

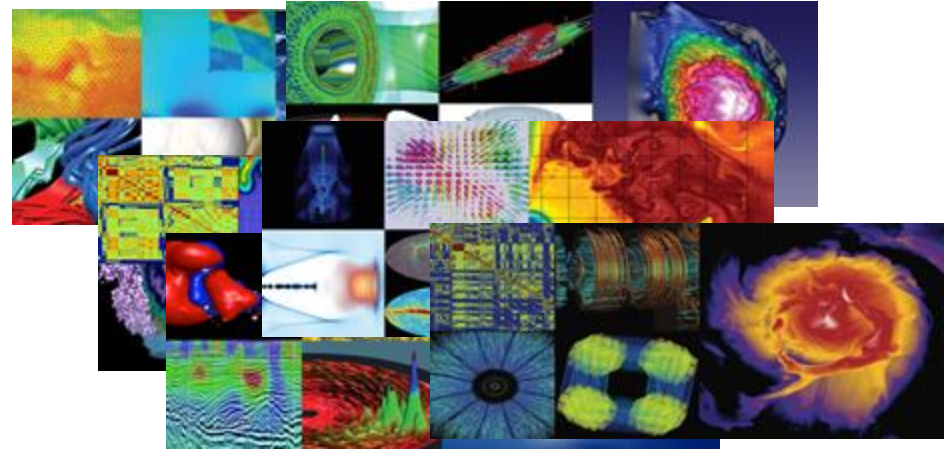
AI Software Usage at NERSC



NAIRR Software Workshop
Dec 3rd 2024

NERSC AI Services Team
Steven Farrell, Shashank Subramanian,
Wahid Bhimji

NERSC: **Mission** HPC for the Dept. of Energy Office of Science



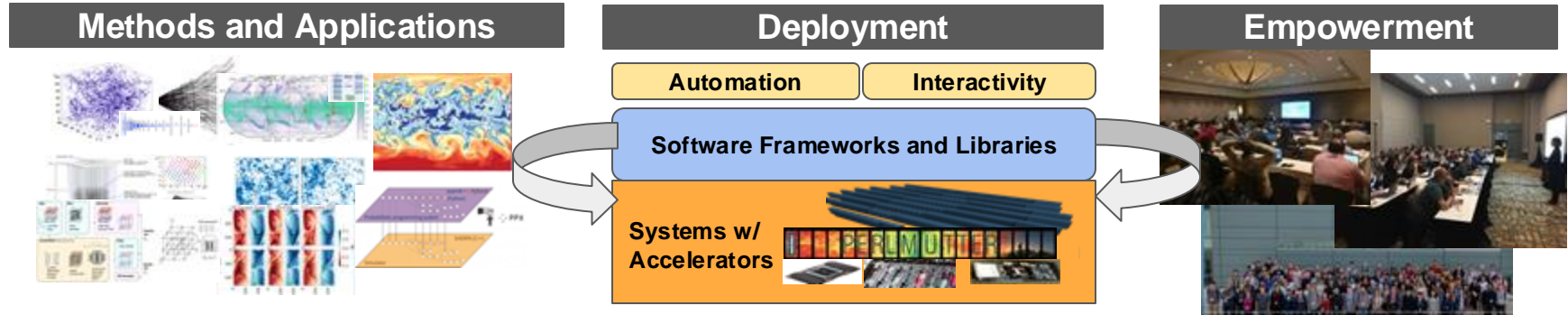
Large compute and data systems

- Perlmutter: ~7k A100 GPUs
- 128PB Community Filesystem

Broad science user base

- > 10,000 users,
- 1000 projects,

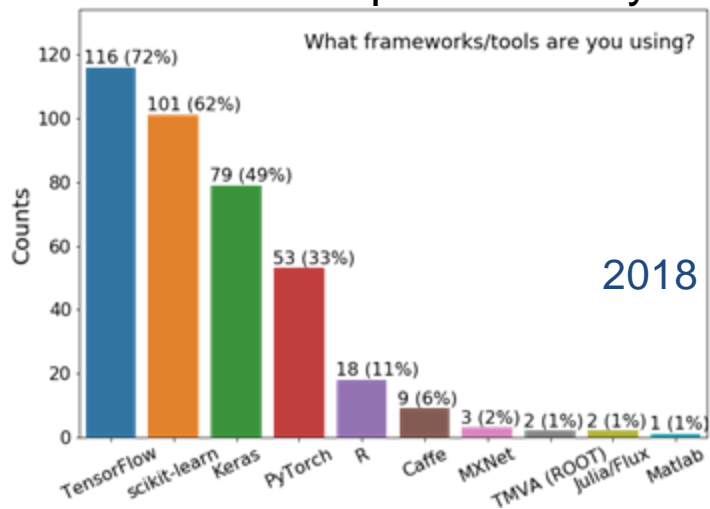
Building a Scientific AI ecosystem



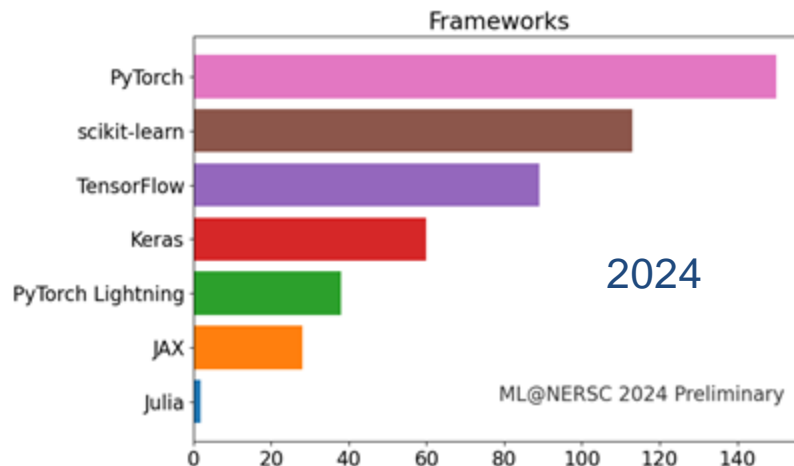
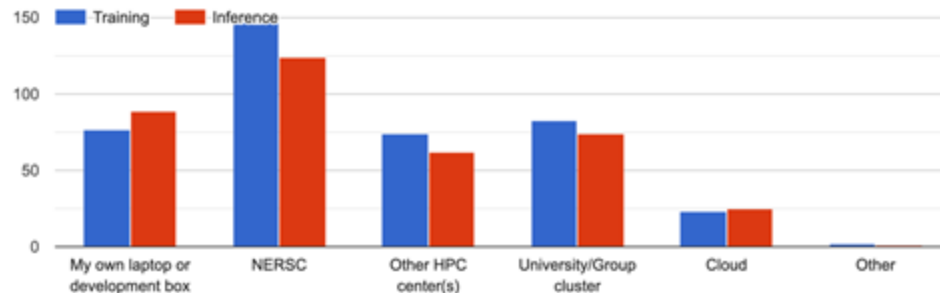
- **Deploy** optimized hardware and software systems
 - Work with vendors for optimized AI software (e.g. NCCL on Slingshot)
- **Apply** AI for science using cutting-edge techniques
