University of Glasgow | School of Computing Science

THE AWARDS 2020 | UNIVERSITY OF THE YEAR

‘Fairness’ in Machine learning for Healthcare Applications (Part2)

Dr. Fani Deligianni,

fani.deligianni@glasgow.ac.uk

Lecturer (Assistant Professor)

Lead of the Computing Technologies for Healthcare Theme

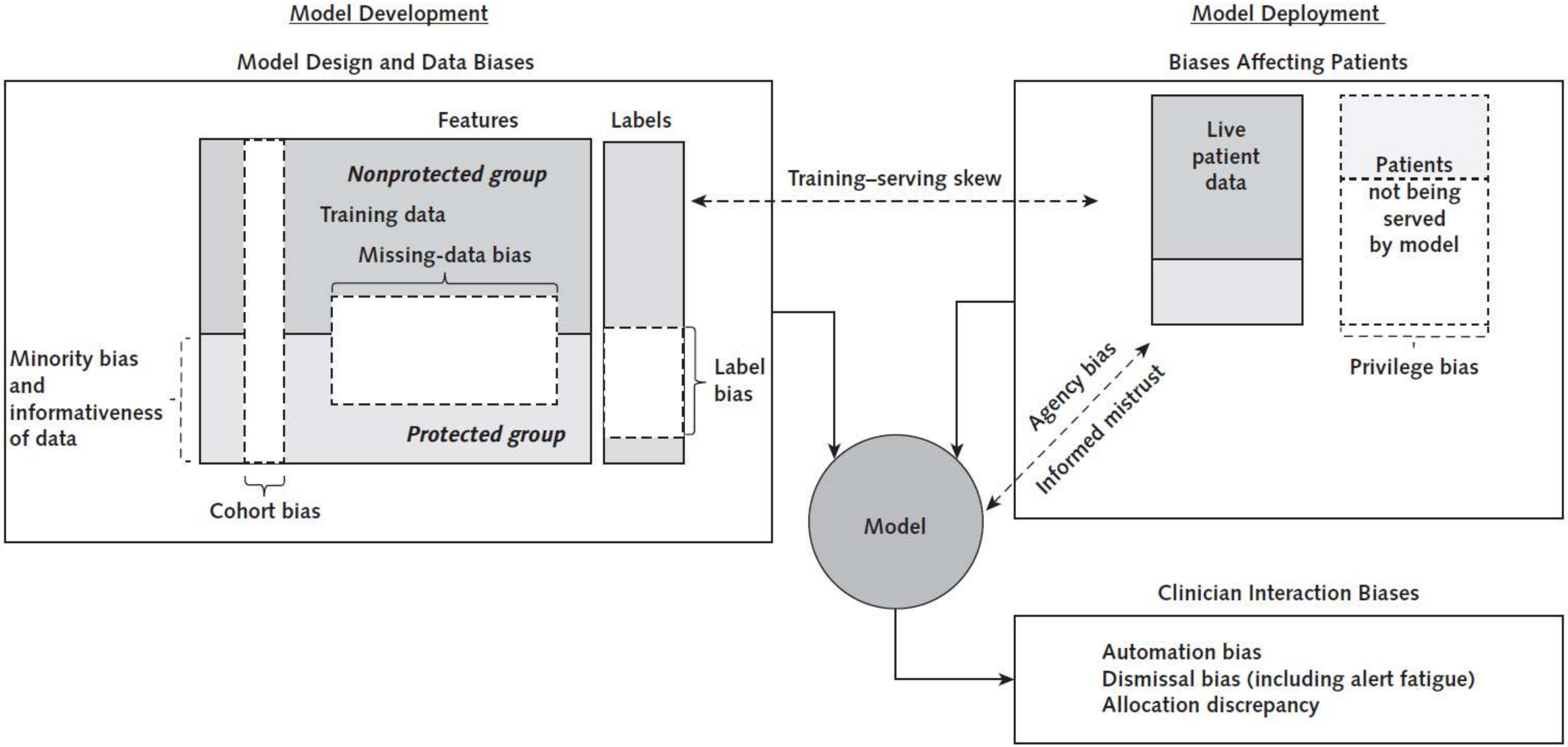
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

