

11-891 Final Report

First Last 1* **First Last 2*** **First Last 3*** **First Last 4***
{ID1, ID2, ID3, ID4}@andrew.cmu.edu

Abstract

In this project we introduce ... to tackle... we
show that ...

1 Introduction and Problem Definition (1-1.25 pages) [1 point]

This will be about 1 page, and lays out the contributions.

*Everyone Contributed Equally – Alphabetical order

2 Related Work and Background [1 point]

Integrate your previous report here

3 Task Setup and Data [1 point]

Integrate your previous report here

4 Baselines [1 point]

Integrate your previous report here

5 Proposed Method (>1 page) [5 points]

This is not a copy and paste from R2, it should be updated to reflect the final proposed model. Explain if there are multiple versions (e.g. your model has possible extensions, different amounts of exemplars, etc). Please also include an easy to read model diagram here.

The below are example subsections – yours may vary!

5.1 Loss functions

5.2 Changes to training data

5.3 Hyperparameters and their effects

Methods	Dev	
	Accuracy ↑	Interaction Turns ↓
Baseline 1 ()		
Baseline 2 ()		
Comparison Method 1 (if applicable) ()		
Comparison Method 2 (if applicable) ()		
Proposed Method (v1)		
Proposed Method (v2)		
Proposed Method (final)		

6 Results (1 page) [5 points]

Replace columns with the correct metrics for your task (extrinsic). Include multiple versions of your final model. You do not need to run on the test set but are encouraged to try if you have nice results on Dev.

Methods	Dev	
	Oracle Accuracy \uparrow	Candidates Sampled \downarrow
Baseline 1 ()		
Baseline 2 ()		
Proposed Method		

7 Analysis (2 pages)

This section should include plots. For example, how key metrics vary with a specific hyperparameter, task complexity, etc.

7.1 Intrinsic Metrics (1 page) [5 points]

These are not the task itself but indicative of how well the approach is doing on tasks related to the downstream, extrinsic metrics. For example, if you are using a reranking method, you could compute the oracle accuracy that chooses the best candidate from the sampled set of candidates.

7.2 Qualitative Analysis and Examples (1 page) [5 points]

We want you to have illustrative examples from your data here. The best way to structure this is likely a full page table that includes: Example input, comparison of model outputs, theory about reason for failure. But, you can also have it be primarily text, if you support each of your points in the text with real examples.

8 Future work and Limitations (0.5-1 page) [3 points]

Please thoroughly discuss the limitations of the proposed model. Where are some places it does not work, phenomena it does not capture, places that ideally it can be improved.

For future work, if you were to extend this project over another 4 or so months of work, what would you be interested to do?

9 Team member contributions

Member 1 contributed ...

Member 2 contributed ...

... contributed ...

References