NLU-DL Project Guidelines

Page Limit

- **6 pages (excluding appendix).**
- Follow the Overleaf Template (click *Menu* → *Copy Project* to duplicate).
- Refer to sample project reports from Stanford's CS224N course for style and structure.

Good LaTex practices

- Use booktabs package for tables
- Use biblatex for citations

Title

Title: A working title. This can be changed later.

Authors: Names of all project partners. If you have a mentor (other than me), you should also indicate them as a co-author.

Role: Under each author name, mention the role of the author (e.g. student or mentor)

Affiliation: Indicate which department and university of each author.

Abstract (200 words):

Indicate in 200 words what the project is about, why it is important, and what is novel. What are/will be the findings of the project?

1. Introduction

Briefly motivate your project and define the problem.

- Start with a broader NLP trend, then narrow to your specific question or hypothesis.
- Highlight what is novel or unexplored.
- Explain why this problem matters and what insights you expect to gain.
- Include a small illustrative figure if possible.
- End with a short summary of your proposed approach and contributions.

2. Related Work

Situate your project within the existing literature.

- Discuss **aspects** of the problem (e.g., modeling choices, data, evaluation) and how prior work approached them.
- Explain **how your work differs or builds on** these approaches.
- Focus on clarity, not quantity, only include relevant, influential papers.
- Avoid making unsupported or dismissive claims about others' work.

3. Data and Environment

Describe the datasets and tools used.

- **Data:** Sources, examples, and key statistics (train/validation/test splits, number of labels, length distributions, etc.).
- **Evaluation metrics:** Define and justify your chosen metrics.
- **Environment:** Mention frameworks (e.g., PyTorch), libraries, and any computational setup.

4. Methods

Explain your models and hypotheses.

- State the **research hypothesis** clearly (e.g., “Word order improves cross-lingual alignment”).
- Include at least two model types:
    - **Baselines:** Describe standard or prior methods you compare against. Use the same data/resources for fair comparison.
    - **Proposed Model(s):** Explain your main idea, architecture, and key innovations.
- Provide high-level math and diagrams where helpful.
- Explain *why* your model should outperform baselines.

5. Experiments and Results

Summarize your experimental setup and findings.

- **Training details:** Hyperparameters, number of runs (recommend ≥3), epochs, and hardware.
- **Example Results Table could look like:**

Model	Validation	Test	Notes
Baseline (reported)	mean ± std	mean ± std	from paper
Or Baseline (reimplemented)	mean ± std	mean ± std	your run
Proposed Model	mean ± std	mean ± std	your main result

- **Analysis:**
    - Discuss why results look the way they do.
    - Highlight where your model improves or fails.
    - Provide qualitative examples if useful.

6. Discussion

Interpret your findings.

- What do the results say about your hypothesis?
- Any unexpected observations?
- Limitations or open questions?
- Possible future directions.

7. Conclusion

Summarize your contributions and findings.

- State main takeaways and next steps.
- Include your **GitHub repository** link.

Appendix (Optional, up to 1 page)

- Extended dataset samples or statistics.
- Additional experimental results.
- Implementation details not central to the main discussion.

Team Contributions

If working in a group, list each member's role clearly. Example:

Aditi: Designed experiments and implemented models.
Alex: Conducted data analysis and wrote the Methods section.
Siva (mentor): Provided guidance and feedback.

Publication and Beyond

Outstanding projects may be developed into conference or workshop papers (ACL, NAACL, COLING, etc.).