21-366 Final Project Guidelines

SPRING 2025

1.1. **Final project.** The final project is an opportunity for students to apply what they have learned in the course to tackle real-world problems or explore advanced topics. The final project will be a group project, with students choosing between the following two options:

a) Expository paper
- Write a paper on an extension of a course topic or a new topic not covered in class.
- The paper should be 10–12 pages long, professionally written in LaTeX, with clear exposition and a detailed bibliography.
- A sample LaTeX template is available on the course repository.
- The paper should include:
  - An introduction
  - Discussion of history of the problem/subject and/or related work
  - Summary of the main results
  - Statement-proof style exposition leading up to the main results
  - Bibliography
- Here is a good example of a well-written paper on neural networks
- Please read Writing a Math Phase Two Paper by Steven L. Kleiman and On Writing by Terence Tao for tips on writing mathematical papers.

b) Coding project
- Use methods learned in class to analyze real-world datasets, build a quantitative model to solve a specific problem, or implement ideas from a research paper.
- Host the project on GitHub and include:
  - A detailed GitHub repository summary
  - A separate 1-2 page write-up in LaTeX explaining the mathematical foundations of some methods used, with a bibliography of the references used
  - (Optional) Slides for the final presentation and/or video recording of the project
- Here is a good example of a well-documented GitHub repository
- Google's Python style guide may be useful for you to maintain clean and readable code

If you would like to have your final paper hosted or GitHub repository linked to on the course repository, please let the instructor know.

1.2. **Project goals.** The goal of the final project is to give students an opportunity to explore a topic of their interest and to add to their academic and professional resume. The project allows students to showcase their strong analytical background, collaborative skills, and ability to communicate complex ideas clearly and effectively to a general audience.

1.3. **Key dates.**
- **February 1:** optional project pitch meeting (via Zoom)
- **February 8:** team formation and short project proposal due. The short proposal is just an opportunity for the instructor to provide feedback and guidance
- **March 1:** one-page progress report due
- **April 12:** draft of the final project due
- **May 1 (tentative):** final project submission deadline
- **Final exam week:** project presentations in class

Department of Mathematical Sciences, Carnegie Mellon University
E-mail address: jkoganem@andrew.cmu.edu.
Date: January 31, 2025.

1.4. **Requirements.** In both cases, the topics or the methods being used should be related to topics covered in class. Students will work in groups of 2-4 and are required to:

- Produce a video recording (7-10 minutes) summarizing their project
- Prepare slides and present their work to the class during the week of the final exam

Please submit the project proposal, progress report, and draft of the final project as a .pdf file to the instructor via email. The final project should be submitted as either as a link to a GitHub repository or a .pdf file of the expository paper. The video recording should be submitted as a link to a Google Drive or Dropbox file. The final presentation slides should also be emailed to the instructor before the final presentation day.

1.5. **Group formation.** Students are encouraged to use this as a learning opportunity and form groups based on their collective interests and skills. Ideally, each team should have a mix of students with different backgrounds and strengths.

Once the groups are formed, please join the course Slack channel and create a private channel for your group. This is where you can work collaboratively, ask questions, share ideas and resources with your teammates. You can also message the instructor directly on Slack if you have any questions.

1.6. **Resources and possible topics.**
Resources:

- Public datasets
  - Kaggle
  - Harvard Dataverse
  - Data Is Plural
  - UC Irvine Machine Learning Repository
- Overleaf
- GitHub
- Google Colab
- Streamlit: a Python library that makes it easy to create custom web apps for ML projects

Possible topics to explore:

- Dimensionality reduction
  - A survey of dimensionality reduction techniques
  - Robust Principal Component Analysis?
  - Theoretical Foundations of t-SNE for Visualizing High-Dimensional Clustered Data
  - UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction
  - An Introduction to Johnson–Lindenstrauss Transforms
  - The Johnson-Lindenstrauss lemma and Linformer
  - Linformer: Self-Attention with Linear Complexity
- Large Language Models and Natural Language Processing
  - Word2Vec: Efficient Estimation of Word Representations in Vector Space
  - BERT: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
  - SBERT: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks
  - SentenceTransformers Documentation
  - Retrieval-Augmented Generation (RAG): Retrieval-Augmented Generation for Large Language Models: A Survey, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks
- Topological data analysis: An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists
- Algebraic graph theory: Spectral and Algebraic Graph Theory
- Google's PageRank algorithm: The PageRank Citation Ranking: Bringing Order to the Web
- Applications of neural networks in physics
  - Machine Learning for Fluid Mechanics
  - Physics-informed neural networks
  - Physics-informed machine learning
  - Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations
  - Physics-informed neural networks (PINNs) for fluid mechanics: A review
  - Accurate computation of quantum excited states with neural networks
- Strum-Liouville theory

