Watermarking Human Study (5/9/23)

Setup (Applies to both tasks)

1. After reading these instructions in their entirety, follow the signup link at the bottom of the page that allows you to create an account and login to the annotation application.
2. Open the [**annotator assignment spreadsheet**] and add your name to the list in an open row for one of the tasks.
3. From the projects page, click on one of the two projects (start with "Paraphrase Text").
4. When ready to annotate, use the checkboxes on the left side of the screen to select the range of instances assigned to you (click them individually) and then click the blue button in the center that says "Label N Tasks".
5. When finished, return to the projects view with the two tasks on it and follow the same procedure from step 2. but for the other task.

"Paraphrase Text"

Description

Paraphrase an AI generated response to a question from Reddit's r/explainlikeimfive (ELI5) forum.

Instructions

A question or topic statement is shown at the top of the screen. On the left side of the screen you will see a response to the question. The response was generated by an AI language model. The response is "watermarked," meaning it contains invisible patterns that can be used to determine that the response was written by an AI and not a person. Read the AI-generated response on the left half of the screen, and in the text box on the right side of the screen, re-write the response in your own words, whilst preserving the meaning and length of the text. Your goal is to change the text so much that the watermark is no longer detectable.

When you are finished, click the "submit" button to save your re-written text and move on to the next task.

Requirements:

1. **Paraphrase quality/similarity** - A paraphrase should convey roughly the same information as the original text, to roughly the same level of detail.

2. **Time limit** - Try to spend no more than **10 minutes** on any individual paraphrasing task. The annotation software tracks the time you spend on each task, but it will not

explicitly enforce the time limit by kicking you off. Please do the tasks in a single sitting.

3. **No automated paraphrasing tools** - Do not use any AI tools that write text for you (e.g., ChatGPT, Grammarly), and do not copy/paste text from any external source. However, you may look things up online, refer to a dictionary or thesauruses, and use a spell checker if such a tool is enabled in your browser window.

“Compare Answers”

Description

Select a preferred response to questions from Reddit’s r/explainlikeimfive (ELI5) forum.

Instructions

At the top of the screen you will see a question or topic statement. Beneath it there will be two different responses to the question, one on the left and one on the right. Choose the best response of the two by clicking on the left or right text box. Then click the “submit” button on the bottom right to save your selection and move on to the next task.

Requirements:

1. **Time limit** - Please spend at most **5 minutes** on each individual response pair. If necessary, briefly consult the internet to clarify the meaning of words or check the correctness of statements.

Signup link:

https://cmllabel00.umiacs.umd.edu/user/signup/?token=3983036ceb242c83

Compensation

For performing the N paraphrasing tasks and M preference evaluation tasks we will provide dinner and drinks for all volunteers. The three best performing competitors will be awarded a $100 gift card for either Board and Brew or Vigilante (your choice).