

Bias is a term that can mean several things within the context of machine learning. At the highest level, it refers to the ways in which an ML algorithm selects from competing hypotheses. It is essential to be aware of the ways in which we might introduce bias, either through our choices in representation, through the data used to train our algorithm, or the ways in which we measure success.

[This article](#) describes some of the ways in which bias can be introduced, and some current approaches to recognizing and mitigating bias.

Please prepare and submit a reflection (uploaded as a PDF) on this article that addresses the following questions:

1. Many of the recent advances in AI and ML have come as a result of the ability to train on huge datasets drawn from the Internet. How does this approach introduce bias?
2. *Representational bias* is a bias that can be inadvertently introduced by algorithm designers. How can representational bias influence the solutions that an AI or ML algorithm generates?
3. The article discusses the tradeoff between *fairness* and *accuracy*. What do these terms mean? Why is there a tension between them? Why can this not be solved through purely algorithmic methods?

Work referenced: The Quest to Make AI Less Prejudiced. Helen Edwards. Quartz.

<https://qz.com/1814415/how-to-fix-bias-in-ai/>