 - “NESAP” and strategic projects - leverage lessons learned for broad ecosystem
- **Empower** and develop workforce through seminars, training and schools as well as staff, student intern and postdoctoral programs
 - Over 20 DL@Scale tutorials (e.g. SC18-23), 1000s of total participants

NERSC ML Survey

- Survey sent to all NERSC users
 - captures their use both at NERSC and elsewhere
- Every 2 years since 2018
 - >200 responses this year



Where do you run your ML models?



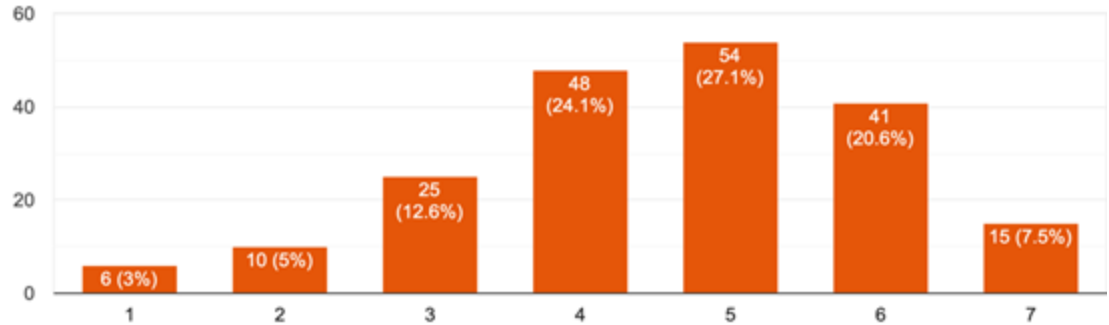
Maturity

- Some shift towards more mature applications since 2022
- Still a lot of R&D compared to in-production
- 37% claim to have AI “in production” for science

How do you rate your expertise in Machine Learning?

