

Para saber mais: viés nos dados

Como discutimos em aula, a inteligência artificial é uma das áreas com maior potencial de transformar o modo como vivemos, da mobilidade à saúde. Com o impulso e progressos da IA nos últimos anos, foi possível o avanço tecnológico dos hardwares e do aumento exponencial de dados disponíveis para treinamento dos modelos de aprendizagem de máquina.

Como os modelos mais promissores na área dependem muito dos dados que produzimos, esses modelos podem reproduzir estereótipos da nossa sociedade, impulsionando ainda mais a desigualdade social. Desenvolver modelos que não reproduzam os aspectos negativos da sociedade é importantíssimo para garantir um futuro mais justo a todos, por isso áreas de pesquisa como ética na IA vem crescendo consideravelmente nos últimos anos.

Gostaria de deixar como recomendação para quem se interessar este [artigo](https://www.weforum.org/agenda/2019/01/ai-isn-t-dangerous-but-human-bias-is/) (<https://www.weforum.org/agenda/2019/01/ai-isn-t-dangerous-but-human-bias-is/>) apresentado no encontro anual do Fórum Económico Mundial, escrito por Richard Socher, cientista chefe da Salesforce, que aborda de maneira muito interessante os perigos envolvidos na reprodutibilidade de estereótipos humanos em sistemas de IA.

Outra ótima referência é o [capítulo 6](https://web.stanford.edu/~jurafsky/slp3/6.pdf) (<https://web.stanford.edu/~jurafsky/slp3/6.pdf>), seção 6.11 - Bias and Embeddings, do livro Speech and Language Processing escrito por Dan Jurafsky e James H. Martin. Nesta seção eles abordam o tema viés, além de trazer referências com técnicas para reduzir este problema no embedding de palavras.