

For the final project, you will design and conduct a multi-week project on an NLP topic of your choice.

You are required to typeset your midway report and final report in the Association for Computational Linguistics (ACL) paper format. You are highly encouraged to use LaTeX to do this, as essentially every NLP project published in the field is typeset with LaTeX. You are not required to use it, but it generally makes typing equations and formatting figures and tables much easier.

You can use this linked template to typeset your materials.¹ If you strongly prefer to use Word, follow this link to the official ACL Word template.²

There are two types of final project you may do:

1. **Empirical project.** You will implement an NLP method of your choice. You may either try to replicate an existing paper's results by reimplementing their method, or you may perform some original experiment(s) to get novel results.
2. **Survey.** You will write a paper analyzing and synthesizing a selection of recent literature on an NLP topic of your choice. *Only select this option if you are comfortable reading research papers.*³

Final Project Proposal (5 points; due Mar. 31)

First, you will come up with a project idea. The project proposal should be a 1–2-page paper (11–12 point text, but otherwise in any format) containing each of the following:

- State explicitly whether you plan to do an empirical project or a survey.
- If you are doing an empirical project:
 - State the specific NLP problem or task you will be working on. Why is it interesting or important? Formally define the task; you may find math or figures helpful for describing it.
 - An outline of your project. What **datasets**, pre-existing **models** (if applicable), and **software** you will use? What methods will you implement?
 - An experimental design. How will you know if you have effectively solved your problem? You should clearly identify your **evaluation** metrics. Also describe at least one simple and easy-to-implement baseline you could compare against, like a pre-trained LM with no fine-tuning, an existing statistical method, or a very simple classifier.
- If you are doing a survey:
 - Identify a coherent research area. This could be organized around a specific NLP task like question answering, a technical challenge like retrieval-augmented generation, or a coherent set of methods for solving NLP problems. This should be an area that has not been recently surveyed.
 - An initial set of research papers that have looked into the area you have identified. You should cite at least 3 related research papers and include them in a References section.⁴ Feel free to use the papers linked in the syllabus (under optional readings on the course schedule) if they're

¹<https://www.overleaf.com/latex/templates/association-for-computational-linguistics-acl-conference/jvxskxpznfj>

²https://2023.aclweb.org/calls/style_and_formatting/

³This is not necessarily the easier option. Surveys will involve a lot of reading of advanced papers, and some original analysis and synthesis of this work.

⁴References do not count toward the page limit.

relevant. Note that you are responsible for the references you put in your paper; many points will be deducted for hallucinated or fabricated sources (which LLMs love to generate)!⁵

- An outline of the themes you expect to uncover, and questions you hope to learn the answers to.

The project proposal is a contract. If we give you full credit, we expect you to implement it and analyze the results, and we will give you full credit for the final report if you do so (assuming you follow all formatting instructions). If you turn in a weak proposal, you may revise and resubmit it before moving forward. But the longer you take to define your project, the less time you will have to do it, so do the best you can before the proposal deadline.

One member of your team will upload your proposal as a PDF to Gradescope. Be sure to include a title and the names of all of your team members in the PDF, and to add your team members to your Gradescope submission.

Midway Report (15 points; due Apr. 21)

The midway report is a progress report of 2–4 pages. It must be in Association for Computational Linguistics format. It should be written in the style of a (very) short research paper. Doing well on this part of the project will benefit you greatly, as it will mean you have less work to do to get a good grade on the final report. We will also be able to provide better feedback if your report is well-written.

The report should be written in such a way that another student in the class could read it using the content we've covered in class so far. This means that if you're using methods or terminology outside of what we've covered in class, you will need to describe and/or define it. That said, you don't need to explain concepts we've covered in class, like what perplexity is. Because the page limit is so short, you will need to be concise!

