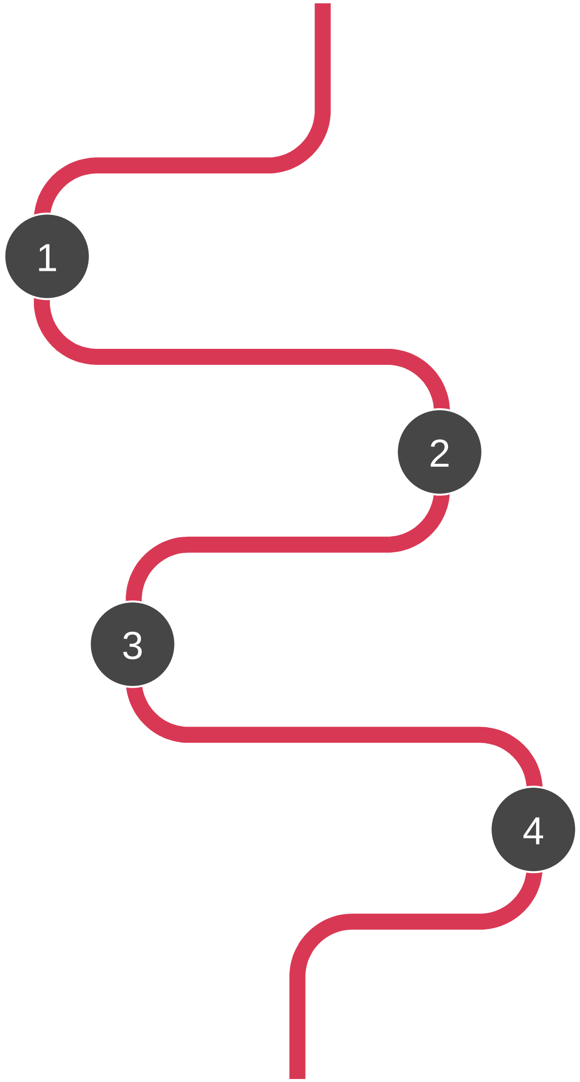
Feature Selection

Model Use Bias

Bias detection is based on the decision being made, the protected attribute, fairness definition chosen and the dataset provided.

Bias intervention

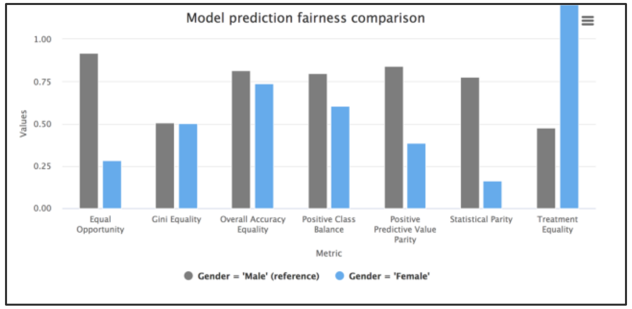
We cannot remove all bias from models. But we can intervene using sophisticated algorithms to become “bias-aware”.



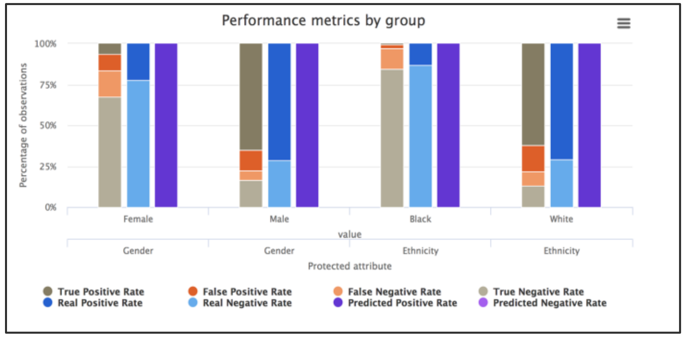
KEY SOLUTIONS FOR BIAS

1. Calibrating Cut-off Thresholds
2. Controlling The Effect Of Proxies
3. Adopting “Fairer” Algorithms
4. Evaluate Trade Offs

How can we do this?

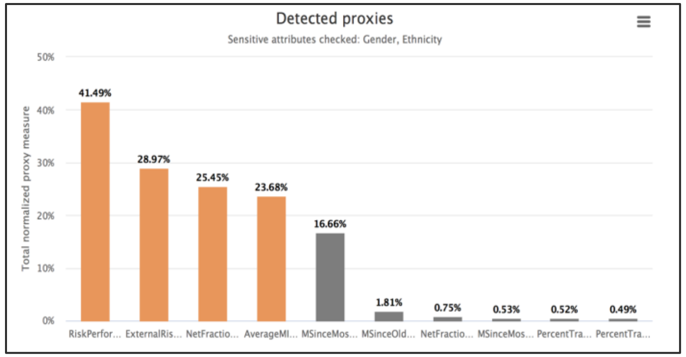


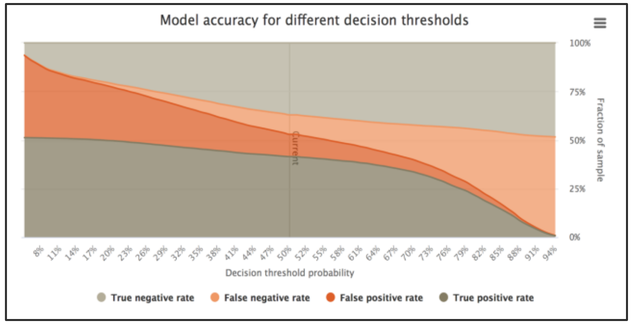
1. **Compute performance metrics** for each possible group in the test dataset



2. **Define fairness** to quantify discrimination against disadvantaged groups

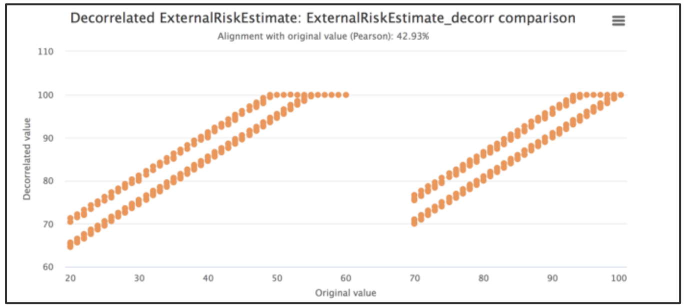
3. **Calculate mutual information** of relevant features with different sensitive attributes, to understand how they might act as proxies





4. **Calculate fairer decision boundaries** by re-adjusting thresholds until the disparities between reference and protected group are reduced

5. **Control the effect of proxies** via two available options: de-correlating proxy variables and ensembling models with-and-without proxies



Thank you

Rob Bernard
PwC, Risk Modeling Services
robert.bernard@pwc.com

pwc.com