# COURSE PROJECT[1]

## 10-423/10-623 GENERATIVE AI (SPRING 2024)
http://423.mlcourse.org

# 1 Overview

The course project for 10-423/623 *Generative AI* provides a chance to explore a real-world application of generative AI techniques. Successful completion of the project will require the following steps: (1) Careful identification of a task and real-world dataset for a generative modeling problem. (2) The establishing of a baseline approach with appropriate metrics. (3) Design of an experimental setup that involves the comparison of multiple algorithms in a controlled setting. (4) Summarizing and reporting of your findings via the following milestones:

1. Team Formation [3 people] (§2.1) – due Tuesday, Apr 02, 2024 at 11:59pm

2. Project Proposal [2-3 pages] (§2.2) – due on Friday, Apr. 5, 2024 at 11:59pm

3. Midway Executive Summary [3-4 pages] (§2.3) – due on Thursday, Apr. 18, 2024 at 11:59pm

4. Final Poster (§2.4)

    - Final Poster PDF Submission – due on Wednesday, May. 1, 2024 at 11:59pm

    - Final Poster Presentation – in-person attendance required Friday, May. 3, 2024

5. Final Executive Summary [5-6 pages] (§2.5) – due on Friday, May. 3, 2024 at 11:59pm

6. Code Upload (§2.6) – due on Friday, May. 3, 2024 at 11:59pm

Below, we begin with some discussion of various aspects of the project (§1.1–§1.7). Then we detail these milestones (§2). The grading breakdown for these components is as follows: Team Formation 2.5%, Project Proposal 15%, Midway Executive Summary 20%, Final Poster 20%, Final Executive Summary 32.5%, Final Code Submission 10%.

## 1.1 On Choosing Task / Dataset

For most projects, the task will be a generation problem. If your task is not a generation problem, there should be a clear generative modeling component in your approach.

If you already have a specific domain of interest, that's great. However, if you are brainstorming, we recommend starting with the *Browse State of the Art*[2] page from PWC. This page features a long list of high-level categories ranging from *Computer Vision* to *Natural Language Processing* to *Audio* and *Code*. You can then explore the wide variety of tasks for a chosen category: for example the *Computer Code* category includes tasks such as *Code Generation* and *Source Code Summarization*. Each of those tasks includes a list of common benchmarks. (A benchmark is just a dataset/task pair.) Each benchmark page includes a (usually very incomplete) list of paper's, their performance on some metric(s) for that task, and (sometimes) a link to code for the paper. The *Dataset* tab for a benchmark usually takes you to the paper for that dataset.

---

[1]Compiled on Monday 1st April, 2024 at 15:06
[2]https://paperswithcode.com/sota

## 1.2   On Choosing Methods

The field of GenAI is changing so rapidly that new methods are constantly arising. This is great fun because it likely affords you the opportunity to do something completely novel, but it's also challenging to keep up with all the latest work. The best strategy here is to read papers. As you read one paper, it will cite dozens of others, and so on. Reading is fundamental to research, and a great skill for you to develop if you want to keep abreast of the latest trends in this area.

## 1.3   Starting from Scratch vs. from an Existing Implementation

Whether to choose a blank project as your starting point or an existing implementation is up to you. Both of these are encouraged as they will lead to different learning outcomes.

Starting from scratch will likely expose you to much of the details of the dataset and task itself, since getting a dataset and baseline properly prepared can take a substantial effort. As well, you'll dive deep into the details of the method without guidance. By contrast, starting from an existing implementation will expose you to another researcher's code; this has pluses and minuses depending on the code quality, but in many cases you will be able to learn a lot of having to read through and understand the code in order to make changes to it.

When we are grading, your starting point will be taken into account: For example, one group may receive high marks for a re-implementation with limited experimentation because they started from scratch, whereas another group may receive similarly high marks for a smaller amount of code but a more substantial set of experiments. Either way, you should carefully document how your time was spent in the executive summary.

## 1.4   Am I going for state-of-the-art?

You *will not* be judged on the performance of your final results relative to state-of-the-art. As such, when selecting a baseline implementation, it is recommended that you consider other aspects of the implementation itself (besides the quality of the model), such as the speed of the code, the simplicity of the code layout and training scripts, whether or not it is using familiar libraries such as PyTorch, etc.

