

Read the document carefully and follow as instructed. The behavioral survey (Link at the end of the document) is to be filled once the task is completed.

Machine Learning Roles (Chat GPT or LLM for Infrastructure Projects Database)

About the Task:

Important Points:

- (1) Please use separate .py files to harvest ONLY the data from API or scrap from websites.
- (2) Firstly, collect data from diverse and large datasets listed below.
- (3) Use Large Language Models and pre-trained ChatGPT4 or Jurassic and any others to do Q&A, Tagging for Keyword Extraction or Named Entity Recognition (NER) but integrating and customizing the results with some of the sources listed below.
- (4) Provide a report with succinct visualization or a simple website with custom Q&A bot with results and all your different .py scripts (class object oriented good scripting practices) and final Python notebook.

Please submit a **PDF of google slides** or a **document** presenting your findings. Upload the PDF onto your Google Drive and share the link as follows:

Google drive link with general access set as 'Anyone with a link' and role set as 'Editor'

Share the link in the Behavioral Survey.

Your evaluation criteria is partially technical and partially the ability to explain meaningful results in a presentable manner.

Natural Language Understanding Trial Task

- Data Sources: Combine different data sources from the below list to (i) get data and (ii) use Python harvester to scrape or download data from the sources. Evaluation will also be based on the diversity of the data chosen.

World Bank Projects
SAM.gov
Open Contracting
Tenders Electronic Daily
Open Tenders
GI Hub Pipelines
California Tenders
Florida Tenders
Texas Tenders
FDOT
TxDOT
Caltran Tenders

Asian Development Bank (ADB) Projects
African Development Bank (AfDB) Projects
Asian Infrastructure Investment Bank (AIIB) Projects

- Learn more about Projects & Tenders standard data structure
 - <https://developer.taiyo.ai/api-doc/ProjectsandTenders/>
 - <https://www.open-contracting.org/data-standard/>

Modeling and Report

1. **Customize a GPT-3 Chatbot.** A chatbot for Q&A for custom dataset
<https://openai.com/blog/introducing-chatgpt-and-whisper-apis>
<https://www.ai21.com/blog/announcing-ai21-studio-and-jurassic-1>

More explanation on prompt engineering question: Find data sources for road projects and tenders in the state of California. Prompt Engineering and Intelligent Question and Answering customized to the datasets mentioned above to answer questions such as How many projects (or tenders) in region X? Which company is winning most XXX (Sector like Transport or Road) project in YYY (Place like California or USA).

Evaluation is based on the following parameters:

Stages	Evaluation Criteria
1	Extensiveness of the dataset and understanding of projects and tenders data structure
2	Modular, DRY Code
3	Config Params, Unit Tests & Logging Standards
4	Data Visualization presentation of results and understanding of the problem statement
5	Explainability, the ability to experiment, do creative relevant problem solving, and structure your analytical thinking with documentation

[Behavioral Survey](#)