

# The Machine Learning Reproducibility Checklist (v2.0, Apr.7 2020)

For all **models** and **algorithms** presented, check if you include:

- ☐ A clear description of the mathematical setting, algorithm, and/or model.
- ☐ A clear explanation of any assumptions.
- ☐ An analysis of the complexity (time, space, sample size) of any algorithm.

For any **theoretical claim**, check if you include:

- ☐ A clear statement of the claim.
- ☐ A complete proof of the claim.

For all **datasets** used, check if you include:

- ☐ The relevant statistics, such as number of examples.
- ☐ The details of train / validation / test splits.
- ☐ An explanation of any data that were excluded, and all pre-processing step.
- ☐ A link to a downloadable version of the dataset or simulation environment.
- ☐ For new data collected, a complete description of the data collection process, such as instructions to annotators and methods for quality control.

For all shared **code** related to this work, check if you include:

- ☒ Specification of dependencies.
- ☒ Training code.
- ☒ Evaluation code.
- ☒ (Pre-)trained model(s).
- ☒ README file includes table of results accompanied by precise command to run to produce those results.

For all reported **experimental results**, check if you include:

- ☐ The range of hyper-parameters considered, method to select the best hyper-parameter configuration, and specification of all hyper-parameters used to generate results.
- ☐ The exact number of training and evaluation runs.
- ☐ A clear definition of the specific measure or statistics used to report results.
- ☐ A description of results with central tendency (e.g. mean) & variation (e.g. error bars).
- ☐ The average runtime for each result, or estimated energy cost.
- ☐ A description of the computing infrastructure used.

This version (v. 2.0), is renamed the Machine Learning Paper Paper Reproducibility Checklist, and is designed to be used simultaneously with the ML Code Submission checklist ([ADD LINK](#)). This change clearly acknowledge that the paper and the code are two separate research artefacts, each with their own checklist.

A few practical suggestions for using the checklist during a paper submission process (conference or journal):

- For each question, have 2 answer fields.
  - The first one is categorical, with choices: {Yes, No, Not applicable}.
  - The second one is for a free-form comment.
- The main role of the second field is to allow users to provide more detailed information, when it is not a clear yes/no distinction, or for the reasons why it is not applicable.
- If the checklist is used during paper submission for a conference, given the time pressure that often accompanies such deadlines, it is recommended to require an initial submission of the checklist at an earlier date, such as with the abstract submission (e.g. 1 week earlier), while allowing authors to update the checklist up to the paper deadline.

For a detailed analysis of the use of version 2.1 of the ML reproducibility checklist at NeurIPS 2019:

- [Add arXiv link.](#)

For previous versions of the checklist, see:

<https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf>