

AI for Science and Engineering

Geoffrey C. Fox, Gregor von Laszewski

OVERVIEW

There exists a need for students to be familiar with **cyberinfrastructure**, **Science applications**, as well as **artificial intelligence**. Instead of learning all these concepts separately, it is important that a holistic approach is chosen that integrates these concepts. The outcome will be a well-educated workforce useful for research, government, and industry.

Due to the rapid change in any of the areas, it is important that the **teaching material can and is updated** regularly to address the newest developments. We have implemented such an approach over several years and focused on particular subtopics to arrive at a state that allows us to teach AI for science.

Focused topics include:

- Introduction of Python for AI for Science
- Cloud Computing,
- HPC Computing,
- Big Data Applications and Analytics,
- Artificial Intelligence/Machine Learning.

We utilize open-source technologies such as GitHub to manage the growing list of educational material. Teaching could be conducted in courses but has been found most effective in research experiences.

Meetups and hackathons enhance these activities.

Material can be reused in other educational activities as our contributions to nanoHub and FAMU REUs demonstrated.

Participants are encouraged to improve the content of the material.

All material is publicly available under Open-Source License.

Cyberinfrastructure Training 90% CI 10% AI4Sci



MLCommons AI4Science And HPC Benchmarks

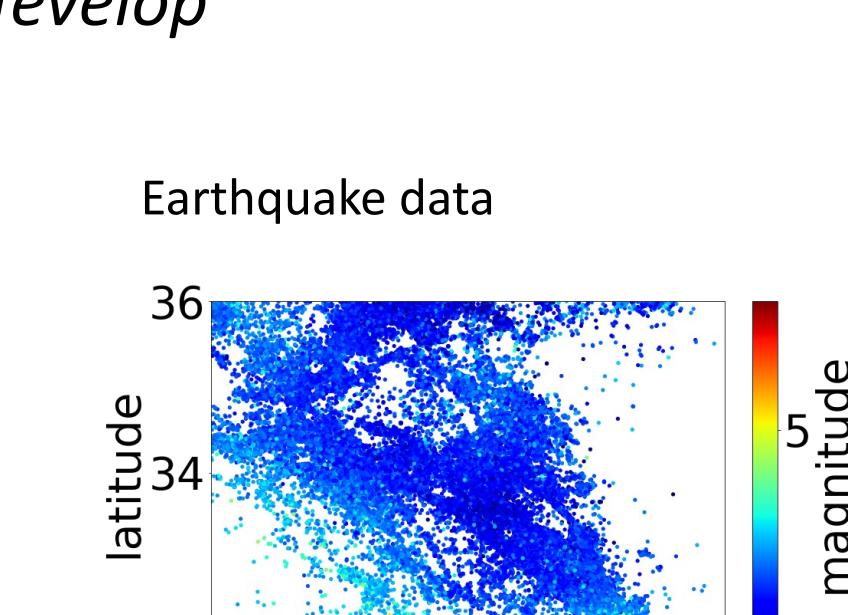
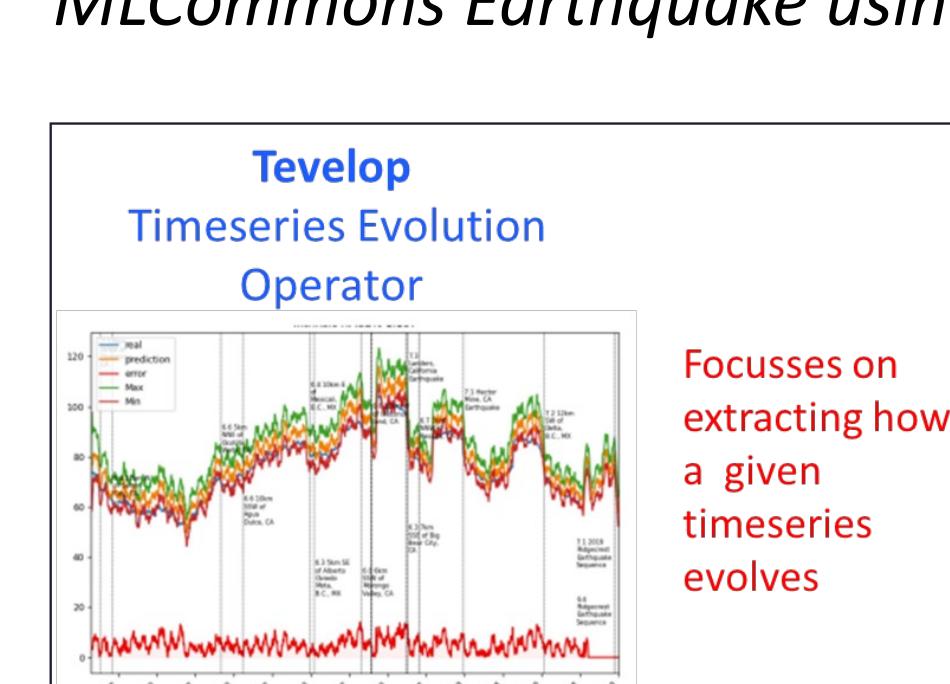
2000 pages of material in the focus topics