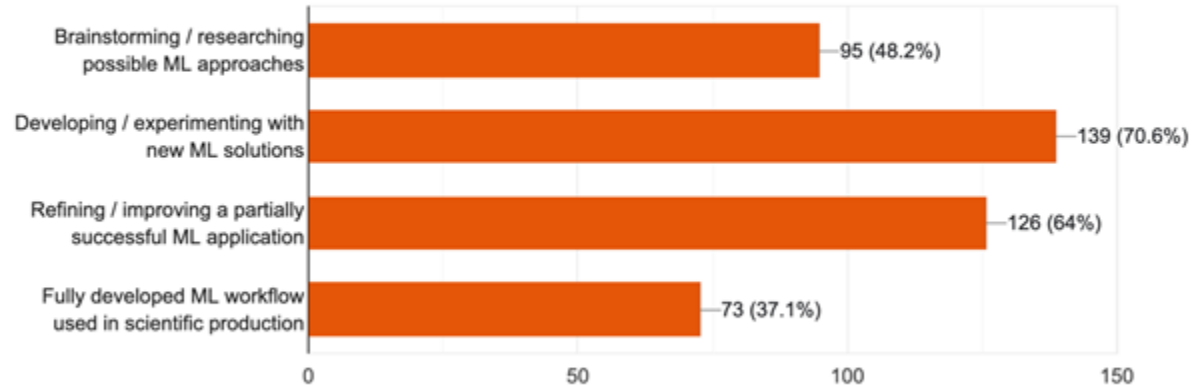
2024

199 responses

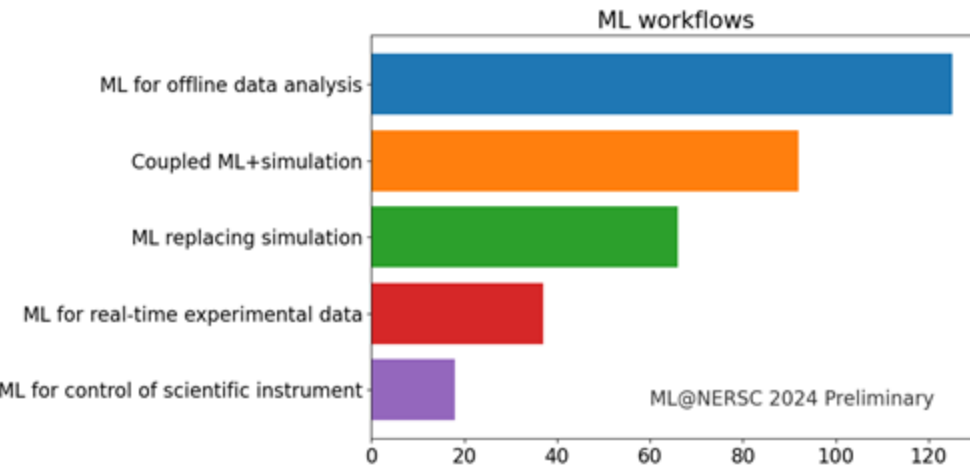
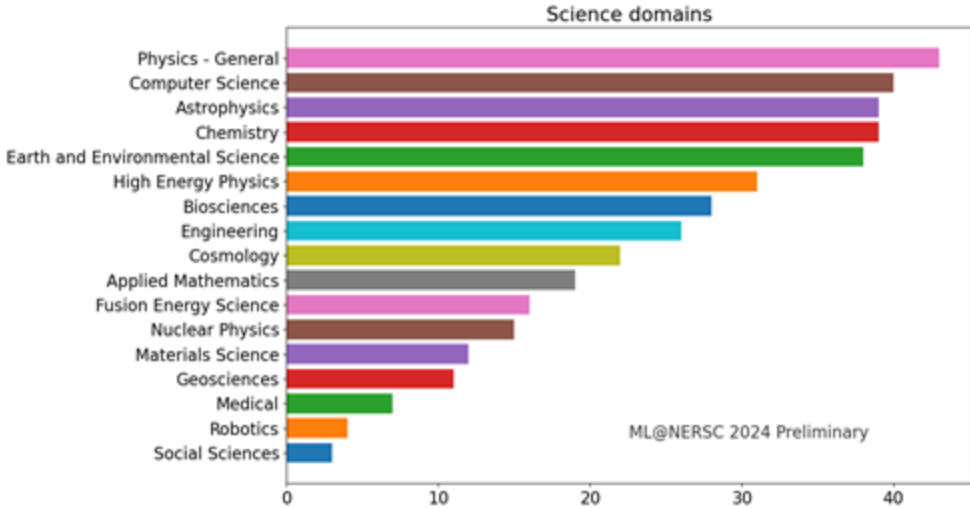
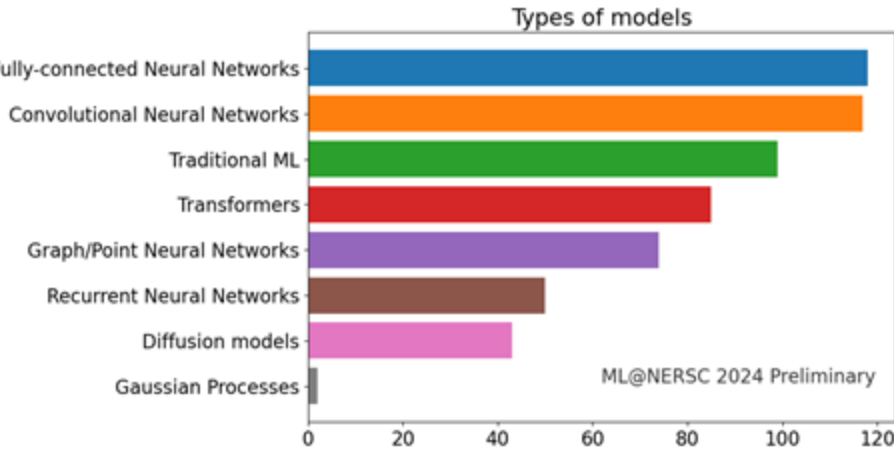
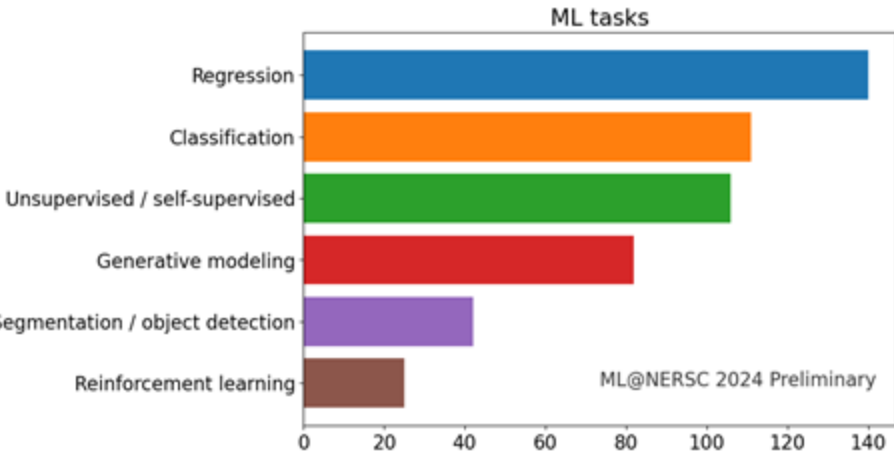


