MSAI-337, Spring 2025
Final Project: Proposal
Due Date: Monday, May 5th @ 11:59PM
Total Points: 5.0

Description: For the Final Project, your group will implement a deep learning natural language processing project of your choice. The project can explore almost any natural language processing task using deep learning. You are responsible for selecting the topic, dataset(s) and language model architecture. You may use a pre-trained language model or can train the language model from scratch. If you elect to use a pre-trained language model, you must, at minimum, have access to its output logits so that you can customize it to your task. You are expected to implement your own code or substantially modify an existing codebase. All projects need to be approved. I am available to meet individually with each group to review the project scope and offer guidance.

In general, simple, well-executed projects are preferred to projects with expansive scopes that may be difficult to complete in a class project. Replicating existing research or applied projects is acceptable, with the understanding that your group project should have some novel component. The proposal should be self-contained and not have any links to external papers, resources, datasets, etc.

Assignment: The project proposals should be a single PDF file about one-page in length and cover the following four points:

- (0.5 points) What task will you address, and why is it interesting? This can be as simple as a couple of sentences.
- (1.0 points) How will you acquire your data? This element is intended to serve as a check that your project is doable -- so if you're inventing a new data set, be as specific as possible here.
- (1.0 points) Which features/attributes will you use for your task?
- (2.5 points) What will your initial approach be? What data pre-processing will you do, which language model architecture will you use, and how will you evaluate your success (Note: you **must** use a quantitative metric)? Think about how you will organize your model outputs to calculate these metrics. You **must** also include a naïve baseline or one constructed from another machine learning technique.

What to Turn In: Submit your PDF file to Canvas. Note: You will likely invest a significant amount of time on this project, so think of something that will be fun, be of interest to the group collectively, and serve as a topic of discussion during internship/job interviews.

Good luck, and have fun!