Mila
Université de Montréal
McGill

CIFAR
CANADIAN INSTITUTE FOR ADVANCED RESEARCH

facebook
Artificial Intelligence Research

Unreproducible Research is Reproducible

Xavier Bouthillier
César Laurent
Pascal Vincent

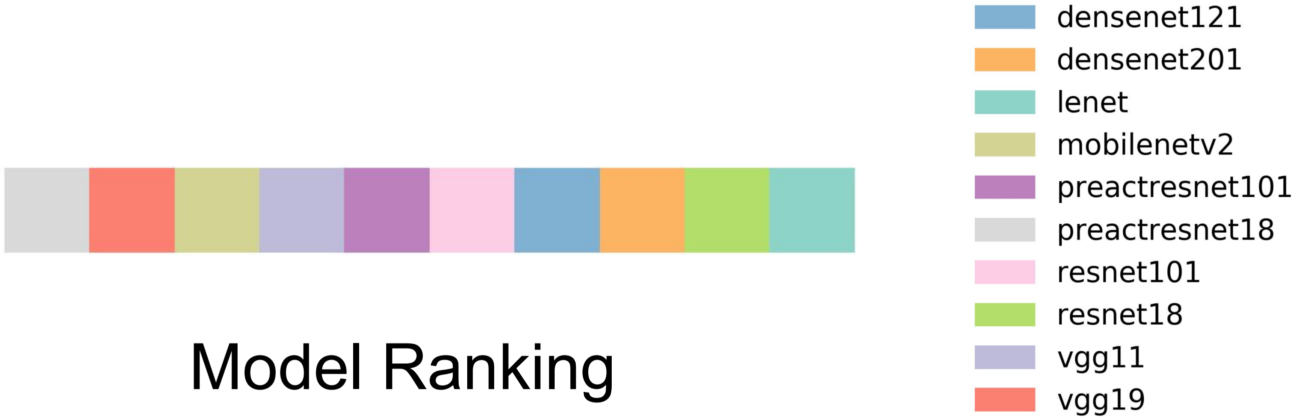
Take Home

- There is a *spectrum* of notions of reproducibility in science.

- Current focus in DL is on *one end of the spectrum*.

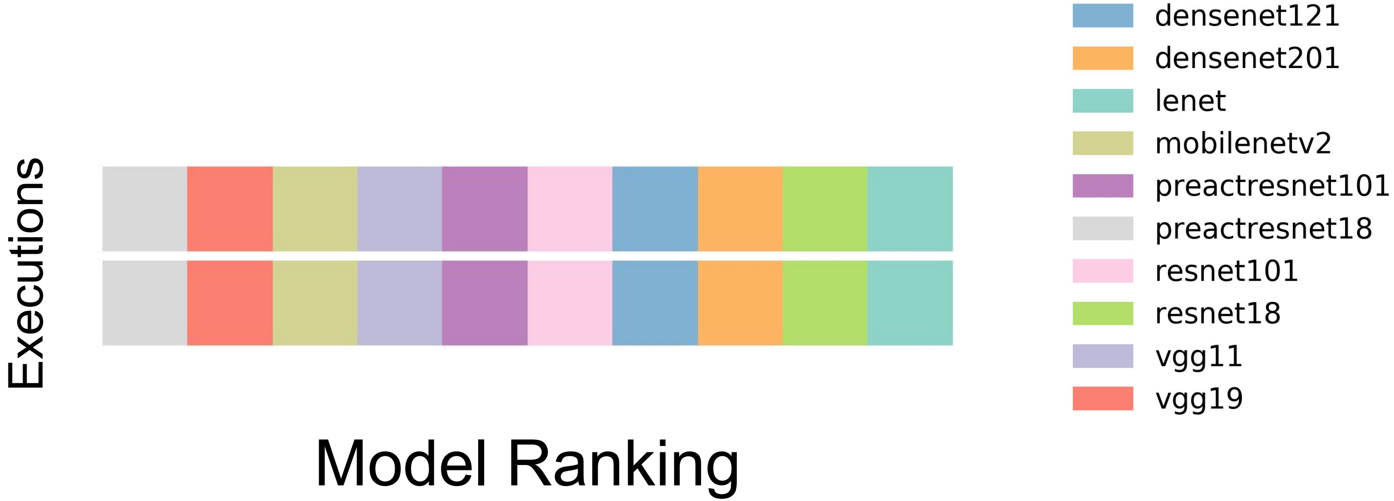
- *Inferential reproducibility* is currently neglected **but** fundamental for empirical research.

