README *for* The Sound of Populism

This README file contains three sections. The first section, *General Description*, describes the contents in this replication folder. The second section, *Hardware Requirements*, describes what hardware and software is used for running the Google Colab scripts (ipynb). The third section, *Instructions*, reports on how to run the scripts.

1 General Description

The replication package contains the following files:

- README.pdf: overview of the replication package and instructions.

- **Data and scripts for model training and evaluation (Table 2):**
  - labelled_data.csv: German-language text used to train German BERT; used in Table 2.
  - GBERT.ipynb: script for fine-tuning German BERT; used in Table 2.
  - gpt_translations.csv: English-language text used to train RoBERTa; used in Tables 1 and 2.
  - RoBERTa.ipynb: script for fine-tuning RoBERTa; used in Table 2.

- **Model inference, summary statistics, and regression analysis (Tables 3–6):**
  - Inference.ipynb: script for generating estimated probabilities for each populism variant.
  - regression_data.csv: dataset containing estimated probabilities from `Inference.ipynb` and LIWC features; used in Tables 3–6.
  - Analysis.ipynb: script for generating regression results (Tables 3–6) and summary statistics.
  - summary_statistics.pdf: summary statistics for regression analysis generated from `Analysis.ipynb`.

2 Hardware Requirements

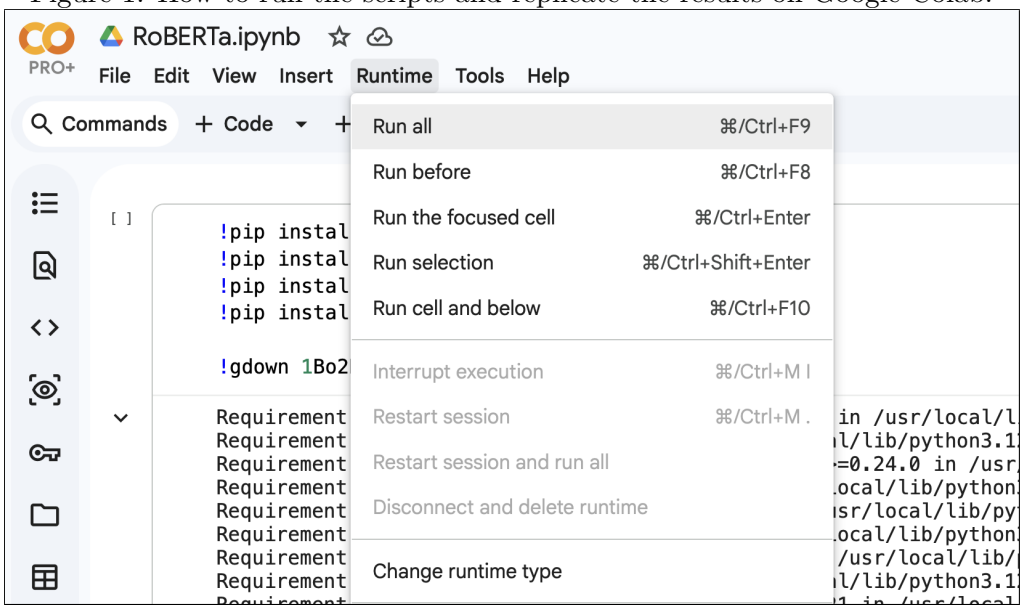
The replication package has been tested for running on A100 GPUs on Google Colab using Python 3.12.[1] Exact performance metrics would vary to some extent depending on the type of GPUs used, but they would not affect the key results. Software packages (e.g., transformers, datasets, and peft) have been pinned to specific versions in the scripts.

3 Instructions

Execution

Each of the scripts above can be run independently. In Google Colab, we only need to hit the **Run all** button under **Runtime** (Figure 1).[2] All the key results will be printed out during execution.

Figure 1: How to run the scripts and replicate the results on Google Colab.



[1] Note that usage of A100 GPUs on Google Colab is not free. Usage of some other types GPUs (e.g. T4) is free. Results based on other Python versions, for example 3.10, might differ slightly.

[2] We can also run the scripts interactively if readers so prefer.