Proceedings of the ASME 2021 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
Virtual Conference: August 17-20, 2021

IDETC/CIE 2021 - 74760

Towards Fairness-Aware Design Decision-Making

Sumaiya Sultana Tanu
Department of Mechanical Engineering
University of Arkansas
Fayetteville, AR

Lu Zhang
Department of Computer Science and Computer Engineering
University of Arkansas
Fayetteville, AR

Dinesh Gauri
Department of Marketing
University of Arkansas
Fayetteville, AR

Zhenghui Sha[1]
Department of Mechanical Engineering
University of Arkansas
Fayetteville, AR

As technology and businesses evolve and the outcomes of an engineering design impact society's lives, a design project's constraints move from strictly technical and economical to social, environmental, and ethical conditions. Among these, one important consideration is to incorporate the fairness dimension in the decision-making process. For example, businesses apply machine learning technologies and big data analytics to study the customers, their needs, and preferences based on vast market data and social media data. This helps enterprises better understand their current market systems and make improved design decisions to cater to customers' needs. However, these data can be embedded with biases towards protected or sensitive classes, such as race, gender, and age, as shown in Figure 1. This often leads to unintended consequences and potentially causes discrimination, inappropriately hurting personal feelings. There is a need to incorporate the concept of fairness in decision-based design (DBD) to support inclusive design decisions.

This study aims to introduce the definitions and statistical measures of fairness such as group fairness, demographic parity, equalized odds, calibration testing, and fairness through unawareness to the engineering design research community [1-6]. From these definitions, we approach two methods to quantify fairness in a dataset. The first method is the Disparate Impact (DI) analysis [7], commonly used in legal domains to evaluate whether a decision-making system is free of disparate impact (indirect discrimination). The higher the DI index value, the lower the chance of bias. The second method uses fairness testing of calibration scores based on sensitive attributes, predicted probability and outcome from machine learning models, and the actual outcome. In our case study, we use a benchmark dataset, the Adult Income dataset, in a binary classification setting to quantify the presence of unfairness using two supervised learning models: Logistic Regression (LR) and CatBoost (CB) classifiers. This data contains demographic attributes about individuals and their income. The income variable is the target variable (Y),

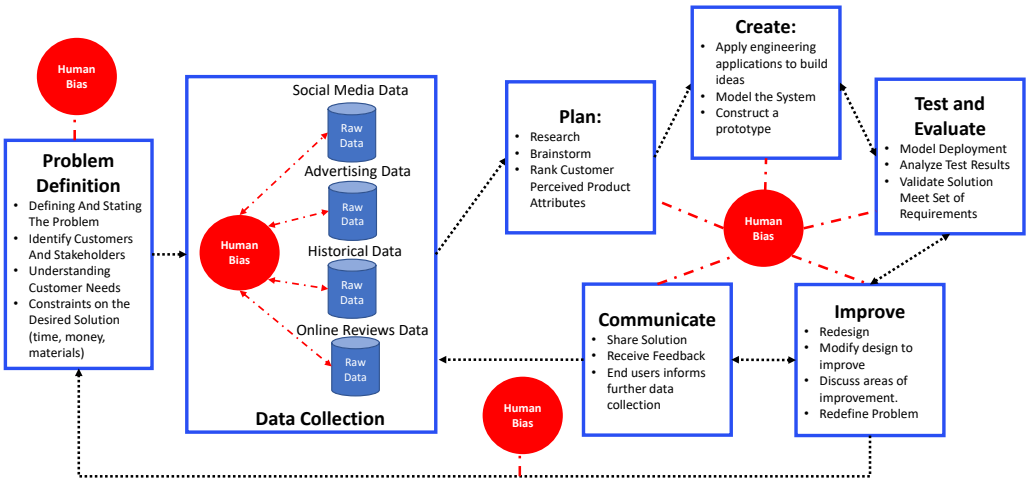


Figure 1: The Potential Bias that can be Introduced in the Engineering Design Process