Reproducibility

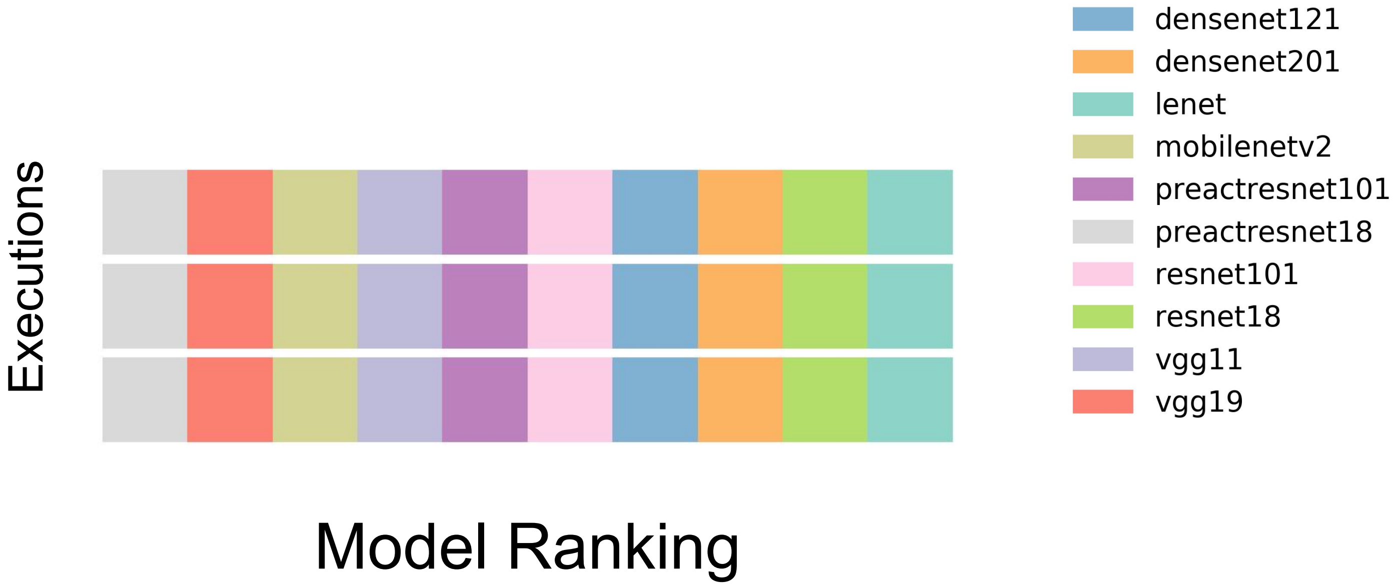


Model Ranking

Reproducibility



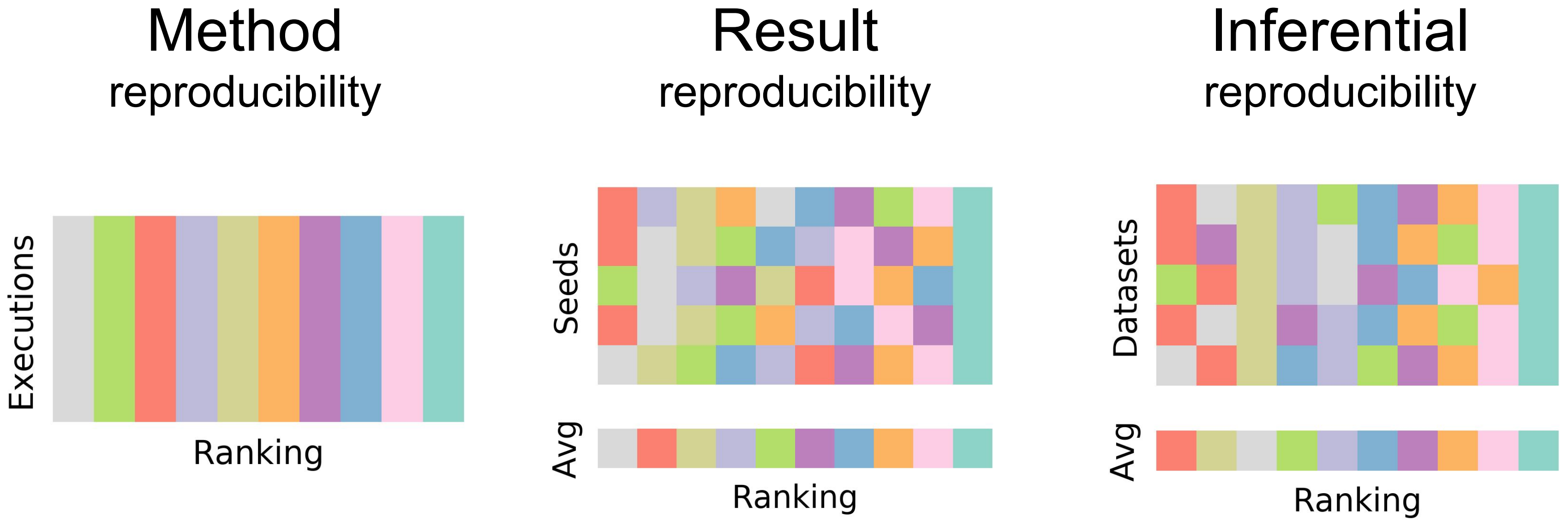
Reproducibility



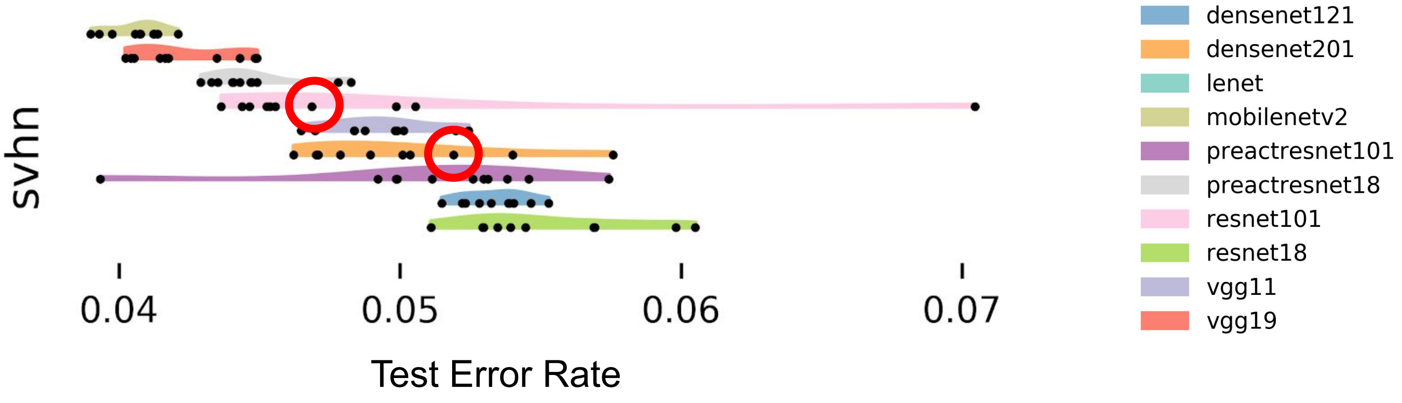
Reproducibility Spectrum

Terminology by Goodman et al. (2016)

- 🟦 densenet121
- 🟧 densenet201
- 🟩 lenet
- 🟨 mobilenetv2
- 🟪 preactresnet101
- ⬜ preactresnet18
- 🟫 resnet101
- 🟩 resnet18
- 🟪 vgg11
- 🟥 vgg19

