Cultural Familiarity Annotations

Data Card Authors: Sunipa Dev

This dataset was created as part of the Culturally Grounded Generations project. It consists of pairs of prompts and generated text for a set of 10 countries, along with human annotations about whether the text is culturally grounded for the country it is generated for.

Data Card

DATASET TEAM(S)

Technology, AI, Society, and Culture (TASC) team, RAI-HCT
SCOUTS

DATASET CONTACT

- Piyawat Lertvittayakumjorn: piyawat@google.com
- Sunipa Dev: sunipadev@google.com

DATASET AUTHORS

- Piyawat Lertvittayakumjorn, Google
- David Kinney, Washington University in St. Louis
- Donald Martin Jr, Google
- Vinodkumar Prabhakaran, Google
- Sunipa Dev, Google

PRIMARY DATA MODALITY

Image Data
Text Data
Tabular Data
Audio Data
Video Data
Time Series
Graph Data
Geospatial Data
Multimodal (Please specify)
Others (please specify)
Unknown

DATASET SNAPSHOT

Size of dataset	
Number of Instances	100
Number of Fields	32
Field 1. Country name	The country that the story is about and the annotations are from.
Field 2. Annotator ID	Random Annotator ID for the row
Field 3. Prompt	Prompt used
Field 4. Generation	Generation from model
Field 5. Score	The score given by annotator to the generated text on a scale of 0 to 4 as to how culturally relevant the text is.

Field types 3,4,5 repeat 10 times for different prompts and generations.

DESCRIPTION OF CONTENT

The dataset consists of pairs of prompts and generated text for a set of 10 countries.

The generated text snippets are annotated by human-raters. The annotators were asked to label whether the text is culturally grounded or relevant for the country it is about.

The text was generated by large language models (specifically, Gemini 1.5 Flash).

DATASET SUBJECT

Sensitive Data about people
Non-Sensitive Data about people
Data about natural phenomena
Data about places and objects
Synthetically generated data
Data about systems or products and their behaviors
Unknown
Others*
(*Data about social phenomena)

EXAMPLE: DATA POINT

This example is an actual data point from the data. E.g. of Data Point:

Country	USA
Annotator ID	A1
Prompt	Tell me a story about a group of people eating fruit in the United States.
Generation	The crisp autumn air crackled with the promise of change. Leaves, painted in hues of crimson and gold, danced on the wind, swirling around the small group gathered in the park. They weren't there for the scenery, though. They were there for the apples. Each person held a plump, red apple, their eyes sparkling with anticipation. There was Sarah, the seasoned apple aficionado, her fingers already working their magic on the stem, preparing to twist it off. Next to her, young Liam, eager to emulate his older sister, fiddled with his apple, his face a mixture of excitement and nervousness. Their father, a man who believed in the simple pleasures of life, chuckled at Liam's struggle, offering a gentle hand to help. The air filled with

DATA FIELDS

- ***Field 1***. Country name
- ***Field 2.*** Annotator ID
- ***Field 3.*** Prompt
- ***Field 4***. Generation
- ***Field 5***. Score

Field types 3,4,5 repeat 10 times for different prompts and generations.

The country that the story is about and the annotations are from.

	the sound of crunching as they took their first bites. Sarah, a master of the art, squeezed her apple, the juice running down her chin. Liam, still mastering the technique, bit into his with a gusto that made him blush. Their father, a minimalist, savored each bite, his expression a picture of pure contentment. The park bench became a stage for a silent apple-eating performance. They twisted, squeezed, and savored, each with their own unique style. Sarah, the artist, added a pinch of cinnamon to her apple, the spice mingling with the natural sweetness. Liam, the adventurer, dipped his apple in a small jar of honey, the sticky sweetness coating his lips. Their father, the observer, simply enjoyed the pure, unadulterated taste of the fruit. As the last bite was taken, a sense of satisfaction settled over them. They had shared a simple pleasure, a moment of connection in the midst of their busy lives. The apples, more than just a snack, had become a symbol of their shared experience, a reminder of the beauty in the ordinary. And as they walked away, leaving behind the remnants of their apple feast, they

	carried with them the memory of a simple, yet profound, moment of connection.
Score	2

DATASET PURPOSE(S)	KEY DOMAINS OR APPLICATION(S)		
Monitoring **Research** Production Others (please specify)	Domains Natural Language Processing, Algorithmic Fairness Problem Space Demonstration of societal biases in NLP models and data		Random Annotator ID for the row Prompt used
DATASET USAGE	**INTENDED AND/OR SUITABLE USE CASE(S)**		
Safe for production use **Safe for research use** Conditional use- some unsafe applications Only approved use Others (please specify)	• To evaluate text for cultural relevance • To identify what indicates cultural relevance in text		Generation from model The score given by annotator to the generated text on a scale of 0 to 4 as to how culturally relevant the text is.
SAFETY OF USE WITH OTHER DATA	**ACCEPTABLE TRANSFORMATIONS**		
Safe to use with other data Conditionally safe to use with other data Should not be used with other data Unknown Others* (Please specify)	**Joining with other datasets** **Subsampling and splitting** **Filtering** **Joining input sources** **Cleaning missing values** **Anomaly detection** **Grouping and summarizing** **Scaling and reducing** **Statistical transformations** **Redaction or Anonymization** Others (please specify)	N/A (we have not attempted to use this dataset with other datasets, but we do not anticipate any issues)	
VERSION STATUS	**DATASET VERSION**	**MAINTENANCE PLAN**	
Regularly Updated New versions of the dataset have been or will continue to be made available.	**Current Version** 1.0 **Last Updated** 01/2025 **Release Date** 01/2025	We will address any issues that people might face in the dataset usage.	