What is the level of maturity of ML in your research? (mark all that apply to your projects)

197 responses

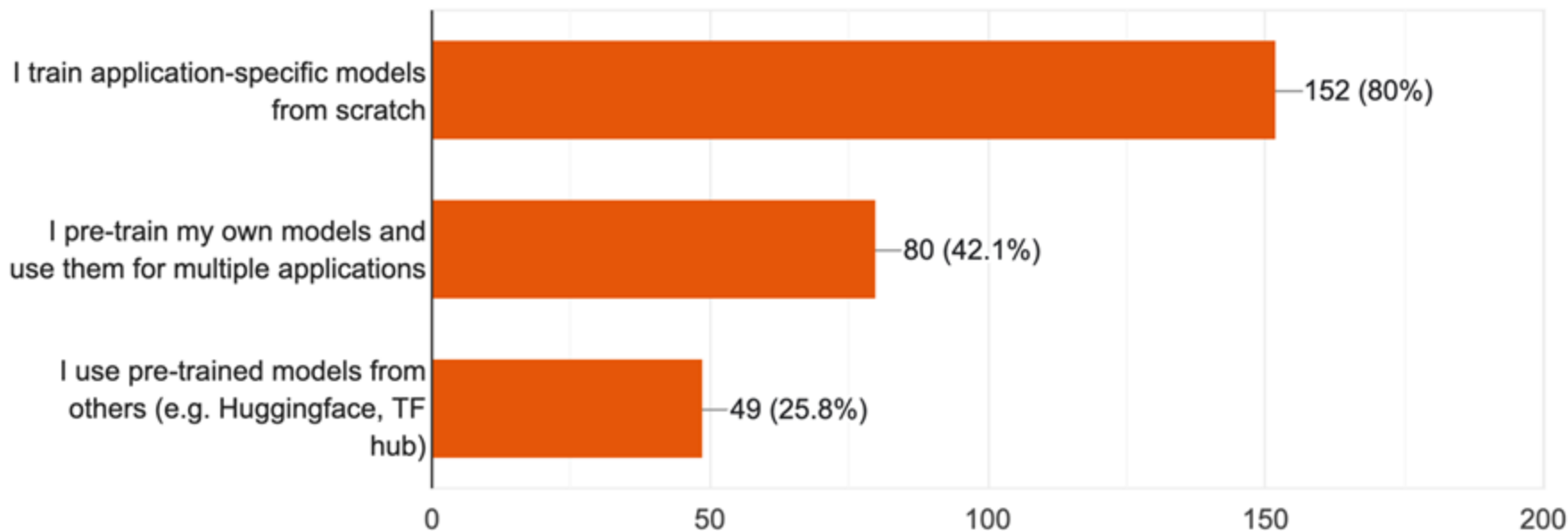


Range of applications



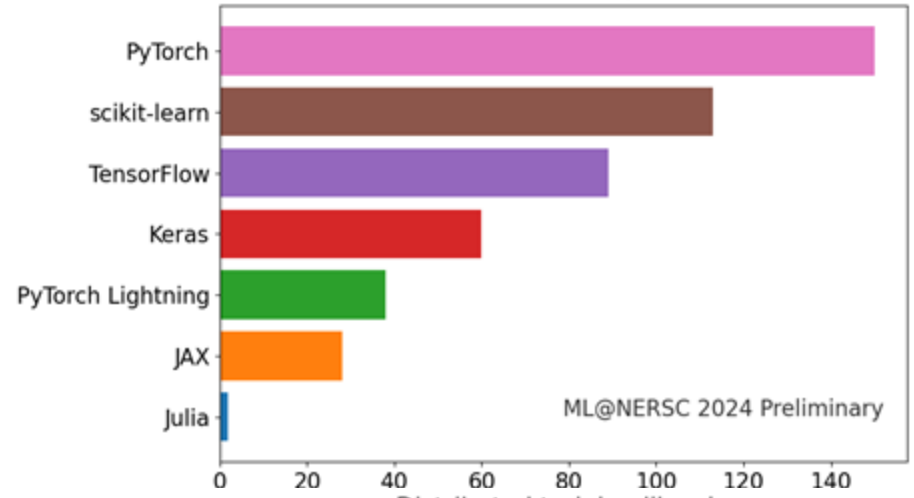
Do you use pre-trained models?