WORLD CHANGING GLASGOW

Eliminating Discrimination Bias

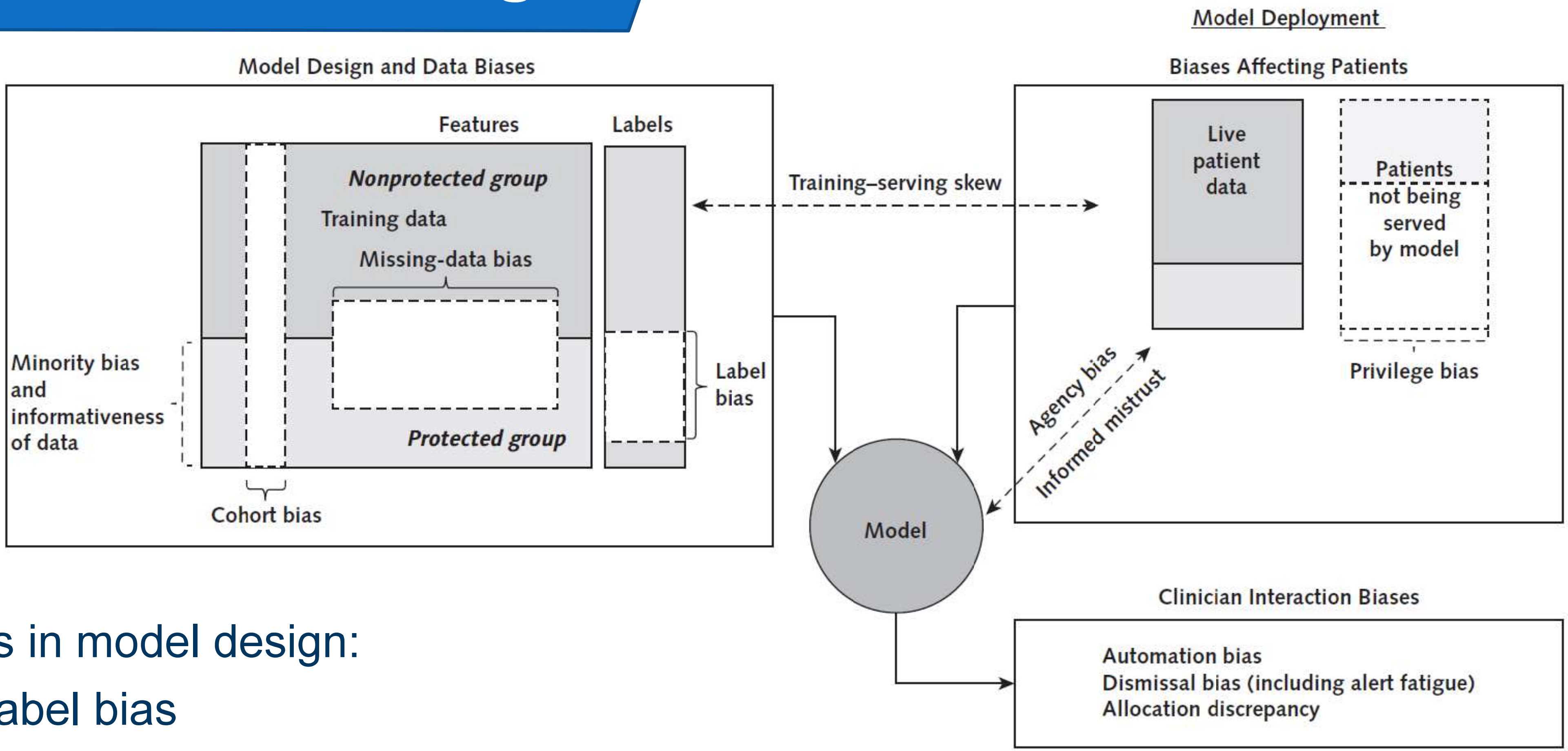
- Algorithmic Accountability Act empowered the Federal Trade Commission
- Large-scale proprietary software is challenging to be accessed and checked
- Corrective measures that alter the results of the AI algorithm is difficult to be explained and justified

