

# 人工智能系统公平性作业

---

## 谁来为AI的公平性负责？ 如何让AI变得更加公平， 真正地服务于人。

---

在数据进入模型之前AI公平性应该由企业保证， 辅助以一部分法律的强制监管， 让数据更好地趋于公平的选择， 平衡各种属性。 在数据被输入模型进行训练之前需要人为检查， 除了匹配数据的分布程度， 数据本身的标签也需要检查----因为标签的比例直接影响了训练结果。 在训练过程中， 训练所用的算法和矩阵乘法加法运算， 都需要更平衡地分配， 不能倾向于某个参数或者输出向量。 可以选择让算法工程师对一些输出向量进行限制， 防止过拟合， 也可以人工加入对抗训练数据， 防止生成相互联系属性。 在得到结果之后， 用户使用的过程中也需要面向更广泛的用户群体， 防止用户带来的偏见对模型， 对AI 公平性的影响。

这些说法都是比较泛泛的说法， 说一些我自己的想法： 什么是公平， 简单的对于两性公平， 这个地球上本来两种人的数量分布就不平衡， 特别是在职业的选择中， 确实某些职业会有很多的性别倾向性， 可是这是事实。 对于计算机来说遵守事实， 预测事实才是它的工作， 这些“不公平”只是数量上的不平衡， 如果人为改造了数据集， 那么得到的预测输出还能否按照我们想的贴合实际， 这是需要打一个问号的， 所以我认为， 对于AI 的公平性， 首要解决的， 是人类对于特定场景的矛盾对立产生的偏见， 而不是数据， 算法的选择。