190 responses

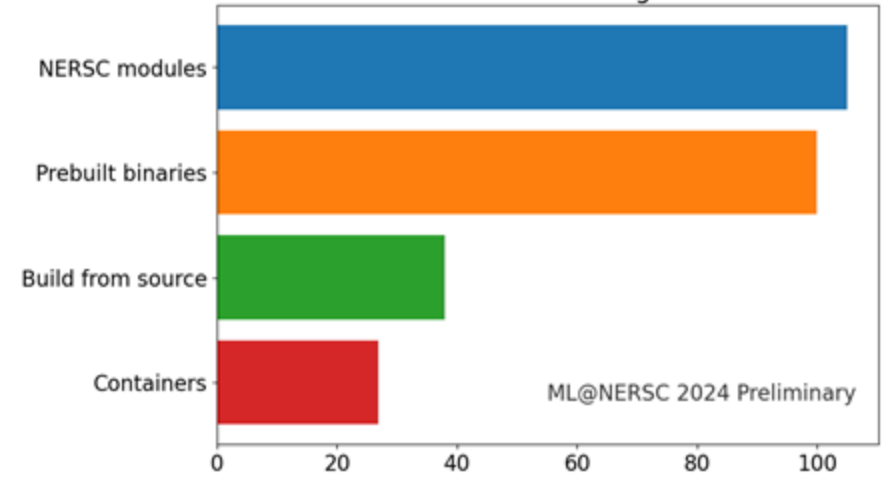


AI Software

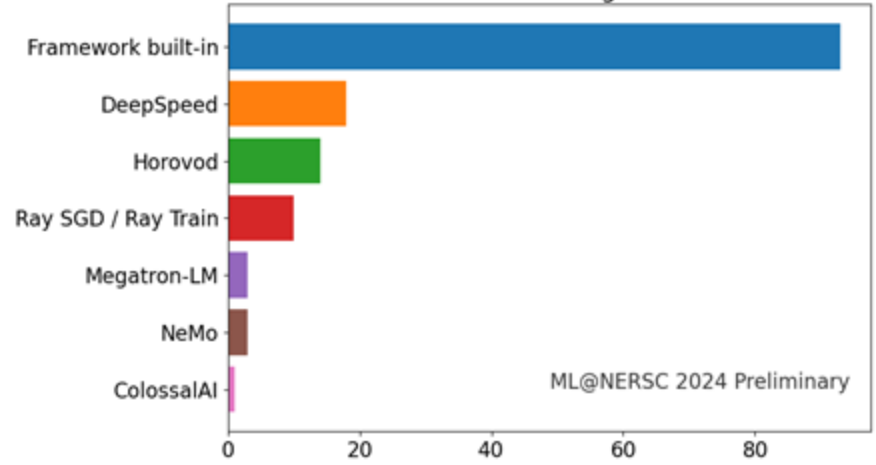
Frameworks



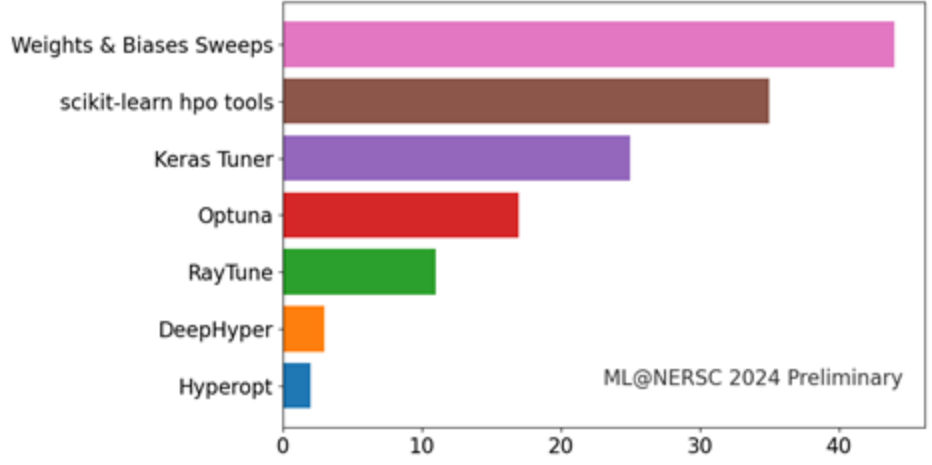
ML software usage



