The following document outlines specific instructions for your class projects, including the types of contributions, schedules, evaluation criteria, and possible project types. We aim to give you a taste of the full pipeline of NLP research, including annotation and understanding of existing data, model replication, experimentation, and error analysis. In addition to addressing the same NLP problem with your teammates, this project is an excellent chance to communicate with NLP researchers. You can make your project an extension of your homework, but it must have novel extensions and research contributions beyond your homework.

Novel contributions to your project can be made in a variety of ways. The problem definition section of your proposal and report should clearly state what types of contributions your project will make:

- Critical analysis of existing model/dataset (default project),
- New research results judged suitable for acceptance to an NLP or ML workshop,
- Collection of your own dataset on new problems or adversarial datasets that can fool the existing systems,
- An in-depth literature survey on emerging topics,
- Interactive demonstration (e.g., Chrome Extension, Flask) or visualization of existing systems,
- Applying NLP tools to your own domain of research (e.g., psychology, chemistry, robotics),
- New open-source repository or dataset with a high impact on the community
- Others (consult instructors if you wish to do other types of projects).

Project Deliverables and Due Dates

Your project takes up 25% of your class grade. Below is the list of your deliverables with due dates:

- Proposal report (5 points, due: Mar 13)
- Midterm office hour participation (5 points, due: Apr 10)
- Poster presentation (5 points, due: Apr 24/26)
- Final report (10 points, due: May 5)

Proposal report Your project proposal should not exceed three pages. For this report, you must use the unofficial ACL-style template (link). In case you haven't used LaTex for scientific writing, this is a great opportunity to learn how scientists and researchers write their manuscripts using this typesetting tool called LaTex. Here is a tutorial for LaTex with Overleaf. Please upload your PDF report on Canvas before the deadline. The following items should be included in your proposal:

- Team name, members, and role assignment
- Title
- Motivation
- Literature survey
- Problem definition
- Proposed idea
- Broader impact

You will be assigned a project mentor (DK, Risako, or Deb) with feedback on your report within two weeks of submitting your proposal.

Midterm office hour participation Your team must schedule an office hour meeting with your assigned mentor (15 to 20 minutes) before the due date (Apr 10) to discuss your intermediate results

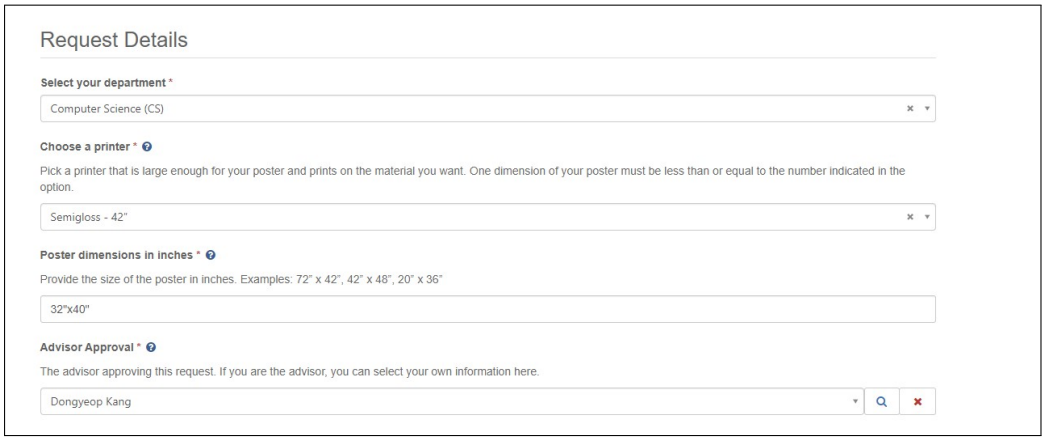
and progress. The mentor expects you to give an update on your progress, ask questions, and consult with your plan until the final presentation and report.