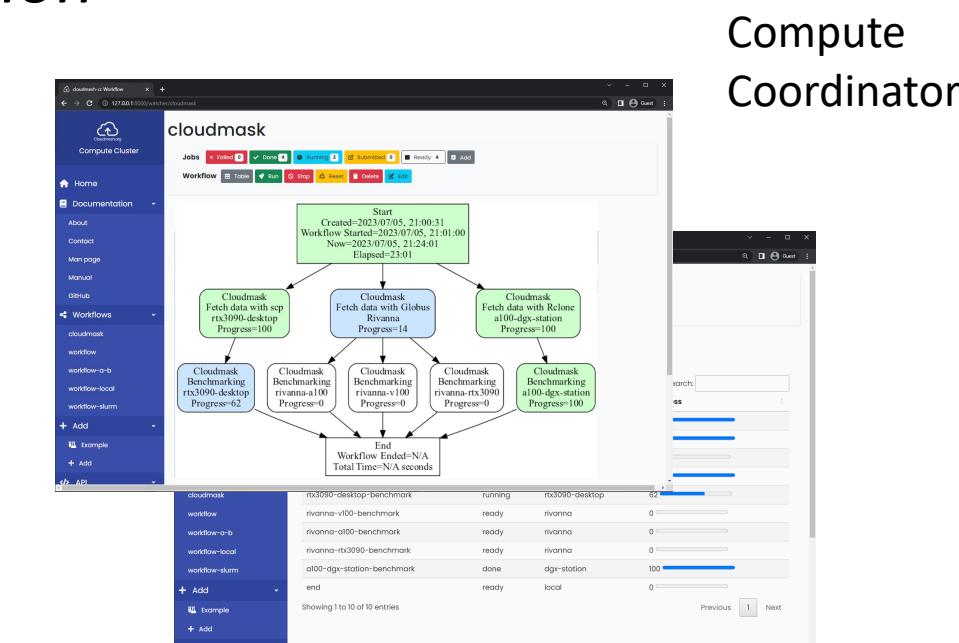
MLCommons Cloudmasking Using Cloudmesh Experiment Execution



