On one hand, machine learning is such an exciting discipline because of its rapid pace of innovation, with new technologies and capabilities emerging at every turn. On the other hand, mastery of machine learning is so challenging because it evolves so rapidly that any frozen skill set (however masterful) risks becoming outdated. In such a dynamic environment, what skills will stand the test of time and continue to serve you, long after many of the precise methods taught in this class go out of fashion?

Whether you plan to become a builder in industry or a scholar in academia, perhaps the most important meta-skill to acquire is to be able to think like a researcher. Thinking like a researcher means (i) accessing new knowledge from a messy scientific literature, pulling out the useful and correct bits from papers while discarding (or correcting) mistakes; (ii) applying new methods and ideas to real problems that you are trying to solve.

To get hands-on experience with these skills, you will complete a course project with the following specifications:

1. You must work on the course project in groups of 2 or 3; **you may not work on the course project alone.**
2. There are two options for the project: a partially pre-designed option and a build-your-own option. Both are described in more detail below. We suggest that groups with less research experience should choose the pre-designed option, but all students are allowed to choose either option.
3. As part of the proposal process, you will decide whether you plan to use GenAI, and for what parts. GenAI is allowed, but the scope of the project should be larger if a significant amount of it will be automated. If your usage of GenAI changes over the course of the project, you must tell us. The pre-designed project does not use GenAI.
4. There are four deliverables associated with the project
    a. **A 1-2-page proposal, due Friday, March 13th at 11:59 PM** — these will largely be graded on completion, although you may be asked to resubmit unsatisfactory work; their primary purpose is to catch and correct any early misconceptions about the project task/scope. The proposal is worth 2% of your final grade.
    b. **A 4-page check-in, due Wednesday, April 8 at 11:59 PM** — these will also largely be graded on completion, and serve as a way for us to ensure that you're making sufficient progress on the project. The intent is that the check-in report can serve as a starting point for writing your final report. The check-in is worth 2% of your final grade.
    c. **A 4-minute video showcase, due Monday April 20 at 11:59 PM** — to be shown in-class on **Friday, April 24** during the recitation timeslot. The final video showcase will be worth 4% of your final grade.
    d. **An 8-page final report, due Wednesday April 22th at 11:59 PM** — the final report is worth 12% of your final grade.
5. The page limits for the deliverables defined above are strict, with the exception of pages only containing references/citations, which will not count against this limit.

6. Each group will be assigned a course staff member as their project mentor: they will be your primary point of contact for any questions that arise during the course of the project. You will have two 15-30 minute meetings with your TA mentor over the semester:
    a. After submitting your proposal, your group must schedule a meeting with your project mentor where you will collectively review your proposal. Based on this meeting, your group will either receive approval to move forward or be asked to resubmit your proposal, incorporating feedback from the meeting. **The deadline for the first mentor meeting is Friday, March 20th.** You must submit your project proposal before meeting with your mentor.
    b. Likewise, after submitting your project check-in you will be expected to meet with your TA mentor again. This will be an opportunity for you to receive low-stakes feedback on the current trajectory of your project and address any potential issues before the project deadline. The deadline for the second mentor meeting is **Wednesday, April 8**. We recommend that you submit your check-in prior to meeting with your mentor.

General tips and recommendations:

- A good place to find papers is in the proceedings of prestigious conferences, e.g. ACL, EMNLP, NeurIPS, ICLR, ICML. If you have further questions about finding literature, please consult your mentor.
- We do not recommend including related works, current progress, or data and code availability sections in the video showcase
- A good tool for recording your video showcase is the SlidesLive Recorder app. Here are some additional tips for your video recording.