Causes of discriminatory bias



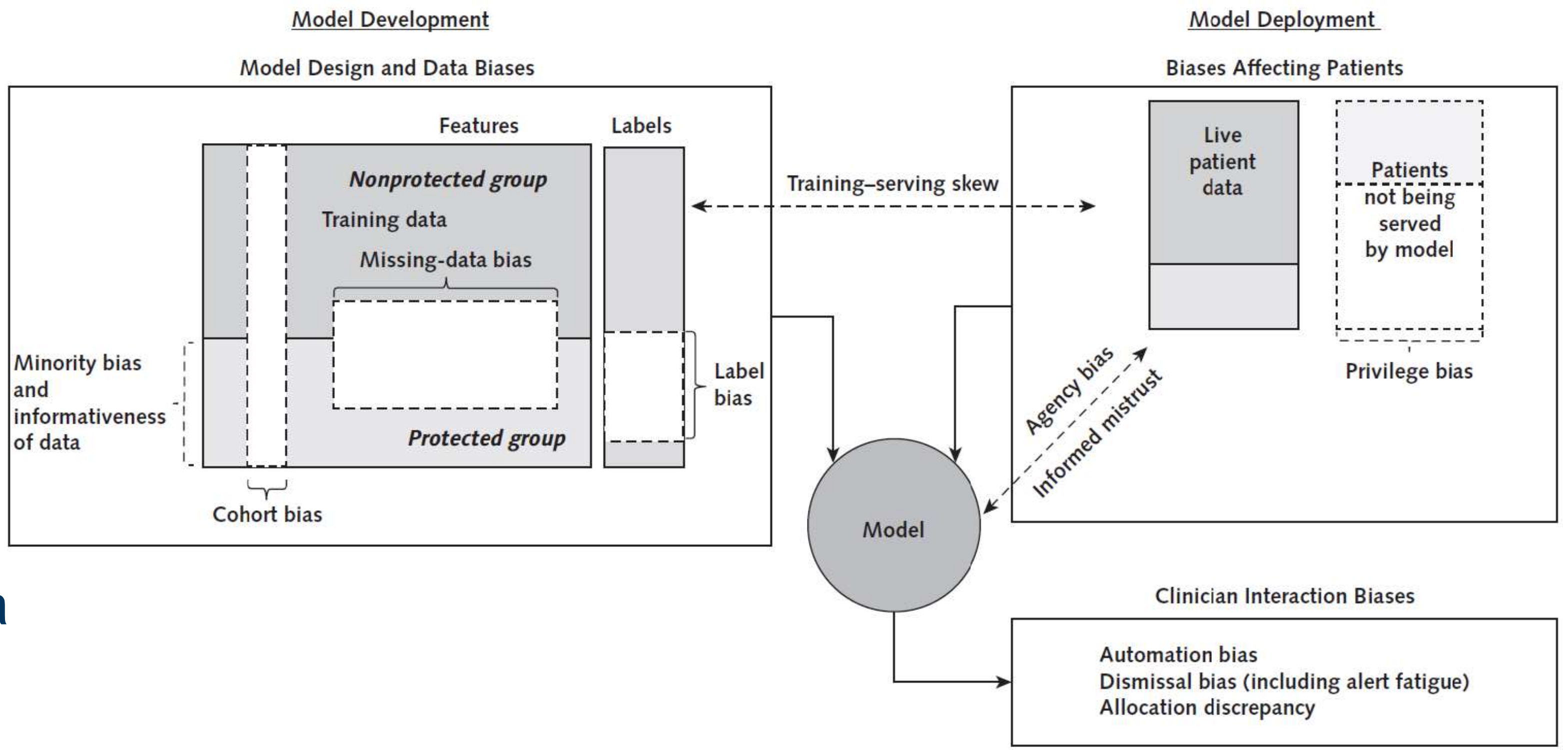
Rajkomar et al. ‘Ensuring Fairness in Machine Learning to Advance Health Equity’, Annals of Internal Medicine, 2018.

