General Guidelines for Making Empirical AI Research Reproducible

Version: 1.0 Nov 10, 2019
Authors: Odd Erik Gundersen and Yolanda Gil

Experiment:
- Are the key aspects of the experiment clearly described?
- Is the motivation of choosing the performance metrics and statistics used for reporting the results stated concisely?
- Is the configuration of the experiment described clearly, including parameters, their ranges and how they were selected?
- Is the computing infrastructure used for running the experiment described, such as hardware used and software dependencies (frameworks, libraries, operating system)?
- Are error bars with mean and variance of the results presented?

Code:
- Is the code implementing the main methods shared in a code repository and properly cited?
- Is the ancillary code implementing the experiment (such as configuration, pre-processing of data, hyperparameters tuning, etc.) shared in a code repository?
- Is the code used for analyzing and visualizing the results shared in a code repository?
- Does the code include a license?
- Is the version of the code indicated and shared using a persistent identifier?
- Is the code properly documented?

Data:
- Is the data shared in a public repository and properly cited?
- Is it stated which samples are used for training, validation and testing?
- Are the execution results produced by the experiment shared?
- Is the data shared using a persistent identifier?
- Is a license describing how the data can be used by others included?
- Is the data documented with metadata?

Paper (re-read the paper and check that it contains the following):
- A statement describing the problem that is being investigated.
- Explanations of the assumptions that are made.
- Description of the approach and claims that are being investigated.
- A conceptual outline of the AI method, if relevant.
- Pseudo code describing the AI method, if relevant.
- Statements about why the experimental design rigorously tests the claims.
- Statements about how the results substantiate the claims.
- Concise description of the overall findings.

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