Group Name – **Nerdsbutstupid**

MEMBERS – Arun Prasad T D

Eshwanth Karti T R

Nithish Ariyha K

Shruthika R

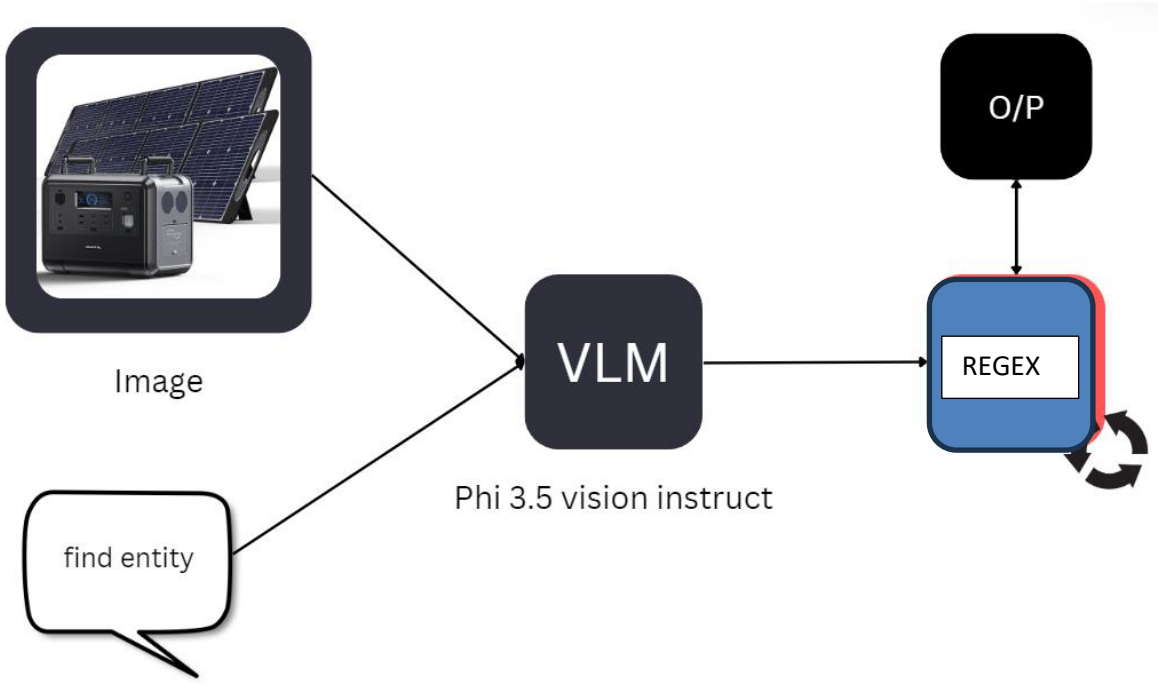
Problem Statement

To make a pipeline which can extract the key findings from an image given the required value. This process is also known as Named Entity Recognition.

Approach to tackle this problem

After many trials and error on how to tackle this problem with a lot of consistencies like corrupt image files and duplicate images. We decided that Using Phi-3.5-vision-instruct tried *finetuning* this model but it leads to a model with degraded capability so decided to go with a multi modal approach of getting our described interest without changing param for VLM.

Phi 3.5 Vision Instruct is a multimodal model that integrates both vision and language capabilities, enabling it to process and understand inputs in the form of both images and text. This model is designed to follow detailed instructions related to visual content, making it ideal for tasks such as visual question answering, image captioning, and object detection. By combining the ability to interpret images with natural language processing, Phi 3.5 Vision Instruct allows for interactive AI applications where users can provide both images and textual queries or commands. This capability is useful for us in named entity recognition.



Regular expressions (regex) are patterns used for matching specific sequences of characters within text, making them valuable for tasks like searching, extracting, or replacing text. In the context of Named Entity Recognition (NER), regex can be particularly useful for identifying entities that follow consistent patterns, such as dates, emails, or specific symbols. In our case we used regex on the VLM's output giving us the required output.

Timeline Summarization