

## LAIRHUB – Lab Meeting

March 07, 2025

**Participants:** Sunny, Paris, Aravind, Divya, Nibras

### Action Items

Project	Project Associate	Updates / Action Items	Resource/ Notes
ETTA News	Ruochen/ Nibras	<p>Workflow for Document Representation and Clustering</p> <ol style="list-style-type: none"><li>1. Preprocessing the news Corpus (after ETL)<ul style="list-style-type: none"><li>- Collect documents (title + abstracts).</li><li>- Remove stop words from the text.</li><li>- One challenge might occur in deciding the vector length, we have specific data-type to manage the scenario.</li></ul></li><li>2. Develop a document representation table</li><li>3. Feature Extraction: TF-IDF Weighting<ul style="list-style-type: none"><li>- Compute Term Frequency-Inverse Document Frequency (TF-IDF) for each term in the corpus.</li><li>- Store TF, IDF, and TF * IDF values in the document representation table.</li></ul></li><li>4. Defining a Threshold for Term Selection<ul style="list-style-type: none"><li>- Set up a threshold to differentiate between low-weighted, mid-weighted and high-weighted terms.</li><li>- We will conduct an EDA to decide the threshold.</li></ul></li><li>5. Query Handling and Alternative Workflow Decision<ul style="list-style-type: none"><li>- If a query term exists in high-weighted words, proceed with clustering.</li><li>- If a query term does not exist in high-weighted words, consider an alternative workflow.</li></ul></li><li>6. Our final goal is to use document representation table for retrieving results, clustering and using GA.</li></ol>	

### ETL & Document Distribution Analysis - Compute and demonstrate

- Our current ETL process
- ER diagram
- Average document length
- Maximum document length
- Minimum document length
- Distribution of document length
- top 10 (varies) high frequency words
- top 10 (varies) low frequency words
- In addition, we will find out whether our corpus follows Zipf's law.
  
- Generate following analysis - **(Later)**
  - Average number of tokens per clusters
  - Get the best tokens and then send them to llama (text preprocessing)
  - Per cluster - tariff & cluster
    - How long does it take to generate
    - How many tokens did I send
      - Show the tokens we are sending to understand quality the labels that are coming back
      - RQ: These tokens we are sending, many of the labels are not coming back as cluster labels. Later we will find out why it's happening.
    - Use stop word list
    - We can perform higher level of filtering
      - TF-IDF (next steps)
    - Remove the query terms (tariffs & cancer) from the llama response
      - We can drop it and ask for six labels from llama
    - Perform sanity check?
      - Remove the other outputs from llama

		<ul style="list-style-type: none"> <li>■ We will focus on time and number of tokens (after using stop word list or TFIDF)</li> <li>● Make short videos on <ul style="list-style-type: none"> <li>○ Incorporated features.</li> <li>○ How do you perform a basic search with connections?</li> <li>○ Zoom-in and zoom-out features.</li> <li>○ Cluster size definitions and color.</li> <li>○ How to get help.</li> <li>○ How we constructed the database.</li> </ul> </li> <li>● MISC: <ul style="list-style-type: none"> <li>○ Attach askit4.info with CRM.</li> </ul> </li> </ul>	
<b>ezee.chat</b>	Sunny/Nibras	<ul style="list-style-type: none"> <li>● User workflow</li> <li>● Paper outline</li> <li>● Identify the procedures for publishing the bot (as a research and private entity) on Meta's WhatsApp AI platform.</li> <li>● Develop a detailed workplan <ul style="list-style-type: none"> <li>○ Use cases for newly-admitted students</li> <li>○ MDL based WhatsApp bot</li> <li>○ Technology &amp; Research project</li> <li>○ Publish &amp; present</li> <li>○ Translation - investment &amp; funding</li> </ul> </li> <li>● Business number for ezee.chat</li> <li>● Index for dialogue</li> </ul>	<b>Existing Educational Bots</b> <a href="https://www.comm100.com/engage/chatbots-higher-ed/">https://www.comm100.com/engage/chatbots-higher-ed/</a>
<b>Grant Submission</b>	Sunny	<ul style="list-style-type: none"> <li>● Grant Proposal Document Society - OVPI Horizon Europe</li> <li>Keep tracking-NRFK (National Research Foundation of Korea)</li> </ul>	
<b>showme.news</b>	Aravind	<ul style="list-style-type: none"> <li>● <b>Meeting set on 14 March, 2025 with Tobii.</b></li> <li>● will test the calibration mode of the trial version of Tobi</li> <li>● will assess the precision of Tobi and pygaze</li> <li>● will be setting up a meeting soon with a Tobi representative in Toronto (we can do this via Teams or Zoom). Please work with Nadia to set up the meeting.</li> <li>● Share the paper on comparing a large number of current GA</li> </ul>	

		<p>platforms/tools (share with Professor).</p> <p>The question we have to figure out:</p> <ul style="list-style-type: none"> <li>• If we purchase only Tobii Pro Spark (and DO NOT purchase Tobii Pro Lab) would we be able to use Tobii Pro Spark and perhaps other development environments (e.g., Tobii Pro Sdk, PyGaze, etc) to ensure our Etta Gaze will work well with minimal calibration complexity (some amount of calibration of course cannot be avoided) and high precision?</li> <li>• If we can use Tobii Pro Spark only to move us forward, Professor recommends purchasing it soon.</li> </ul>	
<p><b>Information Addiction</b></p> <p><b>Wellbeing &amp; Thriving</b></p>	Paris	<ul style="list-style-type: none"> <li>• Wait for the feedback for ARIST</li> </ul>	
LookupMed	Paris/ Divya	<ul style="list-style-type: none"> <li>• <b>Check the recordings for details (7 March, 2025)</b></li> <li>• Followings are the broader concepts <ul style="list-style-type: none"> <li>○ Gerontology</li> <li>○ Medication compliance</li> <li>○ Interventional methods   implementation</li> </ul> </li> </ul>	<p>Documentation URL:</p> <p><a href="https://support.google.com/googleplay/android-developer/answer/13628312?hl=en">https://support.google.com/googleplay/android-developer/answer/13628312?hl=en</a></p> <p><a href="https://support.google.com/googleplay/android-developer/answer/10841920#zippy=%2Corganization-account">https://support.google.com/googleplay/android-developer/answer/10841920#zippy=%2Corganization-account</a></p>