N Ravi *et al.*, FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy, arXiv:2207.00611

Argonne NATIONAL LABORATORY 75 1946–2021

PRACTICAL FAIR PRINCIPLES FOR AI MODELS

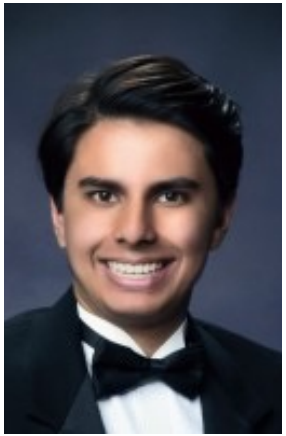
ELIU HUERTA
Lead for Translational AI
Data Science and Learning Division, Argonne National Laboratory
Department of Computer Science, The University of Chicago

Parsl & funcX Fest 2022
14 September 2022

TEAM

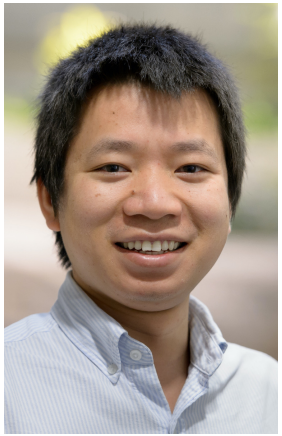
Nikil Ravi

Pranshu Chaturvedi



Eliu Huerta

Zhengchun Liu



Ryan Chard

Ari Scourtas

KJ Schmidt

Kyle Chard

Ben Blaiszik

Ian Foster

WHAT

Define what FAIR means for AI models

WHAT

Define what FAIR means for AI models

WHAT

Define what FAIR means for AI models

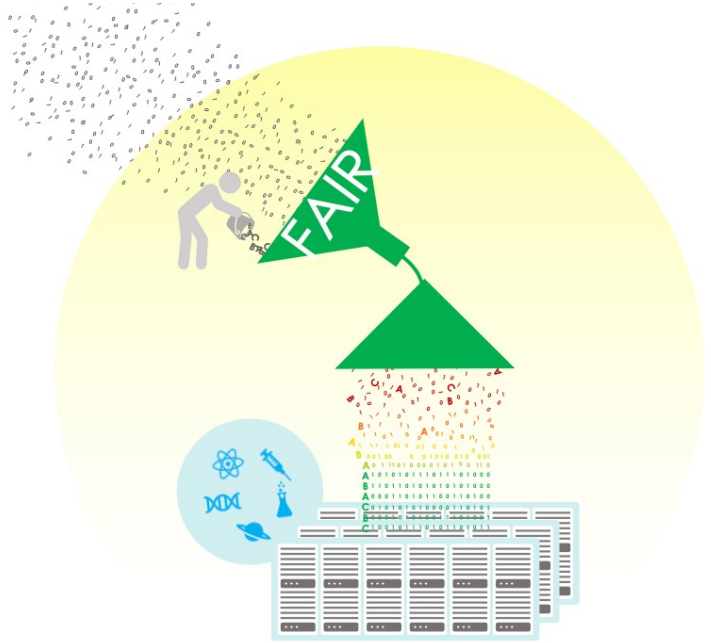
FAIR and AI-ready datasets

Suitable format (HDF5/ROOT/etc.) to leverage modern computing environments

Include Jupyter notebooks to visualize datasets, and explore data type, shape and size

Ready to integrate with APIs for AI research and to enable accelerated training and AI-inference