- Recap: Motivation, Literature Survey, Problem definition (based on your proposal)
- Detailed proposed ideas
- Novel contribution compared to prior work
- Preliminary results and comparison to the baseline performance
- Plan until the end of the semester

Summarize the feedback you receive from your mentor and explain how you plan to address it. By the deadline, your response must be uploaded on Canvas.

Poster presentation Everyone on your team should present their work at your assigned poster session on April 24 (Group A) or April 26 (Group B). Your group assignment will be provided a week in advance. You should stand by your poster during class (4:00pm to 5:00pm) and communicate with instructors and other students (and possibly external guests). If you are a completely UNITE team, please upload your poster to the shared drive (for other students to be able to view it). A PDF or PowerPoint file of your poster must also be uploaded on Canvas before your presentation.

Your poster presentation will be held in Shepherd 164 (aka Drone Lab). Around ten aisles will be set up for you to attach your poster for each poster session. Prepare your presentation and attach your poster 10 minutes before you arrive. Your 32" x 40" poster must be printed by you. Please carefully read the printing instructions and request poster printing through the form. Remember, CSE-printing guarantees posters submitted 2 business days in advance, but **do not** work on the weekends. Some details on how to fill out the initial fields in the request form are given below.



Both Instructors and peers will review your poster presentation. For peer group review points, every group is assigned a random team on their session day to review based on a rubric provided by instructors. If you are a UNITE team, you can complete your review for your assigned team online through a form that we will make available. You can optionally (no points for this) vote for the best poster. The team winning "best poster" will be given extra credit as a reward.

- Recap: Motivation, Literature Survey, Problem definition, Proposed ideas, Novel contribution
- More experimental results and main findings
- Plan for the final report

Final report There should be a maximum of eight pages in the final report with unlimited references and appendices. You can submit a PDF report and zipped code (or link to your github) on Canvas.

We are not evaluating if you succeeded or failed at accomplishing what you initially decided to do. It is totally fine if your results are negative and not significant. The essential criteria we evaluate for your project are whether you put in a reasonable effort, deeply understand the nature and challenges of your problem, set a reasonable hypothesis to tackle the challenges, and can clearly communicate them with others. For detailed assessment guidelines, please see below or visit this link. Here is our rubric for project evaluation:

```
Rubrik (100 points) for Final Report

Introduction / Background / Motivation:
(5 points) What problem do you try to solve? Describe your objectives clearly
↪  without using any technical jargon.
(5 points) How is it done today by other researchers? What are the limitations
↪  and challenges of current practice?
(5 points) Who might be interested in your work? What kinds of impact can you
↪  make?

Approach:
(10 points) What did you do exactly? How did you solve the problem? Why did
↪  you think it would be successful? What is your hypothesis? What is the
↪  scientific novel of your approach?
(5 points) What challenges did you anticipate and/or encounter during the
↪  development of your approach? Did the very first thing you tried work?

Experiments / Results / Error Analysis:
(10 points) How did you measure success? What research questions do you want
↪  to validate? What evaluation metrics and experiments were used? What were
↪  the results, both quantitative and qualitative? Did you succeed? Did you
↪  fail?
(10 points) No matter you succeed or fail, why? Which data points are
↪  incorrectly predicted by yours but previous models can't, or vice versa?
(10 points) Are there still some failure cases? Why can't your approach
↪  address them? Any potential solutions?
(5 points) Are the ideas/problems/results presented with appropriate
↪  illustrations?

Discussion points:
(5 points) Replicability: How easily are your results able to be reproduced by
↪  others?
(10 points) Datasets: Did your dataset or annotation affect other people's
↪  choice of research or development projects to undertake?
(10 points) Ethics: Does your work have potential harm or risk to our society?
↪  What kinds? If so, how can you address them?
(10 points) Discussion: What limitations does your model have? How can you
↪  extend your work for future research?

At least some of these questions and others should be relevant to your project
↪  and should be addressed in the final report. You do not need to address
↪  all of them in full detail. Some may be irrelevant to your project and
↪  others may be standard and thus require only a brief mention.
```

Tips for successful class projects

Don't be ambitious. Please keep in mind that you only have a few months to work on this project, while you are taking other classes. I don't expect you to produce a publishable research outcome. Again, we are aiming to ensure you have gone through the entire NLP research pipeline, from data selection/creation to model replication/building to experiments and error analysis. You will receive a good grade if your final report shows that you have done a reasonable amount of work on the problem you set along the way.