Causes of Bias - Design



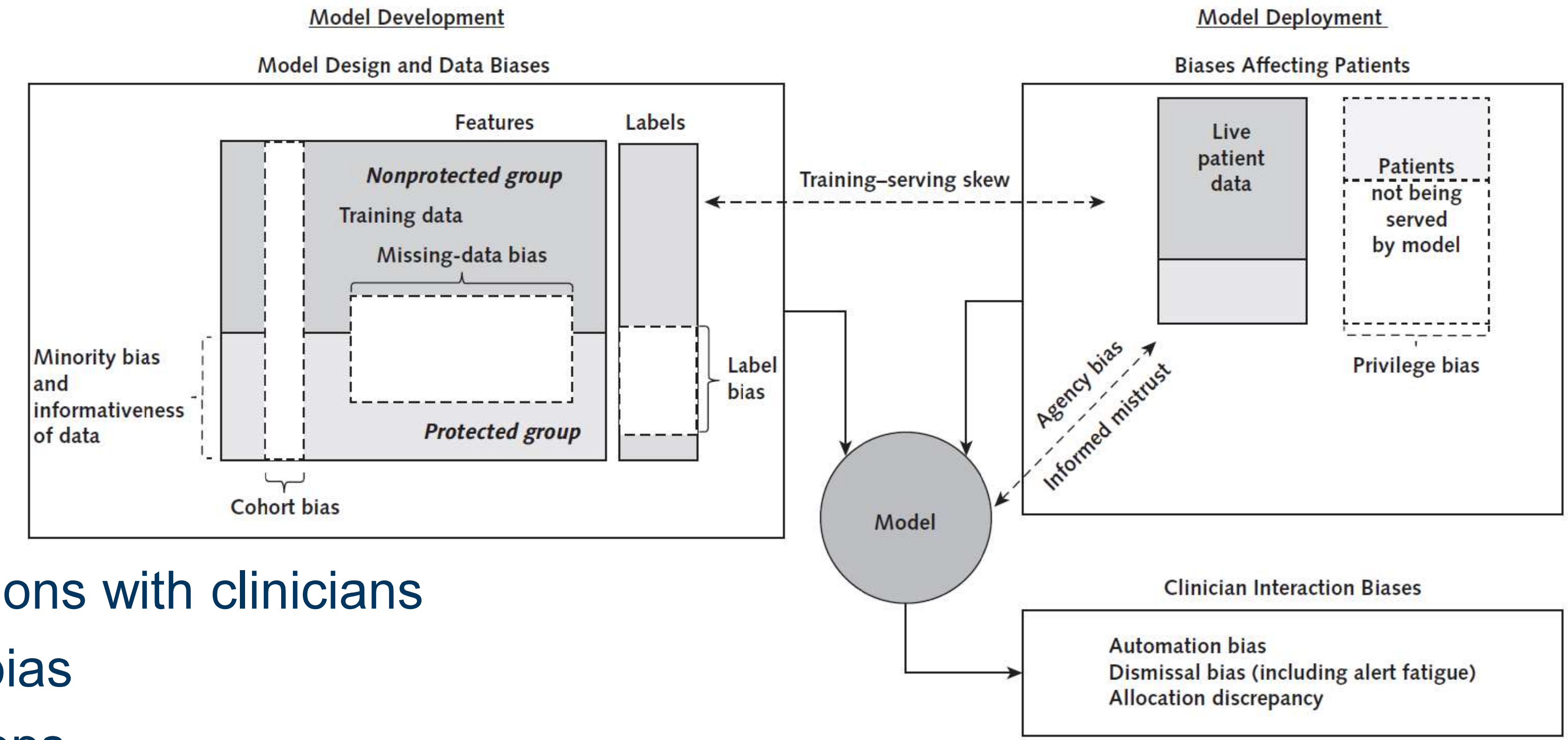
- Biases in model design:
  - Label bias
  - Cohort bias

