Project Title: (R)oute (R)adius (R)esponse: Geographical Information Query System using Generative Pre-trained Transformers

Team members:

Full name	ASU ID
Khadyothan Choudari Dasari	1229729607
Chandra Kishore Reddy Gurram	1229572515
Saikrishna Reddy Singireddy	1229590234
Kowshik Nandi	1230306157

Problem statement:

Our project aims to address the current lack of user interaction through speech with maps during travel, particularly in tasks like navigation. We seek to enable users to inquire about nearby places, such as hospitals or restaurants, without manual interaction with the system. Our solution involves building a language-driven agent that harnesses the capabilities of Large Language Models (LLM) to dynamically reason and respond to user queries expressed in speech during mobility. This involves leveraging external tools and APIs to enhance the overall user experience in real-time navigation scenarios.

Related work:

In a recent paper published on February 5th, the authors discussed the concept of grounding virtual intelligence in real life, as documented in "Grounding Virtual Intelligence in Real Life"[1]. The paper introduces V-IRL agents, which utilize real-world geospatial information and street view imagery to navigate urban terrains, perform complex tasks, and engage in real-time scenarios. The agents demonstrate capabilities ranging from recommending relevant destinations to evaluating city infrastructure, and even collaboratively providing and following verbal directions. The authors present V-IRL as a flexible platform for researchers to harness extensive global data, enabling the creation and testing of diverse autonomous agents [1]. While the implementation of V-IRL is still in progress, a related study focused on using the Google Maps API to acquire desired data based on user inputs. This implementation was detailed in a Medium article titled "Enhancing User Interaction in Navigation Using Google Maps API" [2]. The article explores how leveraging the Google Maps API can address challenges in real-time interaction with maps during travel, aligning with the goals of our project.

References:

[1] Yang, J., Ding, R., Brown, E., Qi, X., & Xie, S. (2024). V-IRL: Grounding Virtual Intelligence in Real Life. *arXiv preprint arXiv:2402.03310*.

[2] https://medium.com/@yashwantltce/a-simple-way-to-find-places-along-the-route-using-google-maps-api-4237fb452ec2

Initial hypothesis:

Can pre-trained Large Language Models (LLMs) effectively execute domain-specific tasks with precision and relevance after fine-tuning? In real-time scenarios and dynamic environmental changes, can these models dynamically adapt and reason utilizing domain knowledge? We expect initial results to be of high accuracy in responding to user queries, but continual fine-tuning with evolving domain knowledge will be undertaken to enhance performance in response to ongoing environmental changes.

Dataset(s):

In this case, the dataset source is the Google Maps API, which provides real-time data. The number of instances and features are dynamic as they depend on user queries and the information obtained from the API. Since the data is obtained dynamically, there might not be a fixed class distribution or predefined dataset splits. Preprocessing steps would involve real-time processing of the data obtained from the API to make it suitable for input to the Large Language Model (LLM).

Dataset source (link and reference)	Google Maps Places & Directions APIs
Number of instances	Dynamic
Number of features	Dynamic

Class distribution (# instances in each class, if applicable)	Dynamic
Dataset splits	Dynamic
Preprocessing steps	NLP techniques to generate Google Direction’s API request body

Method(s):

1. Speech-to-Text and Text-to-Speech:
    a. Utilize Google's Speech-to-Text API with Transformer and Deep Neural Net models for accurate and natural language conversion.
2. Intent Recognition:
    a. Apply NLP techniques including POS Tagging, NER recognition, and RAKE algorithms to discern user intent from input.
3. End-to-End Model Implementation:
    a. Novelty lies in building an end-to-end model for real-time user interaction with maps, addressing a gap in previous work.
4. Libraries & Tools:
    a. Employ Google Maps Places & Direction APIs, NLTK, PyTorch, scikit-learn, and consider advanced language models like Gemma, Llama, or open-source GPT variants.
5. Comparison to State of the Art:
    a. Evaluate the performance of open-source GPT variants in domain expert scenarios relevant to our project, assessing their ability to comprehend and respond to queries related to maps, navigation, and location-based services.

Evaluation:

In the context of our selected problem, we will employ Word Error Rate (WER) to assess the performance of the Speech-to-Text and Text-to-Speech stages. For intent evaluation, F1 scores will be utilized, ensuring that the Google Maps API receives accurate intent-based queries for subsequent processing by the large language model. The final evaluation of the system's response to user input will leverage users' relevance feedback through reinforcement learning, involving assigning positive or negative scores to outputs and fine-tuning the model to enhance its performance. User experience will be gauged through feedback sessions, concentrating on the system's responsiveness and the natural conversational flow.

Management plan:

For the effective execution of our project, (R)oute (R)adius (R)esponse, we have devised a comprehensive management strategy that ensures clear division of labor, accountability, and seamless communication among team members. The project has been segmented into distinct areas: Intent Recognition, End-to-End Model Implementation, and API Integration. Specific tasks have been allocated to team members according to their skills and interests:

- ***Chandra Kishore Reddy Gurram and Khadyothan Choudari Dasari*** will collaborate on the Intent Recognition component. Their combined expertise in natural language processing and artificial intelligence will be crucial for accurately understanding and processing user intents.

- ***Saikrishna Reddy Singireddy*** will lead the integration of Google APIs, a crucial component for accessing real-time geographical data and interfacing with our system. His role is critical in ensuring that our system has access to accurate and up-to-date information.

- ***Kowshik Nandi*** is responsible for the integration of the LLM API, which involves working closely with large language models to process and respond to user queries. His work will ensure that our system's responses are relevant and contextually appropriate.

To ensure accountability and monitor progress, we will implement a weekly checkpoint system. During these meetings, each member will present their achievements, discuss encountered challenges, and set objectives for the forthcoming week. This framework promotes a supportive and accountable team environment.

This management plan is designed to harness each team member's strengths, promote accountability, and ensure effective communication. By adhering to this structured approach, we are confident in our team's ability to deliver a robust solution that enhances the user interaction with maps through speech during mobility.