## 1.5   Scaling Up

Many of the most interesting real-world datasets and models tend to be large: that is, the datasets consist of many training examples and the models consist of many model parameters. Naturally training in these cases can require high-performance computing hardware (e.g. the latest GPUs) and time (e.g. many hours of compute, often parallelized across multiple devices).

As such, you may find you need to work with a subset of the dataset, either by working with a subset of the train/val/test samples or by working with a subset of the labels. Taking this approach is fine so long as you clearly document your experimental design decisions.

You might instead consider working with a smaller model. For example, some models are small by design so that they can run on more limited hardware. Again, making this choice is fine so long as you clearly document why you made the decision to do so.

Scaling up the size of your dataset/model is not required. However, if you choose to focus on aspects of scale, then your project outcomes may look different than those who wanted to, say, focus more on modeling aspects. This is fine and you should simply document the scope of work that went into setting up appropriate computing environments, handling long running experiments, etc.

## 1.6   Novelty and Replication

You are encouraged but not required to produce a novel solution to a problem. Reimplementing an existing method if one is not already available can be your main contribution. However, you are encouraged to try

out a setting that, to your knowledge, hasn't been tried before in the literature.

Novelty could be achieved by introducing a new technique or by applying an existing method to a task that it hasn't been used on before. (Two other directions are inventing a new task altogether or a new theoretical result. Please check in with the course staff if you select one of these two directions.)

After the proposal, you may find out that something similar to your project direction *has* been done before. In this case you have two options (both are fine): (1) change course slightly to preserve your novelty or (2) stay the course and focus on being thorough in your implementation and experiments. Either way, be sure to clearly document the related work and when you found out about it.

## 1.7 Replication

We *do* require that you replicate at least one prior result from the literature. Or, if you plan *not* to do so, you should discuss why not with the course staff ahead of time, and put that justification into writing in each of the milestones.

The reason we want some result to be replicating prior art is that it shows that you are starting from a reasonable baseline. That is, you should have a table that includes both the result from the prior paper, and your attempt at replicating it. Note that sometimes, replication does not mean you get exactly the *same* result as what was reported in the paper. For example, the original paper might report a metric on some benchmark of 92.3, but the best your replication gets is only 91.9. This, unfortunately, is common and could indicate that the authors of the original paper did not provide the exact code to *truly* replicate their work, or that changes in libraries/hardware led to some shift.

Replication and reimplementation are not the same. The former is about reproducing a result, the latter is about writing new code that implements a method from a paper.

## 2 Milestones

### 2.1 Team Formation

Each team will consist of 3 people. Teams must be specified in advance of the proposal deadline.

The deliverable for this milestone will be a single correct **group submission** of a PDF containing the names/Andrew IDs of the team members to Gradescope. To receive full credit, the group must submit only one PDF associated to the usernames of all the members of the group. See Gradescope instructions for adding group members at the link below.

https://help.gradescope.com/article/m5qz2xsnjy

(We **do not** want, for example, multiple PDFs submitted by the same group, or a single submission that leaves off one or more members of the group—both of these would be considered incorrect submissions.)

## 2.2   Project Proposal

The guidelines for the project proposal are as follows:

1. **Overview:** The proposal will describe the task, dataset, proposed methods, and related work. The proposal should be 2-3 pages (excluding your references/bibliography).

2. **Contents:** Your proposal should be organized as follows:

   - *Title box:* Project title and list of group members' names and Andrew emails.

   - *Section 1: Introduction* Very succinctly summarize all aspects of your project proposal.

   - *Section 2: Dataset, Task.* Describe the dataset you will use, a precise definition of the task, and what metric(s) you will use for evaluation.

   - *Section 3: Related Work.* Provide a short literature survey of 4 or more relevant papers. These papers should be relevant to your approach; they do not need to be relevant to the particular task/dataset that you chose.

   - *Section 4: Approach.* Describe the generative AI methods that you plan to implement and compare. Identify your baseline method and which result(s) you plan to replicate. Be sure to include a clear identification of whether your key contribution will be a novel technical approach, a novel application of an existing method to some task, or a from-scratch reimplementation.

