# Scaffolding Reproducibility for the Machine Learning Classroom

Mohamed Saeed
v-mohamsaeed@microsoft.com
Microsoft
Cairo, Egypt

Fraida Fund
ffund@nyu.edu
NYU Tandon School of Engineering
Brooklyn, New York, USA

## ABSTRACT

The importance of reproducibility in machine learning is increasingly recognized, leading to greater interest in training for reproducibility. Effective teaching of reproducibility skills prepares students for sustainable industry careers, provides students with a deeper understanding of research processes, and enhances the reproducibility ecosystem. However, instructors aiming to integrate reproducibility tasks into machine learning courses may find it challenging due to students lacking the necessary skills and experience needed for reproducing research. To address this issue, we have developed learning materials to prepare machine learning students to engage in reproducibility assignments, irrespective of their past research involvement. We present two interactive "reproducibility case studies" that guide students in replicating machine learning results. We have used this approach in an introductory machine learning class and found that it helped clarify to students how to approach the reproducibility assignment.

## KEYWORDS

reproducibility, education, machine learning

**ACM Reference Format:**
Mohamed Saeed and Fraida Fund. 2024. Scaffolding Reproducibility for the Machine Learning Classroom. In . ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

With increasing recognition of the importance of reproducibility in computational science, has come greater interest in *training* scientists and engineers to engage in reproducible research. Toward this end, a number of experience reports have discussed efforts to integrate reproducibility content into undergraduate or graduate coursework [2, 3, 5–7, 9–16]. In machine learning specifically, [9] and [16] have demonstrated the benefits of integrating reproducibility content into machine learning courses, with students reporting having a more critical perspective of results and a deeper understanding of the research process. Teaching these skills effectively can prepare students for industry careers where they will be asked to produce work with consideration to long-term sustainability and benefits the overall reproducibility ecosystem by encouraging the utilization of existing artifacts [5].


*Conference'17, July 2017, Washington, DC, USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

While reproducibility can be integrated in the classroom in a variety of ways, one popular approach is to use a reproducibility assignment, where students are tasked with reproducing a published result related to the topic of the course. However, incorporating reproducibility into a course, especially at the introductory level and especially in the form of a reproducibility assignment, can be a difficult task for instructors. This is primarily due to the students' lack of necessary skills and experience needed for reproducing research. In the literature, a key factor for success was the materials used to scaffold the learning process. These may include conventional homework assignments that equip students with the skills to engage with and reproduce research results [7]; homework and lab sessions to prepare students for a more challenging project involving best practices for reproducibility [11]; or weeks of discipline-specific instruction and discussion about research methods and critically reading papers in the field [9]. Similarly, in our own experience using a reproducibility assignment in an introductory machine learning course at the graduate level, many students struggled without extra scaffolding. The process of reproducing a published result, which involves dissecting a paper into its key claims, identifying supporting evidence, and detailing the computational experiments that yield this evidence, can be overwhelming for students without prior research experience.

To address this problem, we introduce two "reproducibility case studies" aimed at graduate students without prior research experience, that are designed to walk students through the steps involved in reproducing a published result in machine learning. We have used these materials in an introductory machine learning class and students reported that it helped clarify to them how to approach the reproducibility assignment. We make these open source materials available to the broader community as a source code repository, and as an artifact that can be played back directly on the Chameleon Cloud [8] testbed. It is our hope that by integrating reproducibility into the curriculum, we can equip students with the necessary skills to critically evaluate and reproduce scientific findings, thereby contributing to the integrity and robustness of scientific research.

## 2 OUR APPROACH

To help students understand how to approach the task of reproducing a research paper, each set of materials walks through the following steps in the context of a specific published result:

- Identify the specific, falsifiable claims in the paper.
- Find the evidence (figures, tables) that supports these claims.
- Determine the experiments (including all details necessary to reproduce the experiment) to obtain these results.
- Find artifacts (author code, third party code, pre-trained models, data) that can be used to run these experiments.
- Implement and execute the experiment, and compare the results to the published finding.

The material includes discussion questions which prompt students to reflect on what they have learned. At the end of each case study, we suggest additional experiments to validate more claims, with minimal or no modifications to the existing code.

# 3 CASE STUDY: WARM STARTING

Our first case study reproduces results from "On Warm Starting Neural Network Training" [1]. This work examines the distinction between training a model with randomly initialized weights versus using weights from a model previously trained model.

This paper was selected because it does not require advanced knowledge of neural networks, making it suitable for an introductory level course in machine learning. By engaging with this material, students learn how to identify the various claims in the paper and locate the corresponding experiments that verify these claims. Additionally, students acquire practical skills in utilizing open-source code and available data to replicate these experiments.

# 4 CASE STUDY: VISION TRANSFORMER

The second case study focuses on "An Image is Worth 16x16 Words" [4], which introduces the Vision Transformer. This paper is highly influential in the field of computer vision, which makes it relevant and interesting to many students.

In contrast to the first case study, students use pre-trained models rather than training models from scratch. Furthermore, students will not be able to reproduce some findings that rely on a private dataset or a model that has not been publicly released. By engaging with this material, students will better understand how the use of private data affects the reproducibility of a result.

# 5 USE IN THE CLASSROOM

We have used these materials as a precursor to a reproducibility assignment in a graduate-level introductory machine learning course at NYU. While we have not conducted any formal study, our subjective experience has been that students asked fewer questions about identifying claims, evidence, and experiments, than in previous semesters where these materials were not available, and that the student submissions showed a better understanding of how to engage with published work in this context.

