

# ML Reproducibility Challenge 2021

Koustuv Sinha<sup>1,2, ID</sup>, Jesse Dodge<sup>6</sup>, Sasha Luccioni<sup>5, ID</sup>, Jessica Zosa Forde<sup>4, ID</sup>, Sharath Chandra Raparthy<sup>3, ID</sup>, Joelle Pineau<sup>1,2,8, ID</sup>, and Robert Stojnic<sup>7,8, ID</sup>

<sup>1</sup>School of Computer Science, McGill University, Montreal, Canada – <sup>2</sup>Mila - Quebec AI Institute, Montreal, Canada – <sup>3</sup>Université de Montréal, Canada – <sup>4</sup>Brown University, USA – <sup>5</sup>Allen Institute for AI, USA – <sup>6</sup>HuggingFace, USA – <sup>7</sup>PapersWithCode, USA – <sup>8</sup>Meta AI, USA and Canada

**Edited by**

Nicolas P. Rougier

**Reviewed by**

Anonymous Reviewers

**Received**

19 May 2022

**Published**

23 May 2022

**DOI**

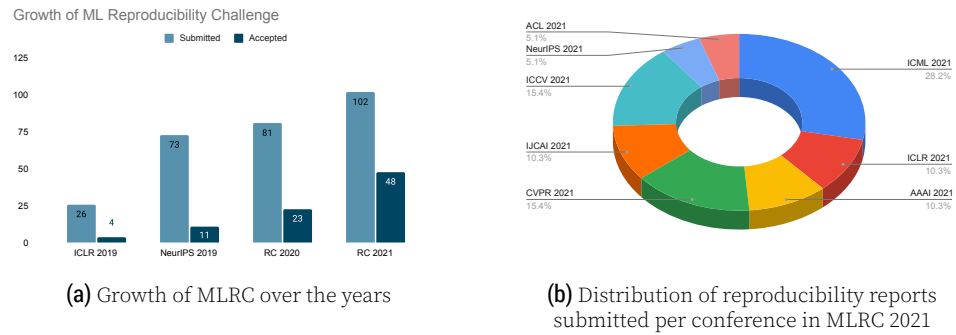
10.5281/zenodo.6574723

## 1 Introduction

The importance of reproducibility in science cannot be overstated. It is one of the key mechanisms in place to enforce the high standards of scientific discoveries, and a key ingredient for an impactful scientific discovery, allowing future practitioners to build on the shoulders of prior work. Reproducible science also promotes open and accessible research, allowing the scientific community to quickly integrate new findings and convert ideas to practice more seamlessly. In the spirit of promoting a culture of reproducible science in the Machine Learning community, we have hosted the fifth iteration of the ML Reproducibility Challenge in 2021. Following the trend of inclusivity and breadth, this iteration involves a challenge to reproduce papers published in nine top conferences in Machine Learning, including NeurIPS, ICML, ICLR, CVPR, ICCV, ACL, EMNLP, AAAI and IJCAI. An important objective of this challenge is to contribute toward improving the understanding of the central claims of the papers published in these top conferences, by inviting participants to run reproducibility study on them. In this special issue of ReScience C Journal, we are proud to present the peer-reviewed accepted papers of the 2021 ML Reproducibility Challenge.

## 2 Challenge

The goal of the challenge was to reproduce the central claims of papers published in top Machine Learning conferences of the year. Participants were invited to work on either all claims, or partial claims, depending on the complexity of the project. Participants were also free to reuse authors' code when available, while being encouraged to explore beyond simply running the code provided to verify reproducibility. Unlike the previous challenge, in this iteration we removed the "Claim paper" step. This step, which was previously used, involved participants pre-registering which paper they wanted to reproduce, in order to encourage early commitment, narrow down the claims they wish to explore in the paper, and covering a larger number of papers. However, we received feedback that this step was not useful for participants, which was also reflected by the low percentage of the number of reproducibility reports submitted vs papers claimed. Removing this step also reduced the complexity of participating in the challenge. As in the last iteration, participants were free to claim multiple papers, and multiple teams could claim the same paper. In this iteration, we observed a jump of reproducibility report submissions to 102, compared to 82 from last year (Figure 1). Reproducibility