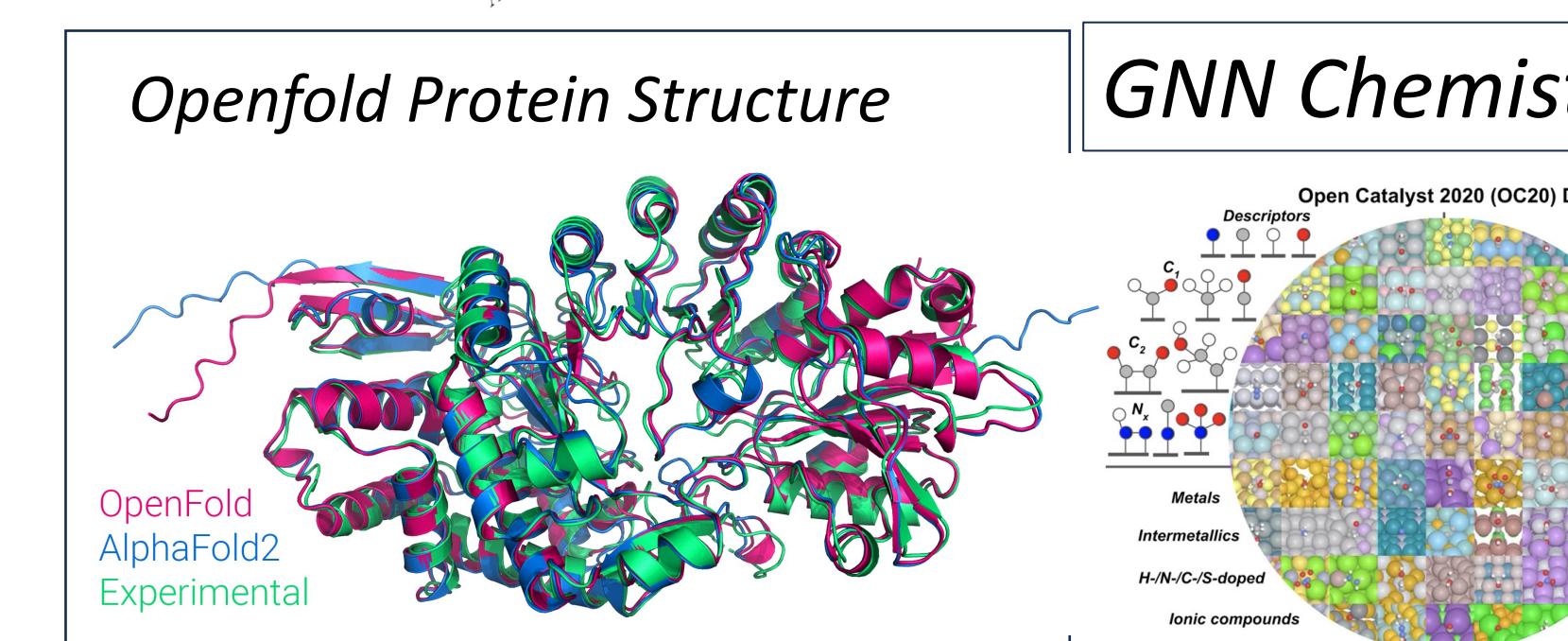
MLCommons Earthquake using Tdevelop



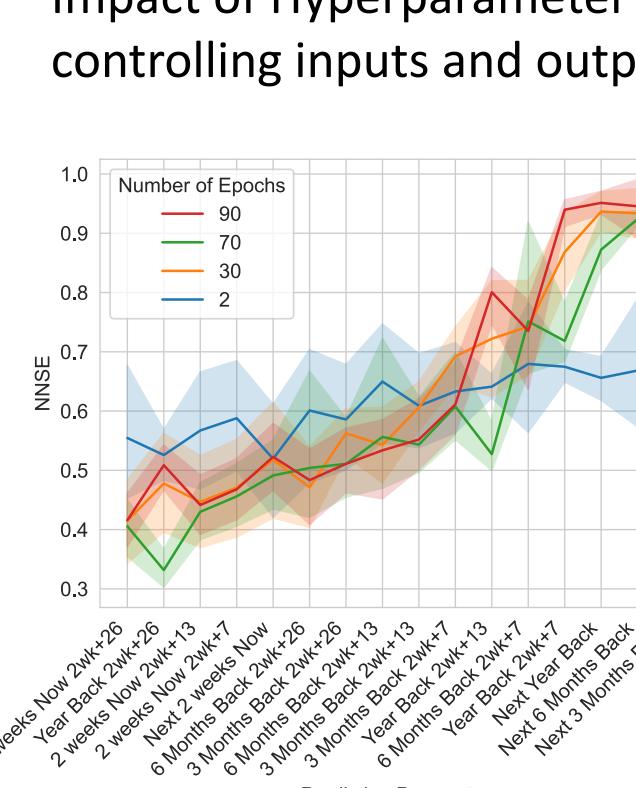
Architecture



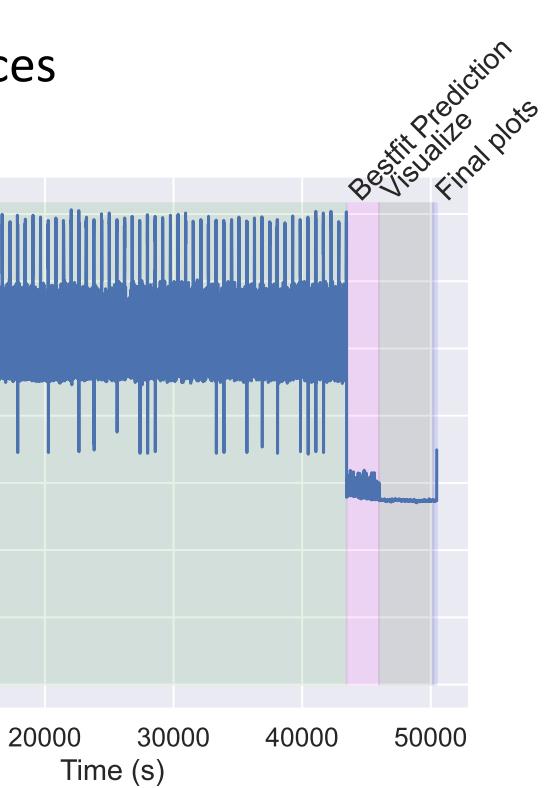
Cloudmesh Compute Coordinator



Impact of Hyperparameter controlling inputs and outputs



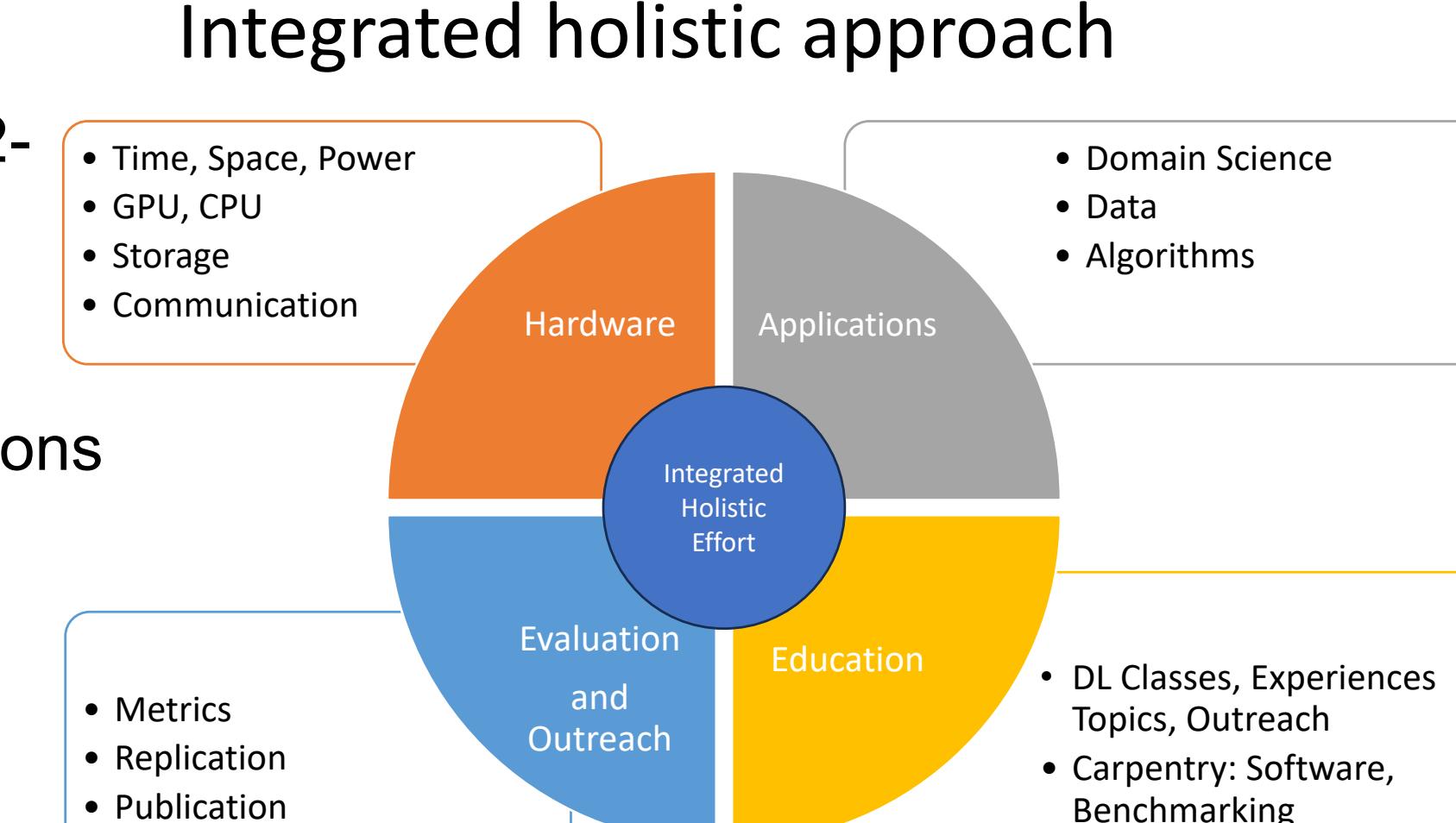
Energy traces



Highlights

- > 200 undergraduate applied Fall 22–Summer 23; > 50 applied Fall 2023
- > 2000 pages educational material
- > 170 repositories
- Interaction of MLCommons applications
- Significant technology development
 - Hybrid cloud computing interface
 - Benchmarking
 - GPU use
 - Experiment management
 - Compute resource coordination
 - Online and PDF material publication framework
 - Individually customizable training material selection and production

