Reproducible Machine Learning Workflows for Scientists Workshop 2025

University of Wisconsin-Madison

August 12th, 2025

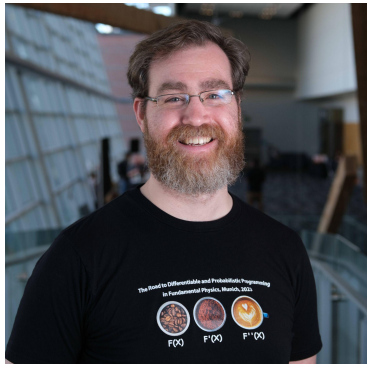
US Research Software Sustainability Institute

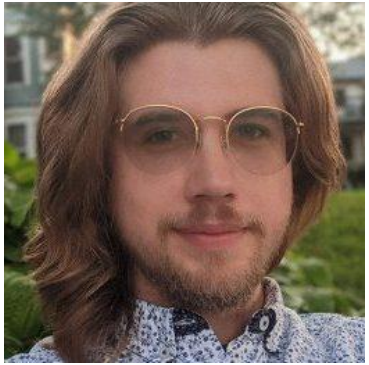
Data Science Institute
UNIVERSITY OF WISCONSIN–MADISON

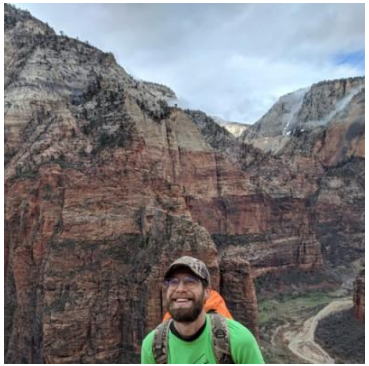
Welcome to the Wisconsin Institute for Discovery

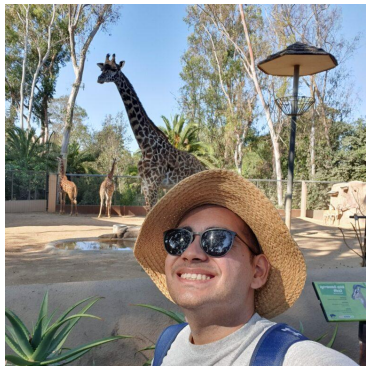
- The workshop will take place this week in the Orchard View Room (Tuesday, Wednesday) and the H.F. DeLuca Forum (Thursday)
- Workshop Venue Indico page for more details
- We're **guests in this space** (thank you Data Science Hub) so please make sure to clean up the space before you leave
- Take **bags when leave for lunch** (not locked)
- Need to leave **at 16:00 each day**

Instructor Team









Matthew Feickert

Research Scientist

Data Science Institute

Christopher Endemann

Research Cloud Consultant

DoIT Research Cyberinfrastructue

Ryan Bemowski

Data Science Facilitator

Data Science Hub

Raheem Hashmani

Ph.D. Candidate

Department of Physics

Sarah Stevens

Director

Data Science Hub

Support and funding

Logistics (all on Indico)

- Internet access
- Meals (on your own, except dinner Wednesday)
- Pre-workshop survey (still fill it out if you haven't)
- Workshop setup (do this ASAP if you haven't)
- OSPool temporary accounts (do this ASAP if you haven't)
- Wednesday night dinner order (do this **before lunch today**)
  - Go to your Indico registration and fill out the *added* "What Chipotle burrito do you want ordered for you for dinner on Wednesday?" question
  - If you have a dietary requirement that conflicts with burritos please make the "Conflicting dietary requirement (email Matthew for alternative)" option *and email Matthew*

Pre-workshop survey so far

Describe your role in your research group related to software

- Mix of students, researchers, self described script writers, research software engineers, managers, and industry experts

What are your motivations for participating in this workshop?

- Pixi, reproducibility research, learning about machine learning, HTC and HPC

Programming languages

- Mostly: Python, Bash, R
- Some: Rust, Java, JavaScript, C++

Pre-workshop survey so far

Most are using:

- Version control and GitHub/GitLab
- Linux containers (at least once)
- AI/ML software in research
- Conda packages (at least sometimes)

Most aren’t yet using:

- Hardware accelerators (GPUs)
- Batch computing (HTCondor, SLURM)
- Pixi

Pre-workshop survey so far

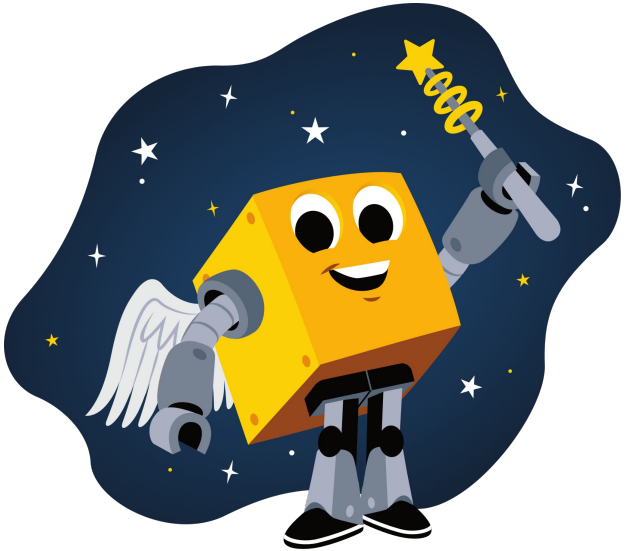
You're installing software by/using:

- building from source
- pip (PyPI)
- conda (conda-forge?)
- R scripts (CRAN)
- operating system package manager (brew, apt)
- "the hard way" 😂

Find it difficult to install:

- Variety of domain specific tools (you're in good company)
- CUDA accelerated tools (TensorFlow, PyTorch, JAX)
- Tools that have no packages (source installs)

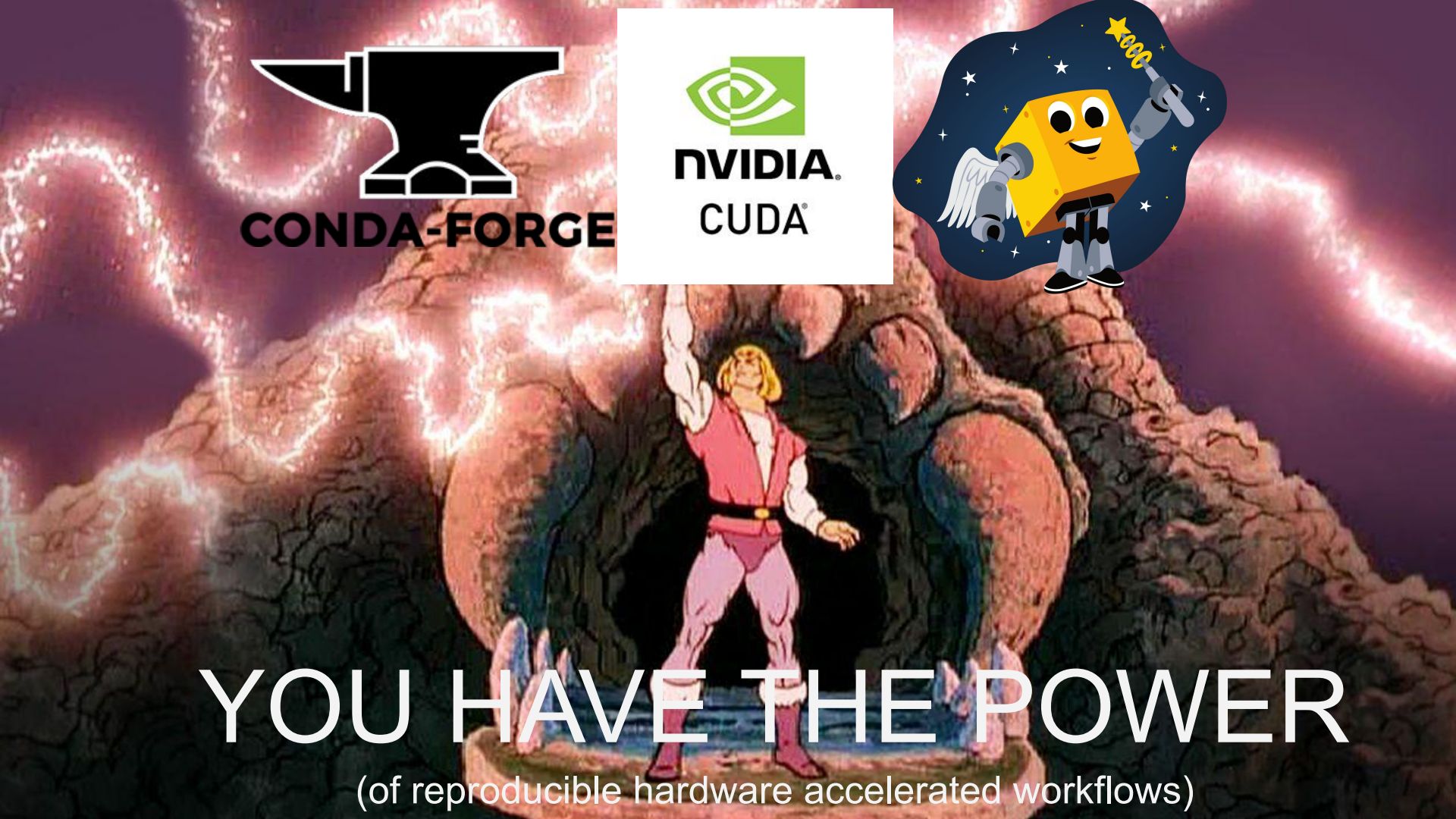
Auspicious times



2015
(SciPy 2015)

2022
(CUDA v12)

2023


CONDA-FORGE
NVIDIA
CUDA
YOU HAVE THE POWER
(of reproducible hardware accelerated workflows)