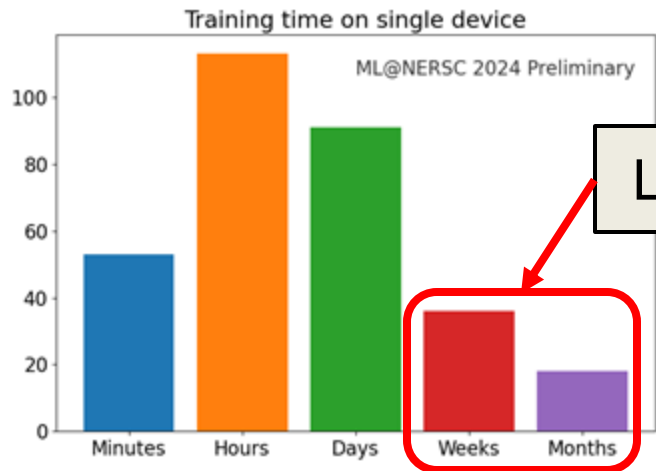
Distributed training libraries



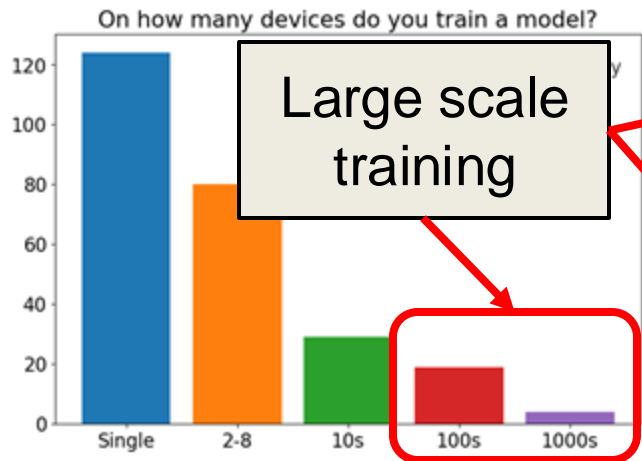
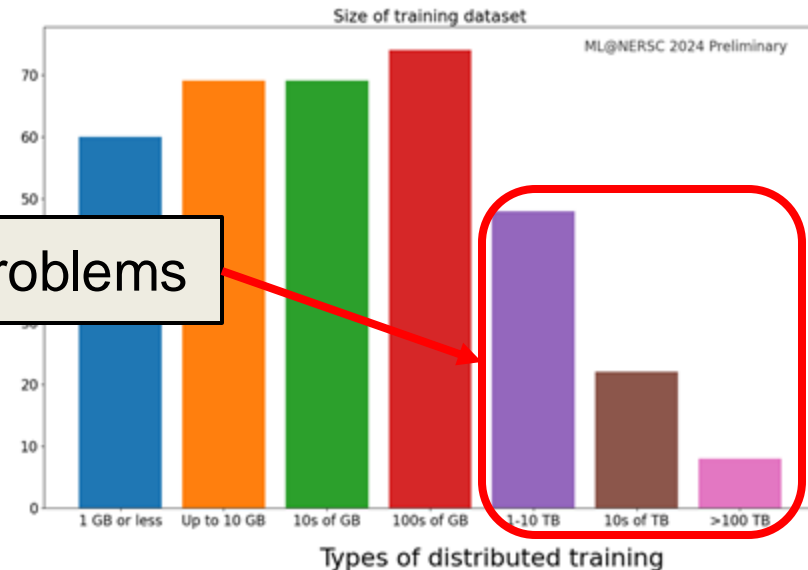
HPO tools



