# Project Report - v2

## Overview

In teams of up to 4 people, you will be working on a final project. This is the second part of the final project: the project report.

My hope is that, through this project, you carry out original research or take a deeper dive into the literature. The rubric is intended to give you a guideline, and my goal is for this report to be a manageable amount of work. If you are feeling overwhelmed by the rubric or this report, please talk to me. I'm here to help you find your way!

## Deliverables

The following must be submitted through MarkUs on the due date.

1. Project report in `.pdf` format.
2. If you used large language models, then you must submit all chat transcripts directly related to the preparation of the assignment in `.txt` format.

## Due Date

9 April 2025

## Project Report Details

**Collaboration** You may work in teams of up to 4 people. If you are having trouble finding a group of people to work with, please email the instructor. Only one report needs to be handed in per group.

**Length** The report should be 4 to 8 pages, not including appendices or bibliography. You are encouraged to keep the text short and informative. Illustrative figures are a great way to convey intuition, and we encourage you to think of creative ways to convey the main idea. You can include as many proofs, experiments, or additional details as you want in the appendix.

**Format** Please use 10pt, 11pt, or 12pt font with standard margins and submit your proposal as a `.pdf`. You may use any typesetting that you wish, but it should be readable. It should also have easily identifiable sections.

**Content** The requirements of this project are intentionally flexible. I hope this gives you room to be creative. There are a number of categories of projects that you could explore. Here are some categories together with example projects.

- **Theoretical:** Perform a theoretical analysis of scaling laws in a simple model that captures a key feature of large model training.
- **Algorithmic:** Develop and compare tokenization strategies for non-language data modalities.

- **Model Architecture:** Develop a more efficient attention mechanism for processing long sequences and analyze its complexity-performance tradeoff.
- **Review/Analysis:** Connect the ideas of this course to another field like physics or mathematics.
- **Dataset:** Develop a methodology for detecting and filtering out contaminated training examples that could lead to memorization.
- **Benchmark/Evaluation:** Design a targeted evaluation framework to measure a specific capability of large language models (e.g., multi-step reasoning or constraint satisfaction).
- **Empirical:** Design and conduct a systematic study to investigate a specific aspect of model behavior (e.g., how different hyperparameters affect training stability).

Parts of the marking scheme assume you will pick a project within one of these 7 broad categories. If your project idea doesn't fit in these categories, that's OK! Come discuss with your instructor, if you'd like some feedback.

**Resource requirements** You do not need to use vast computational resources to get a perfect mark for this project. You can design the project around your resource constraints and you're welcome to use Google Colab. While we do not expect you to run computationally intensive experiments for this part, they should be carefully executed.

Toy data or toy problems are OK and even encouraged! The point is to help the reader understand why or when we would want to use one approach over another, or to understand something better.

## Tips and Tricks

Here are some tips on writing.

- You are encouraged to attach your code.
- Axes on all plots should be labeled.
- Use a vector graphics format for your figures (e.g., pdf, eps, or svg) so that your figures don't look blurry.
- Try to use at most two font sizes.
- If you are proving theorems, long mathematical derivations are best put in the appendix, unless your proofs are concise or very informative. You can state the result in the main text, and defer the proof to the appendix.

## Marking Scheme

This marking scheme is worth 60 marks and is based heavily on Prof. Duvenaud's marking scheme and we refer you to his wonderful recommendations for more details.

- **Abstract** (**4 marks**) Summarize the main ideas of the project and its contributions.

  - Include the abstract at the top of the paper.
  - This should be clear and concise.
  - Do not describe every detail of the project.
  - Describe one or two main ideas.

- **Introduction** (**10 marks**) Describe the context of your paper and the key ideas.
  - The introduction is a critical part of any paper. Most people do not read past the introduction of a paper, so you should put effort into making the introduction approachable and clear.
  - Set the context: if you are trying to solve a certain problem in this project, describe why it is worth solving. If you are trying to prove a certain theorem, describe why it is worth proving.
  - You should also describe the main outline of your paper, including the key ideas and milestones.

- **Visualization** (**4 marks**) Visualize a main idea.
  - You should include at least one visualization of the main idea. The goal is to make the work you've done more accessible. You should create a new figure.
  - You may draw it by hand, as long as it's clear and readable.
  - We recommend OmniGraffle, which has a 14 day free trial.
  - We also recommend matplotlib in Python.

