

## Documenting AI research

IFCC WG-AIGD (Kricka LJ, Baudhuin LM, Fortina P, Ertel A, Hope T, McCudden C, Park JY, Polevikov S, Satchkov D)

Concerns over various aspects for AI-based research and publications (e.g., credibility, reproducibility, documentation of the research, algorithm validation) have been raised by several authors (D'Amour et al; Ross 2021). A particular concern relates to documenting AI research so that the results are reproducible (Gundersen 2019, Gundersen and Kjensmo, 2018; Gundersen et al 2018; Halbe-Kains et al 2020; Isdahl and Gundersen 2019).

Gundersen and colleagues have formulated an author checklist for documenting AI research that is based on best practices proposed by various scientific organizations, scholars, and publishers (Gundersen et al, 2018). The 23-item checklist of recommendations includes five recommendations for data, five recommendations for source code implementing AI methods and experiments, three recommendations for AI methods, and ten recommendations for experiments described in publications (Table 1).

Implementation of this checklist would be valuable for documentation of AI research by authors, and also for reviewers of AI research submitted for publication.

### **Table 1. Recommendations for data, source code, AI methods, and experiments in AI publications (adapted from Gundersen et al 2018)**

#### **A. Data mentioned in a publication should:**

1. Be available in a shared community repository, so anyone can access it
2. Include basic metadata, so others can search and understand its contents
3. Have a license, so anyone can understand the conditions for reuse of the data
4. Have an associated digital object identifier (DOI) or persistent URL (PURL) so that the data is available permanently
5. Be cited properly in the prose and listed accurately among the references, so readers can identify the datasets unequivocally and data creators can receive credit for their work

**B. Source code used for implementing an AI method and executing an experiment should:**

6. Be available in a shared community repository, so anyone can access it
7. Include basic metadata, so others can search and understand its contents
8. Include a license, so anyone can understand the conditions for use and extension of the software
9. Have an associated digital object identifier (DOI) or persistent URL (PURL) for the version used in the associated publication so that the source code is permanently available
10. Be cited and referenced properly in the publication so that readers can identify the version unequivocally and its creators can receive credit for their work

**C. AI methods used in a publication should be:**

11. Presented in the context of a problem description that clearly identifies what problem they are intended to solve
12. Outlined conceptually so that anyone can understand their foundational concepts
13. Described in pseudocode so that others can understand the details of how they work

**D. Descriptions of experiments in a publication should:**

14. Explicitly present the hypotheses to be assessed, before other details concerning the empirical study are presented
15. Present the predicted outcome of the experiment, based on beliefs about the AI method and its application
16. Include the experimental setup (parameters and the conditions to be tested) and its motivation, such as why a specific number of tests or data points are used based on the desired statistical significance of results and the availability of data
17. Present the results (i.e., measures and metrics) and the analysis
18. [Include] an explicit indication of whether the analyses support the hypotheses
19. Justify why the datasets used are appropriate for the experiment, why the chosen empirical design is appropriate for assessing the hypothesis, and why the metrics and measures are appropriate for assessing the results
20. Be described as a workflow that summarizes how the experiment is executed and configured

21. Include documentation on workflow executions or execution traces that provide parameter settings and initial, intermediate, and final data
22. Specify the hardware used to run the experiments
23. Be cited and published separately when complex, so that others can unequivocally refer to the individual portions of the method that they reuse or extend.

## References

- D'Amour A, Heller K, Moldovan et al. Underspecification presents challenges for credibility in modern machine learning. [arXiv:2011.03395v2](https://arxiv.org/abs/2011.03395v2) [cs.LG].
- Gundersen OE, Gil Y, Aha DW. On reproducible AI: towards reproducible research, open science, and digital scholarship in AI publications. *AI Mag* 2018;39: <http://doi.org/10.1609/aimag.v39i3.2816>
- Gundersen OE, Kjensmo S. State of the art: Reproducibility in artificial intelligence. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, 2018.
- Gundersen OE. Standing on the feet of giants - Reproducibility in AI. *Ai Magazine* 2019;December. DOI: [10.1609/aimag.v40i4.5185](https://doi.org/10.1609/aimag.v40i4.5185).
- Halbe-Kains B, Adam GA, Hosny A et al. Transparency and reproducibility in artificial intelligence. *Nature* 2020;586:E14–6. <https://doi.org/10.1038/s41586-020-2766-y>.
- Isdahl R, Gundersen OE. Out-of-the-Box Reproducibility: A Survey of Machine Learning Platforms. September 2019 IEEE 15th International Conference on e-Science and Grid Computing. DOI: [10.1109/eScience.2019.00017](https://doi.org/10.1109/eScience.2019.00017).
- Ross C. Machine learning is booming in medicine. It's also facing a credibility crisis. *Stat News*. 2021;June 2.