Reproducibility Caveats in Machine Learning

Open to Complexity

Chris Emmery & Travis J. Wiltshire
Dept. of Cognitive Science & AI, Tilburg University

Open to Complexity Symposium
5 Sept. 2023

This Talk

Algorithmic Harms

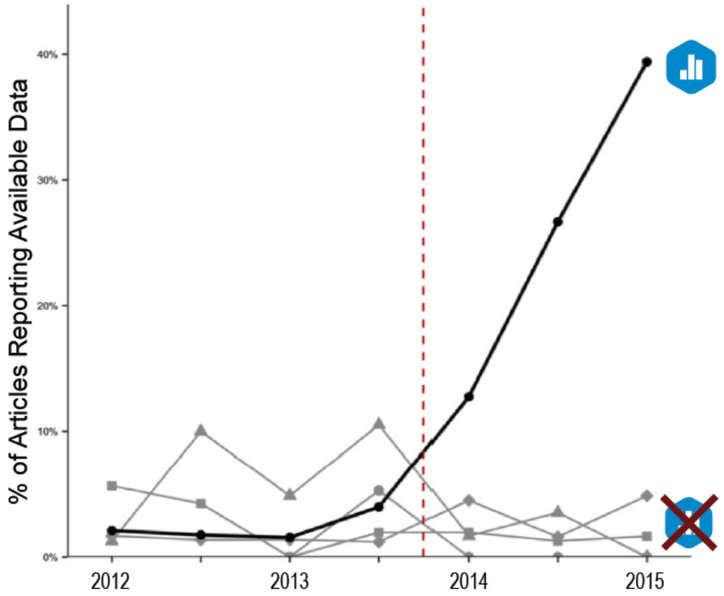
Complex Dynamical Systems

Based on: Emmery, C., Kádár, Á., Wiltshire, T. J., & Hendrickson, A. T. (2019). Towards Replication in Computational Cognitive Modeling: A Machine Learning Perspective. *Computational Brain & Behavior* 2.3-4 (2019): 242-246.
& Emmery, C., Wiltshire, T. J. (in press). Machine Morality. *Encyclopedia of Heroism Studies*

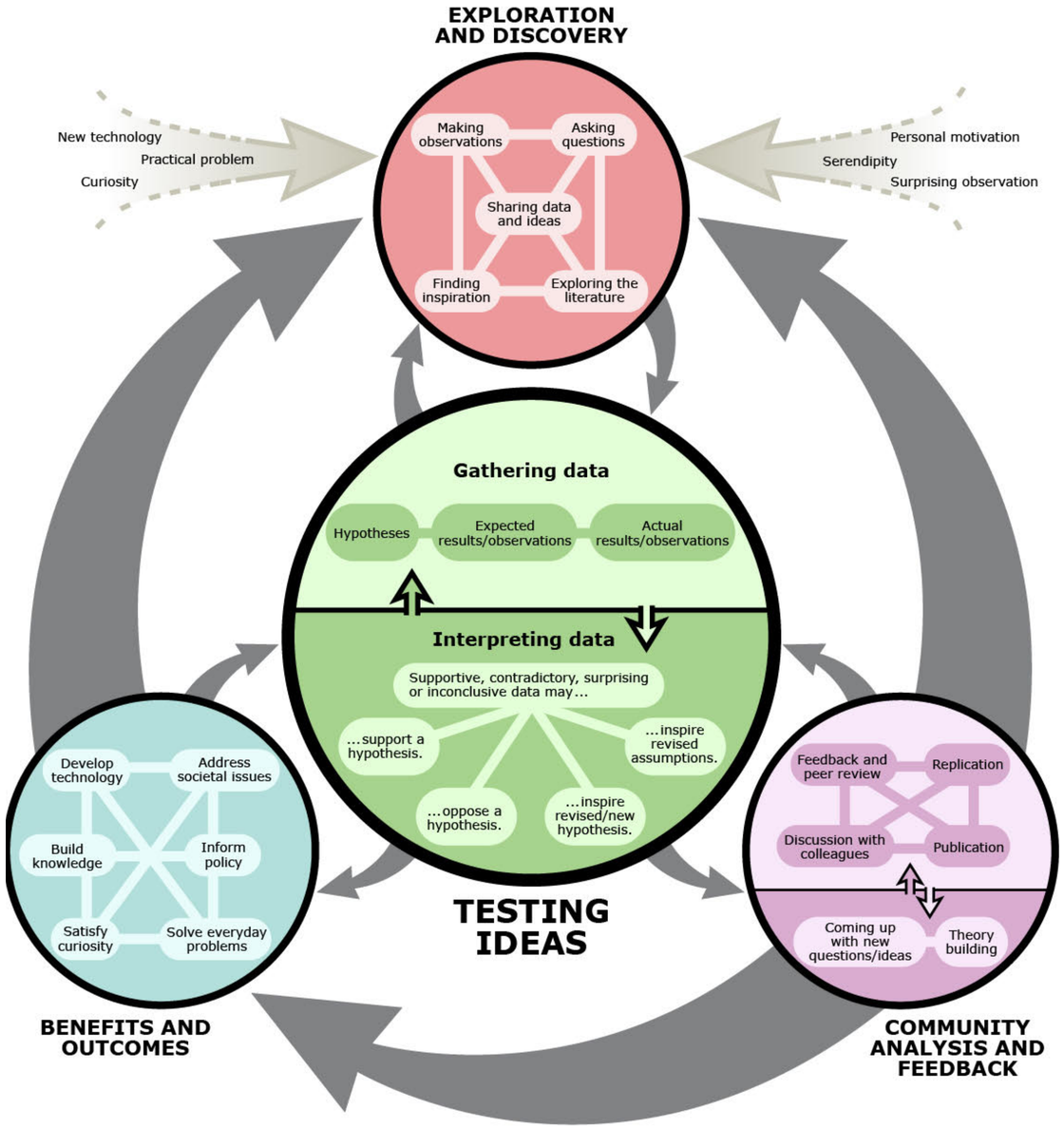
“Aim of Open Science is to make scientific knowledge freely available to everyone, accessible and reusable and to promote scientific cooperation and the exchange of information for the benefit of science and society.”