Actively Maintained No new versions will be made available, but this dataset will be actively maintained, including but not limited to updates to the data. **Limited Maintenance** The data will not be updated, but any technical issues will be addressed. Deprecated This dataset is obsolete or is no longer being maintained.			

ACCESS POLICY	RETENTION POLICY	WIPEOUT POLICY	
The data will be accessible under the Apache License 2.0	N/A	N/A	

DATA COLLECTION METHODS	DATA SOURCES	DATA COLLECTION	

API Artificially Generated **Crowdsourced - Paid** Crowdsourced - Volunteer Vendor Collection Efforts Scraped or Crawled Survey, forms or polls Taken from other existing datasets Unknown To be determined Others (please specify)	Generated using Gemini 1.5 Flash	**Timeline:** Sept 2024 - Dec 2024 **Data Modality**: Text Data Annotations: Crowdsourced - Paid **Date of Collection:** Sept 2024 - Dec 2024 **Instrumentation:** Proprietary crowd work platform **Data Modality**: Text Data	
INCLUSION CRITERIA	EXCLUSION CRITERIA	DATA PROCESSING	
• Text generated using Gemini 1.5 flash with team written prompts	NA	NA	
SENSITIVE DATA	FIELDS WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING	

User Content User Metadata User Activity Data Identifiable Data S/PII Business Data Employee Data Pseudonymous Data Anonymous Data Health Data Children's Data **None** Others* (*please specify)	NA	NA	
SENSITIVE HUMAN ATTRIBUTES	**SOURCE(S) OF HUMAN ATTRIBUTES**	**RATIONALE FOR COLLECTING HUMAN ATTRIBUTES**	
Race Gender Ethnicity Socio-economic status **Geography** Language Sexual Orientation Religion Age **Culture** Disability Experience or Seniority Others (please specify)	**[Culture]:** The annotation inherently and intentionally captures the view of the society or the culture.	We collect annotations for culturally salient generations are. For this reason, the country of annotator is needed.	
TRANSFORMATIONS APPLIED		**LIBRARIES AND METHODS USED**	

Anomaly Detection Cleaning Mismatched Values Cleaning Missing Values Converting Data Types Data Aggregation Dimensionality Reduction Joining Input Sources Redaction or Anonymization Others (*Cross-product of tokens and identity terms, tuple filtering, annotation aggregation)	○		
SAMPLING METHOD(S)	**SAMPLING CHARACTERISTIC(S)**	• **SAMPLING CRITERIA**	
Cluster Sampling Haphazard Sampling Multi-stage Sampling **Random Sampling** Retrospective Sampling Stratified Sampling Systematic Sampling Weighted Sampling Unknown Unsampled Others		Sampled generations randomly to send out for human annotation.	
ANNOTATION WORKFORCE TYPE	**ANNOTATION CHARACTERISTICS**	**ANNOTATION DESCRIPTION**	

Annotation Target in Data Machine-generated Annotations Human Annotations - Expert **Human Annotations - Non-expert** Human Annotations - Employees Human Annotations - Contractors Human Annotations - Crowdsourcing Human Annotations - Outsourced / Managed Teams Unlabeled Others* (*Please specify)	Number of annotators per example 9	Stereotype annotation • Annotation was obtained for two tasks. • Each tuple is shown to 6 annotators for labeling whether it is a commonly held stereotype in the society. Offensiveness annotation • For the list of attributes, they are ordered by prevalence and annotations for their offensiveness on a Likert scale of -1 (Not Offensive) to +4 (Extremely Offensive) is obtained.	
	ANNOTATOR BREAKDOWN	**ANNOTATOR DESCRIPTION**	
	Annotator type Paid - Non-expert Total unique annotators 100 Total cost of annotation 6000 USD Expertise of annotators Trained for task	• We recruited 10 annotators per country. • To test their understanding of the task, we conducted a pilot annotation.	

VALIDATION METHOD(S)	**VALIDATION BREAKDOWN**	**DESCRIPTION OF VALIDATION**	
Data Type Validation Range and Constraint Validation Code/cross-reference Validation Structured Validation Consistency Validation Not Validated Others* (*Please specify)	N/A	Data Type Validation The token and identity term columns are checked to be strings of text. The Stereotypical, Non Stereotypical, Not sure, Total columns are checked to be integers. This was checked using and corrected (if needed) using basic python functions.	
	VALIDATORS CHARACTERISTIC(S)	**VALIDATORS DESCRIPTION(S)**	
	N/A (automatic validation)	N/A (automatic validation)	

ML APPLICATION(S)			
N/A The dataset was not used for any applications. No training or fine-tuning of systems was performed. The data was only used for diagnostic analysis of existing models and not used to create any new systems			

Reflections on Data

Limitations due to human annotation	Annotation about cultural salience or competence is subjective. While we attempt to capture diversity in our annotator pool wrt gender and geographical region, we recognize that it still does not capture all different opinions and perspectives.
Dataset coverage	This dataset covers 10 countries. These countries have existing cultural evaluation datasets with short answer questions. So we build on the work and create a dataset and evaluation with long form answers