**Citi Bridge – Edition 8 – Hackathon – Problem Statement**

**"Unlocking Efficiency: Document and Email Summarization with Gen AI"**

**Introduction**

"In the fast-paced world of communication, individuals and businesses face the challenge of sifting through overwhelming amounts of textual data contained in documents and emails. This task is not only time-consuming but also hampers productivity. Traditional methods of manual summarization are slow and prone to errors, while existing automated solutions often lack accuracy and coherence.

**Problem statement**

In this hackathon challenge, participants are tasked with developing innovative solutions that leverage the power of AI, particularly Generative AI models, to automatically summarize documents and emails.

**Goal**

The goal is to create efficient and accurate summarization systems that can distill the key information from lengthy texts into concise summaries, enabling users to quickly grasp the main points and make informed decisions.

**Approach**

Participants are encouraged to explore creative approaches that not only generate accurate summaries but also ensure coherence and relevance.

**Scope**

The winning solution will not only streamline the process of information extraction but also enhance productivity and decision-making for individuals and businesses alike."

**Evaluation Criteria**

**Accuracy**: The extent to which the generated summaries accurately capture the key information and main points of the original documents/emails.

**Coherence**: The clarity and logical flow of the generated summaries, ensuring that the information is presented in a coherent manner that is easy to understand.

**Conciseness**: The ability of the summarization model to condense lengthy documents/emails into concise summaries without omitting important information.

**Relevance**: The relevance of the generated summaries to the original documents/emails, ensuring that they focus on the most important and relevant content.

**Language Quality**: The quality of the language used in the summaries, including grammar, syntax, and vocabulary, to ensure readability and fluency.

**Robustness**: The robustness of the summarization model across different types of documents/emails and domains, ensuring consistent performance.

**Scalability**: The ability of the solution to scale effectively to handle large volumes of documents/emails efficiently without compromising performance.

**User Experience:** The overall user experience of interacting with the summarization system, including ease of use, speed, and accessibility.

**Innovation**: The degree of innovation demonstrated in the approach taken to tackle the problem, including the use of advanced techniques or novel algorithms.

**Impact**: The potential impact of the solution on improving productivity, decision-making, or information retrieval processes for individuals or businesses.

**Resources:**

Participants will have access to a below dataset for training and evaluation purposes. Additionally, they may use open-source libraries and frameworks for natural language processing and machine learning to develop their solutions.

Email Dataset – JSON is attached as a zip file to the mail

Document dataset is also attached as a zip file to the mail.

**Prizes:**

Prizes will be awarded to the top-3-performing teams based on the evaluation criteria mentioned above. Participants will also have the opportunity to showcase their solutions to Senior leadership team at Citi.

**Conclusion:**

By participating in this hackathon, you will have the opportunity to contribute to the development of cutting-edge document summarization technologies that can revolutionize the way we extract insights from textual data. Join us as we work together to unlock efficiency and productivity with AI-powered document summarization.