Goal: automate data management to support and enable discovery and innovation



WHAT

Define what FAIR means for AI models

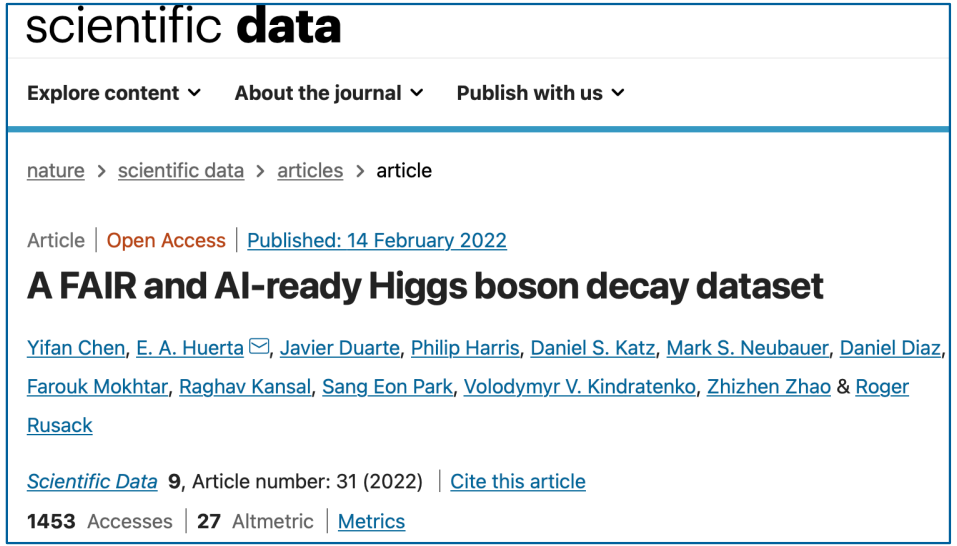
FAIR and AI-ready datasets

Suitable format (HDF5/ROOT/etc.) to leverage modern computing environments

Include Jupyter notebooks to visualize datasets, and explore data type, shape and size

Ready to integrate with APIs for AI research and to enable accelerated training and AI-inference

Goal: automate data management to support and enable discovery and innovation



WHAT

Define what FAIR means for AI models – setting expectations

Linked to FAIR and AI-ready datasets used for model development and testing

Linked to open source repositories that provide scientific software to recreate AI models

Include Jupyter notebooks that explicitly show how to use them, and describe input data type and shape, and output data type and shape

Containerized and ready to use in modern computing environments

Include clear and well known uncertainty quantification metrics

Goal: enable reproducible, accelerated and trustworthy, autonomous discovery

WHAT

Define what FAIR means for AI models – setting expectations

Key goals:

Enable, accelerate and sustain innovation and scientific discovery

North star: autonomous discovery

Data, computing and AI fabric: integrate, consolidate and disrupt

Exemplar: high energy diffraction microscopy

HOW

FAIR and AI-ready Datasets

Big Data Bags,
Minimal Viable Identifiers
Materials Data Facility

Design, Train, Validate & Test AI models with FAIR and AI-ready experimental data from Argonne's **Advanced Photon Source**

PyTorch Model
Baseline Model

NVIDIA TensorRT Model
Accelerated Inference

SambaNova Model
Accelerated Training

AI models trained, validated, tested and **containerized** leveraging **ThetaGPU** and **SambaNova Reconfigurable Dataflow Unit**™ at **Argonne Leadership Computing Facility**

Containerized AI models published in the **Data and Learning Hub for Science** (DLHub)

Published AI models include **DOI**, **tutorial** to invoke them, **sample test set** with **uncertainty quantification metrics**, and **rich and descriptive metadata** for reproducibility purposes

DLHub

Connect published AI models in **DLHub** with **Materials Data Facility** and **ThetaGPU** and quantify reusability, interoperability and reproducibility using **funcX** as a universal distributed computing service

funcX

HOW

MATERIALS DATA FACILITY CHiMaD NIST

BraggNN: Training Dataset

Ravi, Nikil; Liu, Zhengchun; Sharma, Hemant; Chaturvedi, Pranshu; Huerta, E.A.; Scourtas, Aristana; KJ, Schmidt; Chard, Ryan; Blaiszik, Ben

