CS 839: Project Information

Instructions: This document contains some instructions and useful information for the class project for the Fall '25 offering of CS 839: Foundation Models.

The goal for the final project is to use what you've learned in class (or in the presentations) to study, extend, or solve new problems in our class area. This could mean one of several things:

- Extending the results from an existing work to new settings, or crafting a new approach to making it work for a case that was not tackled before (caveat: compute limitations might limit some of the possibilities along these lines),
- Proposing a new approach or a variation or combination of one or more techniques we have used and trying it out, reporting and interpreting the results (positive or negative),
- A principled comparison of multiple existing approaches, potentially with some discussion of the implications,
- Other ideas: you can suggest additional possibilities, but do clear them with me.

Our expectation is that the projects have something to do with our course: there should be a setting that involves the use of foundation models or pretrained models. We expect your project to be at a reasonably high level; you might want to make it the initial backbone that will become a future paper.

Note: if you select a project that is more ambitious than the 1 or 1.5 month timeline for our class, that is all right! We will accept (without penalty) project reports that include not-yet completed components (but have expected or partial results discussed) in cases where such projects are broader than the timing or resources permit.

Deliverables: There are two items that must be turned in.

First, a project proposal/pitch: a one-page pitch that describes what you plan to do, due **Nov. 13**. The one page pitch must include

- The overarching idea you're exploring and motivation: why is this an interesting question and why are you interested in it?
- Your basic hypothesis that you are evaluating,
- Methodology: what tools, datasets, models, and infrastructure are required to do what you propose to do, and how you plan to tackle each of these,

Second, the final report: a report of up to five pages. The report should include:

- An introduction that explains your basic setting, motivation, and hypothesis,
- Related work that you have explored,
- A methodological (theoretical or experimental) development that helps confirm or reject your hypothesis,
- Conclusions and future work.
- Conclusions and future work.

The due date is **Dec 9th**. The grading rubric simply reflects (i) having each of the above items for the final report included, and (ii) meeting a minimum standard for each of them. If in doubt about whether what you're doing meets the bar, ask me!

Team Sizes Your group sizes should be 3-6 (but there is some flexibility). You're welcome to have the same group as for your presentations.

1 Frequently Asked Questions

- Must the project only use techniques that were taught in the class?
  A: No, you may use other techniques if they relate to foundation models. If there is any question about the fit, definitely ask me!
- Can I combine the project with other class projects?
  A: Yes, if the project is done this fall (that is, you cannot submit an old project). The project should also still be relevant.
- Do I need to run empirical experiments?
  A: No, you can do an entirely analytical project. If you aren't certain what's involved, ask me.
- If I do experiments, do I need to use real datasets?
  A: No, you can use entirely synthetic data if you need to.
- Can I use pretrained models, libraries, open-source repositories, frameworks, etc?
  A: Yes, you may use anything, but you must clearly explain which part of the project involves your work and which is using existing implementations. Everything used should be cited. If your project is predominantly empirical, we expect you to do some implementation work. That is, you should work on some aspect of the pipeline. Just re-running an existing paper's code and models without any new component is not sufficient.
- Will there be computing resources for training on GPUs?
  A: There is a cluster that you can obtain access to, but it is likely to be very busy (this is shared with all other classes). If you need to do fine-tuning or training, Google Colab and similar free resources might come in handy as a first step.