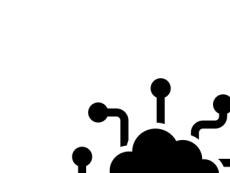
Integrated holistic approach



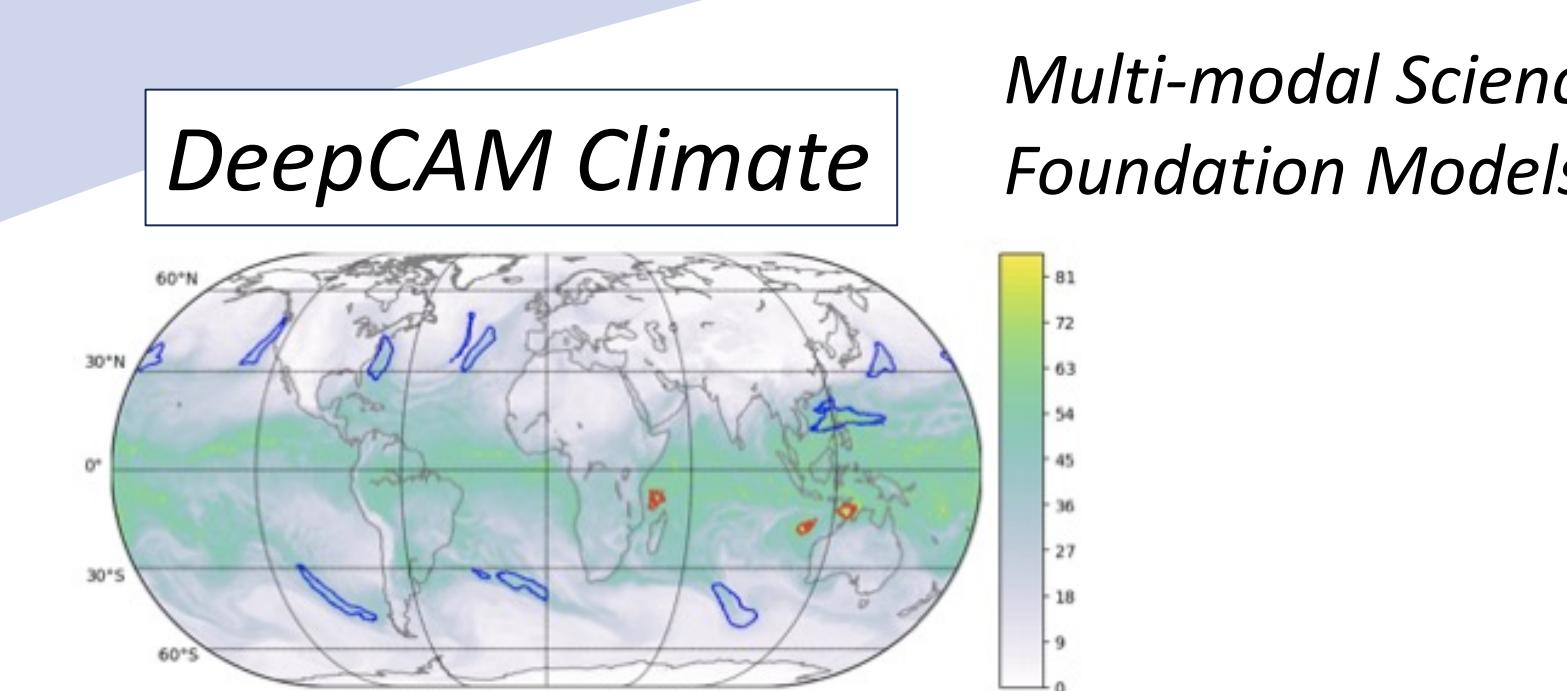
AI for Science Training 10% CI 90% AI4Sci