Be sure to include the following in your report:

1. **Abstract:** A concise summary of your report. This should be less than 250 words. It should summarize the problem and highlight your main findings so far. I often find that it's easiest to write abstracts *last*, once I've written and thought through the content of the entire paper.
2. **Introduction:** This section introduces the reader to the problem area you're trying to solve or survey. Give a specific example and a technical definition of the problem. Elaborate on the following:
 - Why is your problem interesting or important? If you could solve this problem, would you be able to answer a scientific question or solve a real-world problem? Would it improve the efficiency, performance, or usability of existing systems?
 - Why is the problem hard, and how do obvious solutions fail? For empirical projects, you could answer this question by implementing and/or running a simple and obvious baseline, and (if you have time) analyzing failure cases. For surveys, you should locate and cite relevant evidence in the literature.
3. Description of progress. For empirical projects, include the following sections:
 - **Methods:** Describe the method you have implemented. It should be clear and give enough technical detail that someone who wasn't on your team could reimplement your approach and get similar results. This could include the names of the LMs you based your method on (if

⁵I've given models like Claude a fully specified list of citations and asked it to reformat them. Even in this scenario, it ended up inserting new fake references! Use LLMs with significant caution when building and formatting your bibliography. This has burned some top researchers recently, too: <https://gptzero.me/news/neurips/>

applicable), the algorithm(s) you used, and possibly some formal equations or illustrations if helpful.

- **Experiments:** Describe the dataset(s) you're using and give citations for each. Also describe the evaluation metrics or methods you're using to measure how well your methods and baselines perform. Be sure to include any other important details like key hyperparameters, training time, how long your method took to run, etc.
- **Results:** This is where you describe any quantitative or qualitative results that you've obtained so far. Use tables and plots; do not put numbers in the main text. Do your findings look like what you expected to see so far? Why or why not? If you've done some error analysis or qualitative analyses, what do they tell you about how your method could be improved?

If you're doing a survey, your outline will be a bit more flexible:

- **Overview:** Give a high-level summary of the papers you've read. Also describe the **taxonomy**, or sorting system, that you will use to categorize the papers you have read. The remaining sections of your survey paper should follow from your proposed taxonomy. Your taxonomy should contain at least 3 categories (ideally more), and make it clear what the main trends in the field are. See this paper's Figure 1 for an example.⁶
 - **Technical summaries:** For each category in your taxonomy, describe in technical detail what solutions have been tried for the problem you are surveying. You may find it helpful to use equations or illustrations to describe these ideas. Aim for at least 3–5 references per category of your taxonomy. Including even more will make writing your final report easier. Again, you are responsible for the references in your paper; significant points will be deducted for hallucinated or fabricated sources!
 - **Discussion:** Provide your opinions on the state of the field in your chosen area. This section should contain your own original synthesis of the papers you've cited, and indicate what you think the most promising and untapped directions are for future work in this area.
4. **Project plan:** This is where you describe your plan for the rest of your project. Empirical projects should include a list of experiments or analyses, and survey papers should include a list of topics/specific papers that you plan to read. This should be clear, concrete, and reasonable given the time left before the final report deadline. It's totally fine if this deviates from your project proposal! Just note how your plans have changed and why.
 5. **References:** These should ideally be compiled with BibTeX. Note that references do not count toward the page limit; you can include as many as you like. Empirical projects should include at least 3 references, including the model(s), dataset(s), and most related method(s) to the method being implemented. Survey papers should aim for at least 3–5 references per taxonomy category, and at least 10–15 references total at this stage.

Grading will be based on the following criteria:

- **Level of progress:** Has there been reasonable progress made so far? Ideally, you should be at least halfway done with running experiments and/or reading papers by this point, but this isn't a hard rule. Good progress will depend on the kind of project you're pursuing, and we will take this into account when grading.

⁶<https://aclanthology.org/2025.findings-emnlp.750.pdf>

- **Quality of writing:** Are important technical details clearly presented? Is the motivation clear? Are the results or taxonomy presented in a way that makes them easy to find and easy to understand, even before the reader looks at the text? Is the requested information provided to an appropriate amount of detail (given the page limit)? Is everything in ACL format, and does it adhere to the page limit?
- **Clear trajectory:** Do you have a **concrete** plan for the rest of the project? Is the plan reasonable given the timeframe?

