

DEEP LOGIC AI- Coding Assignment: AI Engineer

Intern Summer 2024

Part 1: Document Conversion, OCR, and Preprocessing

Task: I was told to make a model that can convert the content in a pdf document to get the contents inside the document. Then I was instructed to preprocess the contents retrieved from the document.

Implementation:

- As the first step, I installed and imported the libraries which are necessary for implementation like PyPDF2, re, nltk, pytesseract,...
- Then I made my model to read the contents of the pdf document by analysing the text of pdf and images of the pdf.
- I used the 'pytesseract' for OCR to the document.
- After that, I got information through text and information through images. I compared the length of both variables and took the better content variable.
- Then I pre-processed the data by removing extra spaces and removing unwanted symbols.
- I used the `sent_tokenize` from nltk to divide the data obtained into sentences and then used the `word_tokenize` to divide the sentences into words.

Part 2: LLM-Powered Understanding and Actions

Task: After the implementation of task 1, I was instructed to integrate a LLM to extract information from the data which was word_tokenized and translate the data generated/extracted into required language.

Implementation:

- As done in the first task, I installed the transformer module and imported the pipeline.
- I used the bert llm for this Information extraction and classification. By defining the model pipeline, I used the llm for text-generation and zero-shot-classification.

- I was not able to use bert for translation, so I used marian- a light weighted model only for translation and translated the content generated.

Conclusion:

Firstly, I would like to thank Deep Logic AI for accepting my application and giving me a chance to work on such an interesting task. Due lack of time and the busy schedule of the academic as my exams are scheduled next month, I was only able to work on the assignment for 1 day(Sunday). Though my exams are in the next month, I am willing to dedicate the time specified by the company if I am told to work from next month and do my best and fulfil my duties. If I could have given more time, I would have improved the results of the model. I hope you understand my situation and evaluate accordingly.