# Assignment 1 – Bias in NLP models

## Assignment Instructions

In this assignment we will work on understanding and evaluating bias in pre-trained language models. The assignment will be completed in the Google Colab notebook.

In the Google Colab notebook, we will:

1. Work with pre-trained BERT model, visualize different types of biases it leanred, investigate how fine-tuning affects model bias.
2. Learn how to estimate model fairness with WEAT test.

For the best experience, change the runtime to use a GPU accelerator. You can use a free GPU on colab by selecting: **Runtime → Change runtime type → Hardware Accelerator: GPU**

This assignment will be completed with PyTorch and HuggingFace Transformers libraries.

Link to assignment Bias in NLP Models

## What To Submit

After completing the assignment in Colab, students are required to save their files in PDF and ipynb file formats. The files should be saved as lastname-assignment1.pdf / lastname-assignment1.ipynb.

Saved files should be uploaded on D2L.

**Assignment Developed by Anastasia Razdaibiedina|Reviewed by Yinka Oladimeji**

**Instructor: Sayyed Nezhadi | Course TAs: Anastasia Razdaibiedina |Rishav Raj Agarwal**

**| Course Director: Shingai Manjengwa (@Tjido)**