

# Reproducibility Report Instructions for CSE 481N

List Authors Here

Spring 2024

## Reproducibility Summary

This summary section should be less than one page.

### Motivation

Briefly explain why the original paper’s research question is worthy of a reproducibility attempt. Consider, for example, public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors and impacts of the resulting technology in these contexts.

### Scope of Reproducibility

State the main claim of the original paper you are trying to reproduce. We recommend picking the central claim of the paper.

### Methodology

Briefly describe what you did and which resources did you use. E.g., did you use author’s code, did you reimplement parts of the pipeline, how much time did it take to produce the results, what hardware you were using and how long it took to train/evaluate.

### Results

Start with your overall conclusion—where was your study successful and where not successful? Be specific and use precise language, e.g., “we reproduced the accuracy to within 1% of reported value, that upholds the paper’s conclusion that it performs much better than baselines.” Getting exactly the same number is in most cases infeasible, so you’ll need to use your judgment to decide if your results support the original claim of the paper.

### What was Easy

Describe which parts of your reproduction study were easy. E.g., was it easy to run the author’s code, or easy to reimplement their method based on the description in the paper. The goal of this section is to summarize to the reader which parts of the original paper they could easily apply to their problem.

### What was Difficult

Describe which parts of your reproduction study were difficult or took much more time than you expected. Perhaps the data was not available and you couldn’t verify some experiments, or the author’s code was broken and had to be debugged first. Or, perhaps some experiments just take too much time/resources to run and you couldn’t verify them. The purpose of this section is to indicate to the reader which parts of the original paper are either difficult to reuse, or require a significant amount of work and resources to verify.

### Communication with Original Authors

Briefly describe how much (if any contact) you had with the original authors.

## **1 Introduction**

A few sentences placing the work in context. Limit it to a few paragraphs at most.

Since your report is on reproducing a piece of work, you don't have to heavily motivate that work, but you should explain why the original paper's research question is worthy of a reproducibility attempt. Consider, for example, public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors and impacts of the resulting technology in these contexts.

Make it clear what the original paper is about and what its contributions are.

## **2 Scope of Reproducibility**

Explain the claims from the paper you picked for the reproduction study and briefly motivate your choice. We recommend picking the claim that is the central contribution of the paper. To find what this contribution is, try to summarize the most important result of the paper in 1–2 sentences, e.g., “This paper introduces a new activation function X that outperforms a similar activation function Y on tasks Z, V, and W.”

Make the scope as specific as possible. It should be something that can be supported or rejected by your data. For example, this scope is too broad and lacks precise outcome (what is “strong performance?”): “Contextual embedding models have shown strong performance on a number of tasks across NLP. We will run experiments evaluating two types of contextual embedding models on datasets X, Y, and Z.”

This scope is better because it's more specific and has an outcome that can be either supported or rejected based on your work: “Finetuning pretrained BERT on SST-2 will have higher accuracy than an LSTM trained with GloVe embeddings.”

### **2.1 Addressed Claims from the Original Paper**

Clearly enumerate the claims you are testing:

1. Claim 1
2. Claim 2
3. Claim 3

## **3 Methodology**

This section is to explain your approach—did you use the author's code, did you aim to reimplement the approach from the paper description? Summarize the resources (code, documentation, GPUs) that you used.

### **3.1 Model Descriptions**

Describe the models used in the original paper, including the architecture, learning objective and the number of parameters.

### **3.2 Datasets**

Describe the datasets you used and how you obtained them.

### **3.3 Hyperparameters**

Describe how you set the hyperparameters and what was the source for their value (e.g., paper, code, or your guess).

### **3.4 Implementation**

Describe whether you use the existing code or write your own code, with the link to the code and which language/packages were used. Note that the github repo you link should be public and have a clear documentation.

You'll likely use a command like `includegraphics` to include images here.

Figure 1: Consider using figures and tables to present information as clearly as possible.

You'll probably use a `tabular` environment here.

Table 1: Consider using figures and tables to present information as clearly as possible.

### 3.5 Experimental Setup

Explain how you ran your experiments, e.g. the CPU/GPU resources and provide the link to your code and notebooks.

### 3.6 Computational Requirements

Provide information on computational requirements for each of your experiments. For example, the number of CPU/GPU hours and memory requirements. Mention both of your estimation before running the experiments and actual resources took for reproducing the experiments. You'll need to think about this ahead of time, and write your code in a way that captures this information so you can later add it to this section.

## 4 Results

Start with a high-level overview of your results. Does your work support the claims you listed in section 2.1? Keep this section as factual and precise as possible, reserve your judgment and discussion points for the "Discussion" section that comes later.

Go into each individual result you have, say how it relates to one of the claims and explain what your result is. Logically group related results into sections. Clearly state if you have gone beyond the original paper to run additional experiments and how they relate to the original claims.

