

A RELATED WORK

Fairness transfer. With the presence of the distributional difference between different datasets, how well fairness can be transferred is as important as performance generalization. **(1) Cross-domain transfer.** Schumann et al. (2019) discussed how fairness can be transferred across domains. A line of works have studied the problem of cross-domain fairness: evaluation of fairness under datasets bias Hinnefeld et al. (2018), adapt trained models to few-shot (samples) downstream tasks Slack et al. (2020), adapt models to unsupervised domains with losing fairness by robust optimization Rezaei et al. (2021) or by invariant casual inference Singh et al. (2021). As a general framework, Madras et al. (2018) proposed to train transferable adversarially-fair representations (LAFTR). A closely work by Zhang et al. (2020) leveraged the large volume of unlabeled data to mitigate unfairness. However, the approach did not address challenges raised from possible distribution shift from auxiliary sets and the scarcity of minority groups. **(2) Fairness in cross-task transfer.** Ramapuram et al. (2021) first showed that self-supervised contrastive learning can benefit downstream fairness. Zhao et al. (2020) studies transfer in class incremental learning. Baldini et al. (2021) showed the varying fairness of language models in downstream tasks and emphasize the importance of fairness mitigation during fine-tuning. These studies highlighted the necessity to consider distributional differences when transferring fairness, but barely took the ill distributed target dataset into account as did in this proposed work. Therefore, extremely scarce minority samples in the target dataset will bias the transfer.

B ETHICS STATEMENT

Our work is propose an unfairness metric to guide the contrastive fairness pre-training. In this case, we believe our proposed methods can effectively mitigate negative social impacts of machine pre-training learning techniques.

C REPRODUCIBILITY STATEMENT

Our experiments are implemented with PyTorch on Linux system. All datasets as well as the code platform we use are public.