1. Need to divide in roles: people for data collection read those papers, so on.
2. Look for existing similar research and models, in papers, on huggingfaces, etc
3. Start collecting urdu dataset resources

* Existing dataset in roman and native urdu give information on length, nature, data acquisition pipeline
* Collect sufficient to data to start with other processes (tokenization etc)
* Can keep augmenting data with richer resources
* Data preprocessing, filtering will be a cyclic continuous process
* I lost him
* Hum paper bhi likh sakte hain(?)

For Naqvi: Data collection work on those who are a bit inexperienced in DL.

Modular approach.  
  
Focus on why, why is this important, why we need to do this work. (need literature review for that)

For proposal, we first need literature review to justify the WHY part. WE need to identify stake holders. And we need to identify the need.