

Final Project Proposal Instructions

April 2025

Each team should submit one project proposal, which is an important part of the final project.

1 Choose your Projects

Each team may choose the default project (building GPT-2) or propose their own custom project. Custom projects are highly encouraged.

A custom project can be:

- closely related to your research or personal interests;
- an implementation of a recent paper, enhanced with your own insights, thoughts, or improvements;
- an application of NLU techniques (such as sequence-to-sequence learning) in other domains (e.g., mathematics);
- an exploration of novel ideas (related to NLU techniques) that ultimately may not succeed but lead to a deeper understanding of the problem.

For instance, recent advances in using reinforcement learning to enhance the mathematical or reasoning abilities of LLMs make for an excellent custom project topic.

2 Content of the Proposal

2.1 Basic information

The proposal should have the following information at the top:

- Title: The title of your project (you can change this later).
- Team member names: List the names and @tsinghua.edu.cn email addresses of all of your team members.
- (Optional) External Collaborators: If you have any collaborators who are not students of this class, list them.
- Custom or Default Project: Indicate which you are doing.

2.2 Research paper summary (max 2 pages)

Select the paper most relevant to your project and provide a summary that includes bibliographical information, background, a summary of the contributions, limitations, and a discussion. You may need to reference a broader range of sources and justify your choice of this particular paper.

2.3 Project description (1-2 pages)

Describe what your plan to do in this project. The description may include following aspects:

- **Main Goal(s):** Describe the scientific question you aim to answer, such as testing model performance, improving a model, or confirming a hypothesis. Motivate why this goal is important, interesting, and feasible. Include any secondary or stretch goals, if applicable. Relate the project to your chosen paper.
- **NLP Task(s):** Clearly describe the NLP task you will address, including examples of input and output, if applicable. You do not need to repeat this if already covered in the paper summary.
- **Data Usage:** Specify the dataset(s) you will use, including size and any preprocessing steps. If collecting your own data, describe the method and timeline.
- **Method(s):** Describe the models/techniques you plan to use, highlighting any differences from the paper's method. Specify which parts you will implement and which parts you will use from existing sources. Mention any original contributions.
- **Baseline(s):** Describe the baseline methods you will use, including whether they are implemented by you, downloaded, or compared from previously published scores.
- **Evaluation Metrics:** Specify the evaluation metrics, including numerical and qualitative measures. Compare your method to existing scores or state-of-the-art performance, and outline expectations.