# PROJECT WORK PLAN

## Project Planning and Setup

### Step 1: Define Project Scope and Objectives

- **Deadline**: Day 1 (2 hours)

- **Action**: Write a project work plan document outlining the project goals, scope, deliverables and timelines.

- **Deliverable**: Project Work Plan Document

- **Assigned To**: Zeynep TUTAR (1 hrs) & Mattia VARAGNOLO (1 hr)

### Step 2: Set Up Project Infrastructure

- **Deadline**: Day 1 (1 hour)

- **Action**: Set up a version control system (e.g., GitHub) and final report template.

- **Deliverable**: Project repository and report template.

- **Assigned To**: Zeynep TUTAR (1 hr)

## Data Collection and Preparation

### Step 3: Identify and Collect SQL Dataset

- **Deadline**: Day 2-3 (8 hours)

- **Action**: Research and collect SQL dataset from repositories like GitHub, Kaggle, and other sources.

- **Deliverable**: A SQL dataset stored in a structured format.

- **Assigned To**: Zeynep TUTAR (4 hrs) & Mattia VARAGNOLO (4 hrs)

### Step 4: Data Cleaning and Preprocessing

- **Deadline**: Day 4-6 (10 hours)

- **Action**: Clean and preprocess the datasets to ensure consistency and quality. This includes removing duplicates, handling missing values, and standardizing SQL syntax.

- **Deliverable**: Cleaned and preprocessed SQL datasets.

- **Assigned To**: Zeynep TUTAR (2 hrs) & Mattia VARAGNOLO (2 hrs)

## Model Selection and Baseline Creation

### Step 5: Select Pre-trained LLM

- **Deadline**: Day 7 (4 hours)

- **Action**: Choose suitable pre-trained LLM that will be fine-tuned for SQL understanding.

- **Deliverable**: Selected pre-trained LLM model.

- **Assigned To**: Zeynep TUTAR (4 hrs)

### Step 6: Establish Baseline Performance

- **Deadline**: Day 8 (6 hours)

- **Action**: Test the selected LLM on a few SQL-related tasks to establish a baseline performance.

- **Deliverable**: Baseline performance outputs.

- **Assigned To**: Mattia VARAGNOLO (4 hrs) & Zeynep TUTAR (2 hrs)

## Fine-tuning the LLMs

### Step 7: Create a Pipeline

- **Deadline**: Day 7 (6 hours)

- **Action**: Create the training pipeline and the data manipulation

- **Deliverable**: Training and evaluation code

- **Assigned To**: Mattia VARAGNOLO (6 hrs)

### Step 8: Fine-tune LLM with SQL Dataset

- **Deadline**: Day 9-14 (24 hours)

- **Action**: Fine-tune the selected LLM using the cleaned and preprocessed SQL dataset. This involves training the model on SQL syntax, query structures, and database interaction patterns.

- **Deliverable**: Fine-tuned LLMs.

- **Assigned To**: Zeynep TUTAR (12 hrs) & Mattia VARAGNOLO (12 hrs)

### Step 9: Validate Fine-tuned Model

- **Deadline**: Day 15-17 (14 hours)

- **Action**: Validate the performance of the fine-tuned models using a separate validation set to ensure the models are learning correctly.

- **Deliverable**: Validation performance outputs.

- **Assigned To**: Mattia VARAGNOLO (7 hrs) & Zeynep TUTAR (7 hrs)

## Testing and Evaluation

### Step 10: Test Model Performance

- **Deadline**: Day 18-19 (8 hours)

- **Action**: Conduct testing of the fine-tuned model to ensure it correctly interprets natural language inputs and generates appropriate SQL queries.

- **Deliverable**: Test performance outputs with identified issues and fixes.

- **Assigned To**: Zeynep TUTAR (8 hrs)

### Step 11: Evaluate Model Performance

- **Deadline**: Day 20-21 (10 hours)

- **Action**: Evaluate the overall performance of the fine-tuned models using metrics like accuracy, query completion rate, response coherence, and user satisfaction.

- **Deliverable**: Overall model evaluation.

- **Assigned To**: Mattia VARAGNOLO (10 hrs)

## Documentation

### Step 12: Document the Project

- **Deadline**: Day 22-23 (10 hours)

- **Action**: Prepare comprehensive documentation covering all aspects of the project, including data collection, model fine-tuning, testing, and evaluation. Also prepare the powerpoint presentation.

- **Deliverable**: Project documentation and presentation.

- **Assigned To**: Zeynep TUTAR (8 hrs) & Mattia VARAGNOLO (2 hrs)

### Total Project Hours

- Zeynep TUTAR: 49 hours

- Mattia VARAGNOLO: 48 hours