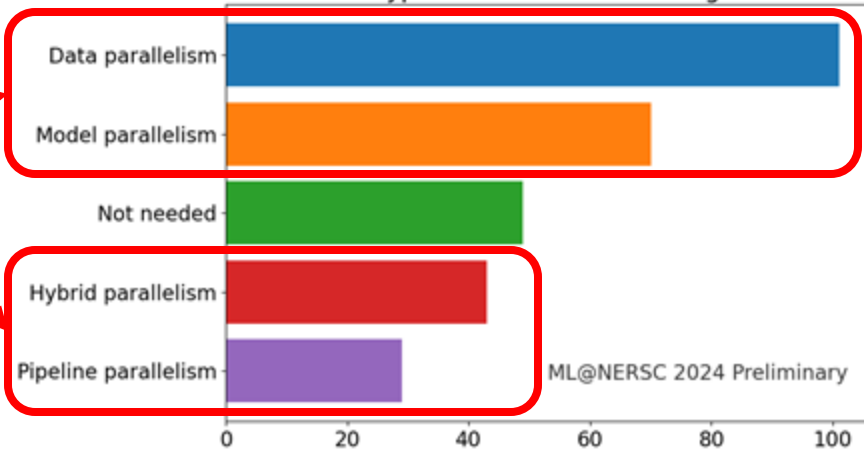
Need for AI at scale



Large problems



Large scale training

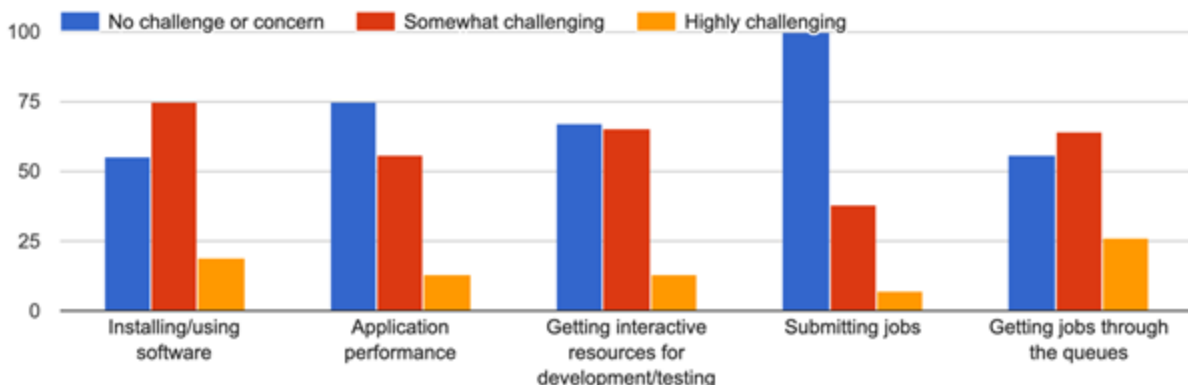


User challenges

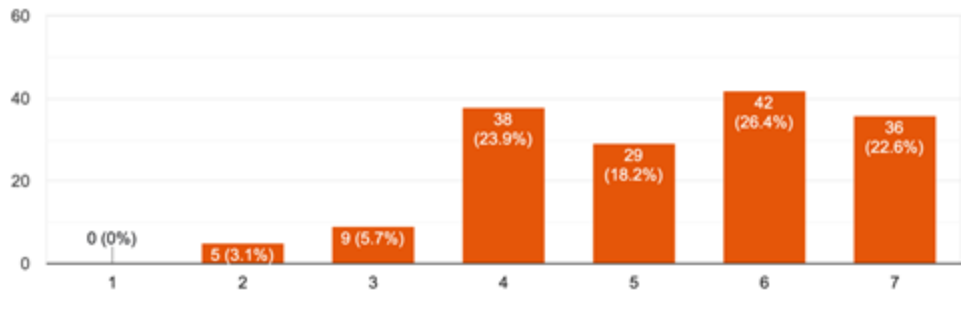
Some free-text comments [paraphrased]:

- Need for more resources
- Desire for comprehensive software stacks
- Guidance on how to deploy models
- Entry level training material (inc on containers; batch submission; performance..)

For your ML workloads at NERSC, what are your biggest pain points? Please rate your level of concern or challenges with the following:



How well does the ML software stack at NERSC satisfy your needs?
159 responses



We also directly monitor user workload

- Instrument user [python imports](#)
 - Catches user imports whether they use our modules or not (some data is missed like jobs that hit wall-clock time)
 - Unique users >30x since 2017

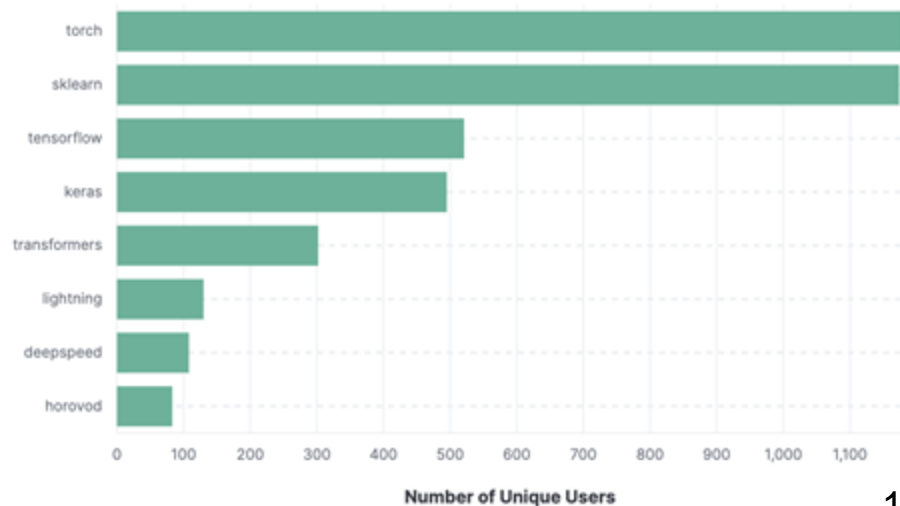
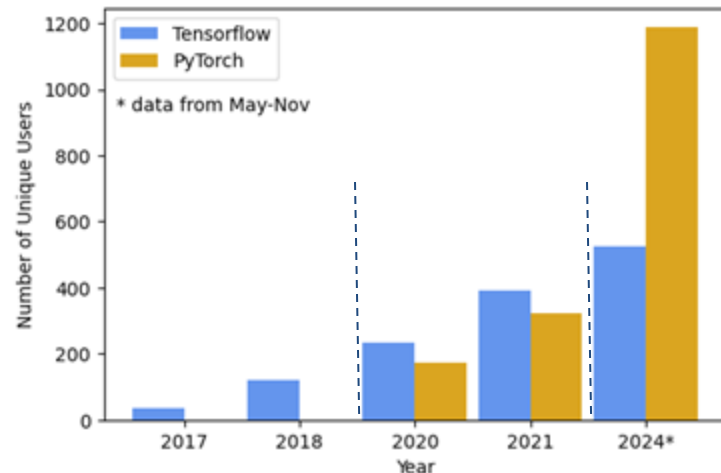
- May-Nov 2024 Data:

>1k Unique PyTorch users

Twice as popular as Tensorflow

SciKit-learn remains popular

Significant usage of distributed DL



Conclusions

- NERSC has seen growth of AI in broad science user-base over the last decade
 - Now >1000 unique users of Deep Learning libraries
 - Shift in maturity and expertise towards production, but still need for experimentation and demand for introductory training resources
- Standard frameworks (e.g PyTorch) and open source tools dominate
 - These frameworks have changed drastically over this period and others like Jax are rising fast
 - Growth in using pre-trained models, generative AI, transformers etc. but majority of use-cases (from survey results) involve other model architectures, and training models from scratch.
 - Also significant need for interfacing with other scientific software/simulation.
- Opportunities for further analysis from direct workload monitoring