CyberTraining for AI for Engineering and Science 50% CI 50% AI4Sci

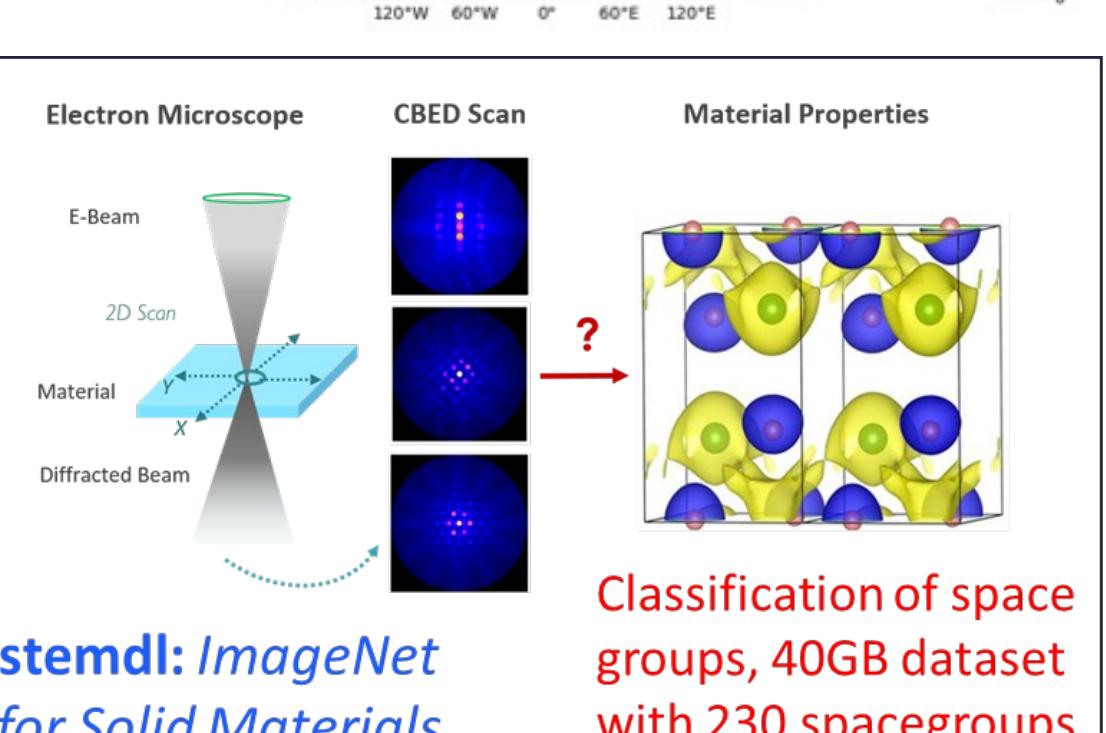
Science Analysis Patterns Common API



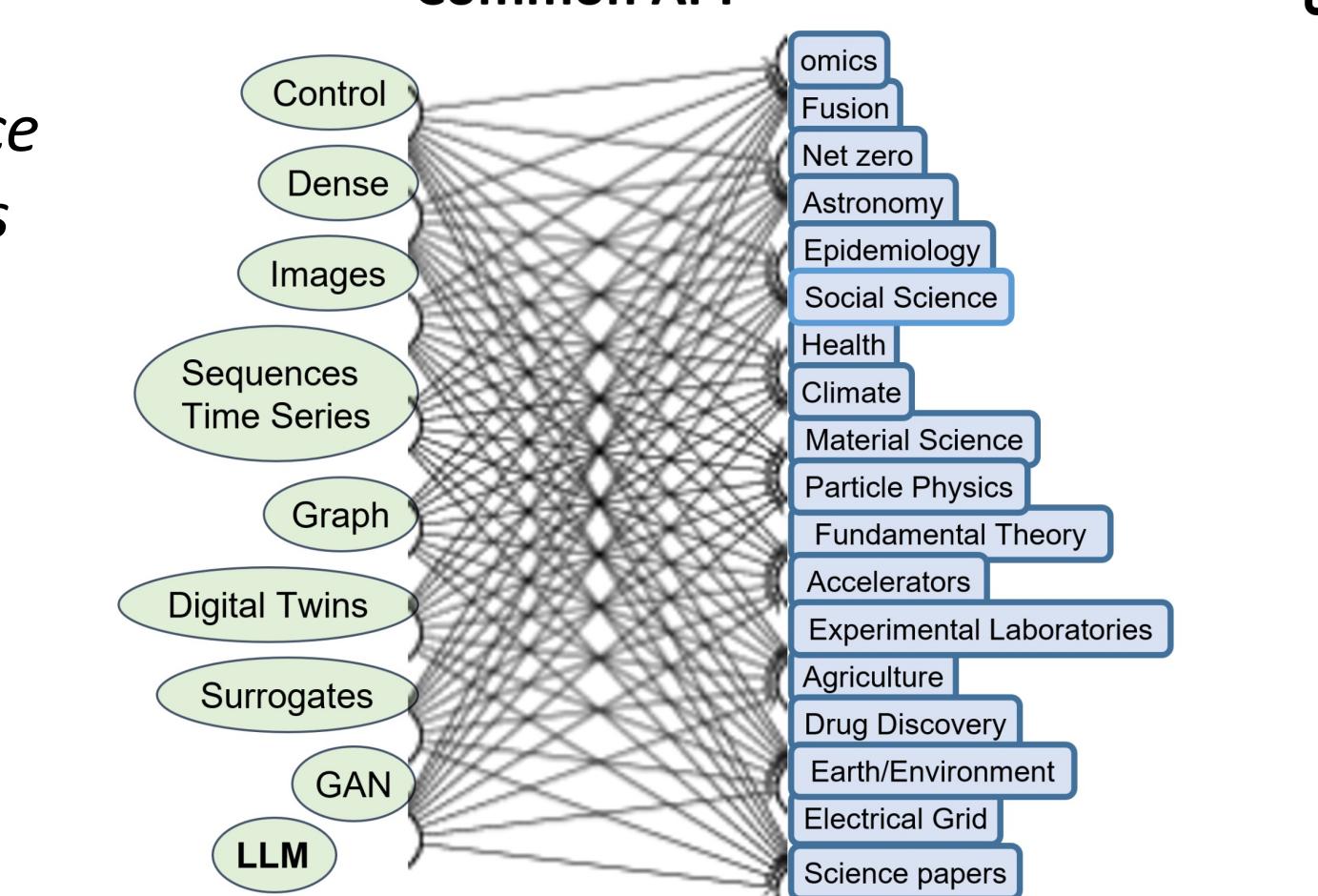
Multi-modal Science Foundation Models



Classification of space groups, 40GB dataset with 230 spacegroups



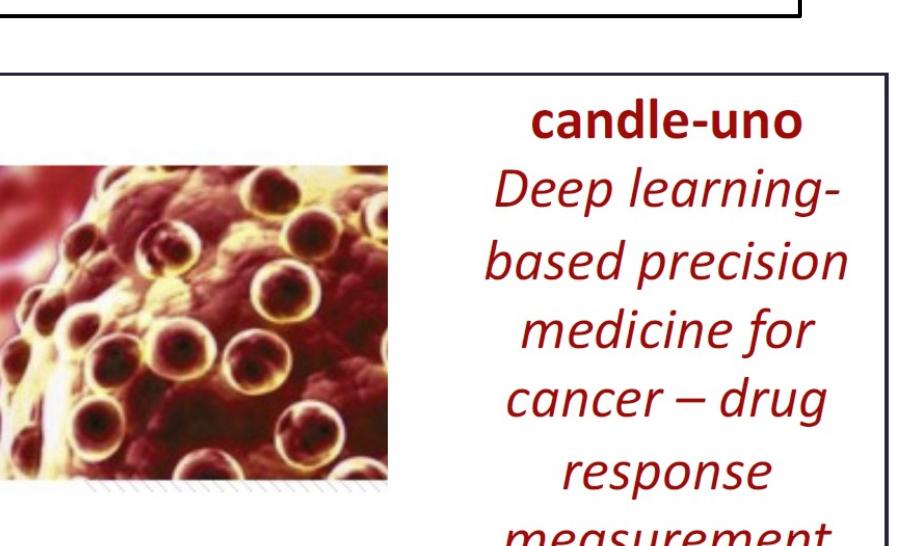
Applications



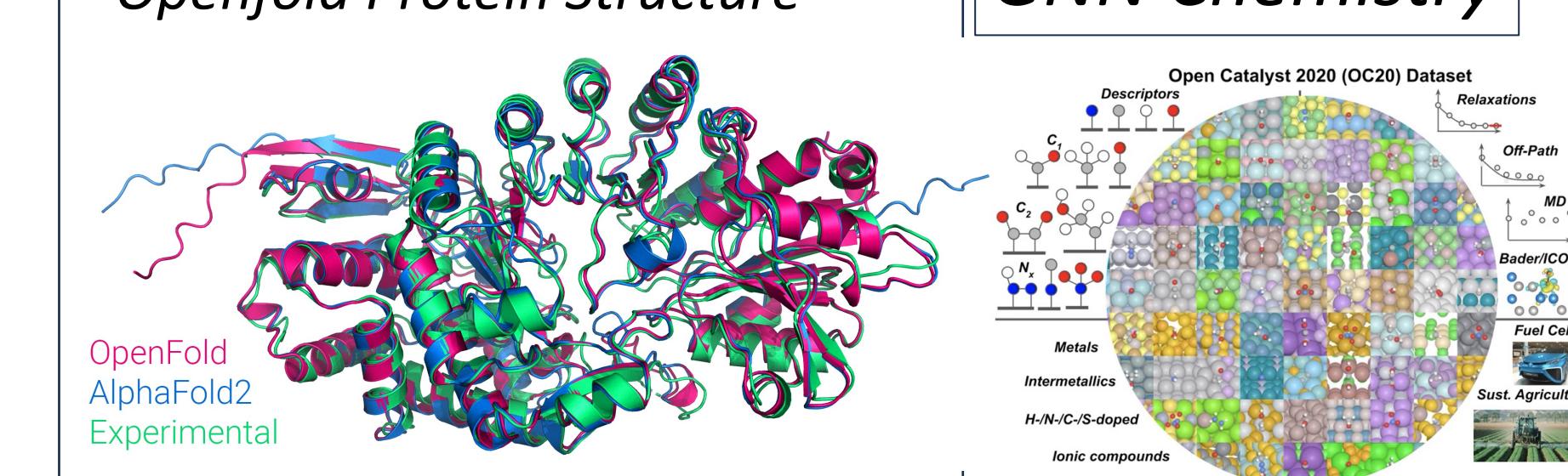
Cosmoflow Universe



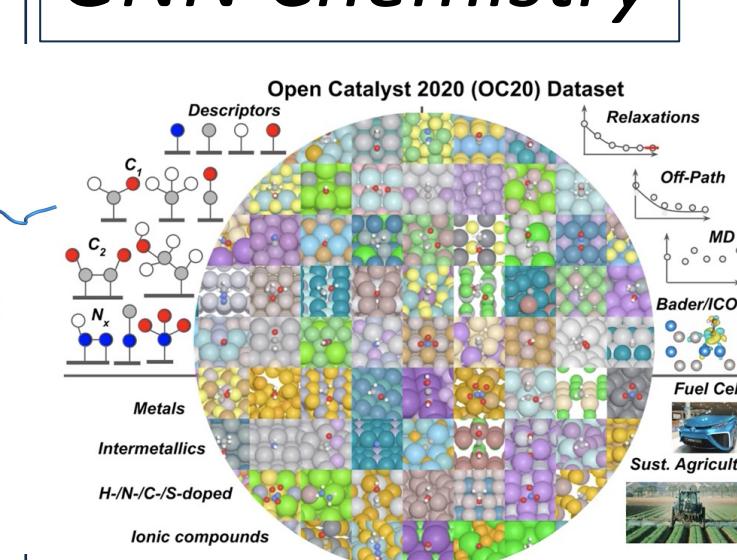
Candle-Uno Drug Discovery



Openfold Protein Structure



GNN Chemistry



Selected References

1. G. von Laszewski, J.P. Fleischer, Geoffrey C. Fox, et.al. **Templated Hybrid Reusable Computational Analytics Workflow Management with Cloudmesh, Applied to the Deep Learning MLCommons Cloudmask Application**. In: eScience'23. Second Workshop on Reproducible Workflows, Data, and Security (ReWorDS 2022). Limassol, Cyprus, Oct. 2023.
2. G. von Laszewski, J.P. Fleischer, Robert Knuuti, Geoffrey. C. Fox, et.al. **Opportunities for Enhancing MLCommons Efforts while leveraging Insights in High-Performance Big Data Systems Gained from Educational MLCommons Earthquake Benchmarks Efforts**. submitted. Sept. 2023.
3. J. P. Fleischer, G. von Laszewski, et.al. **Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory**. In: Algorithms 15.7 (2022). issn: 1999-4893. doi: 10.3390/a15070230. url: <https://www.mdpi.com/1999-4893/15/7/230>. <https://github.com/cyberaide/paper-cloudmesh-cc-ieee-5-pages/blob/main/vonLaszewski-cloudmesh-cc.pdf>.