Organizations

MDF Open

DOI

10.18126/iftp-twz1 View on Datacite

Year

2022

Source Name

ravi_braggnn_training

Tags

machine learning microstructures experiment

HOW

MATERIALS DATA FACILITY CHiMaD NIST

BraggNN: Training Dataset

Ravi, Nikil; Liu, Zhengchun; Sharma, Hemant; Chaturvedi, Pranshu; Huerta, E.A.; Scourtas, Aristana; KJ, Schmidt; Chard, Ryan; Blaiszik, Ben

Organizations

MDF Open

DOI

10.18126/iftp-twz1 View on Datacite

Year

2022

Source Name

ravi_braggnn_training

Tags

machine learning microstructures experiment

SambaNova BraggNN

kj.schmidt913_gmail/BraggNN_SN3 Copy

Zhengchun Liu; Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Python static method

Input

Image map
Type: ndarray
Shape: ['11', '11']

Output

list of Bragg peak positions
Type: ndarray
Shape: ['1', '2']

Run with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
X = get_my_data() #replace this
dl = DLHubClient()
dl.run('kj.schmidt913_gmail/BraggNN_SN3', X)
```

Get More Info with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
dl = DLHubClient()
dl.describe_servable('kj.schmidt913_gmail/BraggNN_SN3')
```

DLHub SDK Installation

```
pip install dlhub_sdk
```

DLHub SDK documentation

HOW

MATERIALS DATA FACILITY CHiMaD NIST

BraggNN: Training Dataset

Ravi, Nikil; Liu, Zhengchun; Sharma, Hemant; Chaturvedi, Pranshu; Huerta, E.A.; Scourtas, Aristana; KJ, Schmidt; Chard, Ryan; Blaiszik, Ben

Organizations

MDF Open

DOI

10.18126/iftp-twz1 View on Datacite

Year

2022

Source Name

ravi_braggnn_training

Tags

machine learning microstructures experiment

SambaNova BraggNN

kj.schmidt913_gmail/BraggNN_SN3 Copy

Zhengchun Liu; Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Python static method

Input

Image map
Type: ndarray
Shape: ['11', '11']

Output

list of Bragg peak positions
Type: ndarray
Shape: ['1', '2']

Run with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
X = get_my_data() #replace this
dl = DLHubClient()
dl.run('kj.schmidt913_gmail/BraggNN_SN3', X)
```

Get More Info with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
dl = DLHubClient()
dl.describe_servable('kj.schmidt913_gmail/BraggNN_SN3')
```

DLHub SDK Installation

```
pip install dlhub_sdk
```

DLHub SDK documentation

HOW

BraggNN: Training Dataset

Ravi, Nikil; Liu, Zhengchun; Sharma, Hemant; Chaturvedi, Pranshu; Huerta, E.A.; Scourtas, Aristana; KJ, Schmidt; Chard, Ryan; Blaiszik, Ben

Organizations

MDF Open

DOI

10.18126/iftp-twz1 View on Datacite

Year

2022

Source Name

ravi_braggnn_training

Tags

machine learning | microstructures | experiment

SambaNova BraggNN

kj.schmidt913_gmail/BraggNN_SN3 Copy

Zhengchun Liu; Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Python static method

Input

Image map
Type: ndarray
Shape: ['11', '11']

Output

list of Bragg peak positions
Type: ndarray
Shape: ['1', '2']

Run with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
X = get_my_data() #replace this
dl = DLHubClient()
dl.run('kj.schmidt913_gmail/BraggNN_SN3', X)
```

Get More Info with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
dl = DLHubClient()
dl.describe_servable('kj.schmidt913_gmail/BraggNN_SN3')
```

DLHub SDK Installation

```
pip install dlhub_sdk
```

DLHub SDK documentation

HOW

globus

MATERIALS DATA FACILITY CHiMaD NIST

BraggNN: Training Dataset

Ravi, Nikil; Liu, Zhengchun; Sharma, Hemant; Chaturvedi, Pranshu; Huerta, E.A.; Scourtas, Aristana; KJ, Schmidt; Chard, Ryan; Blaiszik, Ben