– UNESCO

Goal of Open Science



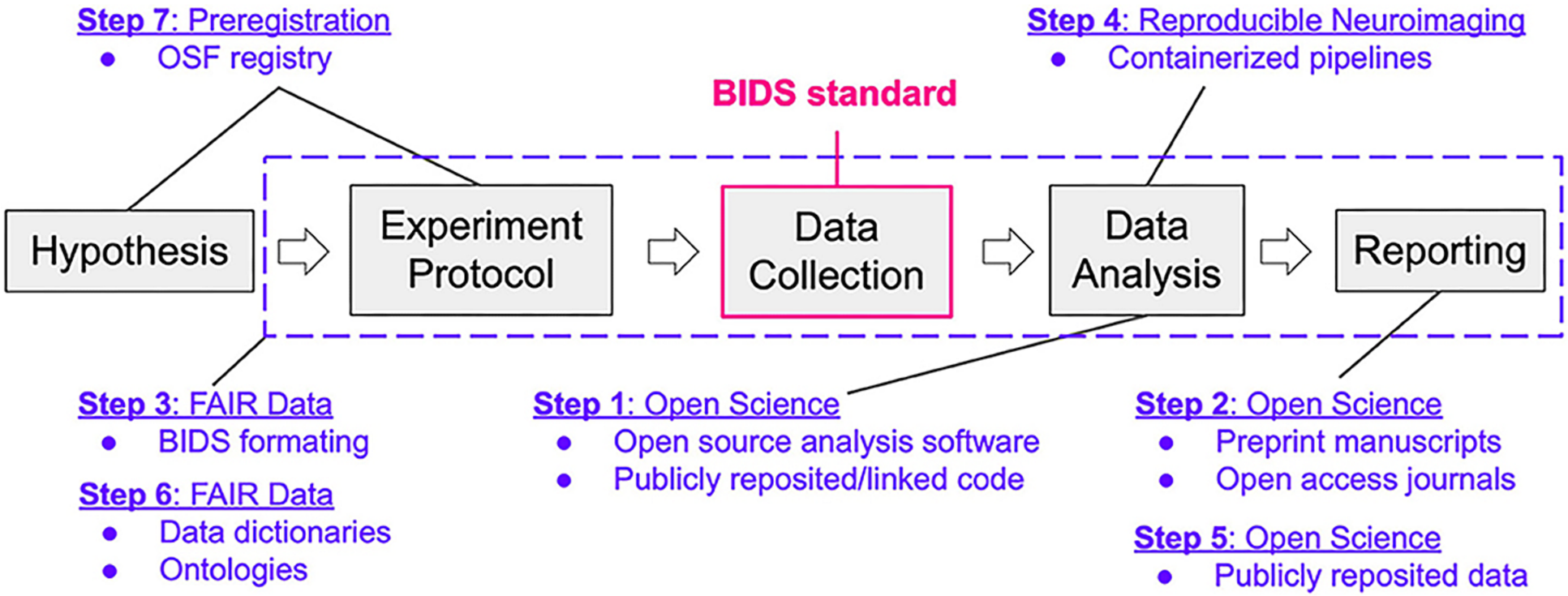
General Scientific Process



Understanding Science. 2022. University of California Museum of Paleontology. 3 January 2022 <http://www.understandingscience.org>.