4. G. von Laszewski, J.P. Fleischer, and Geoffrey C. Fox. **Hybrid Reusable Computational Analytics Workflow Management with Cloudmesh**. Tech. rep. arXiv, Oct. 2022. doi: 10.48550/ARXIV.2210.16941. url: <https://arxiv.org/abs/2210.16941>.
5. Gregor von Laszewski, Geoffrey C. Fox, et.al. **Using GAS for Speedy Generation of HybridMulti-Cloud Auto Generated AI Services**. In: IEEE COMPSAC 2021: Intelligent and Resilient Computing for a Collaborative World45th Anniversary Conference. Status = Added in NSF-PAR. All Virtual: IEEE, July 2021. url: <https://laszewski.github.io/papers/vonLaszewski-openapi.pdf>
6. Gregor von Laszewski and Geoffrey C. Fox. **Proceedings of the CyberTraining REU2022 CyberTraining for Students and Technologies from Generation Z**. Tech. rep. <https://github.com/cloudmesh-community/pub/raw/master/docs/vonLaszewski-reu2022.pdf>.
7. Gregor Laszewski,, Geoffrey C. Fox, et.al. **Proceedings 2021 FAMU REU supported by Cybertraining-DSC**. Online Books Digital Science Center. Bloomington, IN: Online, Indiana University, Aug. 2021. <https://cybertraining-dsc.github.io/docs/pub/reu2021.pdf>.

Selected Technologies

1. **Bookmanager** to create custom collections of educational material from GitHub documents. <https://github.com/cyberaide/bookmanager>