Project Proposal

On the next page, we have provided you with an example project proposal to review. **If you are unsure of what to do for your project, we suggest that you implement this exact proposal** (this is the so-called pre-designed option)**.** If you choose not to implement the example project (opting instead for the build-your-own option), you must have your proposal approved by your TA mentor at your project proposal meeting. This proposal could be a minor variation on the pre-designed option, a totally different project designed from scratch, or anything in between. The primary purpose of requiring mentor approval of project proposals is to ensure that the scope of your project is appropriate. It is often the case that students propose outstanding projects that would be suitable for a capstone project, industry internship, PhD thesis, personal project, or other venue but which are unfortunately not suitable for a 10-701 course project.

Example Project Proposal

Task and Dataset:

In this project, we propose to explore machine translation, a common natural language processing task. We plan to use this Kaggle dataset, consisting of English sentences and their French translations; see **sections 1 through 3** of the associated paper for more information about the dataset. Our goal is to use this dataset to train machine learning models that take an English sentences as input and return as output the same sentence in French.

Methods:

To solve this task, we will use three *distinct* methods for performing machine translation, each implemented from scratch. These methods will be coded entirely by our group, with no reliance on existing models, large or otherwise. That being said, we do intend to use Python packages including PyTorch where appropriate. For our purposes, methods that correspond to the same underlying model but with different hyperparameters (e.g., feed-forward neural networks with differing numbers of hidden layers) do not count as distinct. However, different classes of neural architectures (e.g., RNNs vs. feed-forward neural networks) are considered distinct. At least one of our methods will include some component (e.g., an architecture, embedding, optimizer, etc …) first published in a contemporary research paper on machine translation, i.e., something published in a top machine learning conference in the past 3 years.

All implementations will be completed in Python; at the end of the project, we will submit all code along with our final report, following good coding practices and thorough documentation.

Evaluation:

To standardize comparison across methods, we propose to reserve the last 10% of the dataset as a test set and report each method's performance using this test set.

We will evaluate each model using the BLEU metric, using pre-built methods to compute this score where applicable (e.g., PyTorch's Torchtext implementation).

Project Timeline:

- By the proposal deadline, we will have finalized the three methods we intend to implement from scratch. We will have a list of 3-4 papers we believe to be relevant to bring to our TA mentor for review.
- By the check-in deadline, we will have initial results for all from-scratch implementations (we plan to continue iterating on these until the final report deadline). We will further have critically reviewed and summarized 2-3 relevant papers in the field.
- By the final report deadline, we will have finished reviewing 4 works for the literature review / prior works section, completed all implementations, compiled our major findings from each approach, and produced a thorough analysis of the results.

Project Check-in Specifications

The following requirements must be met:

a. The check-in can be at most 4 pages, single-spaced. Pages containing only references/citations do not count against this limit.

b. You must typeset the writeup in LaTeX using the provided template, found here: https://www.overleaf.com/latex/templates/style-and-template-for-preprints-arxiv-bio-arxiv/fxsnsrzpnvwc

c. The writeup must have a descriptive title and contain the names and AndrewIDs of all group members who at some point contributed to the project, regardless of whether or not they are still enrolled in the course.

d. The writeup must be submitted as a group on Gradescope; you should have one submission per group.

Your check-in should contain the following components. If you are using the pre-designed project, you can refer to the project description as appropriate (e.g., instead of giving a long description of the problem and dataset).

1. Title and Author List
2. Problem and Dataset: See final write-up for details.
3. Related Works**:** See final write-up for details.
   Unlike the final write-up, include and critically review **at least 2 relevant papers in the field**. For each paper, you should present a takeaway on what the paper is about and how it relates to your project. We expect that **at least one paper is from the last three years**.
4. Methods**:** Clearly outline what you have done and plan to do. At this point, you should have a good idea of what a reasonable baseline is, as well as a vague idea of your proposed methodology. We expect at least one baseline method and at least one proposed method. Please be as specific as possible about your methodology: what dataset will you use? What libraries will you use? How will you evaluate the results? Will there be any human/qualitative evaluation?
5. Experimental Results: See final write-up.
   If you do not have all of your experiment results yet (and you likely will not!) please include empty skeleton plots and tables where your results will go.
6. Current Progress**:** Describe where you currently are in the project process. This could include (1) a checklist of completed and remaining tasks, (2) a breakdown of who on your team is working on each task, (3) a timeline with specific milestones, and/or (4) one or two paragraphs describing current progress. We also ask that you **share the GitHub link to your repository in this section**.
7. Data and code availability: See final write-up for details.
8. References. Does not count towards page limit.