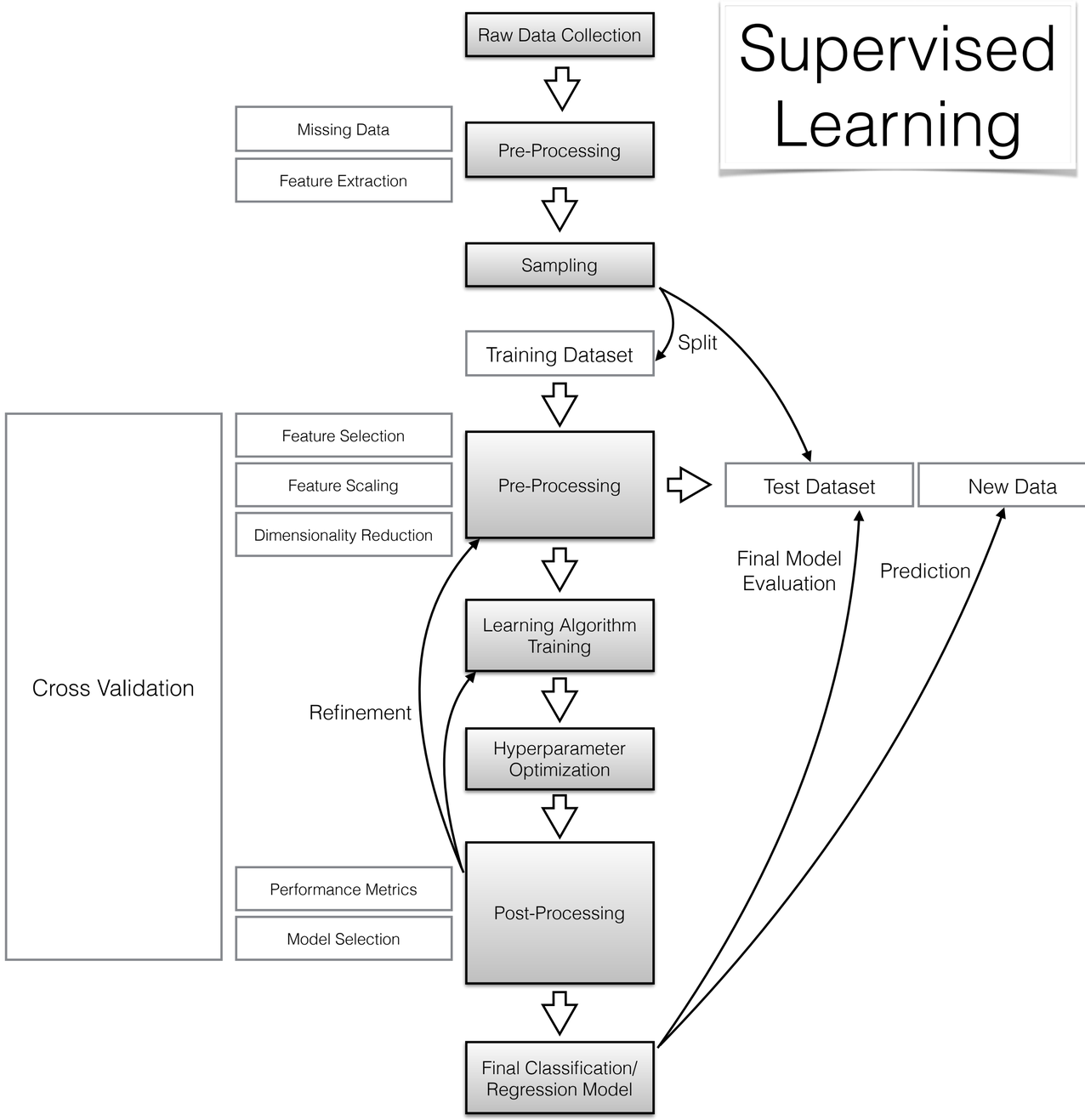
Open Workflow

For A Cognitive/Neuroscience



Bush, K. A., Calvert, M. L., & Kilts, C. D. (2022). Lessons learned: A neuroimaging research center's transition to open and reproducible science. *Frontiers in big Data*, 82.

