

Intention

COG411 (Introduction to Neural Networks) is a course intended to provide students with a broad overview of neural networks, including computational neuroscience, connectionism, and machine learning. As such, it is an ideal course to introduce students to the role bias plays in machine learning. In previous iterations of the course, the four elements of the Cognitive Science DEI learning objectives had been addressed as a brief discussion topic, wherein students wrote a short written response after watching a NeurIPS keynote speech by Kate Crawford on the topic of bias in Artificial Intelligence. The intent this semester was to expand on this discussion to further support student engagement with these learning objectives. To this end, I added additional written elements to the discussion, requiring students to find an example of bias in a current machine learning area, as well as analyze and discuss potential methods for addressing or preventing bias. I also added an additional day of class discussion in which I demonstrate how bias is introduced to an algorithm for sentiment analysis. Finally, I moved these discussions to the beginning of the semester in order to encourage students to do their course research paper on bias in machine learning.

Implementation

The implementation of these changes was largely successful, and student responses to these changes were generally positive. Students readily engaged with the discussion of bias and produced excellent written responses to the Kate Crawford keynote, finding additional examples and providing nuanced analysis of some of the strategies used to reduce or address bias. Students were moderately engaged with the sentiment analysis demonstration and discussion. In the future, I would like to convert it to a lab activity, but the large file size of the word vector files makes it difficult to provide the complete activity to students in an accessible way. At least four students wrote their final paper at least in some part about bias in machine learning, an increase from previous semesters.