**University Grenoble-Alpes**<br>
**INTERNSHIP PROJECT**

**2021**

**GIPSA-lab, GSTATS Data Science Chair**
**Romain COUILLET**

**Office D1178**

# Fairness in AI: A Random Matrix Analysis

## Project context and application

The advent of artificial intelligence and the ubiquitous nature of machine learning in all areas of science and the society at large raises new concerns regarding the question of "fairness" of machine learning algorithms. In particular, while machines ought not to be biased in their decision making (as they rigorously follow instructions), algorithms are still coded by human beings and, most importantly, exploit data selected by human beings, sometimes imperfectly so, and quite often prone to unfair features.

This realization has triggered first experts in law and in ethics to raise the question of fairness in AI. Recently though, this question has turned into a mathematical object constraining datasets and algorithms to be "more fair".

The objective of the internship will be twofold: (i) at first, it will aim at understanding the various aspects of the new field of "fairness in AI", identify the open directions, the main contributions and the main bottlenacks; then (ii) based on the recent advances in the GAIA team on large dimensional statistics for an improved understanding of AI algorithms, the objective will be to revisit some simple learning mechanisms under the length of fairness. More specifically, item (ii) will consit in devising explicit "fairness constraints" on existing learning mechanisms and observe its theoretical and practical consequences : for instance by studying a 2x2-class classification problems in which 2 classes are unwanted discriminatory features.

## Main steps

- Review of the literature on fariness in AI, and latest works of the GAIA team on random matrix theory for AI.
- Theoretical analysis of simple methods to induce fairness in algorithms.
- Application to the algorithms studied in the GAIA team (for instance in unsupervised learning).

**Associated domains:** Fairness in ML, andom matrix theory, machine learning.

**Requirements:** Good coding skill in Matlab or Python, knowledge of the basics of machine learning and statistics/probability theory.

**Location:** The internship will take place at GIPSA-lab, University of Grenoble-Alpes, in the Grenoble area.

## References

[1] Course material MoSIG : Romain Couillet, « Fainess in ML »