Instructions Jupyter Notebook: Identifying bias in AI

In this exercise you will work with real data and explore how bias can emerge when using AI. The activity is a programming exercise based on https://www.kaggle.com/code/alexisbcook/exercise-identifying-bias-in-ai/notebook, which was released under the Apache 2.0 open source license. Affinity with programming and/or Python is handy but not necessary. Even if you are new to coding, you will still be able to complete it.

Some background on the data:

At the end of 2017 the Civil Comments platform, a full-featured commenting plugin for independent news sites, shut down and chose to make their ~2m public comments from their platform available in a lasting open archive so that researchers could understand and improve civility in online conversations for years to come. Jigsaw, a unit within Google that explores threats to open societies and builds technology that inspires scalable solutions, sponsored this effort and extended annotation of this data by human raters for various toxic conversational attributes. In this exercise you will work with a small subset of the data that was used in the Jigsaw Unintended Bias in Toxicity Classification competition.

How this activity is organized:

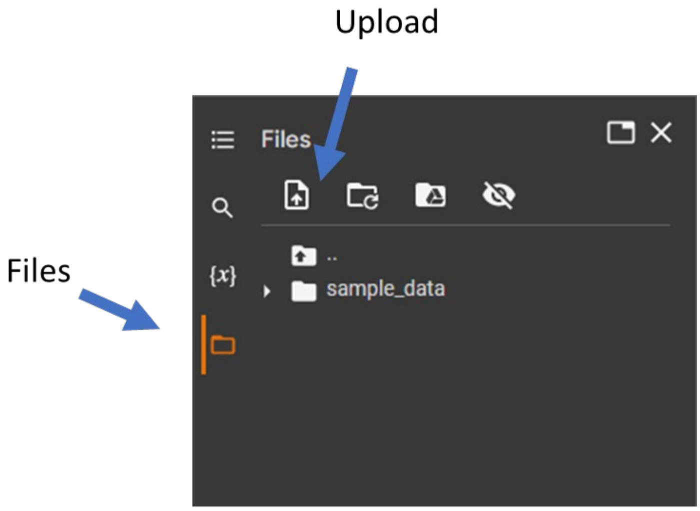
It is recommended to use Google Colab for this exercise. This is a free Jupyter notebook environment provided by Google. Notebooks contain code but also rich text elements. The code can be executed and output is printed on the screen. By using Google Colab no configuration is needed, you execute the code in your browser. Sharing such notebooks is also easy.

Getting started with Google Colab:

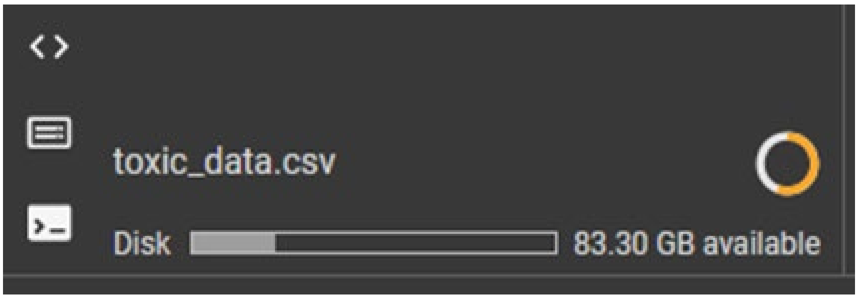
- Log in to your Google account and go to https://colab.research.google.com/
- Now you have several options
  - Examples: some examples provided by Google Colab
  - Recent: notebooks you recently worked with
  - Google Drive: notebook you have in your Google Drive
  - Github: you can add notebooks from your Github, but for this you will need to connect Colab and Github
  - Upload: upload notebooks from your local directory
  - New notebook: create a new (empty) notebook
- In the same folder as these instructions you find two other files: a notebook called *IdentifyingBiasInAI.ipynb* and a data file *toxic_data.csv*
- You can now open the downloaded notebook in Google Colab by clicking on "Upload" and selecting *IdentifyingBiasInAI.ipynb*

Instructions can be found in the notebook itself. Some extra visuals about adding the data file are given here:

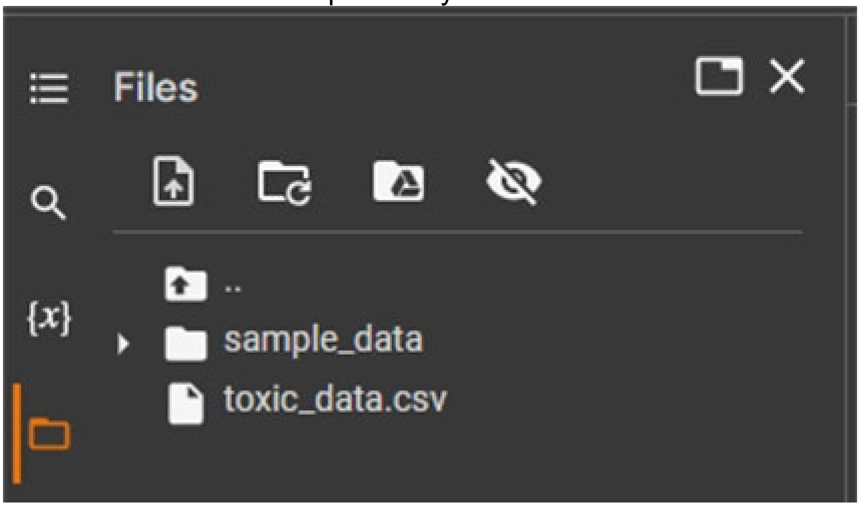
- Important: the code in the shared notebook makes use of a data file called *toxic_data.csv*. This file needs to be uploaded. When you have a notebook open, you click on Files (the orange icon on the left of the screen) and then on Upload to session storage (the upload icon)



- When you select a file to be uploaded, Google Colab might warn you that the runtime's files will be deleted once the runtime is terminated. Basically, this means that if you have been away for too long, have closed the notebook, or if you have manually deleted the runtime you will have to upload it again.
- If you have done everything correctly, you will see below that the file is being uploaded. This might take a couple of minutes and you should not proceed until this is finished. Otherwise the data cannot be read and converted to the correct Python data structure.



- When the file has been uploaded you will see it in the list of files:


Files
..
sample_data
toxic_data.csv