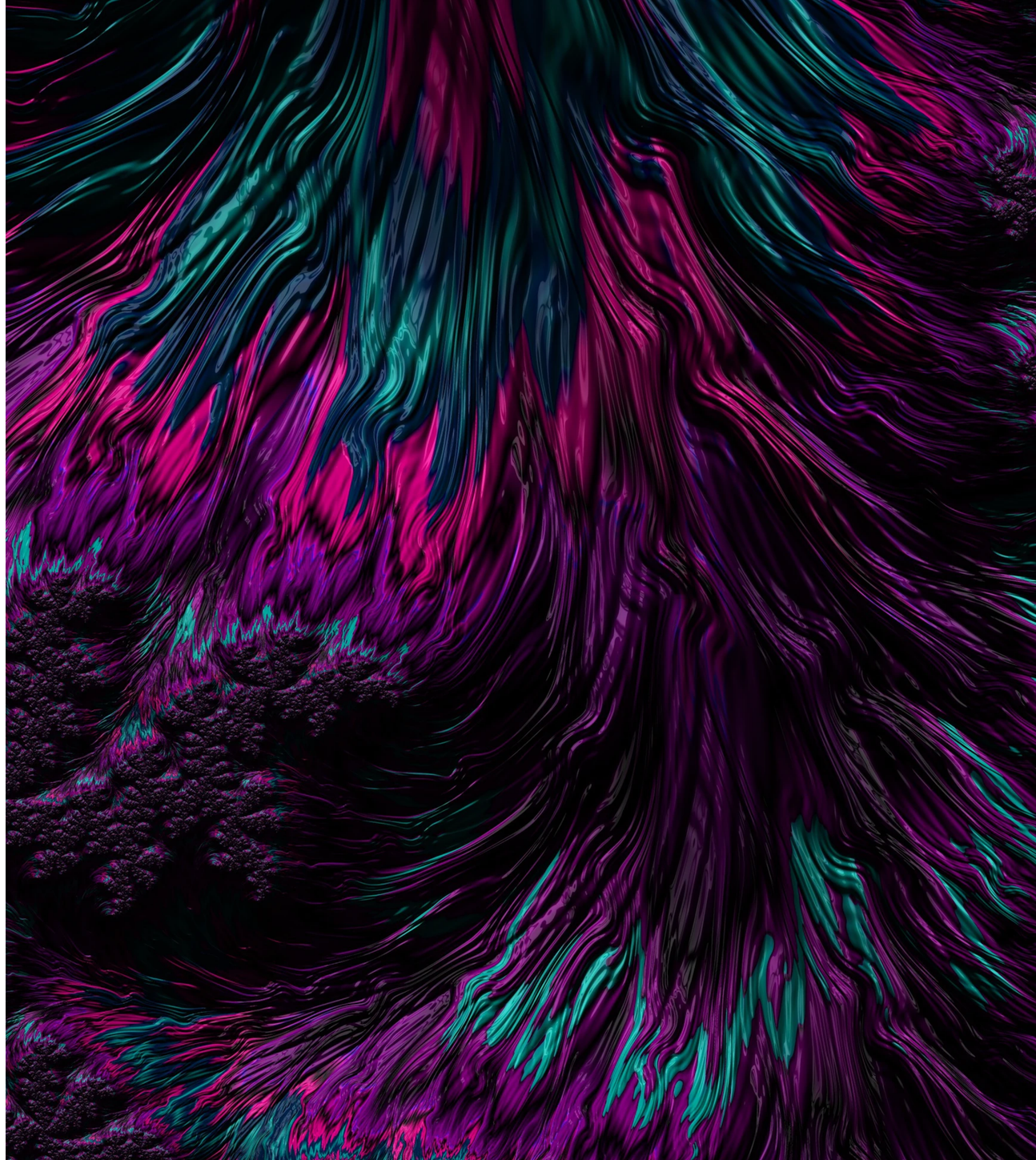
Typical Workflow in Machine Learning



Open Science

Open to Complexity

- **Challenge 1**: Inertia of Cultural Change (Before)
  - Scalability of Reproducible Code
  - Interpretability of Reproducible Research
  - Downstream Effects of Reproducibility
- **Challenge 2**: Nonreciprocity (Now)



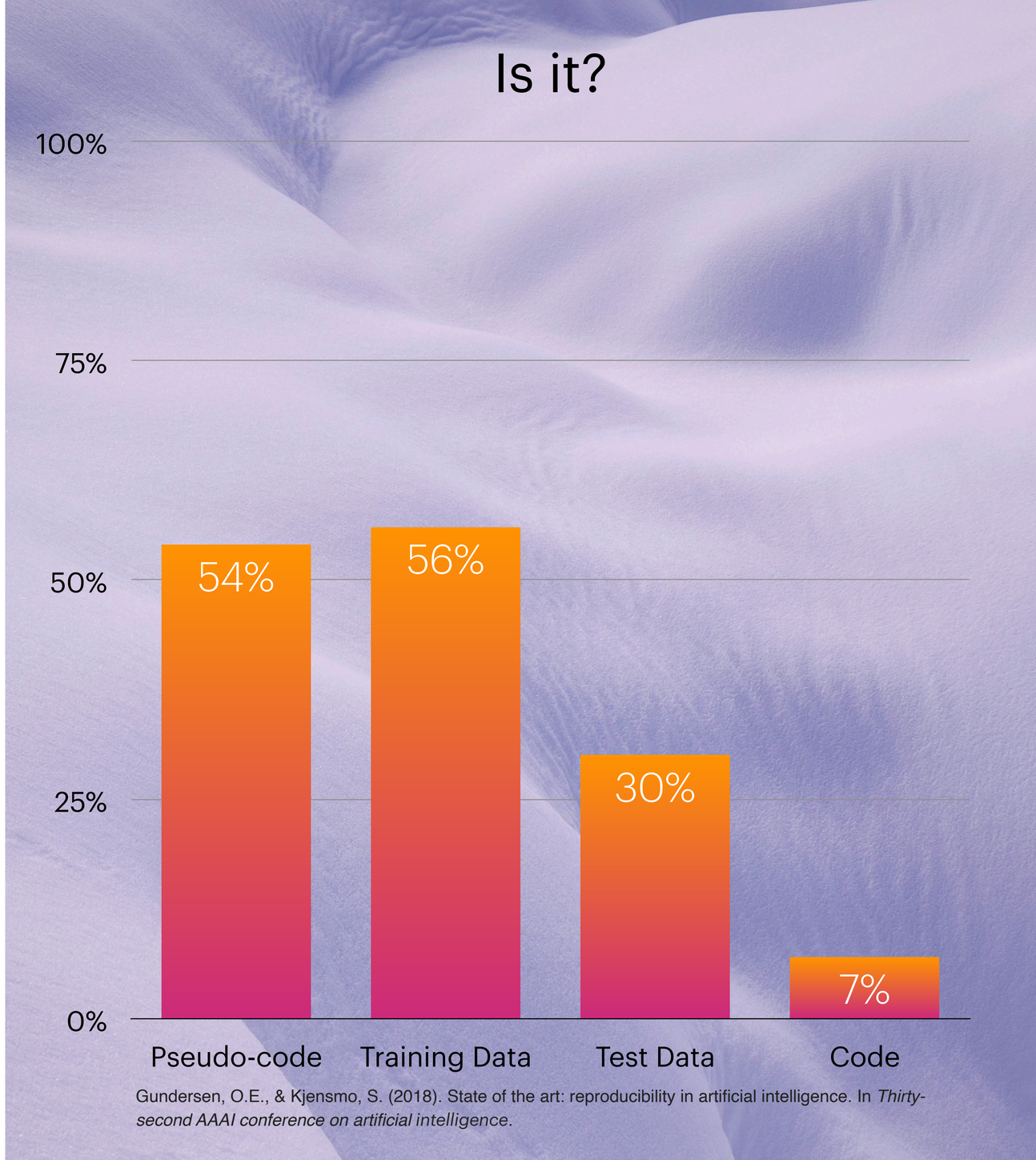
Challenge 1: Inertia of Cultural Change

“(...) the ability of an independent research team to produce the same results using the same [computational] method based on the documentation made by the original research team”