Causes of bias - Data



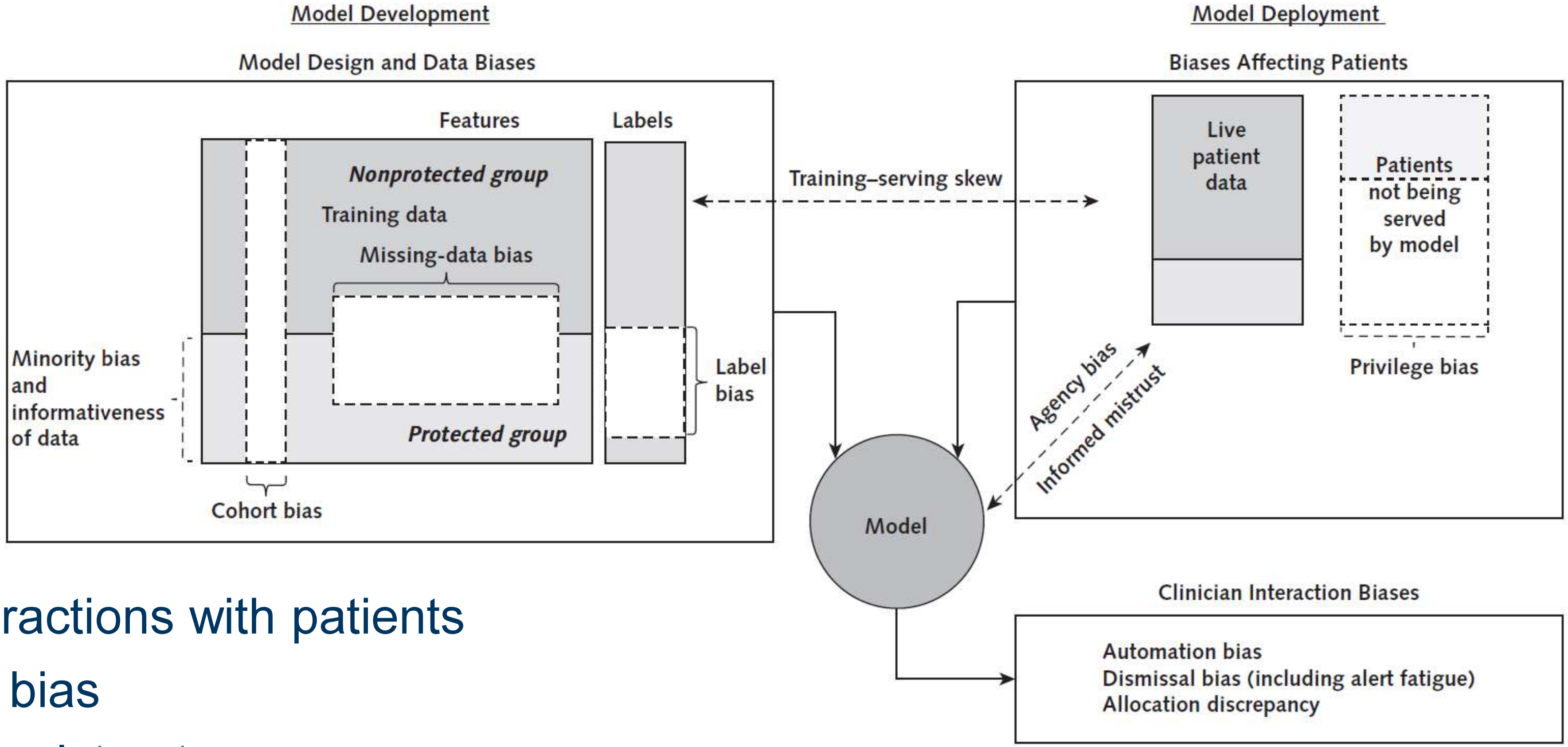
- Biases in training data
  - Minority bias
  - Missing data bias
  - Informative bias
  - Training-serving skew

Causes of bias – Model Interaction



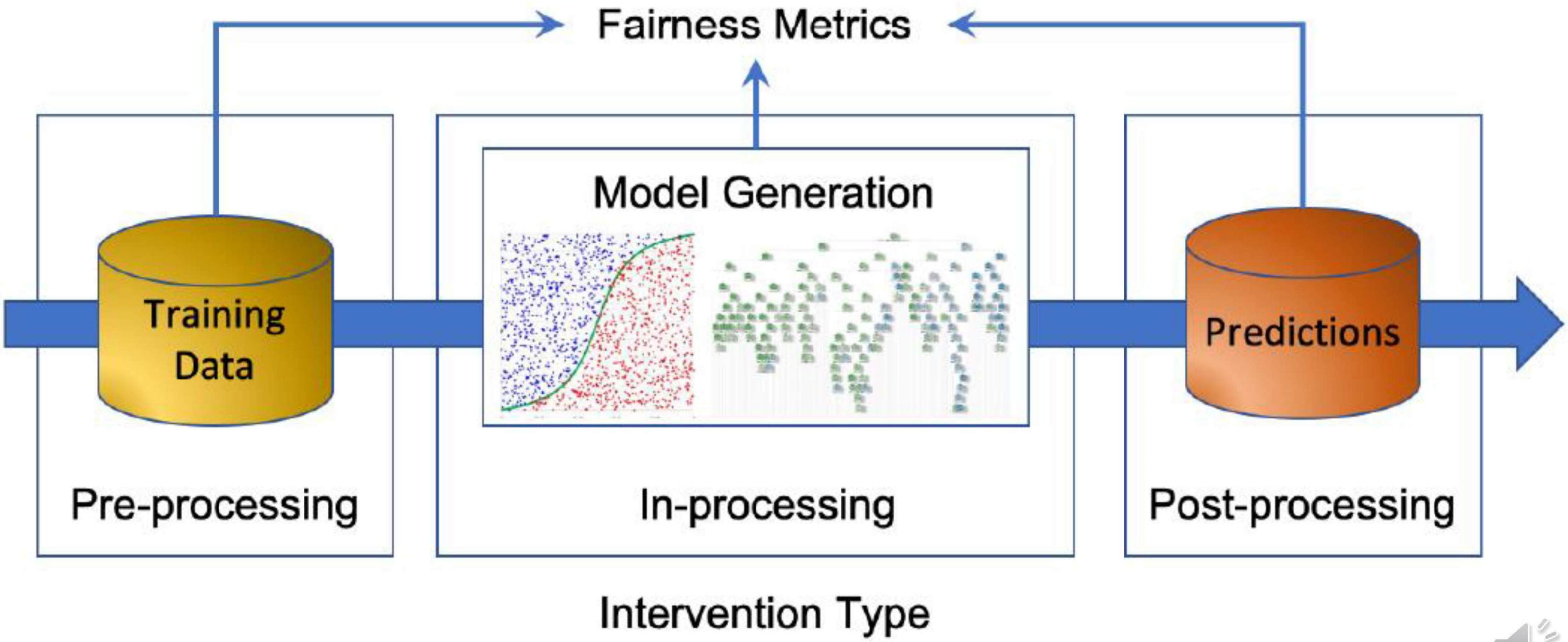
- Biases in interactions with clinicians
  - Automation bias
  - Feedback loops
  - Dismissal bias
  - Allocation discrepancy

Causes of bias - Patients



- Biases in interactions with patients
  - Privilege bias
  - Informed mistrust
  - Agency bias

Fairness Metrics



Caton et al. Fairness in Machine Learning: A Survey, arXiv:2010.04053, 2020

Guarantees Against Discriminatory Bias

- ***Calibration within groups:*** Calibration of algorithmic bias (statistical parity)

$$E[Y|R,W] = E[Y|R,B]$$

- ***Balance for the negative class:*** The average score received by people that are positive with relation to the outcome Y, should be the same in each group
- ***Balance for the positive class:*** The average score received by people that are negative with relation to the outcome Y, should be the same in each group

Metrics for fairness

$$E[Y|R,W] = E[Y|R,B]$$

Independence: $R \perp S$

Separation: $R \perp S|Y$

Sufficiency: $Y \perp S|R$

Guarantees Against Discriminatory Bias

- Not all conditions can be satisfied in the general case
- The trade-off between these conditions is not well understood
- A trade-off between guarantees does not have a scientific/clinical base but it is simply an estimate when the base rates differ between two groups
- Clinical usefulness via (ie. decision curves) should be also taken into consideration

Check points for ‘fair’ decisions

- Equal patient outcomes
- Equal performance
- Equal allocation

Causes of bias - Recommendations

Design:

- Define the goal of a machine-learning model and review it with diverse stakeholders
- Decide what groups to classify as protected
- Investigate if historic data are affected by healthcare disparities

Causes of bias - Recommendations

Design:

- Define the goal of a machine-learning model and review it with diverse stakeholders
- Decide what groups to classify as protected
- Investigate if historic data are affected by healthcare disparities

Data:

- Assess whether the protected group is represented adequately in terms of number of features
- Collect and document training data

Causes of bias - Recommendations

Design:

- Define the goal of a machine-learning model and review it with diverse stakeholders
- Decide what groups to classify as protected
- Investigate if historic data are affected by healthcare disparities

Data:

- Assess whether the protected group is represented adequately in terms of number of features
- Collect and document training data

Training:

- Train a model taking into account the fairness goals.

Causes of bias - Recommendations

Evaluation:

- Measure important metrics and allocation across groups.
- Check generalization of the model in deployment
- Assess usefulness of the models
- Identify/Interpret factors behind model decisions’

Causes of bias - Recommendations

Evaluation:

- Measure important metrics and allocation across groups.
- Check generalization of the model in deployment
- Assess usefulness of the models
- Identify/Interpret factors behind model decisions’

Monitor:

- Systematically monitor data and re-assess model through deployment
- Continue evaluation of users’s interaction and trust to the model
- Consider clinical trial design to assess outcome

Summary

- Quantifying calibration bias in terms of statistical parity and balance in terms of positive and negative classes is an important step to identify issues
- Further research into ‘fairness’ is required to understand how to eliminate bias along with maximizing clinical usefulness

References

- Rajkomar et al. ‘Ensuring Fairness in Machine Learning to Advance Health Equity’, Annals of Internal Medicine, 2018.
- Kleinberg et al. ‘Inherent Trade-Offs in the Fair Determination of Risk Scores, Proceedings of Innovations in Theoretical Computer Science, 2017.
- Caton et al. Fairness in Machine Learning: A Survey, arXiv:2010.04053, 2020