Final Video Showcase Specifications

The following requirements must be met:

a. The final video showcase can be at most 4 minutes long. We might stop videos early and/or deduct points for videos that exceed this time limit.
b. The video must be submitted as a single video file (recommended: .mp4). The submission must play correctly and have clearly audible audio; any text/plots/tables shown in the video must be legible when viewed full-screen on a moderate-sized monitor.
c. The video must have a descriptive title and contain the names and AndrewIDs of all group members who at some point contributed to the project, regardless of whether or not they are still enrolled in the course.
d. The video must be submitted as a group on Gradescope; you should have one submission per group.

Your final video showcase should contain many of the same components as the check-in and final report. However, the video showcase should provide less detail and more intuition. E.g., you might focus on a subset of the methods or a subset of the experiments to home in on the points you want to communicate.

Grading Assessment and Rubric

The following rubric will be used to assess the final video showcase submissions:

● Completeness (20 pts) - all of the required components are present in your video i.e., title/author list, clear problem statement and dataset description, methods overview (including baseline(s) and proposed method(s)), evaluation protocol/metrics, results, and discussion/analysis.

● Technical Soundness (25 pts) - the methods and evaluation protocol are described accurately and follow rational machine learning principles/best practices as covered in the course; comparisons between methods are fair.

● Results Quality (20 pts) - results are communicated with clear, well-labeled plots/tables and are sufficient to support the main claims; the video highlights the most important comparisons and includes qualitative examples when appropriate.

● Discussion and Insight (15 pts) - the presentation goes beyond listing numbers and includes interpretation of results, explanations for observed behavior, and meaningful limitations/failure cases and next steps.

● Clarity (10 pts) - this portion of your grade will assess the quality and organization of your video; it is crucial that you present your work in a clear and understandable way given the 4-minute constraint.

● Formatting and Delivery (5 pts) - the video respects the 4 minute limit, audio is intelligible, and text/figures are readable when viewed full-screen; the submission adheres to the guidelines we have established above and includes author identification.

● Performance (5 pts) - finally, a small portion of your group's grade will be based on how well your methods work. Crucially, this is not a cross-group competition: many groups are exploring fundamentally different methods that render comparisons meaningless. Any set of reasonable results will receive the majority of the credit for this rubric item, with a small portion being reserved for truly exceptional performance.

Final Report Specifications

Your final report will consist of three deliverables. There will be a separate Gradescope submission for each of the following items:

1. A writeup: the details for what should go into the writeup can be found below. The following requirements must be met:
    a. The writeup can be at most 8 pages, single-spaced. Pages containing only references/citations do not count against this upper limit.
    b. You must typeset the writeup in LaTeX using the provided template, found here: https://www.overleaf.com/latex/templates/style-and-template-for-preprints-arxiv-bio-arxiv/fxsnsrzpnvwc
    c. The writeup must have a descriptive title and contain the names and AndrewIDs of all group members who at some point contributed to the project, regardless of whether or not they are still enrolled in the course.
    d. The writeup must be submitted as a group on Gradescope; you should have one submission per group.
2. A statement of individual work: each group member must independently write a short paragraph describing their contributions to the project and submit them individually to Gradescope. These will not be graded and will only be referenced in the (unlikely) event that we need to assign different grades to separate group members. These statements should not be included in the writeup and do not count in the page total for the report.
3. All code written for the project: every group must submit all the code they wrote to Gradescope. You may submit as many files as you need. Each file must have a meaningful name so that your project mentor can easily identify its purpose. If you wrote code in the form of Python notebooks, please convert those to .py files before submitting them to Gradescope. Submit your code as a group; you should have one code submission per group.

