Final Project

Step 1: Choose an NLP problem, not restricted to the following:

- reconstruction (auto-correct)
- document classification
    - plagiarism detection
    - author identification
    - sentiment analysis
    - ...
- token classification
    - part-of-speech tagging
    - named entity recognition
    - word sense disambiguation
    - ...
- language modeling (auto-complete)
- machine translation
- ...

Step 2: Identify or invent a solution based on a language model. Describe the model in detail.

Some aspect of the problem and/or solution should be novel: either a new problem, an (ever so slightly) new solution, or an existing solution applied to an existing problem to which is has not been applied before.

Step 3a: *[for probabilistic/generative models only]* Train and apply your solution to synthetic data that aligns with the assumptions made by the model.

Evaluate the results quantitatively and qualitatively: highlight examples where the model performs well and poorly. Any unusual/unexpected results require explanation (and frankly, probably debugging).

Step 3b: Train and apply your solution to "real" data acquired legally.

Evaluate the results quantitatively and qualitatively: highlight examples where the model performs well and poorly.

Step 4: Discuss pros and cons of the model. Consider:

- quality/correctness
- data, time, and computational requirements
- interpretability
- ...

Projects that do not strictly follow this format are possible with approval.

Timeline

- Oct 29: project proposal due (10%)

    One paragraph outlining the chosen problem and model.

- Nov 24: project report draft due (60%)

  It should be complete.

- Dec 8: final project report due (30%)

  Incorporate feedback from draft.

Notes

Your descriptions of methods should be sufficient for anyone who has taken this course, e.g. your classmates. Note that this means you do not need to repeat description of methods presented in class.

You may work in groups of up to 3. Please submit your report in PDF form along with your code as a ZIP file. Include with your code a README file if it requires any special setup, e.g. extra Python packages or datasets.

Prompt engineering is out-of-scope for this course and project.

Please include a table of all results. Please do not include screenshots of terminal output - copy the text into the document.