Seshat Global History Databank Text Dataset and Benchmark of Large Language Models' History Knowledge

Dataset links

Dataset Repository (Github) Croissant Metadata (Github)

Usage

This dataset can be used to benchmark LLMs on their expert level history knowledge.

License/Responsiblity

This dataset is licensed under CC by 4.0. The authors bear all responsibility in case this dataset contains violations of rights.

Dataset preservation/DOI

Upon acceptance we will publish the dataset on Zenodo with a respective DOI, as to ensure long term preservation of the dataset.

Loading the dataset

using Python and Pandas:

```python
import pandas as pd
main = pd.read_parquet("Neurips_HiST-LLM.parquet")
ref = pd.read_parquet("references.parquet")
```

Results reproduction

Model APIs

Model APIs were been queried using the `openai_run_batch.py` and `togetherai_run.py` python scripts.

Plots and Tables

Plots and Tables can be re-generated by running the `process_res1.py` script first then the other R and gnuplot scripts can be executed

Dataset metadata

Dataset metadata documented in the `croissant.json` file.

Model Fingerprints

When model fingerprint are available we created extra columns for each model fingerprint. These columns are named via the following pattern `<model-name>_<model-fingerprint>`.

Column Descriptions

additional$_{\text{review}}$

Boolean This column describes whether datapoints underwent additional expert review. See section 3.2 of the Paper.

Q

The multiple choice question.

A

The expected completion of the prompt.

polity old id

ID for polity according to Seshat ids.

start year str

String for when polity started existing (in BCE/CE format).

end year str

String for when polity stopped existing (in BCE/CE format).

start year int

Int for when polity started existing (in BCE/CE format).

end year int

Int for when polity stopped existing (in BCE/CE format).

name

Polity name.

nga

Natural Geographic Area for Polity.

category

Immediate parent category of fact from Seshat codebook.

root cat

Root category of fact from Seshat codebook.

value

Value of fact in Seshat.

variable

Variable of fact in Seshat.

id

Request id for openai batch requests.

region

World region for fact.

description

Description provided by RAs for fact.