MoM (Review Meeting 2)

Date and Time: 14th February 2024, 8 pm to 9.30PM IST.

**Meeting Attendees**:

* David de Hilster, Hugo Watanuki, Dr Shobha G, Dr Jyothi Shetty, Dr. Vishalakshi Prabhu, Nihar Mandahas, Skanda P R, Pratheek Rao, Manvith L B, Arya Hariharan, Eshan Mathur, Prashant Ronad, Nikhil Vasu.

**Agenda 1**: Present the findings and work done in the past few weeks related to the Legal Assistance Tool and understand future steps.

*Discussion/Demonstration Done:*

1. Legal NER model’s working was demonstrated, and source of the model and dataset was discussed.
2. Searching and sorting of relevant legal documents was demonstrated using ECL queries.

*Recommendations*:

1. The team should validate the model and the dataset and fully understand the behavior, working, and downfalls of the model and dataset.
2. A means has to be found to regularly update the legal corpus to ensure its up to date with new cases.
3. Timeline must be updated and followed.
4. Discussion with project mentors should be done before proceeding with implementation of a certain idea/feature.

*Actions to be Taken:*

1. Update the timeline.
2. Validate the model and dataset and clearly outline its behavior, working and drawbacks/downfalls.
3. To create a git repository and share it with mentors.
4. To record the changes from the original proposal along with the justification for the changes

**Agenda 2**: Present the dataset related to “Building NLP pipeline for Electronic Health Records” and understand the future steps.

*Discussion/Demonstration Done:*

1. There was problem in finding valid data for the project. The one that’s collected from hospital is presented. To begin with, we have 200 patient’s brain EHR.
2. Alzheimer 80,000 image in JPG format which falls into 4 categories is discussed.
3. Compression of brain MRI from DICOM to JPG is very challenging for real time data.

*Recommendations:*

1. Hugo said that there are two final solutions, one based on NLP for brain electronic records and other based on CNN for brain MRI.
2. David said he needs some time to go through medical terms.
3. Suggested to create GitHub link for discussion and sample solutions.
4. Hugo suggested to start with smaller dataset on PC initially and then spray on HPCC (after experiment). They also said different cluster for this project.
5. First validate Proof of Concept and then scale.

*Actions to be Taken:*

1. Stick to the timeline.