

Workshop:

Biases in Artificial Intelligence & Applications to NLP

Jean-Michel LOUBES, **Professor of Mathematics**, Université Toulouse Paul Sabatier; **Chair Fair and Robust Machine Learning** at Artificial and Natural Intelligence Toulouse Institute ([ANITI](#))

05 may 2021, 9:00 a.m. – 11h00 a.m.

In this workshop we will review how Artificial Intelligence's algorithms work and what is the machine learning process. We will also study the impact of certain biases, with a particular focus on Natural Language Processing cases.

Neste workshop, apresentaremos como funcionam os algoritmos de Inteligência Artificial, e em que consiste o processo de aprendizagem de uma máquina. Estudaremos também o impacto de alguns enviesamentos, dando um foco particular a casos de Processamento de Linguagem Natural.

Free attendance, by zoom, <https://videoconf-colibri.zoom.us/j/86750029079>

Organisation: Catarina Vaz Warrot (awarrot@letras.up.pt)
Alexandra Guedes Pinto (mapinto@letras.up.pt)

More informations:

Due to the rapid growth of available data, algorithms have entirely or partially replaced human decision makers in nearly all areas of our life, including such sensitive domains as school admission, credit lending, hiring processes, and even predictive police and automatic judicial decisions to name a few.

This worrying, but inevitable, trend has raised serious concerns about the ethical and moral consequences of such a shift. The central question in this issue is: "Are we sure that these algorithms are not violating modern ethical standards?"

Unfortunately, case by case studies in several domains reveal that the blind use of algorithms does indeed lead to undesirable and, at times, completely immoral decisions. A partial explanation of this phenomenon lies in the fact that the overwhelming majority of algorithms are trained on historical and human-made data, which encode internally discriminatory biases.

Much as the folklore phrase says that "your algorithm is as good as your data", classical and more modern deep learning algorithms often mimic and even amplify discriminatory biases encoded into the sample by reproducing existing disparities to the whole population.

Informações complementares:

Devido ao rápido crescimento dos dados disponíveis, os algoritmos substituíram total ou parcialmente decisões humanas em quase todas as áreas de nossa vida, incluindo domínios sensíveis como a admissão em escolas, a obtenção de créditos, processos de contratação e até mesmo polícia preventiva e decisões judiciais automáticas.

Esta tendência preocupante, mas inevitável, levantou sérias preocupações sobre as consequências éticas e morais de tal mudança. Tal situação levanta uma questão central: "Temos a certeza de que esses algoritmos não violam os padrões éticos modernos?"

Infelizmente, estudos de caso em vários domínios revelam que o uso cego de algoritmos pode conduzir a decisões indesejáveis e, às vezes, completamente imorais.

Uma explicação parcial desse fenômeno reside no fato de que a esmagadora maioria dos algoritmos são treinados a partir de dados históricos e feitos pelo homem, que codificam vieses discriminatórios.

Assim como a frase que diz que "o seu algoritmo é tão bom quanto os seus dados", algoritmos de aprendizagem muitas vezes imitam e até amplificam vieses discriminatórios codificados no corpus, reproduzindo disparidades existentes na população.