

6.806 / 6.864 Project Guidelines

The goal of the class project is to give you hands-on experience with the complete NLP modeling pipeline, from data preparation to analysis of results. Anything that involves language data and machine learning is fair game! A minimal option is to take an existing model from the literature and run it on a standard or newly collected dataset; however, exploring new modeling ideas are enthusiastically recommended. If you're working with an existing model, we'd ask that you (1) reimplement it rather than using existing code (pre-built neural net components for RNNs and transformers are fine—use your judgment); and (2) do a little bit of analysis beyond “accuracy x on dataset y ” (which predictions are especially accurate / inaccurate? what do learned representations capture? etc.). **This is a class about language, and project proposals that involve running an “NLP-type model” on data with no human language component (e.g. nucleotide sequences or astrophysical time series) will not be accepted.**

Projects count for 50% of your grade, and must be done in groups of 2–4. We will not allow individual projects, and will allow groups larger than 4 only by special request. Midway through the semester, we'll post a special Piazza poll to match students to project groups. Proposals and status reports can be typeset however you'd like; for final projects, please use the ACL style files (<https://2021.aclweb.org/downloads/acl-ijcnlp2021-templates.zip>).

The **minimum** length for your project writeup is

$4 \text{ pages} + (1 \text{ page}) * (\# \text{ of group members in 6.806}) + (2 \text{ pages}) * (\# \text{ of group members in 6.864})$

The **maximum** length for your project writeup is 2 pages more than the minimum.

In the course of working on your project, please submit the following components on the dates provided (we'll provide feedback as quickly as possible):

1 Proposal [DUE 15 APRIL]

A 1- or 2- paragraph description of your proposed project. This should include

- What problem you're planning to solve
- How you're planning to evaluate / analyze your solution
- What dataset you're going to use

- What computing resources you need to train your model

Please make sure that you have access to the data and compute you need *before* submitting this proposal! For example, you're probably not going to do any large-scale language model pretraining experiments on your laptop, and you're not going to get access to protected medical records in the next month if you don't have it already.

2 Status report [DUE 6 May]

A 1–2 page summary of project progress so far. What experiments have you already run? How long is model training taking? What information have you learned that you didn't anticipate in your initial project proposal? What still needs to happen for the final report submission?

3 Writeup [DUE 20 MAY]

A complete writeup of the project, formatted as an ACL-style paper. Include references to relevant literature; a complete description of your prediction problem, model, and dataset; and any relevant experimental results and analysis (including figures and tables as relevant).

4 Presentation [18 and 20 May]

We'll be holding a **virtual poster session**. Prepare a *single slide* summarizing your approach and main results. You'll be assigned one day as a presenter and one day as a spectator. As presenter, your group will stand in an assigned location and describe your project to spectators; as spectator, you can walk around and ask your classmates about projects. We'll send around instructions (with assignments to days and locations) the week before presentations.

5 Grading

- 10% project proposal (full points as long as you submit something)
- 15% status report (full points if you include (1) a description of one experiment, even if you don't have final results, and (2) a description of what you're planning to do between now and the final presentation)
- 25% presentation (full points as long as you present at your assigned time with a poster)
- 50% project writeup (assessed for substance, soundness of experimental evaluation, clarity of presentation; *not assessed for* beating the state-of-the-art on your task of choice)