

General Reproducibility Guidelines for AI Research

Version: 1.3 June 25, 2020

Authors: Odd Erik Gundersen, Yolanda Gil, Mausam

For each experiment, check that the following is described:

- How the experimental design rigorously tests the claims.
- The evaluation metrics and the motivation for choosing these metrics.
- All (hyper-)parameters for each model/algorithm, number and range of values tried per parameter, and the criterion for selecting best parameter setting.
- The final parameters for each model/algorithm.
- The computing infrastructure used for running the experiment (hardware and software), such as which software and version (libraries, frameworks, operating system etc), processing units (GPU/CPU), memory and more.
- For each reported result, the number of algorithm-runs it is averaged over and its variance.

For data used in the paper, check the following:

- For closed datasets, describe the dataset.
- For a new dataset, deposit it to a public repository with a description and metadata.
- For a new dataset, release it with a license that allows free usage for research purposes.
- All open datasets are cited.

For all code, check the following:

- All source code required for conducting the experiment is shared and cited.
- The version of the code used for conducting the experiments is specified.
- A license is added with the source code to allow free usage for research purposes.

For the paper, check the following:

- Claims being investigated are stated clearly.
- For theoretical papers, complete proofs are provided (for example in the appendix).
- Assumptions and limitations are identified.
- A conceptual outline and pseudo code describing the AI method is given.
- Statements about how the results substantiate the claims.

These guidelines are based on the following resources: 1) Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications by Gundersen, Gil and Aha, AI Magazine 2018, 2) The ICRM criteria generated by at the 2012 Workshop “Reproducibility in Computational and Experimental Mathematics” and 3) The Machine Learning Reproducibility Checklist (version 1.2).

The latest version of the guidelines: http://folk.idi.ntnu.no/odderik/reproducibility_guidelines.pdf

Support on how to implement guidelines and feedback form: http://folk.idi.ntnu.no/odderik/reproducibility_guidelines_how_to.html