

Report 2

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

This template is a recommended format for the report. You can use a different format if you prefer, but we will assign points following the breakdown given here. In addition to the points by each section, we'll also have 3 points for report clarity.

1 Task and Dataset(s) (less than 1 page) [1 point]

Give us a quick (e.g. 1 paragraph) reminder of the task and dataset(s) you're working with. Do you have all the data you need for the project? If you've been collecting data, give an update on it here. How far are you from completing data collection?

*Everyone Contributed Equally – Alphabetical order

2 Models (1 page) [4 points]

2.1 Baselines

Please explain the baselines you are running. These can be models or approaches from past work, but you need to be running them yourselves so that you can analyze their results and outputs.

1. Baseline 1 is ... We expect this to capture
2. Baseline 2 is ... We expect this to capture

2.2 Comparison Methods

If there are past approaches to this task and dataset that you'll be comparing to, but not running yourself, describe them briefly here.

1. Comparison Method 1 is ... whose key insight is ...
2. Comparison Method 2 is ... whose key insight is ...

Methods	Dev	
	Accuracy \uparrow	Interaction Turns \downarrow
Baseline 1 ()		
Baseline 2 ()		
Comparison Method 1 (if applicable) ()		
Comparison Method 2 (if applicable) ()		

3 Results (1 page) [4 points]

Replace columns with the correct metrics for your task (extrinsic). Give text descriptions of the findings as well.

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

4 Analysis

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

4.1 Intrinsic Metrics (1 page) [4 points]

These are not the task itself but are indicative of how well the approach is doing on related skills.

4.2 Qualitative Analysis and Examples (1 page) [4 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.

5 Updated proposal for your method (1 page) [5 points]

Based on the analysis and findings you've had here, what do you plan to do for the final stage of the report? Make this self-contained, so that we don't need to refer back to Report 1. Use one or more figures as appropriate.

6 Project-related

6.1 Comments and questions

If you foresee any challenges, or have open questions or want feedback on specific things, you can mention them here.

6.2 Team member contributions to this report

Member 1 contributed ...

Member 2 contributed ...

... contributed ...

References