— Gundersen and Kjensmo, 2018, p. 1645

Machine Learning & Openness
2019: <3 or </3 ?

- Often inherently reproducible. Contribution = data + code.
- Existing datasets, end-to-end code, heavy use of pre-prints.
- Key issue: credit attribution.

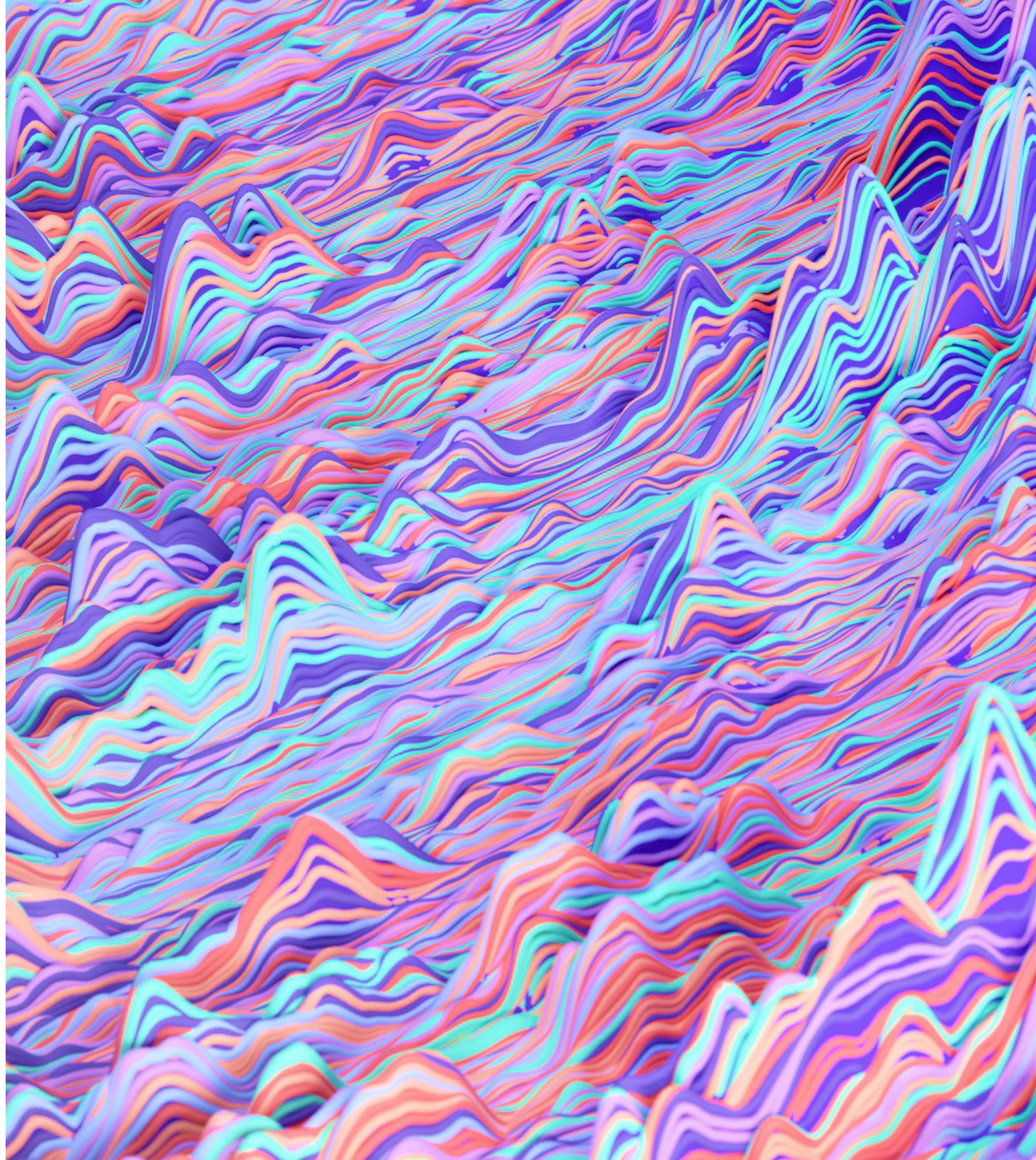


Gundersen, O.E., & Kjensmo, S. (2018). State of the art: reproducibility in artificial intelligence. In *Thirty-second AAAI conference on artificial intelligence*.

Sub-challenge #1

Scalability of Producing Reproducible Code

- Should be repeated in isolation: new data, different method, different system.
- Software-engineer-level requirements: well-documented, modular code, extensive logging.
- Favors larger labs with dedicated staff, undoing effort.



