Redaction System:

Introduction:

In today's data-driven world, the protection of sensitive personal information is of paramount importance. To address this critical issue, we have devised a challenge that calls upon your skills and creativity to develop a robust PII redaction system. This system will play a crucial role in safeguarding sensitive information in various conversations of different domains, from healthcare and finance to e-commerce and beyond.

Components:

PII Detection Engine: Create an intelligent PII detection engine capable of identifying and classifying various types of personally identifiable information such as names, addresses, social security numbers, email addresses, and more.

Redaction Algorithm: Design and implement an efficient redaction algorithm that can redact PII information from the input data without altering the document's structure or readability. The redacted information should be replaced with placeholders (e.g., "NAME1" for names and "ADDRESS1" for addresses).

User Interface: Develop a service that allows users to interact with the system. Users should be able to send text and get back the redacted text as well as a map of placeholders and the original redacted text.

Dataset Description

~800 lines(text) from Customer-Agent conversations with potential PCI and PII information like Name, Address, email, Card information, Passport Number etc.

The eval set will contain 200 text lines from a similar distribution.
Link: redaction_train_set

Expected deliverables

1. Finetune a NER model to identify the PII and PCI entities with the data provided. The provided data does not have entity labels, so it is expected that you generate these labels from an already existing bigger model and then use those predictions as the training data for your NER model finetuning.
    a. Finetuning Pipeline
    b. Label creation pipeline

2. Redaction Module: A working algorithm for redacting and keeping a map of placeholder text to original text.
3. A service endpoint which we can interact with via Postman, the payload will include text and the response should have the redacted text as well as the map for un-redacting.
4. An apt eval pipeline for the finetuned model(use a bigger model to obtain labels, use them as ground truth labels, generate predictions from your model, calculate relevant metrics to compare). This pipeline should be able to take in a csv with text lines as input and add redacted text, redaction map and metric columns.

Evaluation Criteria

1. The overall design of the system.
2. ML fine tuning pipeline.
3. ML eval pipeline and metric selection.
4. Logic for redaction and mapping.
5. Implementation of service and other engineering components
6. Final presentation and working demo.