ML4P Shock Validation Instructions

Description:
The Machine Learning for Peace (ML4P) project collects articles published by a curated sample of more than 350 online newspapers publishing in more than 37 languages and uses large language models (LLMs) to detect the occurrence of political events. We are looking for help monitoring the ability of these models to accurately detect and describe major political events around the world.

Tasks:
- Compare LLM-generated news summaries to original article text and assess their ability to prioritize important events and provide accurate summaries
- Identify articles reporting on events that do not match the identified category or events happening in the wrong country
- Rate the importance of political events reported in news

For this task, you will be provided with a csv spreadsheet. Each row in the spreadsheet represents a specific country-month where our analysis suggests an important event (or events) occurred. Note: while the events described in each row should all belong to the same event category (ex. protest, legal change, etc.), there will often be multiple instances of each event being reported (ex. two distinct protests occurring in the same month). The spreadsheet will contain the following columns:
- Country: The country where we believe the event occurred
- Date: The month in which we believe the event occurred (ignore the day, which is always set to the first day of each month).
- Event: The type of event that we believe happened. A list of event categories and definitions can be found [here](#).
- Num Local: The number of articles reporting on the event category by local/domestic news sources during that country-month
- Num Int: The number of articles reporting on the event category by international news sources during that country-month
- Main Text: The first ~300 characters of text for all articles published by local/domestic news sources. Each bullet-point corresponds to a unique article.
- Events List: A list of the five most important events produced by a GPT model. To produce this list, we asked a GPT model to summarize the five most important events reported in the `Main Text` column. The country name at the beginning tells us where each event occurred. Note: there may be fewer than five events listed if there are not five events reported in `Main Text`.
- Misplaced Countries: The number of events in the list that begin with a country-name other than what is listed in the `Country` column.
- Final Summary: A condensed summary of the `Events List`.

Using this csv file, you will be asked to make the following assessments. For each instruction, please create a new column in the csv file and record your assessment:

- Does the `Events List` accurately capture the country where these events occurred?
  - Read the `Events List` summary of events. Each of the event summaries should begin with a country-name. If the country name accurately describes the country where the event occurred, in a new column titled `country_correct`, please record "Yes". Otherwise, please record the name of the country where the event actually occured.
- Do the events in `Events List` match the definition of the event recorded in the `Event` column.
  - Read the `Events List` summary of events. Read the definition for the event category in the `Event` column (available [here](#)).
  - In a new column titled `event_correct`, record the number of events in `Events List` that do not match the event category in `Event`. If all events in `Events List` match the definition of the event category in `Event`, you should record 0.
- Does the `Events List` accurately capture reports from `Main Text`?
  - Read the `Events List` summary of events. Read through the bullet points in `Main Text` text. In a new column titled `summary_incorrect`, record the number of events in `Events List` that do not match any articles from `Main Text`. If all events in `Events List` match at least one article from `Main Text`, you should record 0.
  - Note: for rows where the number of bullet points in`Main Text` exceeds 10, it may be easier to copy and paste the contents of `Main Text` into a word document and use ctrl+f to search for keywords.
- Are any events in `Main Text` more important than those in `Events List`?
  - GPT is instructed to summarize the most important events from `Main Text` in `Events List`. After reading `Events List` and `Main Text`, were there any events in `Main Text` that were more important than any of the events in `Events List`. Here, we consider the important events to be those that indicate major changes in a country's politics or governance.
  - In a new column titled `importance_missed`, copy and paste any bullet points from `Main Text` that you think are more important than the events in `Events List`. This is subjective, so please only highlight cases where extremely important events have been missed.