Deborah Raji

Written by Ashley Marcus, Bianca Nachmani, and Dana Dyer

Professor Rangita de Silva de Alwis's Spring 2022 AI and Implicit Bias Policy Lab kicked off with a class discussion featuring ethical AI, thought-provoking leader, and computer scientist: **Deborah Raji**.

In 2020, Deborah Raji was coined one of MIT's 35 under 35 innovators through her extensive research on artificial intelligence biases. Her first-authored work was featured in the New York Times, Washington Post, The Verge, Venture Beats, National Post, EnGadget, Toronto Star and won the Best Student Paper Award at the ACM/AAAI Conference for AI Ethics & Society. Raji is also a key figure in Coded Bias, the feature-length film on Netflix, which exposes prejudices and threats to civil liberty in facial recognition algorithms and artificial intelligence. Raji worked closely with the Algorithmic Justice League initiative, founded by Dr. Joy Buolamwini of the MIT Media Lab, on several projects to highlight cases of bias in computer vision. Together, Raji and Buolamwini forced companies to implement changes to their bias recognition systems. Raji was also a mentee in Google AI's flagship research mentorship cohort, working with their Ethical AI team on various projects to operationalize ethical considerations in ML practice, including the Model Cards documentation project. Recently, Raji worked as a research fellow at the Partnership on AI, formalizing documentation practice in Machine Learning through their ABOUT ML initiative as well as pushing forward benchmarking and model evaluation norms.

In class discussion with Penn Law students, Raji describes the time she first became aware of bias in AI programs and its dangerous effects early on in her career. After her third year of college, Raji

took on a project that used computer programs to flag inappropriate images as "not safe for work." But Raji quickly realized that the program flagged images of people of color at a significantly higher rate than that of white people. The source of the problem, as it turned out, was the training data. The program learned to recognize NSFW content based on a combination of images from pornography and stock photos. Porn, however, is more diverse. As a result, the model associated dark skin with inappropriate content. This troubling discovery led Raji to look for bias in other widely used AI training data. Again, she found that dark-skinned people were highly underrepresented in the training data leading to facial recognition systems' inability to recognize dark-skinned faces accurately. Police departments used the same inaccurate systems to detect and apprehend suspects.

Soon thereafter, Raji met and partnered with Dr. Joy Buolamwini on her project Gender Shades. Raji helped prepare data for audits on AI facial recognition programs for large companies, such as Amazon and Kairos. Through Raji and Buolamwini's diligence, they discovered that the programs identified dark-skinned women with significantly less accuracy than they did light-skinned men. Their findings made the front page of the New York Times business section as they called for companies to recognize their facial recognition flaws and better their systems. Since then, Raji has continued to work passionately to ensure external bias auditing of AI.

As Raji continues to reveal the inaccuracies in current AI systems, she told Professor de Silva de Alwis' Policy Lab that she is currently pushing legal reform to protect against pervasive biases and ensure access to external auditors. Raji's initial realization about algorithmic bias sparked much needed change in the field of facial recognition systems. We are very grateful for Ms. Raji's recent visit to the Policy Lab, as she inspired Penn law students to continue to advocate for equitable and ethical change, despite the pushbacks one may receive. To Deborah Raji: ***Thank you for your dedication, leadership, and inspirational journey***.