**Figure 1.** Statistics of the ML Reproducibility Challenge

reports were spread across all nine conferences, with most papers chosen from ICML 2021, and the least from ACL 2021. A majority of the participants were students using the challenge as a part of their machine learning courses from various institutions around the world, including but not limited to: KTH (Royal Institute of Technology Stockholm, Sweden), Queen’s University (Ontario, Canada), Indian Institute of Technology (Gandhinagar, India), University of Amsterdam (Netherlands), University of Southern California, (USA), Indian Institute of Technology (Guwahati, India), Tsinghua University (China), University of Ljubljana (Slovenia), University of Michigan (USA), University of Waterloo (Ontario, Canada), Istanbul Technical University (Turkey), and EPFL (Switzerland).

After in-depth peer review, in this special issue we present the top 47<sup>1</sup> accepted reports, selected from 102 submissions, thus driving up the acceptance rate from 28% last year to 47% this year. This increase is largely due to significant improvements in the quality of the reports & their methodology, which is encouraging to see.

### 3 Best Paper Awards

Starting this year, we are presenting best paper awards to a few select reports to highlight the excellent quality all-round of their reproducibility work. The selection criteria consisted of votes from the Area Chairs, based on the reproducibility motivation, experimental depth, results beyond the original paper, ablation studies, and discussion/recommendations. We believe the community will appreciate the strong reproducibility efforts in each of these papers, which will improve the understanding of the original publications, and inspire authors to promote better science in their own work.

#### 3.1 Best Paper Award

- Piyush Bagad, Jesse Maas, Paul Hilders, Danilo de Goede; *Reproducibility Study of “Counterfactual Generative Networks”*

#### 3.2 Outstanding Paper Award

- Matija Teršek, Domen Vreš, Maša Kljun; *Study of “Counterfactual Generative Network”*
- Ian Hardy; *[RE] An Implementation of Fair Robust Learning*

<sup>1</sup>We accepted 48 reports, but one team did not submit their camera ready version till the time of the publication of this editorial.

- Guilly Kolkman, Maks kulicki, Jan Athmer, Alex Labro; *Strategic classification made practical: reproduction*
- Andrea Lombardo, Matteo Tafuro, Tin Hadži Veljković, Lasse Becker-Czarnetzki; *On the reproducibility of "Exacerbating Algorithmic Bias through Fairness Attacks"*

## 4 Platforms

This challenge is conducted with the support of PapersWithCode<sup>2</sup> and OpenReview<sup>3</sup>. PapersWithCode is an open, collaborative platform to discover latest trending machine learning research papers with their codebases, which enables rapid re-usability and reproducibility of published works. PapersWithCode enabled the challenge organizers to reach a wide audience of students and researchers who participated in the competition. As was the case last year, OpenReview provided crucial logistic support by providing an unique platform to claim and submit reproducibility reports. After submission, all reports went through a thorough peer review process consisting of hundreds of reviewers from the Machine Learning community, and OpenReview provided an easy-to-use platform for managing reviews and administrative processes. Finally, we used a public Github repository<sup>4</sup> to perform the final editorial process of converting accepted papers into ReScience format, and thereby publish 48 high quality reports in this special issue.

## 5 Conclusion

Reproducibility of central claims of papers published in Machine Learning conferences has been a center of considerable attention over the past several years. In recent years, conferences such as NeurIPS, ICLR, AAAI, ICML, EMNLP have routinely included reproducibility workshops and challenges to cultivate the culture of reproducible science in the community. Several conferences have also introduced code submission policies and Reproducibility Checklists to further advance the cause and build momentum of reproducible science. We hope our continued endeavour of hosting annual challenges and publishing high-quality peer-reviewed reproducibility reports will contribute more information about existing published papers, and help strengthen their core contributions in the process, while also promoting open, accessible and sound machine learning research.