- **Formal description** (**10 marks**) Include a precise description of the main idea.
  - Your paper should be on a topic within large language model development, focusing on training, evaluation, deployment, or theoretical understanding. This part of the rubric asks that you describe your contribution clearly and formally.
  - Depending on your project category, include one of the following:
    * For theoretical contributions: Include a formally stated theorem, conjecture, or mathematical framework with clearly stated assumptions and implications.
    * For algorithmic contributions: Include a pseudocode or algorithm box detailing your method's key steps, complexity analysis, and requirements.
    * For model architecture contributions: Include equations describing your architectural modification, attention mechanism, or other model components
    * For review/analysis contributions: Include a formal framework for comparing approaches, with clearly defined metrics and criteria for comparison
    * For dataset contributions: Include a formal specification of your data collection/filtering methodology, including precise definitions of quality metrics and acceptance criteria
    * For benchmark/evaluation contributions: Include a formal specification of your evaluation framework, including precise task definitions, scoring criteria, and evaluation methodology

    - For empirical studies: Include a formal experimental framework including specific hypotheses, controlled variables, and measurement methodology
  - If none of the above are applicable, check with the instructor before handing in the assignment.
  - Additionally, every formal description should:
    - Clearly state assumptions and limitations
    - Define all terms and notation
    - Include an intuitive explanation of your approach
    - Connect to concepts from the course, particularly the Llama 3 technical report

- **Related work** (**4 marks**) Explain exactly how your project relates to the literature.
  - If your project builds on previous work, clearly distinguish what they did from what your new contribution is.
  - It's OK if you do not find all related papers, but do your best to find a few closely related papers.
  - For closely related papers, include 1-2 sentence summaries.
  - Have a proper bibliography section and in-text citations.

- **Experiments or demonstration** (**10 marks**) Demonstrate a key idea.
  - Things to include based on your project category:
    - For theoretical contributions: Include an illustrative example or counter-example demonstrating your result.
    - For algorithmic contributions: Include an implementation on a small scale with a comparison to a standard approach.
    - For model architecture contributions: Include an implementation on a toy dataset with a comparison to the simplest alternative model.
    - For review/analysis contributions: Include a table comparing the properties of different approaches.
    - For dataset contributions: Include an implementation of your data strategy on a small corpus showing improved quality metrics compared to a baseline method.
    - For benchmark/evaluation contributions: A proof-of-concept implementation of your evaluation framework using a small set (around 100) of carefully designed test cases, demonstrating how it captures capabilities not measured by existing benchmarks
    - For empirical studies: A targeted experiment using a small-scale model that isolates and measures your effect of interest with careful controls
  - If none of the above are applicable, check with the instructor before handing in the assignment.
  - Where applicable, include:
    - Dataset preparation methodology
    - Training procedures

    - Resource utilization metrics (memory, time)
    - Hyperparameter choices and rationale
    - Data preprocessing decisions
    - Implementation details and practical considerations

- **Limitations** (**4 marks**) State the limitations of your approach.
  - Most new ideas have limitations and that is OK! In fact, it is expected.
  - Describe some settings in which you would expect your approach to perform poorly, or where all existing models fail.
  - Give some examples of possible extensions, ways to address these limitations, or open problems.
  - If you are doing a review, explain the limitations of the ideas that you are reviewing.
- **Conclusions** (**2 marks**) Conclude your paper.
  - State the results that you obtained in your paper.
  - If you did experiments, repeat the main conclusions of your experiments.
  - Repeat the main ideas in your paper.
- **Novelty** (**10 marks**) Explore something new in your paper! This is a sliding scale.
  - (**0/10 marks**) If you did not include any new ideas, but you wrote a great review of the literature.
  - (**5/10 marks**) If you made some interesting innovations. Examples might include:
    - If you are proving a theorem, you used a somewhat uncommon proof technique or managed to extend some well-known result to a slightly new setting.
    - You introduced an algorithm that is an interesting, but small, twist on existing methods.
    - You used an existing model type in a new setting.
    - You mostly reviewed the literature, but you made some unexpected, detailed, and precise connections between apparently unrelated subfields.
  - (**10/10 marks**) You made very novel contributions. In this case, I would expect that your paper really contributes some kind of new insight.
- **Contributions** (**2 marks**) List the contributions of each team member.

## FAQ

**How broad is the scope for the project? Can I do a project that overlaps with my thesis topic?** It makes sense to pick a project topic that overlaps significantly with your interests, and we strongly encourage you to pick a project that may inform your thesis work. Still, the scope of the project needs to be related to the course content or to large AI model training more broadly.

**Can I reuse work done prior to the course or work that was done in collaboration with people not on the project team?** If your project involves any work that was already done prior to the course or was done by people not on the project team, then your project proposal and report *must* report the nature of this work. Our expectations will be higher compared with completely de novo projects.

**Can I reuse a project from a different class?** No.

**Will your expectations change depending on the number of students in a team?** We aren't modifying the marking scheme depending on the team size.

## Academic Integrity

Because this assignment will involve citing other people's work, it is important that cite properly. In general, you should follow U of T's Code of Behaviour on Academic Matters.