Your Assignment- due 30aug25

- Take POS tag data from NLTK https://www.nltk.org/
- Use HMM, EnCo-DeCo and any LLM of your choice to compare performance
- You can use pre-written/available/LLM-geneated code, but you will have to explain what the code is doing
- Later you will be asked to implement something innovative on/based-on POS tagging which will require you to code

Assignment discussion

Template for

You have to strictly follow this format

Define POS tagging

- Input (caution: you cannot give input inside the code)
- Output

Data downloading and cleaning

- How much of data did you use
- From what source
- Did you use any cleaning? If so, what and how and why
- Did you use the POS tags as such or modify

Our recommendation

- Use 12 universal cross lingual tags (yes 12 only)
- Run
  - *import nltk*
  - *nltk.download('brown')*
  - *from nltk.corpus import brown*
  - *print(brown.tagged_sents(tagset='universal')[0])*
- This returns Brown corpus sentences mapped to universal tags

HMM Based POS tagging

- What did you read for this part of the assignment
- Why is Viterbi linear time?
- Is your program running?
- If yes, give the demo

EnCo-DeCo based

- What did you read for this part of the assignment

- What algo does the decoding phase use?
- Is your program running?
- If yes, give the demo

LLM based

- What did you read for this part of the assignment
- Which LLM did you use?
- Is your code running?
- If yes, give the demo

Compare and contrast

- Give a tabular comparison of Precision (P), Recall (R) and F1 score
- Analyse and explain the observations

Per POS accuracy

- For each of the 12 tags, compare and contrast as in the previous slide