   - *Section 5: Expected Outcomes.* Describe the experiments you will run and the results you expect (or hope) to get at the end. (In the end, it's fine if your results do not match these expectations—this is research after all—but you must articulate a hypothesis.)

   - *Section 6: Plan.* Identify how you will break up the work between each of the team members. (We strongly encourage you to consider pair programming or the like for the most important aspects of the implementation.) Identify what you plan to have completed by the Midway Executive Summary deadline.

   (If the core contribution of your project will be theoretical, you should follow a similar organization to that described above replacing descriptions of methods/experiments with discussion of expected theoretical results.)

3. **Pre-existing Work:** If you wish to rely on pre-existing open source code for your baseline model, you can use that as the baseline against which your implemented algorithms will be compared. Clearly identify the source code you plan to use in Section 2 of your proposal (include a link to the code).

   You are not permitted to use work you began prior to this course as your course project. If you wish to build on something you worked on prior to this course, you should clearly identify which parts of the work were done ahead of time in Section 6.

4. **Submission:** Your project proposal PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

## 2.3   Midway Executive Summary

Below are the guidelines for the midway executive summary milestone.

1. **Overview:** The midway executive summary offers each group a chance to present their progress halfway through the project's duration. The midway executive summary should be 3-4 pages (excluding your references/bibliography). You are welcome to include appendices, which do not count towards the page limit, and refer to them from within the paper.

2. **Contents:** A suggested organization for your midway executive summary is below. However, you are welcome to deviate from this organization, so long as you include all the relevant content.

   - *Title box:* Project title and list of group members' names and Andrew emails.

   - *Section 1: Introduction* Concise overview of the entire executive summary. The introduction should in length be similar to that of a long abstract. It should highlight your motivation, proposed method, and expected results.

   - *Section 2: Dataset, Task.* Detailed description of the task, dataset, and metric(s) for evaluation. This section should be nearly in its final form.

   - *Section 3: Related Work.* A short literature survey of 8 or more relevant papers. This section should be nearly in its final form.

   - *Section 4: Approach.* Description of both (a) your baseline approach and (b) the main methods that you will implement. This section should be nearly in its final form.

   - *Section 5: Experiments.* Precise description of the experiments you will run, and any results if applicable. You are required to present results of a baseline model on your task of interest. You must include skeleton tables/plots—these can be empty at this point. You should also include prose with references to the skeleton tables/plots describing what they will contain.

   - *Section 6: Plan.* Timeline of remaining milestones with dates and who is responsible for each milestone.

3. **Executive Summary Design:** The key point of this milestone is to communicate the motivation and goals to the reader. If you received constructive feedback on your proposal direction from the course staff, this is your chance to course correct and present any changes to your original proposal. As with any research endeavor, we expect that a carefully planned project is much more likely to succeed. If this is the first time you find yourself following instructions to "Write the Paper First", please read Jason Eisner's advice page on the topic here.

4. **Executive Summary Submission:** Your midway executive summary PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

## 2.4   Final Poster

Below are the guidelines for the final poster milestones:

1. **Overview:** It should include all the components detailing your work as well as your now completed results. Students are required to attend the final poster session to participate in peer evaluation.

2. **Poster Printing:** The size of the poster should be 40 inches x 30 inches (the size of the poster boards). If you have an SCS account, you may use SCS poster printing. However, please be sure to read through their guidelines to avoid delays. This service is only available to students with an SCS account: https://computing.cs.cmu.edu/desktop/printing-posters.html. If you do not have access to free poster printing, then you have an option of printing your poster on sheets of paper and arranging them together as a poster. However, please make sure that you still design your poster with a consistent flow in mind (see point 5). For help on doing this, see here.

3. **Poster Submission:** Your final poster PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

4. **Poster Presentation:** You will present your poster in-person during the final presentation timeslot selected by the registrar. We will arrange the timing of the poster session so that you have a chance to present your poster to your classmates as well as view their posters.

5. **You will get 9 minutes to present your poster plus 3 minutes for questions.** We will be looking for the following:

   (a) Motivation: Is the goal of the project clear? Are the task, dataset, and metrics clearly described? Is there a clear understanding of related work?