Please include a title and the names of all of your team members in the PDF. Be sure to remove the “review” argument on Overleaf so that the author names show up. Only one member of your team should upload a PDF to Gradescope; be sure to add your team members to your Gradescope submission.

You will receive feedback on your midway report. This will contain comments on your level of progress so far, feedback on whether your plans for the remainder of the project are too ambitious or not concrete enough, suggestions for methods or related work to include, and feedback on the clarity of the writing.

Final Report (30 points; due May 7)

This should extend your midway report, filling in new results and missing details. This should be 4–6 pages in ACL format. We will make expectations clear in comments on your midway report. In general, empirical projects should aim for at least a couple additional experiments or analyses over the midway report, and survey papers should aim for at least 20 total references by this point (and possibly a new taxonomy category or two).

Grading will be based on similar criteria to the midway report. In addition to those criteria, we will also consider the following:

- The thoroughness of your evaluations or survey
- The originality of your ideas and analysis
- The amount of thought, effort, and creativity put into your project
- The performance of your methods (if you are doing an empirical project)
- The insightfulness and originality of your taxonomy and synthesis (if you are doing a survey)
- The clarity of your writing and analysis, and the presentation of your results
- Whether you incorporated the feedback you received on your proposal and midway report

After your conclusion, we ask that you include a brief summary of each team member’s individual contributions to the project in a section called **Author Contributions**. This should be only 1–2 sentences per person, and will not count toward the page limit. We will generally give each team member the same grade, but if there is a significantly unequal distribution of effort, we may investigate further and potentially give different grades to different team members.

Only one member of your team should upload a PDF to Gradescope. Add your team members to the Gradescope submission. If you ran an empirical project, we will ask you to include a link in your final report to your GitHub repository or to a Google Drive link where we can download your source code.

General Tips and Project Ideas

To get a feel for the style of a good paper, read some of the research papers linked in the syllabus. None of them are perfect, so read attentively: what do you find helpful or not when reading these papers? Copy the style of papers you find readable or enjoyable, and avoid copying the style of papers that are unclear.

How to write well. Here are some great tips on writing good technical papers:

- Write the Paper First, by Jason Eisner
- Tips for Writing Technical Papers, by Jennifer Widom
- Ten Simple Rules for Mathematical Writing, by Dimitri Bertsekas

Empirical project ideas. Here are some ideas for empirical projects:

- Extensions to existing NLP methods
- Analysis, visualization, and/or interpretability of existing NLP systems
- Better preprocessing of existing datasets to improve the performance of an NLP system
- Collecting new data or better data for a specific NLP task
- Post-processing LLM outputs (e.g., spelling correction, removing unnecessary spaces, detecting degenerate outputs)
- Applying/improving an existing NLP system in a new domain
- Extending one of the homework assignments—for example, by applying the same methods to a new task, or improving one of the methods significantly
- Implementing a new evaluation metric for some task/system
- Training a probe to detect some concept in an LLM’s activations
- Replicating an NLP paper’s results from scratch

Survey paper inspiration. Here are some examples of good survey papers. Obviously we will not ask you to cite over 100 papers like some of these do, but do take some cues from the way these papers are organized and presented:

- A survey on LLMs for story generation/creative writing: <https://aclanthology.org/2025.findings-emnlp.750.pdf>
- A survey of ways to measure generalization in NLP: <https://www.nature.com/articles/s42256-023-00729-y>
- A survey of retrieval-augmented generation methods: <https://arxiv.org/abs/2404.10981>

Late Policy

If you have any free late days remaining, you may use them for the project proposal or midway report. If you have none left, you will lose 10% of that milestone’s grade for each late day. As long as you submit the proposal on time, revisions do not count toward your free late day budget. Note that we will not give extensions on the midway report if your proposal takes a while to get right, so do aim to get this right the first time around!

Late submissions will not be allowed for the final report.

AI Policy

You are allowed to use AI tools, but you are fully responsible for what you submit. If your paper is found to contain false information, or if the provided code is found not to have the described functionality, this will result in significant point deductions.