Final Project

(CMPSC 497; modified 4/2/2025)

Project description: Train an open-source LLM to generate an approach for a research problem. (You can build your dataset like the final project-related experiments we have done: research problem is a short piece of text, and the approach is a long piece of text)

Goal:
- ☐ Gain hands-on experience in training LLMs.
- ☐ Practice defining and designing your own NLP problem and building a system to solve it.
- ☐ Additionally, critically assess potential issues in a given problem.

Participants: individually or as a group (<=3 students per group)

Data & Model & Experiments & Evaluation & Analysis: you decide

Requirements: We require "training", so ``prompting existing LLMs to get approaches'' is not acceptable.

What you need to submit:

- ☐ **4/25/2025**: Submit your project report in PDF, with all group members' names.

Submit through Canvas assignments with github url.

Evaluation:

- **PDF quality (80%)**: we score by the following two aspects:
  - Dataset construction (20 points)
  - LLM selection and training details (20 points)
  - Evaluation metric, and experiments (20 points)
  - Your thoughts (e.g., what you learned in this project; what issues you found in this project, etc.) (20 points)

- **(optional) Presentation (20%)**: A default score **18** (out of max 20) for the presentation part if you choose NOT (such as your scores of other parts can secure an "A" for you). If you choose to present, the evaluation criteria include: slide quality, work quality, presentation skills, question answering, etc.)