Tip 1: Be specific and use precise language, e.g. "we reproduced the accuracy to within 1% of reported value; that upholds the paper's conclusion that it performs much better than baselines." Getting exactly the same number is in most cases infeasible, so you'll need to use your judgment to decide if your results support the original claim of the paper.

Tip 2: You may want to use tables and figures to demonstrate your results. Any time you include a figure or table, like Figure 1 or Table 1, make sure it's referenced in the text.

#### 4.1 Result 1

#### 4.2 Result 2

#### 4.3 Additional Results not Present in the Original Paper

Describe any additional experiments beyond the original paper. This could include experimenting with additional datasets, exploring different methods, running more ablations, or tuning the hyperparameters. For each additional experiment, clearly describe which experiment you conducted, its result, and discussion (e.g., what is the indication of the result).

## 5 Discussion

Describe larger implications of the experimental results, whether the original paper was reproducible, and if it wasn't, what factors you believe made it irreproducible.

Give your judgment on whether the evidence you got from your experiments supports the claims of the paper. Discuss the strengths and weaknesses of your approach—perhaps you didn't have time to run all the

experiments, or perhaps you did additional experiments that further strengthened the claims in the paper.

### 5.1 What was Easy

Describe which parts of your reproduction study were easy. E.g., was it easy to run the author's code, or easy to reimplement their method based on the description in the paper. The goal of this section is to summarize to the reader which parts of the original paper they could easily apply to their problem.

Tip: Be careful not to give sweeping generalizations. Something that is easy for you might be difficult to others. Put what was easy in context and explain why it was easy (e.g., code had extensive API documentation and a lot of examples that matched experiments in papers).

### 5.2 What was Difficult

Describe which parts of your reproduction study were difficult or took much more time than you expected. Perhaps the data was not available and you couldn't verify some experiments, or the author's code was broken and had to be debugged first. Or, perhaps some experiments just take too much time/resources to run and you couldn't verify them. The purpose of this section is to indicate to the reader which parts of the original paper are either difficult to reuse, or require a significant amount of work and resources to verify.

Tip: Be careful to put your discussion in context. For example, don't say "the math was difficult to follow," say "the math requires advanced knowledge of calculus to follow."

### 5.3 Recommendations for Reproducibility

Describe a set of recommendations to the original authors or others who work in this area for improving reproducibility.

#### Communication with Original Authors

Document the extent of (or lack of) communication with the original authors. To make sure the reproducibility report is a fair assessment of the original research we recommend getting in touch with the original authors. You can ask authors specific questions, or if you don't have any questions you can send them the full report to get their feedback.

For citations: Make sure you understand the difference between citing work as a noun phrase (`citet`) and using a parenthetical (`citep`). The first one looks like Clark et al. [2018], and the second one is different [Lin et al., 2018].

For the references section: Use bibtex and check its output; manual corrections are often necessary. Put your bibtex references in a file called `refs.bib`. The citation to Press and Smith [2018] exemplifies the style I prefer for arXiv papers; when a paper has been published in a peer reviewed forum, cite that version instead of the arXiv version.

#### References

Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. Creative writing with a machine in the loop: Case studies on slogans and stories. In *Proceedings of the Conference on Intelligent User Interfaces*, Tokyo, Japan, March 2018. URL <http://homes.cs.washington.edu/~nasmith/papers/clark+ross+tan+ji+smith.iui18.pdf>.

Lucy H. Lin, Scott B. Miles, and Noah A. Smith. Natural language processing for analyzing disaster recovery trends expressed in large text corpora. In *IEEE Global Humanitarian Technology Conference*, San Jose, California, October 2018. URL <https://homes.cs.washington.edu/~nasmith/papers/lin+miles+smith.ghct18.pdf>.

Ofir Press and Noah A. Smith. You may not need attention, October 2018. URL <https://arxiv.org/pdf/1810.13409>.

## **A Example Appendix Section**

### **A.1 Example Appendix Subsection**

For completeness, you may optionally include appendices, but don't expect us to read them. Make sure the main body of your paper is complete and meets the rubric requirements on its own. For example, you might show a table of results for a main experiment in your paper, and put tables for additional experiments in an appendix. But you should still qualitatively and comprehensively report on all results that you intend to include in your paper, e.g., "a similar trend holds for the other datasets (full results in §A.1)."

Appendices must be well-organized: separate appendices (with descriptive titles) should be created for separate topics or datasets, and be properly grouped and numbered (e.g., A.1, A.2, A.3, B.1, B.2. . . ). When you reference an appendix, it should be clear what it's about, and you should link to it (e.g., "detailed prompts can be found in §B.1").

### **A.2 Example Appendix Subsection**

## **B Example Appendix Section**

### **B.1 Example Appendix Subsection**