   (b) Details of the proposed methods: Which methods are applied? Are they clearly explained? Is the choice of methods technically sound? How are they adapted to this problem?

   (c) Results: Are the current results reasonable? Are the results explained well? Were the errors analyzed in order to understand why each method behaves as it does?

   (d) Presentation: Were poster/explanations concise? Were poster/explanations clear? Did Q&A demonstrate understanding?

## 2.5   Final Executive Summary

Below are the guidelines for the final executive summary milestone.

1. **Overview:** The final executive summary gives each group the opportunity to fully describe their course project work. The midway executive summary should be 5-6 pages (excluding your references/bibliography). You are welcome to include appendices, which do not count towards the page limit, and refer to them from within the paper.

2. **Contents:** A suggested organization for your final executive summary is below. However, you are welcome to deviate from this organization, so long as you include all the relevant content.

   - *Title box:* Project title and list of group members' names and Andrew emails.

   - *Section 1: Introduction.* Concise overview of the entire executive summary including motivation, proposed method, and results.

   - *Section 2: Dataset and Task.* Detailed description of the task, dataset, and metric(s) for evaluation. Description of the model.

   - *Section 3: Related Work.* A short literature survey of 8 or more relevant papers.

   - *Section 4: Methods.* Description of both (a) your baseline approach and (b) the main methods.

   - *Section 5: Experiments.* Full description of your experimental design, results, and analyses. Whereas a typical experiments section focuses on key metrics for the benchmark task on the full dataset, yours should reflect the research diary nature of this document. For example, if you ran a pilot experiment on a fraction of the dataset (e.g. 1/10), you should include those results here. You should also include results with timing information about your experiments, e.g. an estimate of how long it takes to train the model for one epoch and to convergence, an estimate of how long it takes to validate the model on the val/test sets. Also, describe what hardware you are using (e.g. specs of your CPU or GPU).

   - *Section 6: Code Overview.* Identify specific portions of code that *you* wrote and explain what those section do in prose. You should either identify a portion of code by including a screenshot of it and placing it in a referenced Figure in an Appendix, or by referring to specific line numbers (e.g. lines 234–270) in the code you upload to Gradescope.

   - *Section 7: Timeline.* Succinctly identify how much time you spent on each of the various stages of the project. For example, include a table with hours spent on each of: reading papers/dataset websites/etc, reading code documentation (e.g. from PyTorch or Torchvision), understanding code from an existing implementation, compiling/running existing code, modifying existing code to do something new (or writing new code from scratch), writing scripts to run experiments, running experiments, compiling results, writing this document, etc.

   - *Section 8: Research Log.* Explain the meandering path that you took to arrive at the work that this summary represents. Try to explain what the key challenges were and how (or if) you overcame them. This section is your opportunity to showcase the *work* that you did and is arguably the most important part of this document. For example, if an important line of your results table is empty, you can use this part of the document to explain why it wasn't easy to fill in. If your plan deviated from your proposal, this is your chance to describe the work that you did to realize your plan needed to change.

   - *Section 9: Conclusion.* Summary of your main findings and considerations for future work.

3. **Executive Summary Design:** The executive summary should stand-alone. That is, it should not be considered as an addendum to the final poster. Rather, it should offer the prose narrative that your visual poster lacks. You should include figures and visuals in the executive summary if they enhance the reader's experience.

4. **Executive Summary Submission:** Your midway executive summary PDF should be submitted to Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.

## 2.6 Code Upload

Below are the guidelines for the final code submission.

- **Overview:** The code upload milestone is the simplest of all since you are submitting your code as is for review. Note that while this milestone is required, you should view it as an opportunity to showcase the work that you have done. That is, your code is an important outcome of your work. By uploading your code, you will also be able to refer directly to it in your Executive Summary PDF.

- **What to submit:** The purpose of this code submission is to demonstrate what you have implemented.

  All the code for your project can be uploaded exactly as it was when you last used it—you do not need to clean it up. If you used a large open source library, you can include it all.

- **Submission:** You should upload your code to the Code Upload slot on Gradescope. Each group should only have 1 submission that includes all the members of the team via Gradescope's group submission feature.