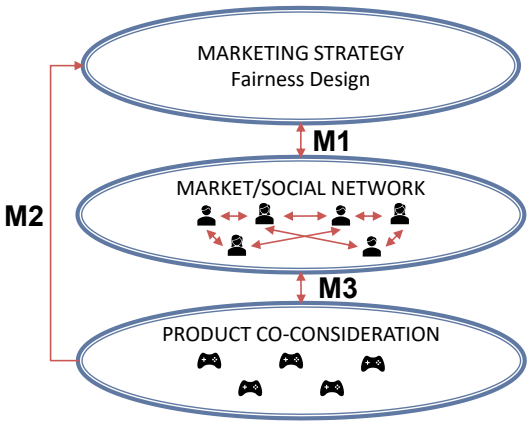
[1] Corresponding author: zsha@uark.edu

i.e., a binary variable indicating an individual's annual income where $Y = 1$ if the income earned is higher than \$50,000 and $Y = 1$ if the income earned is less than or equal to \$50,000. We train the dataset using the two classifiers to predict the income for new individuals. If an individual (new customer or follower) receives a predicted outcome of $\hat{Y} = 1$, he/she is chosen as a target audience. Otherwise, they are rejected.

It is observed that both the classifiers revealed that Gender and Ethnicity are the two critical sensitive attributes that can affect the predictions made in decision models for test data. First, our results indicate that the CB model outperforms the LR model with a prediction accuracy of 83%. This means that for 83% of individuals from the test data, the model correctly predicts whether this individual earned more than 50K or not. Second, to achieve a fair prediction, eligible individuals (with income higher than 50K) from unprivileged groups with the same features as the privileged group members need to receive the binary prediction of 1, irrespective of their sensitive attribute membership. This will classify algorithmic decision-making to be fair and objective. With the disparate impact (DI) analysis, CB turns out to have a higher index value than LR, indicating better performance in fairness in predictions.

Results from the study conclude that quantification of fairness in data is possible using fair ML. However, a thorough understanding is required to bridge the gap between fairness application and DBD to best fit design ecosystems' characteristics. Our case study gives a general idea of how businesses extend their reach of potential customers and improve their experiences through social media data, e.g., by mining consumer demographic information. Many studies have established the connections between consumer behavior, market competitor, and product design [8]. Yet, there exists a knowledge gap to quantitatively understand the effects of enterprise marketing strategies (fair or unfair) within such connections established, as shown in Figure 2.

In the future study, we plan to integrate fairness evaluation methods in the DBD framework to develop a fairness-aware marketing strategy design with positive targeting. We are working on various design solutions to create links between the design of marketing strategies and choice modeling with fairness consideration in M1 and M2 shown in Figure 2. This proposed framework will help develop a data-driven analytical approach to quantitatively predict the effects of enterprises' marketing strategies on customers' choice behaviors by considering the complex customer-product-enterprise relations.



- **M1:** Link between marketing strategy and social network
- **M2:** Link between marketing strategy and product co-consideratio
- **M3 (Existing Literature):** Link between social network, market an product co-consideration

Figure 2: A Framework on Design for Market Systems with Fairness and Positive Targeting

Reference:

[1] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). "Fairness Through Awareness." In Proceedings Of The 3rd Innovations In Theoretical Computer Science Conference, pages 214–226. ACM.

[2] Zafar, M., Valera, I., Rodriguez, M.G. and Gummadi, K. (2015). "Learning Fair Classifiers." ArXiv: Machine Learning.

[3] Verma, S., & Rubin, J. (2018). "Fairness Definitions Explained." 2018 IEEE/ACM International Workshop on Software Fairness (FairWare), 1-7.

[4] Chouldechova, A., (2016). "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," Big Data, vol. 5, no. 2, pp. 153-163, 2017. Available: 10.1089/big.2016.0047.

[5] Chen, Jiahao, et al. "Fairness Under Unawareness." Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, pp. 339–348.

[6] Joseph, M., Kearns, M., Morgenstern, J.H., & Roth, A. (2016). "Fairness in Learning: Classic and Contextual Bandits." In Advances in Neural Information Processing Systems. 325–333. ArXiv, abs/1605.07139

[7] Pessach, D. and Shmueli, E., (2020). "Algorithmic Fairness." ArXiv, abs/2001.09784.

[8] M. Wang, Z. Sha, Y. Huang, N. Contractor, Y. Fu, and W. Chen. (2018). "Predicting product co-consideration and market competitions for technology-driven product design: a network-based approach," Design Science, vol. 4, p. e9, 2018.