

Final Project

The **Final Project** provides you with an opportunity to apply your newly acquired NLP skills to an in-depth application or research problem.

Project Ideas

Your task is to develop machine translation or question-answering systems. You can either utilize the provided dataset or find and use a well-defined dataset suitable for your project.

1. **Machine Translation:** This task involves translating text from one language to another. Consider using one of the following beginner-friendly machine translation datasets:
 - [European Parliament Proceedings Parallel Corpus 1996-2011](#): Explore sentence pairs across a suite of European languages.
2. **Question Answering:** In this task, you'll work with provided sentences or text samples from which questions are asked and must be answered. Here are some good question-answering datasets:
 - [Stanford Question Answering Dataset \(SQuAD\)](#): Focuses on answering questions related to Wikipedia articles.
 - [Amazon question/answer data](#): Engage in question-answering related to Amazon products.

Project Submission

Remember to submit your final report and code by **April 4, 2024**.

Final Project Report Guideline

Each team submits one project report and source code worth **35%** of your overall grade. This document specifies what information you should include in your report.

Report contents (5-8 pages)

Your report should be 5-8 pages (excluding references). We've provided a section structure for you to use.

Key information. Your report should have the following information:

- **Title:** The title of your project.
- **Team member names:** List the names and email addresses of all your team members.
- 1. **Abstract.** An abstract should concisely (less than 250 words) motivate the problem, describe your aims, describe your contribution, and highlight your main finding(s).
- 2. **Introduction.** The introduction elucidates the problem at hand, highlighting its difficulty, significance, and inherent interest. Furthermore, it outlines the fundamental concepts that underlie your approach and the outcomes you have achieved. While the abstract provides a concise summary of the same content, the introduction offers a broader scope for motivation, detailed explanations, and references to relevant prior research, and ultimately, captures the reader's attention.
- 3. **Related work.** This section plays a crucial role in providing the reader with a comprehensive understanding of the research context surrounding your work. It accomplishes this by offering an overview of existing literature in the field.
 - As you delve into this discussion, you can include papers that have inspired your approach, papers that serve as baselines for your work, papers proposing alternative approaches to the problem, and even papers that have applied similar methods to different tasks.
- 4. **Method.** This section details your approach to the problem. For example, this is where you describe the architecture of your neural network(s), and any other key methods or algorithms.
 - You should be specific when describing your main approaches – you probably want to include equations and figures.
 - If any part of your approach is original, make it clear. For models and techniques that aren't yours, provide references.
 - If you're using any code that you didn't write yourself, make it clear and provide a reference or link. When describing something you coded yourself, make it clear.

5. **Experiments and results.** This section contains the following.
 - **Data:** Please provide references and describe the dataset you are utilizing. Additionally, ensure that you clearly explain the associated task.
 - **Evaluation method:** Describe the evaluation metric(s) you use.
 - **Experimental details:** Report how you ran your experiments (e.g., model configurations, learning rate, training time, etc.)
 - **Results:** Report the quantitative results that you have found so far. Use a table or plot to compare results and compare against baselines.
 - **Analysis.** Your report should include a qualitative evaluation using appropriate evaluation metrics and report the results.
6. **Conclusion.** Summarize the main findings of your project, and what you have learned. Highlight your achievements and note the primary limitations of your work.
7. **References.**

Instructions for Final Project Code Submission and Presentation

To ensure a smooth and organized process for the final project code submission and grading, please adhere to the following instructions:

1. **Code Submission Guidelines:**
 1. Create a GitHub repository that includes all relevant data and code for your project.
 2. Add a link to the repository in your written report.
 3. Include all code for preprocessing, training, feature extraction, prediction, evaluation, analysis, etc.
 4. You may use Jupyter Notebooks, but make sure your code is modular, well-commented, and well-structured.
 5. It is crucial that your repository is public by the time you submit the written report.
2. **GitHub README:**
 1. Include a detailed README file in your GitHub repository.
 2. The README should provide step-by-step instructions on how to use your code and reproduce your results.
 3. Mention any dependencies or setup steps required to run your code successfully.
 4. Highlight any known issues, limitations, or areas for improvement in your project.
 5. Include any additional information or resources that may be useful.

3. Results Presentation:

1. Each team must deliver an oral presentation.
2. The presentation time is limited to 8 minutes, followed by a 7-minute Q&A session.
3. All team members must be present in person and contribute to the presentation based on their individual contributions to the project.
4. The presentation should focus on conveying the main message effectively and raising interesting discussion points.
5. Submit the PDF version of your presentation via Brightspace before 11:00 am on the morning of the presentation.