**AI Project Methodology**

**Functional Part:**

1. Use Case: In automobile industry
2. Goal: To understand the level of CO2 emission from a vehicle(car).
3. Customers/Users: Users will be R&D Engineers in automobile industry
4. Dataset: <https://www.kaggle.com/code/mruanova/cars-co2-emissions-multiple-regression-w3schools/data> from Kaggle
5. Team organization: Raphael Abba Ekwo & Surya Teja Sista.

**Technical Part:**

**Part-1:** We have built an industrial ready project based on our chosen dataset.

1. Git: Collaborated our work using Git.
2. Models: Altogether we have 4 models
   1. **Data Preparation**: Our dataset contains 5 columns; one is the prediction column, and the rest are the data columns that are required for the prediction. Out of those 4 columns we removed 2 of them which is the cars column and the model column. We have no null values in our dataset.
   2. **Models** **training**: We have trained our model with 80% of the data.
   3. **Prediction**: Prediction is done on the 20% of the data and with customed data as well.
   4. **Metrics Evaluation**: We have evaluated the metrics such as MAE, RMSE, R2 with the actual and predicted values.

1. Environment: We have used conda environment (AI\_with\_mlflow)
2. Template: We have used a cookie cutter template to structure our project.
3. Documentation Library: We have used Sphinx documentation library.

**Part-2:** Integration of ML flow library to our project.

1. We have installed ML flow library in our python environment
2. The results of parameters and metrics are tracked and displayed in put local ML flow UI
3. Our code has been packaged in a reusable and reproducible format with ML flow projects.

**Part-3:** Integration of SHAP library to our project.

1. Installed necessary shap library.
2. We have built an Explainer and computed the shaplay values.
3. We have displayed the visual explanations for a specific point of our dataset.
4. We have displayed the visual explanations for all data points at once.
5. We have displayed the summary plot on whole dataset for each class.