Procedural Fairness in Machine Learning (Abstract Reprint)

Ziming Wang[1] , **Changwu Huang**[2] , **Ke Tang**[1] and **Xin Yao**[3]

[1] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology
[2] School of AI and Liberal Arts, Beijing Normal-Hong Kong Baptist University
[3] School of Data Science, Lingnan University

Abstract Reprint. This is an abstract reprint of a journal article by [Wang *et al.*, 2026].

Abstract

Fairness in machine learning (ML) has garnered significant attention. However, current research has mainly concentrated on the distributive fairness of ML models, with limited focus on another dimension of fairness, i.e., procedural fairness. In this paper, we first define the procedural fairness of ML models by drawing from the established understanding of procedural fairness in philosophy and psychology fields, and then give formal definitions of individual and group procedural fairness. Based on the proposed definition, we further propose a novel metric to evaluate the group procedural fairness of ML models, called GPFFAE, which utilizes a widely used explainable artificial intelligence technique, namely feature attribution explanation (FAE), to capture the decision process of ML models. We validate the effectiveness of GPFFAE on a synthetic dataset and eight real-world datasets. Our experimental studies have revealed the relationship between procedural and distributive fairness of ML models. After validating the proposed metric for assessing the procedural fairness of ML models, we then propose a method for identifying the features that lead to the procedural unfairness of the model and propose two methods to improve procedural fairness based on the identified unfair features. Our experimental results demonstrate that we can accurately identify the features that lead to procedural unfairness in the ML model, and both of our proposed methods can significantly improve procedural fairness while also improving distributive fairness, with a slight sacrifice on the model performance

References

[Wang *et al.*, 2026] Ziming Wang, Changwu Huang, Ke Tang, and Xin Yao. Procedural fairness in machine learning. *Journal of Artificial Intelligence Research*, 85, 2026.