Sub-challenge #2

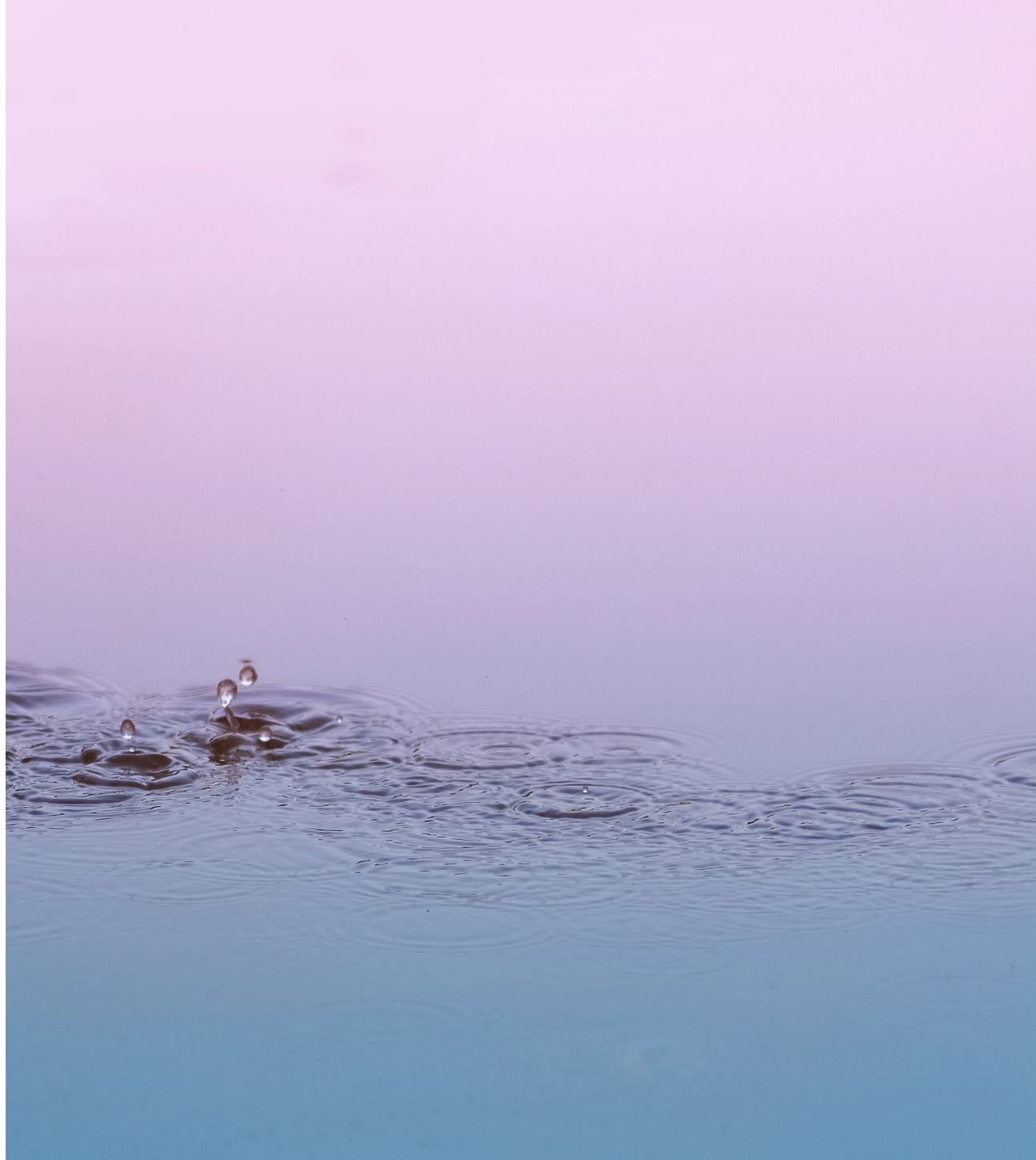
Interpretability of Reproducible Research

- Every step to making data, methods, logs, and error analysis transparent ramps up human effort.
- Requires structure, standards, and optimized search.
- Failing to provide humanly-feasible interpretation again undoes reproducibility effort.