Use existing tools. As part of your homework, you will learn various tools and libraries. Instead of implementing something from scratch, use them for your projects. You might consider using the following tools for your project:

- Tutorials on PyTorch and HuggingFace programming
- Huggingface model and data cards
- The state-of-the-art models in different NLP tasks/datasets in PapersWithCode leaderboard
- Wandb and tensorboard for tracking your training
- Demonstrate your tool using Gradio
- Other python libraries such as transformers, diffusers, timm, datasets, safetensors, accelerate, optimum, tokenizers, evaluate, simulate, and more.
- ChatGPT, GPT3 Playground, and other LLMs inference

Secure computing resources. In order to conduct your experiments, you will need some computing resources. Ensure that your department has GPU computing resources, such as MSI. Additionally, you can use publicly available resources such as Google Cloud/CoLab and Amazon AWS.

The DOs for successful projects

- Clearly divide work between team members for optimal collaboration process
- Start early and work on it regularly every week rather than rush at the end
- Set up workflow – download data, verify data, set up base code on github, communicate via Slack, sharing results on Google spreadsheet, etc
- Have a clear, well-defined hypothesis to be tested (+ novel/creative hypothesis)
- Conclusions and results should provide some insights
- Meaningful tables and plots to display the key results
  - nice visualizations or interactive demos
  - novel/impressive engineering feat
  - good empirical results in both qualitative and quantitative ways.

The Don'ts for successful projects

- Data not available or hard to get access to, which stalls progress
- All experiments run with prepackaged source – no extra code written for model/data processing
- Team starts late – only draft of code up before dues
- Just ran the model once or twice on the data and reported results (not much hyper-parameter tuning and statistical significance test)
- A few standard graphs (loss curves, accuracy) without any analysis
- Results/Conclusion don't say much besides that it didn't work. Negative results are fine, but only with in-depth analysis and justification

Example Project Topics and Projects from Previous Classes

The following are examples of project topics from previous classes:

- Experiment with improving an architecture on a well-defined NLP task
- Case study: apply an architecture to a dataset in the real world that has not been done before
- Compete in a predefined competition (SemEval 20XX, Kaggle, etc)
- Stress test on comparison study of known models/architecture (e.g., when are LSTMs better than Transformers for task XYZ?)
- Design a novel NN layer, objective function, etc on NLP tasks
- Multi-domain/Multi-lingual NLP (RL+NLP, CV+NLP, Social Science + NLP)
- Visualization/Interpretability/controllability study of NN models
- In-depth error analysis on XYZ datasets using the sota model
- Collection of adversarial datasets for XYZ tasks
- Human evaluation on current NLG evaluation metrics
- Collect a new dataset of interesting language variation or cognition
- Capacity of latest large language models (e.g., GPT3, chatGPT) on existing or new tasks

You can access the final reports for previous NLP classes by clicking the link below:

- Understanding Narrative Transportation in Fantasy Fanfiction, by Kelsey Neis and Yu Fang, CSCI 8980 S22 (final report)
- Exploring Episodic Memory through Cross-modal representations, by Abhiraj Mohan, Emily Mulhall, Jayant Sharma, CSCI 8980 S22 (final report)
- Generating Controllable Long-dialogue with Coherence, by Zhecheng Sheng, Chen Jiang, Tianhao Zhang, CSCI 5980 F22 (final report)
- Cross-lingual Transfer Learning for Irony Detection, by Chen Hu, Jiaqi Liu, Keyang Xuan, CSCI 5980 F22 (final report)