## 6 Acknowledgement

We thank the board and program committee of NeurIPS, ICML, ICLR, ACL, EMNLP, CVPR, ICCV, AAAI and IJCAI for partnering with us in this massive initiative and supporting the challenge. We thank the OpenReview team (in particular Andrew McCallum, Parag Pachpute, Melisa Bok, Celeste Martinez Gomez, Pam Mandler and Mohit Uniyal) for their constant support in hosting and building the customized portal used in our challenge. We thank Ana Lucic, Maurits Bleeker and Samarth Bhargav for their feedback on using the Reproducibility Challenge in their course at University of Amsterdam. We thank Robert Stojnic, Ross Taylor and Elvis Saravia from PapersWithCode for hosting and supporting the challenge along with its logistics. We thank the ReScience board (in particular Nicolas Rougier, Konrad Hinsén, Olivia Guest and Benoît Girard) for presenting the accepted reports in their esteemed journal. Finally, we thank all of

<sup>2</sup><https://paperswithcode.com>

<sup>3</sup>[https://openreview.net/group?id=ML\\_Reproducibility\\_Challenge/2021/Fall/](https://openreview.net/group?id=ML_Reproducibility_Challenge/2021/Fall/)

<sup>4</sup><https://github.com/ReScience/MLRC>

our participants who dedicated time and effort to verify results that were not their own, to help strengthen our understanding of the concepts presented in the papers.

## 7 Reviewers

Our reviewers need a special section dedicated to thank them for their tireless efforts in screening and providing valuable feedback to the Area Chairs (Jesse Dodge, Sasha Luccioni, Jessica Zosa Forde, Sharath Chandra Raparthy and Koustuv Sinha) to select the best papers. We were fortunate enough to attract a large pool of reviewers, who spent their precious time to critically review the reports. We would like to specifically acknowledge our Emergency reviewers who responded to our call for help to review some additional reports at the last minute. Starting this iteration, we also introduce Outstanding Reviewer Award to select reviewers for their high quality and timely reviews for the challenge. The selection criteria involved votes from the Area Chairs after careful review of the reviews posted in the challenge. We thank the reviewers for their exceptional effort and hope they will continue to support us in future iterations.

### 7.1 Outstanding Reviewers

Alex Gu	Karan Shah	Pascal Lamblin
Cagri Coltekin	Leo M Lahti	Prithvijit Chattopadhyay
David Rau	Maxime Wabartha	Samuel Albanie
Divyat Mahajan	Maxwell D Collins	Sunnie S. Y. Kim
Frederik Paul Nolte	Olga Isupova	Tobias Uelwer
Kanika Madan	Olivier Delalleau	Varun Sundar

### 7.2 All Reviewers

Abhinav Agarwalla	Hanna Suominen	Martin Klissarov
Akshay Ravindra Kulkarni	Hao He	Matthew Kyle Schlegel
Ali Hürriyetoglu	Harsha Kokel	Matthew Ryan Krause
Andreas Ruttor	Heng Fang	Mayur Arvind
Animesh Gupta	Jiangwen Sun	Melanie F. Pradier
Anis Zahedifard	Jie Fu	Mike Chrzanowski
Brent M. Berry	Jishnu Jaykumar P	Minjia Zhang
Chao Qin	Kaushy Kularatnam	Monjoy Saha
David Arbour	Kiana Alikhademi	Nadia Tahiri
David Krueger	Labiba Kanij Rupty	Nan Rosemary Ke
Dong Gong	Li Erran Li	Nikolaos Vasiloglou
Fan Feng	Lluís Castrejon	Otasowie Owolafe
Felix Gimeno	Mahima Agumbe Suresh	Owen Lockwood
Furkan Kinli	Mahzad Khoshlessan	Pablo Robles-Granda
Gabriel Synnaeve	Maja Schneider	Patrick Philipp
Gang Wang	Mani A	Paul Tylkin
Georgios Leontidis	Marija Stanojevic	Prateek Garg

Praveen Narayanan

Qingzhi Hu

Raj Ghugare

Ramesh Ragala

Razvan Pascanu

Samiran Das

Seohyun Kim

Shiju SS

Simon Kornblith

Stefan Magureanu

Ujjwal Verma

Venkatadheeraj Pichapati

Vibha Belavadi

Wei Han

Wenbin Zhang

Wenhao Yu

Xavier Bouthillier

Xavier Sumba

Xiang Zhang

Xin Lu

Xinggang Wang

Xingrui Yu

Yuntian Deng

Zahra Atashgahi

Zhourong Chen