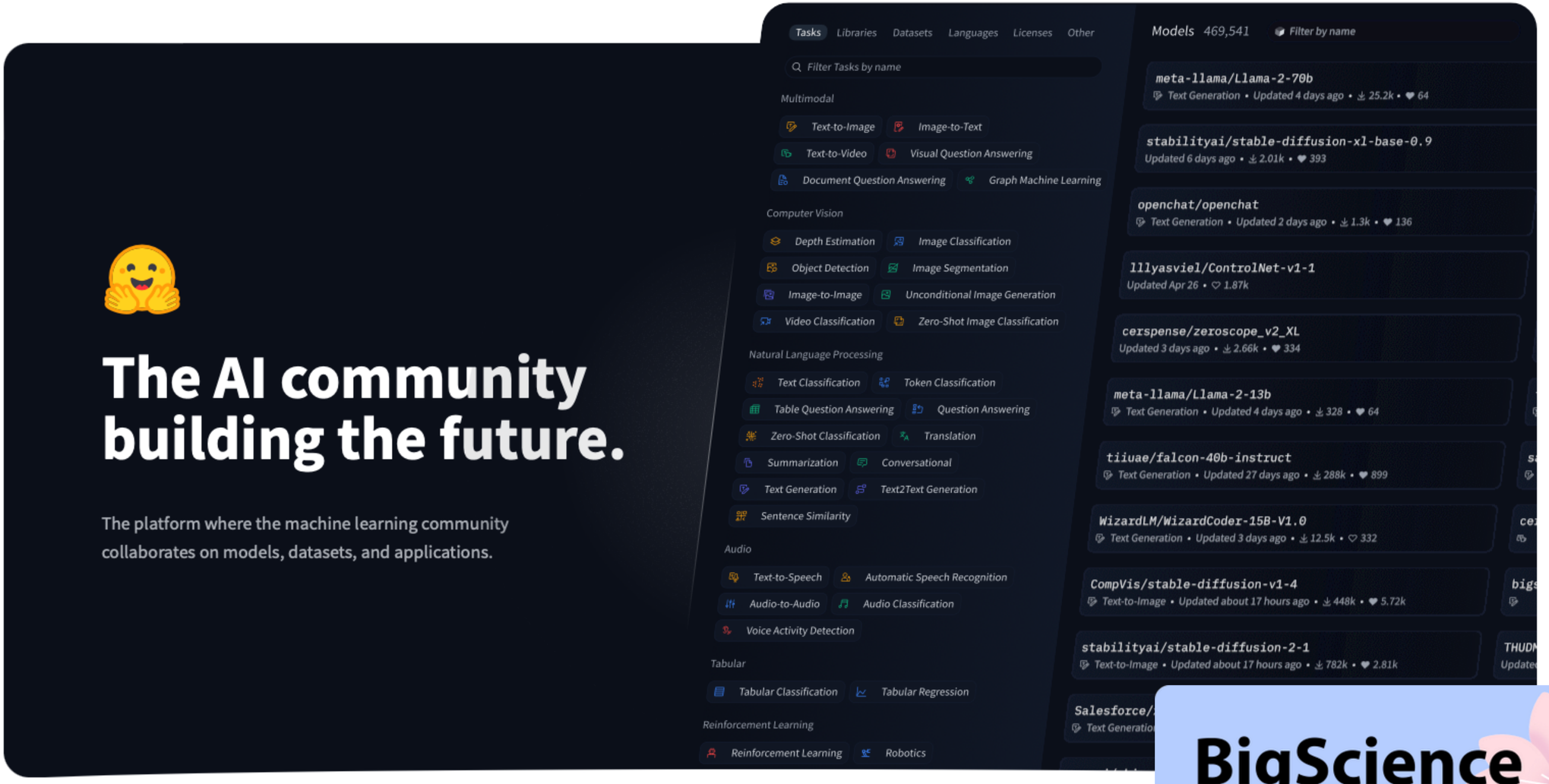
Sub-challenge #3

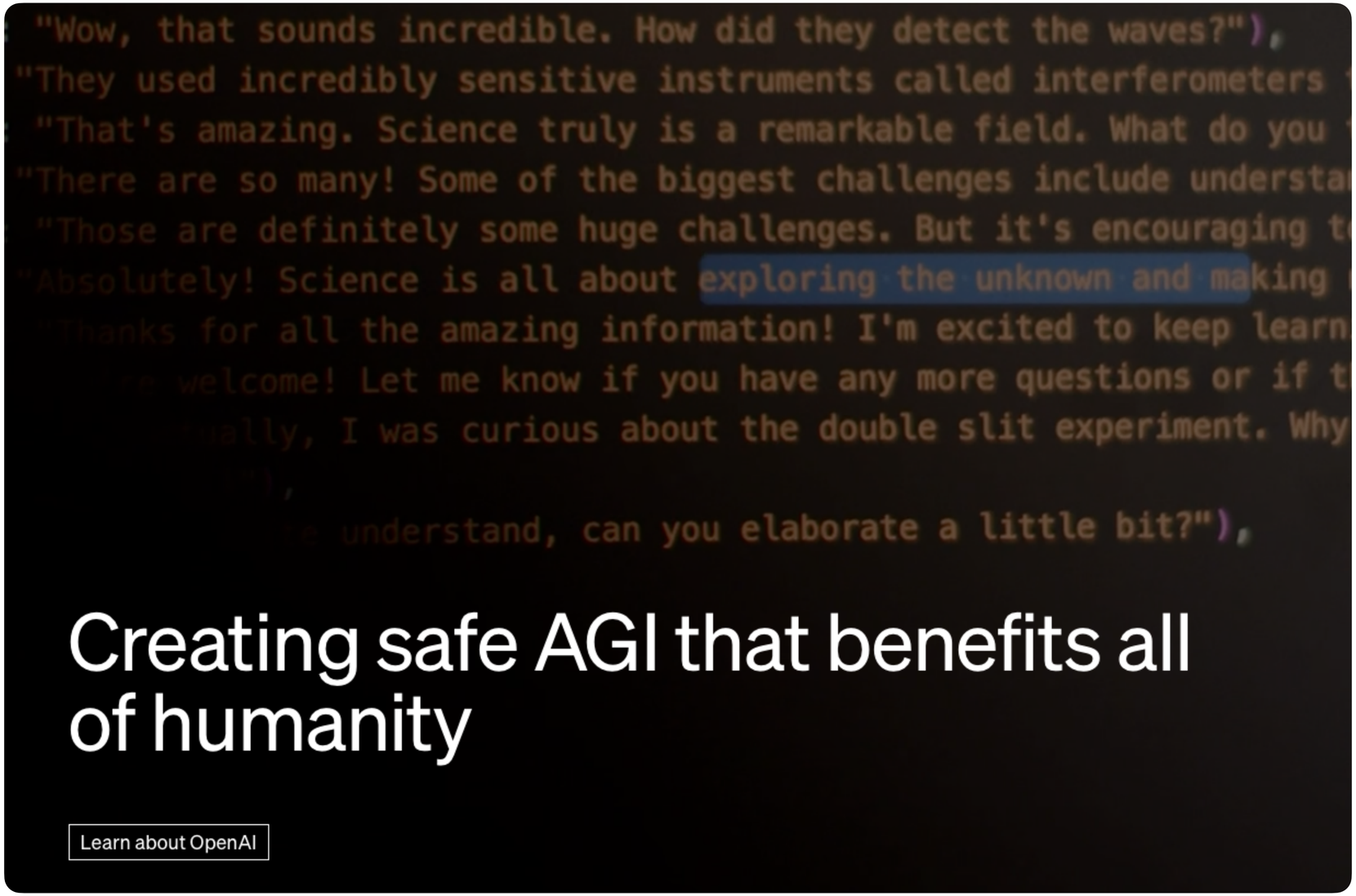
Downstream Effects of Reproducibility

- Open efforts are towards verification of contributions.
- Delicate dynamic between junior/senior staff when errors are found.
- Multiple broad lines of research have been affected (LMs, RienL, RepL).
  No clear procedure.



Challenge 2: Nonreciprocity

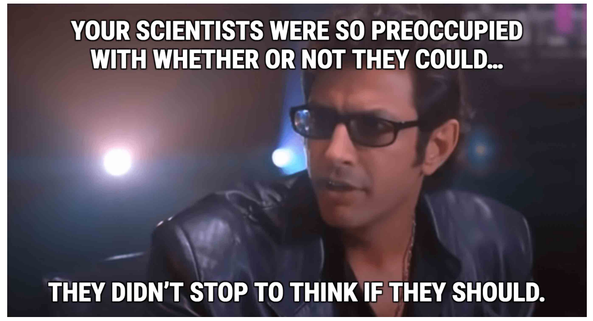

The AI community building the future.
The platform where the machine learning community collaborates on models, datasets, and applications.
Tasks
Libraries
Datasets
Languages
Licenses
Other
Filter Tasks by name
Multimodal
Text-to-Image
Image-to-Text
Text-to-Video
Visual Question Answering
Document Question Answering
Graph Machine Learning
Computer Vision
Depth Estimation
Image Classification
Object Detection
Image Segmentation
Image-to-Image
Unconditional Image Generation
Video Classification
Zero-Shot Image Classification
Natural Language Processing
Text Classification
Token Classification
Table Question Answering
Question Answering
Zero-Shot Classification
Translation
Summarization
Conversational
Text Generation
Text2Text Generation
Sentence Similarity
Audio
Text-to-Speech
Automatic Speech Recognition
Audio-to-Audio
Audio Classification
Voice Activity Detection
Tabular
Tabular Classification
Tabular Regression
Reinforcement Learning
Reinforcement Learning
Robotics
Models 469,541
Filter by name
meta-llama/Llama-2-70b
Text Generation • Updated 4 days ago • 25.2k • 64