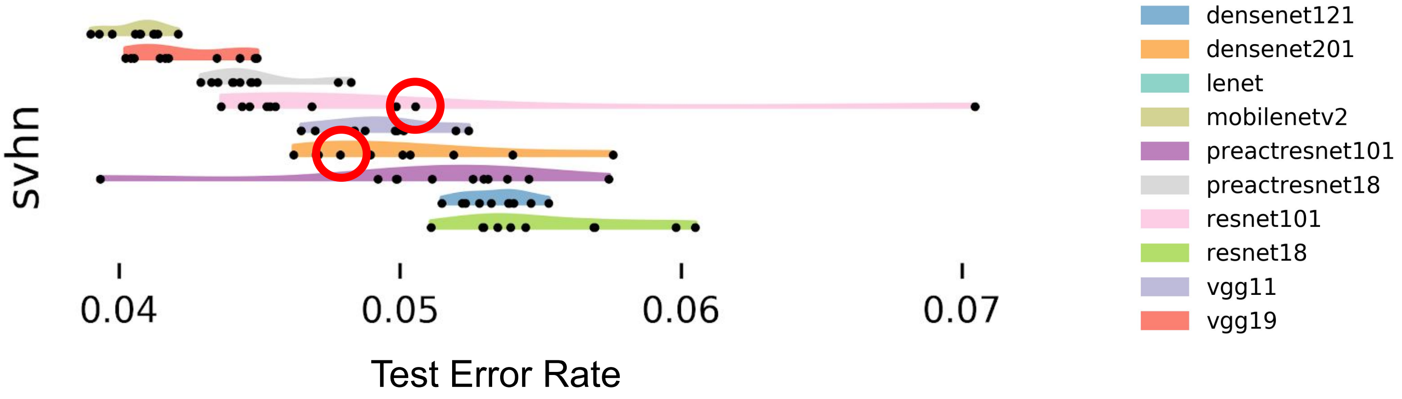


Method Reproducibility



Each black dot can be precisely reproduced

Method Reproducibility



Reproducible method != Reproducible conclusions:

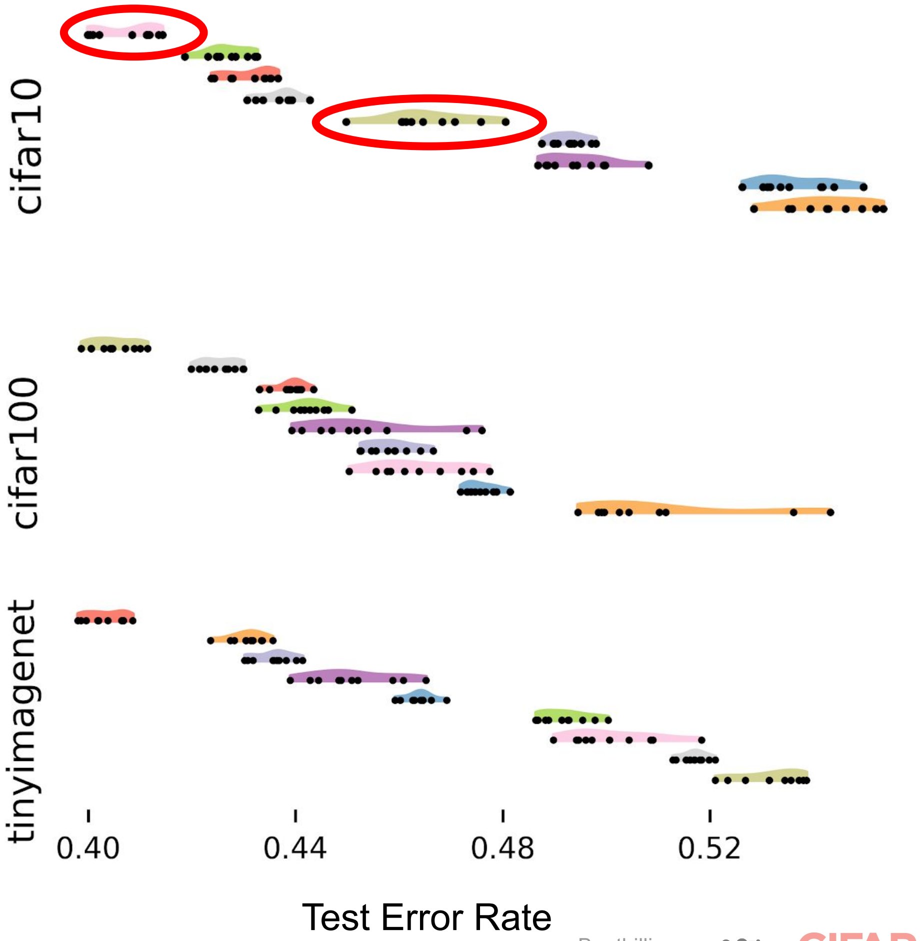
> One cannot conclude that model A is better than model B with only 2 points!