2. **Cloudmesh Experiment Executor** to create AI Hyperparameter permutation experiments <https://github.com/cloudmesh/cloudmesh-ee>.
3. **Cloudmesh StopWatch** to simplify Benchmarking <https://github.com/cloudmesh/cloudmesh-common>
4. **Cloudmesh GPU** to simplify collection and display of GPU energy and temperature traces <https://github.com/cloudmesh/cloudmesh-gpu> [

Selected Web Sites

1. **AI-First Engineering Cybertraining**. <https://cybertraining-dsc.github.io/docs/courses/ai-first>
Most recent new course in DSC Cybertraining with a focus on deep learning
2. **Cloudmesh Code Repositories** <https://github.com/cloudmesh> Repository for Cloudmesh code used in DSC Cybertraining
3. **Cloudmesh Documentation** <https://cloudmesh.github.io> Documentation for core technologies used in DSC Cybertraining
4. **Cloudmesh Pi Cloud and Cluster Web Site** <https://cloudmesh.github.io/pi> The Raspberry Pi component of DSC Cybertraining site
5. **Cybertraining by the Digital Science Center** <https://cybertraining-dsc.github.io>
6. The full Cybertraining cyberinfrastructure course material Summer 2021 Repository for REU Projects with FAMU <https://cybertraining-dsc.github.io/docs/report/2021-reu/>

Online Books

1. Cloud Computing
2. Introduction to Linux
3. Python for Cloud Computing
4. Handbook for Cloud Computing and Big Data
5. Customized book with bookmanager