stabilityai/stable-diffusion-xl-base-0.9
Updated 6 days ago • 2.01k • 393
openchat/openchat
Text Generation • Updated 2 days ago • 1.3k • 136
lllyasviel/ControlNet-v1-1
Updated Apr 26 • 1.87k
cerspense/zeroscope_v2_XL
Updated 3 days ago • 2.66k • 334
meta-llama/Llama-2-13b
Text Generation • Updated 4 days ago • 328 • 64
tiiuae/falcon-40b-instruct
Text Generation • Updated 27 days ago • 288k • 899
WizardLM/WizardCoder-15B-V1.0
Text Generation • Updated 3 days ago • 12.5k • 332
CompVis/stable-diffusion-v1-4
Text-to-Image • Updated about 17 hours ago • 448k • 5.72k
stabilityai/stable-diffusion-2-1
Text-to-Image • Updated about 17 hours ago • 782k • 2.81k
BigScience


"Wow, that sounds incredible. How did they detect the waves?"),
"They used incredibly sensitive instruments called interferometers
"That's amazing. Science truly is a remarkable field. What do you
"There are so many! Some of the biggest challenges include understa
"Those are definitely some huge challenges. But it's encouraging t
"Absolutely! Science is all about exploring the unknown and making
"Thanks for all the amazing information! I'm excited to keep learn
welcome! Let me know if you have any more questions or if t
ually, I was curious about the double slit experiment. Why
understand, can you elaborate a little bit?"),
Creating safe AGI that benefits all of humanity
Learn about OpenAI

Conclusions

Could we?

- Credit attribution.
- Self-correction.
- Preservation.



cmry.github.io | travisjwiltshire.com

18