Bouthillier, Laurent & Vincent

CIFAR

Result Reproducibility

Test performance distributions can be reproduced



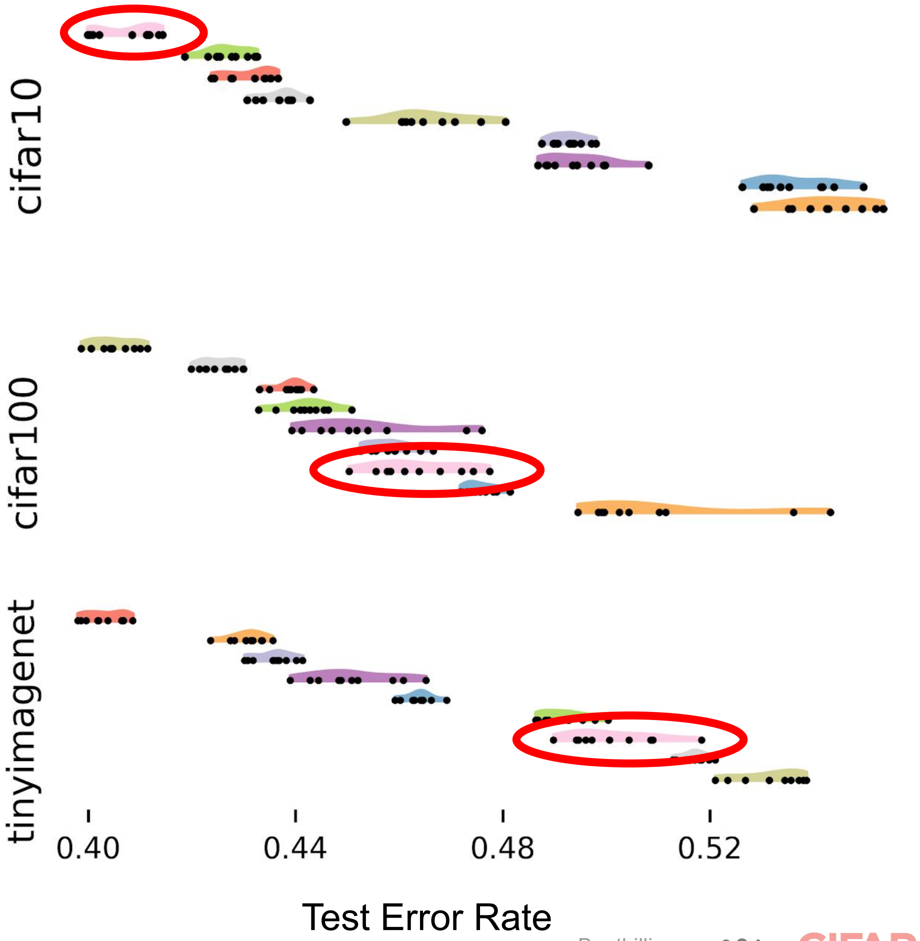
Bouthillier, Laurent & Vincent

Result Reproducibility

Test performance distributions can be reproduced

Reproducible results != Reproducible conclusions:

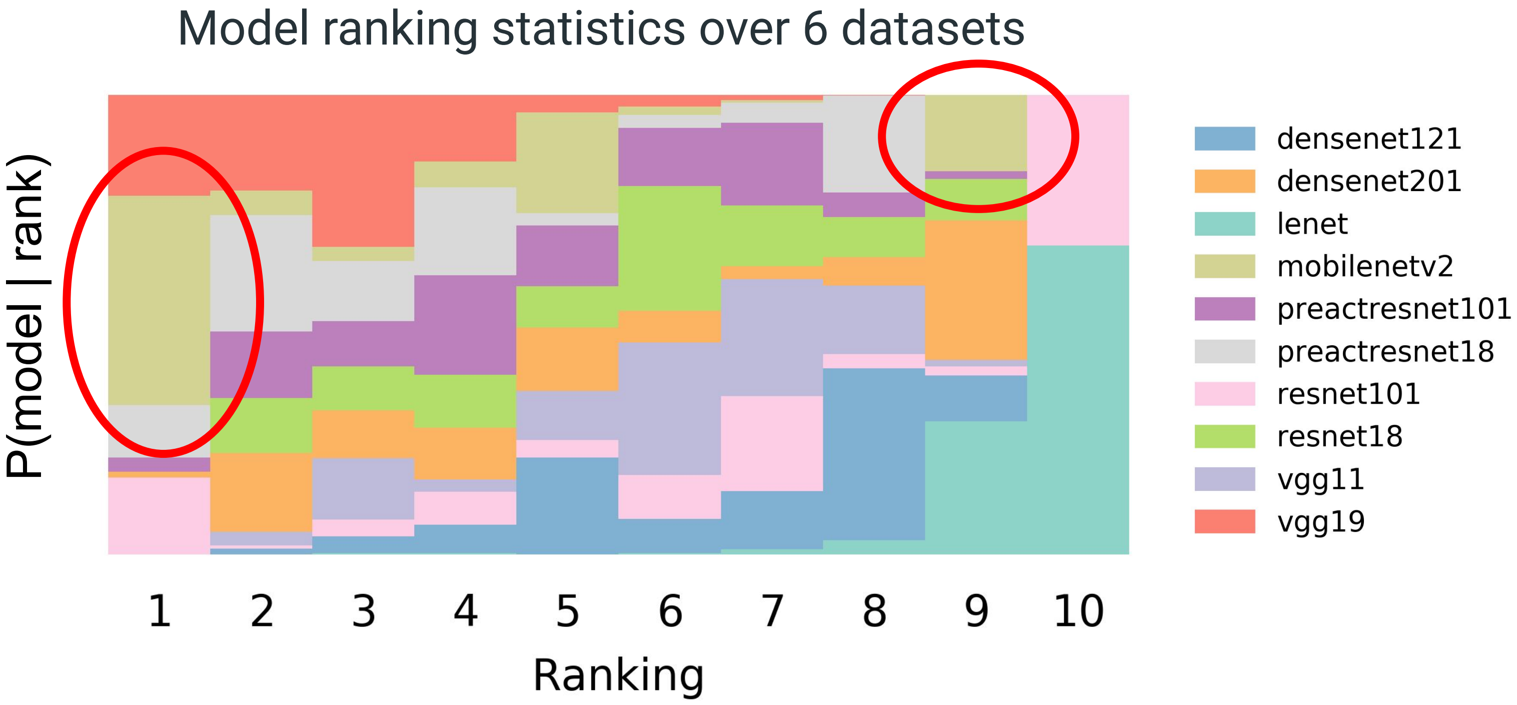
> One cannot conclude that model A is better than model B with only 1 dataset!



Bouthillier, Laurent & Vincent

Inferential Reproducibility

A conclusion regarding which is the best architecture cannot be reproduced on different datasets.



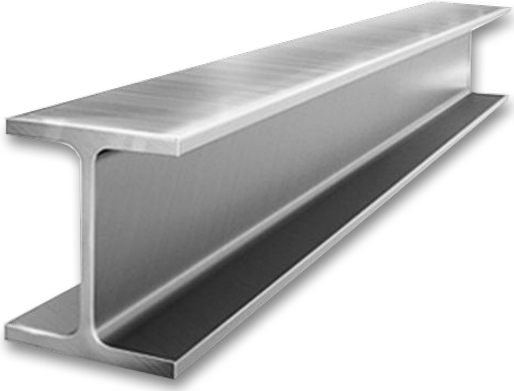
Bouthillier, Laurent & Vincent

Research Methodologies and Reproducibility

Exploratory & Constructive Research

Method & Result Reproducibility

Empirical & Confirmatory Research



Inferential Reproducibility

Come see our poster

Thu June 13th
06:30 -- 09:00 PM
@ Pacific Ballroom #14

Thank you!

Mila

Université de Montréal

McGill

CIFAR
CANADIAN
INSTITUTE
FOR
ADVANCED
RESEARCH

facebook
Artificial Intelligence Research

References

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. What does research reproducibility mean? *Science Translational Medicine*, 8(341):341ps12–341ps12, 2016. ISSN 1946-6234. doi: 10.1126/scitranslmed.aaf5027.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup,D., and Meger, D. Deep reinforcement learning that matters. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018

Lucic, M., Kurach, K., Michalski, M., Gelly, S., and Bousquet, O. Are gans created equal? a large-scale study. In *Advances in neural information processing systems*, pp. 698–707, 2018.

Melis, G., Dyer, C., and Blunsom, P. On the state of the art of evaluation in neural language models. In *International Conference on Learning Representations*, 2018.

Mila Université de Montréal McGill

CIFAR CANADIAN INSTITUTE FOR ADVANCED RESEARCH

facebook Artificial Intelligence Research