

# NLP4CSS: Final Project

*Instructor: Anjalie Field; TAs: Miriam Wanner, Kuleen Sasse, Rohan Allen*

---

**Guidelines.** This assignment is to be completed in groups of **2-4** students. Be sure to comply with course policies on the course website.

**Submission.** This homework has two written components. You will submit them as PDF files on gradescope. For both components, use the ACL style files available here <https://github.com/acl-org/acl-style-files>.

## 1 Introduction

Your final project can be on any topic related to the course material. Most projects will take the form of a small research paper, describing your methods and findings, but feel free to propose an alternate format. The specific requirements are:

- The project must be relevant to the course
- The project should be completed in groups, ideally of 2-4 students
- The project must incorporate a method from the second half of the semester (e.g., neural topic models, masked-LLM usage / metaphor detection, LLMs for social simulations, etc.). The method does not need to be central to the project, for example, it could be used to report preliminary data statistics. It does not need to be included in the proposal.
- There are two concrete components: a written research proposal and a final paper.

We suggest conducting an analysis of a text corpus, but you may chose a different approach. We have compiled a list of datasets available for your use here. You are welcome to use a different data set. However, given the short time-frame of the assignment, we discourage trying to collect new data, and you are expected to have your data on-hand by the proposal deadline.

## 2 Part 1: Proposal (April 7)

Write a short proposal outlining your project and convincing us why it's worth doing. The type of content you should include are:

- Your research questions, hypotheses, or the problem you are aiming to solve
- The implications of your proposed topic and why it is worth studying
- The data, model, and metrics you intend to use
- Some summary statistics of the data you are intending to work with (this is to ensure you were able to obtain the data)

Your proposal should be 1-2 pages in ACL format. If you have figures or tables that exceed the page limit, you may include them in an Appendix. There are two primary goals of the proposal. The first is to ensure the feasibility of your intended topic, such as ensuring you are able to obtain the data you need. The second is to allow course instructors to provide feedback on your proposed topic, such as the relevance to the course and if we expect you to be able to obtain results. The proposal will be 5% of your course grade.

### 3 Part 2: Final Report (May 7)

Most final papers will be in the form of an ACL-style paper of 4-8 pages and adhering to ACL guidelines concerning references, layout, supplementary materials, and so forth. Some sections you might consider including are:

- Introduction
- Related Work
- Methodology
- Results
- Discussion

These section headers are not mandatory, and you are welcome to use whatever structure best presents your work. **Include a brief authorship statement at the end of your paper**, explaining how the individual authors contributed to the project. This statement is required even for singly-authored papers, because we want to know whether your project is a collaboration with people outside of the class.

The final report will be 20% of your course grade. Please note that you cannot use late days for the final report. You will also submit the codebase. If you use existing packages or open source code, make sure to cite it.