Writeup Details

Your writeup must at minimum contain the following components. As above, you don't need to repeat the common information about the pre-designed project.

1. Title and Author List
2. Problem and Dataset: Briefly describe the task and motivate its importance. Then describe the dataset, including where the data is sourced from and any potential limitations, issues or biases the data might suffer from.
3. Related Works: Your writeup should contain a brief (one or two page) review of the academic landscape surrounding your chosen topic. This should be a minimum of 4 relevant papers, at least two of which need to be from the last three years. The review should begin with the historical context: what were some of the pioneering works in the field and how did they influence more recent research? When discussing the current state of research in your chosen topic, **it is crucial that you do more than just list papers and methods**: you should analyze the content of the works that you've read by e.g., drawing connections between different lines of inquiry, comparing and contrasting approaches, finding limitations or weaknesses in one paper that are addressed by another, etc…
4. Methods: The bulk of your writeup should be a thorough, detailed description of all the models your group implemented for each of the approaches. We expect **at least one baseline method and at least one proposed method**. Crucially, you must demonstrate a deep understanding of all the methods you implemented. In addition, you should describe the training procedure(s) and hyperparameter optimization techniques. From your report, a technical reader should be able to replicate your results by following these descriptions, i.e., there should be no ambiguity as to how you implement your model. If applicable, you should also briefly detail any approaches you tried but ended up not working well.
5. Experimental Results: Show plots and/or tables of the performance of your algorithms and interpret what they mean; be sure to label all of your figures and explain the findings. You must also define all performance metrics you used for evaluation. Describe how the results in each of the experiments aligned or didn't align with your expectations. Make sure to provide confidence intervals where appropriate or standard errors when comparing methods. We also ask that you share the GitHub link to your repository in this section.
6. Discussion and Analysis: Finally, analyze your models and their corresponding results. Provide explanations for the relative performances you observed and highlight any limitations/shortcomings of your approaches. Comment on how you would further improve your methods.
7. Data and code availability: Include a link to the dataset you are using, self hosting if necessary. If you cannot publish the dataset (for example, due to privacy concerns) please explain these circumstances.

   Include a link to your GitHub repository in this section
8. References

Grading Assessment and Rubric

The following rubric will be used to assess the final report writeups:

- Completeness (20 pts) — all of the required components are present in your project as described above.
- Literature Review (10 pts) — the most important part of any literature review is synthesis, i.e., drawing connections and identifying trends in the area. The bulk of your literature grade will be determined by the level of analysis your group performs when reviewing the research. A simple list of papers is not sufficient: your writeup must demonstrate a deeper understanding of the research area.
- Technical Soundness (30 pts) — the methods you implement must be described in sufficient technical detail such that your project mentor can properly assess your work; a good rule of thumb is that a well-informed practitioner should be able to recreate your methods entirely from their descriptions. Your methods should follow machine learning principles/best practices as covered in the course, e.g., hyperparameter tuning should be done with a held-out validation dataset or cross-validation and not on the test dataset.
- Implementation Correctness (20 pts) — your code matches the description of your methods in the writeup; for this portion of your grade, we may manually inspect any or all of the code you submit. As such, your group should follow good coding practices, e.g., meaningful variable names and detailed comments. We reserve the right to deduct points if your code is unintelligible (given a good faith effort by your project mentor).
- Clarity (10 pts) — this portion of your grade will assess the quality and organization of your writeup; it is crucial that you present your work in a clear and understandable way.
- Formatting (5 pts) — your writeup must adhere to the guidelines we have established above, e.g., it respects the 8 page limit and uses the correct LaTeX template
- Performance (5 pts) — finally, a small portion of your group's grade will be based on how well your methods work. Crucially, this is not a cross-group competition: many groups are exploring fundamentally different methods that render comparisons meaningless. Any set of reasonable results will receive the majority of the credit for this rubric item, with a small portion being reserved for truly exceptional performance.