- Multilinear algebra
- Fourier analysis
- The Hartman-Grobman theorem and the Stable manifold theorem for hyperbolic equilibrium points
- Numerical Linear Algebra
  - Computing the SVD of a General Matrix Product/Quotient
- Numerical methods for partial differential equations

1.7. **Tips for project selection.**

a) Choose a project that aligns with your interests and goals.
   - If you are more interested in applications, consider working on the coding project that can showcase your programming skills and your machine learning knowledge. If you decide to work with data, try to find an interesting dataset; datasets such as MNIST are somewhat overanalyzed, though they can be useful for learning and benchmarking.
   - If you are more interested in theory, consider working on an expository paper that can showcase your strong theoretical understanding and your ability to break down and communicate complex ideas.

b) If possible, choose a project that includes clear and effective visualizations. This makes it easier to tell a story and to present your work to a general audience.

c) The devil is in the details, so while presentation is important, please also make sure the technical details are checked thoroughly and that you can explain them clearly to others if needed. For this reason, make sure to choose a topic that you are comfortable with.

d) For deep learning projects, PyTorch is generally recommended over TensorFlow as it is more Pythonic, modular, and in my opinion easier to debug. Furthermore, as suggested by this graph here, PyTorch is arguably the preferred deep learning framework at this moment.

   The math department has computational resources that you can utilize if you require significant computational power. You can also try to utilize the Pittsburgh Supercomputing Center's Bridges-2 supercomputer.

1.8. **Tips for video recording.** It is highly recommended that you practice your presentation (maybe even multiple times) before recording it. It is easy to overestimate the amount of time you have, so make sure to focus on the most important aspects of the project.

You may use the same slides for both the final presentation and the video recording. For the video recording, you may use Zoom or any other screen recording software.

Please submit your recording as a link to a Google Drive or Dropbox file and email it to the instructor.

1.9. **Tips for final presentation.** Here are some tips for the final presentation.

1) Tell a story: clearly and concisely communicate to your audience what the presentation is about
2) Have a clear and well-defined problem statement
3) Explain what you did or tried to do to study the problem
4) State the results and what you learned from the project
5) Discuss future directions
6) Summarize and conclude
7) Acknowledge your team members, the resources used, and thank the audience for their attention

Make sure to always keep the audience in mind when you give presentations. Presenting to experts on a subject is different from presenting to a general audience. For the final presentation for this class, you should treat it more like the latter as your audience likely would not have the same level of expertise as you do on the subject.

A few more tips on making slides.

- Be concise: avoid long paragraphs and use bullet points
- Make sure plots are clearly labeled, not blurry, and easy to read
- Include relevant information only, you can provide links to additional resources at the end

The most important tip for giving a presentation is to practice. Practice helps you to be more comfortable with the material, be more confident in your delivery, and also help you manage your time better.

If you would like the presentation to be recorded, please let the instructor know in advance so that the recording can be scheduled.

1.10. **Grading.** The grading for the project will be based on the following components.

a) Progress report submission (10%)
b) Draft submission (10%)
c) Final project (paper or repository) (40%)

- Clarity of exposition and organization (20%)
- Depth of analysis, quality of writing and/or code (20%)

d) Video recording (10%)
- Clarity and organization of presentation (5%)
- Depth of analysis (5%)

e) Final presentation (30%)
- Instructor evaluation (15%) and peer evaluation (15%)

A separate group evaluation form will also be provided to each student at the end to evaluate the contributions of each member on their team. If it is clear that a group member did not contribute to the project or contributed minimally, it will be up to the instructor's discretion to adjust the final project grade for that student accordingly.

1.11. **Final thoughts.** Ultimately, this project is for you to learn something new. Find a project that you are passionate about and devote time to it. If you have any questions or need help, please do not hesitate to reach out to the instructor.

I look forward to seeing your presentations and learning from your projects. Good luck and have fun!