Milestone 1: Statement of Work Revised

Title: **Pourfect AI** – Where Every Pour is Perfectly Yours!

Authors: Sydney Levy (sydneylevy@g.harvard.edu), Michelle Hewson (michellehewson@g.harvard.edu), Soline Boussard (sboussar@g.harvard.edu), Niki Ekstrom (nikiekstrom@g.harvard.edu), Aida York (ayork@g.harvard.edu)

Background and Motivation:

- Developments in natural language processing and machine learning have revolutionized modeling unstructured data
- There's a growing interest in applying these technologies to simplify everyday tasks like crafting personalized cocktails.
- Cocktail creation can be complex, requiring attention to numerous small details such as ingredient combinations and different flavor profiles.
- With the rise of large language models (LLMs), the process can be streamlined by generating cocktail recipes that are tailored to individual preferences.
- By leveraging LLMs, we can revolutionize how people create unique cocktails with ease.

Problem Statement:

Many cocktails and drink recipes available today can overwhelm casual drinkers and cocktail enthusiasts. Many users struggle to find cocktail recipes that fit their personal preferences, available ingredients, or dietary restrictions. Additionally, there is no comprehensive and user-friendly app that allows users to search for cocktails based on specific ingredients they already have or want to use, leading to wasted ingredients and missed opportunities to enjoy personalized drinks.

Our goal is to develop a cocktail app that solves these problems by providing users with an intuitive platform where they can:

- Discover new cocktail recipes based on their personal preferences (taste, glass, type of occasion).
- Input ingredients they already have to find matching cocktail recipes, reducing waste and optimizing ingredient use.

- Access detailed recipe instructions, including alcohol and non-alcohol variations for diverse user preferences.
- Save their favorite cocktails and figure out missing ingredients, enhancing their cocktail-making experience.

Target Audience:

- **Casual Drinkers**: People who enjoy occasional cocktails and are looking for easy-to-follow recipes to make drinks at home, especially for social gatherings, parties, or events.
- **Cocktail Enthusiasts**: Individuals with a strong interest in mixology who want to explore a variety of cocktail recipes, experiment with new ingredients, or refine their bartending skills at home.
- **Beginner Bartenders:** Those new to cocktail-making who want a guide to learn the basics, find recipes, and learn to craft drinks for themselves or others.
- **Ingredient-Conscious Drinkers**: Users who want to make cocktails based on what they already have in their kitchen, minimizing ingredient waste.
    - This group might also include those with dietary restrictions (vegan, low-sugar) or who prefer non-alcoholic variations (mocktails).

Scope and Objectives:

1. Develop a database with various different cocktail/mocktail recipes to cater to a large audience
    a. Mocktails
    b. Cocktails
2. Cloud structure and containerization
    a. Create a scalable and containerized cloud structure to host application
3. Model training
    a. Leverage pre-trained models (such as BERT or other hugging face models) to suggest cocktail recipes based on the user's input (desired taste, strength of beverage, type of alcohol, etc.)
4. User interface and visualization
    a. Design a simple user interface that allows hosts and invitees to interact with the recipes/suggestions that the LLM suggests

Resources:

- We require:
    - Databases with Cocktail / Mocktail Recipes:
        - https://www.kaggle.com/datasets/jenlooper/mr-boston-cocktail-dataset
            - 990 rows
            - Includes cocktail name, category, ingredients, and measurements
        - https://www.kaggle.com/datasets/ai-first/cocktail-ingredients
            - 546 rows
            - Includes both alcoholic and non-alcoholic beverages, ingredients / measurements, and ideal glassware for the cocktail
        - https://github.com/stevana/cocktails/blob/master/data/cocktails.yaml

                - 75 rows in yaml files
                - Includes timing of drink (dinner/pre-dinner), ingredients, preparation, taste
            - https://www.thecocktaildb.com/api.php
                -
    - Cloud services
    - Pre-trained BERT models, available through the HuggingFace library

High-Level Project Stages:

- Obtain a database of cocktail and mocktail recipes
- Create a container and utilize cloud infrastructure
- Fine-tune a pre-trained LLM to suggest cocktail recipes given a user's inputted preferences
- Create app / user interface
- Test the interface on users and further fine-tune the model

Limitations and Risks

- Challenges:
    - Data quality: ensuring the completeness and accuracy of data, and ensuring that each dataset can be adjusted to specific event requirements
    - Model accuracy: achieving a high "accuracy" in recommendations when there is no ground truth
    - Integration of different components of the project
- Mitigation strategies:
    - Implement data cleaning, processing and merging various datasets to adjust to our needs
    - Continuously fine tune and adjust models based on user feedback and performance
    - Regularly test with potential users to incorporate feedback

Milestones

- Data collection: [Oct 10, 2024]
- NLP model development: [Oct 18, 2024]
    - Including Milestone 2 Deliverables
- Backend implementation: [Oct 31, 2024]
    - Including Milestone 3 deliverables
- Frontend development: [Nov 14, 2024]
    - Including Milestone 4 deliverables
- Final deployment and deliverables: [Dec 11, 2024]
    - Including Milestone 5 / Final Deliverables