Although these materials were designed to prepare students for a reproducibility assignment, they can also be used in other contexts. For example, they can help students learn more about the methodology of conducting research in machine learning, or they can be used to teach students about the specific topic of the case study (e.g. to teach about the Vision Transformer architecture).

# 6 ARTIFACT DESCRIPTION

We release the following materials for broader use:

- **Case Study: Warm Starting** as a Github repository[1] and as an artifact to replay on Chameleon[2] .
- **Case Study: Vision Transformer** as a Github repository[3] and as an artifact to replay on Chameleon[4].

Each case study is organized as a series of interactive Python notebooks: an overview notebook, one or more notebooks about the claims in the paper, one or more notebooks with experiments to validate the claims, and a concluding notebook.

# 7 ACKNOWLEDGEMENTS

This material was developed for the Summer of Reproducibility 2023. This material is based upon work supported by the National Science Foundation under Grant No. 2226408.

# REFERENCES

[1] Jordan Ash and Ryan P Adams. 2020. On warm-starting neural network training. *Advances in neural information processing systems* 33 (2020), 3884–3894.
[2] Richard Ball, Norm Medeiros, Nicholas W Bussberg, and Aneta Piekut. 2022. An Invitation to Teaching Reproducible Research: Lessons from a Symposium. *Journal of Statistics and Data Science Education* 30, 3 (2022), 209–218.
[3] Thomas Cokelaer, Sarah Cohen-Boulakia, and Frédéric Lemoine. 2023. Reprohackathons: promoting reproducibility in bioinformatics through training. *Bioinformatics* 39, Supplement1 (06 2023), i11–i20.
[4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
[5] Fraida Fund. 2023. We Need More Reproducibility Content Across the Computer Science Curriculum. In *Proceedings of the 2023 ACM Conference on Reproducibility and Replicability* (Santa Cruz, CA, USA) *(ACM REP '23)*. 97–101.
[6] Yayu Gao, Chengwei Zhang, Xiaojun Hei, and Guohui Zhong. 2019. Learning networking by reproducing research results in an ns-3 simulation networking laboratory course. In *2019 IEEE TALE*. IEEE.
[7] Nestoras Karathanasis, Daniel Hwang, Vibol Heng, Rimal Abhimannyu, Phillip Slogoff-Sevilla, Gina Buchel, Victoria Frisbie, Peiyao Li, Dafni Kryoneriti, and Isidore Rigoutsos. 2022. Reproducibility efforts as a teaching tool: A pilot study. *PLOS Computational Biology* 18, 11 (11 2022), 1–11.
[8] Kate Keahey, Jason Anderson, Zhuo Zhen, Pierre Riteau, Paul Ruth, Dan Stanzione, Mert Cevik, Jacob Colleran, Haryadi S. Gunawi, Cody Hammock, Joe Mambretti, Alexander Barnes, François Halbach, Alex Rocha, and Joe Stubbs. 2020. Lessons Learned from the Chameleon Testbed. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20)*. USENIX Association.
[9] Ana Lucic, Maurits Bleeker, Sami Jullien, Samarth Bhargav, and Maarten De Rijke. 2022. Reproducibility as a mechanism for teaching fairness, accountability, confidentiality, and transparency in artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 12792–12800.
[10] Wolfgang Mauerer, Stefan Klessinger, and Stefanie Scherzinger. 2023. Beyond the Badge: Reproducibility Engineering as a Lifetime Skill. In *Proceedings of the 4th International Workshop on Software Engineering Education for the Next Generation* (Pittsburgh, Pennsylvania) *(SEENG '22)*. ACM, New York, NY, USA, 1–4.
[11] K Jarrod Millman, Matthew Brett, Ross Barnowski, and Jean-Baptiste Poline. 2018. Teaching computational reproducibility for neuroimaging. *Frontiers in Neuroscience* 12 (2018), 727.
[12] Joel Ostblom and Tiffany Timbers. 2022. Opinionated Practices for Teaching Reproducibility: Motivation, Guided Instruction and Practice. *Journal of Statistics and Data Science Education* 30, 3 (2022), 241–250.
[13] Melissa L. Rethlefsen, Hannah F. Norton, Sarah L. Meyer, Katherine A. MacWilkinson, Plato L. Smith II, and Hao Ye. 2022. Interdisciplinary Approaches and Strategies from Research Reproducibility 2020: Educating for Reproducibility. *Journal of Statistics and Data Science Education* 30, 3 (2022), 219–227.
[14] Lars Vilhuber, Hyuk Harry Son, Meredith Welch, David N. Wasser, and Michael Darisse. 2022. Teaching for Large-Scale Reproducibility Verification. *Journal of Statistics and Data Science Education* 30, 3 (2022), 274–281.
[15] Lisa Yan and Nick McKeown. 2017. Learning networking by reproducing research results. *ACM SIGCOMM Computer Communication Review* 47, 2 (2017), 19–26.
[16] Burak Yildiz, Hayley Hung, Jesse H Krijthe, Cynthia CS Liem, Marco Loog, Gosia Migut, Frans A Oliehoek, Annibale Panichella, Przemysław Pawełczak, Stjepan Picek, et al. 2021. ReproducedPapers. org: Openly teaching and structuring machine learning reproducibility. In *International Workshop on Reproducible Research in Pattern Recognition*. Springer, 3–11.

[1] Github: https://github.com/teaching-on-testbeds/re_warm_start_nn
[2] Trovi Artifact: https://chameleoncloud.org/experiment/share/5b5717df-9aa9-470f-b393-c1e189c008a8
[3] Github: https://github.com/teaching-on-testbeds/re_vit
[4] Trovi Artifact: https://chameleoncloud.org/experiment/share/8f0e34c5-d2c4-45be-8425-36686ad57650