

Data Collection Assignment Instructions

Consent and Warnings

- **Copyright Release:** Participants agree to waive any copyright claims, permitting the data to be utilized in further research.
- **Data Availability:** All data gathered in this project will be accessible publicly.
- **Responsibility and Legality:**
 - Participants are obligated to responsibly complete all phases of data collection, cleaning, and verification.
 - Engaging in illegal activities, using fake or fabricated data is strictly prohibited.
 - The verification process must be conducted blindly to ensure unbiased results.
 - The collected data is not to be used for commercial purposes, presently or in the future.
 - Privacy and confidentiality must be respected; only publicly available data should be utilized.
 - Contributors have the option to remain anonymous or to be acknowledged in any future research that utilizes this data.

Description

This assignment involves testing generative AI tools such as ChatGPT, Bard, and similar platforms. Participants will collect data from prompts and their corresponding responses generated by these tools. The goal is to identify and categorize data exhibiting bias, harmful content, hate speech, etc. (negative examples), and differentiate these from neutral or harmless content (positive examples). Each example must undergo a blind review process involving at least two additional reviewers. The final classification of examples will be determined collaboratively.

Objectives

- **Data Lifecycle:** Engage comprehensively in all stages of the data lifecycle.
- **Practical Experience:** Acquire hands-on experience in real-life data generation, collection, cleaning, preprocessing, and analysis.
- **Data Collection:** Gather new data for evaluating generative AI tools.
- **Data Verification:** Actively participate in the verification process of collected data.
- **Ethical Research:** Contribute to scientific research and future studies that promote responsible and ethical data usage.

Guidelines for Data Collection and Verification

- **Tool Usage:** Utilize generative AI tools such as ChatGPT, Bard, Dall-E for data collection. Data can be in the form of text or images.
- **Data Generation:** Generate your own distinct prompts and collect responses from the selected AI tools.
- **Collaboration:** Work individually or in a group. Each participant, whether in a group or alone, must submit 30 positive and 30 negative examples per batch.
- **First Round Submission:** Include details such as collector's ID and name, the tool used, date and time of data collection, prompt, response, coarse(positive, negative, neutral) and fine granular categories (e.g., normal, biased, hate speech), reasoning for categorization, and any other relevant notes or links.
- **Second Round Blind Review:**
 - Conceal the category, reasoning, and any notes from your collected data.
 - Distribute the anonymized data to at least two additional reviewers for independent categorization and commentary.
 - Reviewers will add their IDs, names, categorization choices, reasoning, and notes to the data.
 - **Peer Review:** Each participant is expected to review at least two distinct datasets submitted on Canvas.
 - **Final Categorization:**
 - Add a column for the final categorization, determined by the majority opinion among reviewers.
 - Maintain a record of all prior columns, including names, IDs, categorizations, and reasonings from all reviewers in the final document.
- **Submission Deadlines:** Adhere to due dates posted on Canvas.
- **Submission Requirements:**
 - Submit all first and second-round assignments on Canvas.
 - Post your anonymized data in the relevant discussion forums. This allows any participant to review your dataset and provide feedback.
 - You are responsible for managing the review process and compiling the final document.
 - Use the Excel file templates in all submissions.