**Assignment Objective:**

1. Assess your understanding of Natural Language Processing (NLP) Modeling/Evaluation techniques.
2. Assess your knowledge of the State of the Art (SOTA) NLP models.
3. Assess your strengths in conducting research and executing self-guided development.

**Task: Develop a question rewrite model for the Disfl QA benchmark dataset.**

**Benchmark dataset**: Disfl QA: https://github.com/google-research-datasets/Disfl-QA

**Background**: In any conversational workflow, users input noisy questions due to typing errors/discontinuity in their thinking process, etc. Regardless of input errors, it is important to extract an accurate question out of such inputs, to answer questions correctly and to predict the nature/intent of a given question.

**Task Description:**

1. **Develop a question rewrite model using the train.json dataset:** https://github.com/google-research-datasets/Disfl-QA/blob/main/train.json

   (You may use any data processing/augmentation technique as you see fit.)

2. **Evaluate the model using the valid.json dataset:** https://github.com/google-research-datasets/Disfl-QA/blob/main/dev.json

   (You can use one or more text generation evaluation metrics like accuracy/bleu/gleu/etc.)

3. **Evaluate the model for overfitting and/or overconfidence.**
4. **Document your process (research/development/improvement steps)**

**Submission: (Create and share a GitHub repository)**

1. Code (Model + Evaluation): we should be able to replicate your results.
2. Results (model performance on train and valid datasets)
3. A Report on your research and development process.

**A link to your repo can be emailed to ard_codetest@chata.ai within 2 weeks of receiving the assignment.**

**Judging Criteria:**

1. We will test your model against the given test.json dataset as well as other held-out test (inhouse) datasets.
2. In a subsequent interview process – we will be discussing your submission, and the question/answering would be guided by the solution(s) you would have provided.

We are a research-intensive startup and through this process we are hoping to find individuals who will fit in this environment. We hope to find future teammates and that you will learn something new from this process, regardless of the outcome.

Thank you. We look forward to your submission.