Organizations

MDF Open

DOI

10.18126/iftp-twz1 View on Datacite

Year

2022

Source Name

ravi_braggnn_training

Tags

machine learning | microstructures | experiment

SambaNova BraggNN

kj.schmidt913_gmail/BraggNN_SN3 Copy

Zhengchun Liu; Nikil Ravi; Pranshu Chaturvedi; E.A. Huerta; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Python static method

Input

Image map
Type: ndarray
Shape: ['11', '11']

Output

list of Bragg peak positions
Type: ndarray
Shape: ['1', '2']

Run with DLHub SDK

```
from dlhub_sdk.client import DLHubClient
X = get_my_data() #replace this
dl = DLHubClient()
dl.run('kj.schmidt913_gmail/BraggNN_SN3', X)
```

Get More Info with DLHub SDK

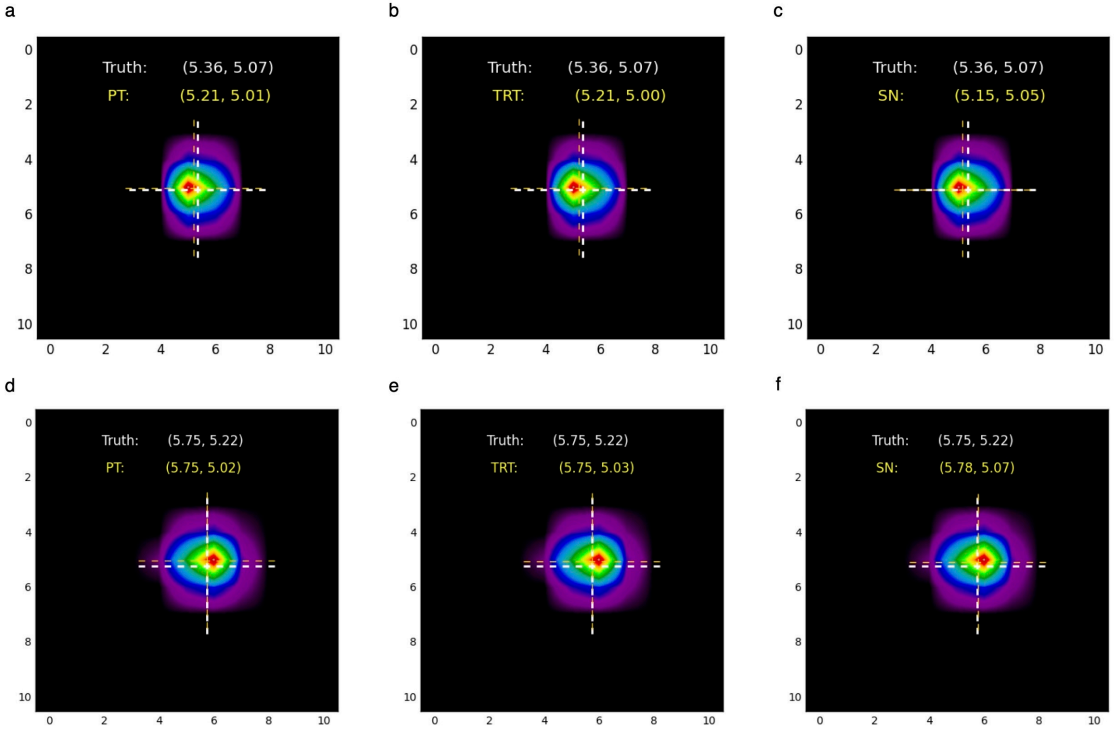
```
from dlhub_sdk.client import DLHubClient
dl = DLHubClient()
dl.describe_servable('kj.schmidt913_gmail/BraggNN_SN3')
```

DLHub SDK Installation

```
pip install dlhub_sdk
```

DLHub SDK documentation

END-USER HAPPINESS



Bragg peak analysis ran natively in ThetaGPU

Bragg peak analysis ran by combining MDF, DLHub, ThetaGPU, funcX, and Globus

END-USER HAPPINESS

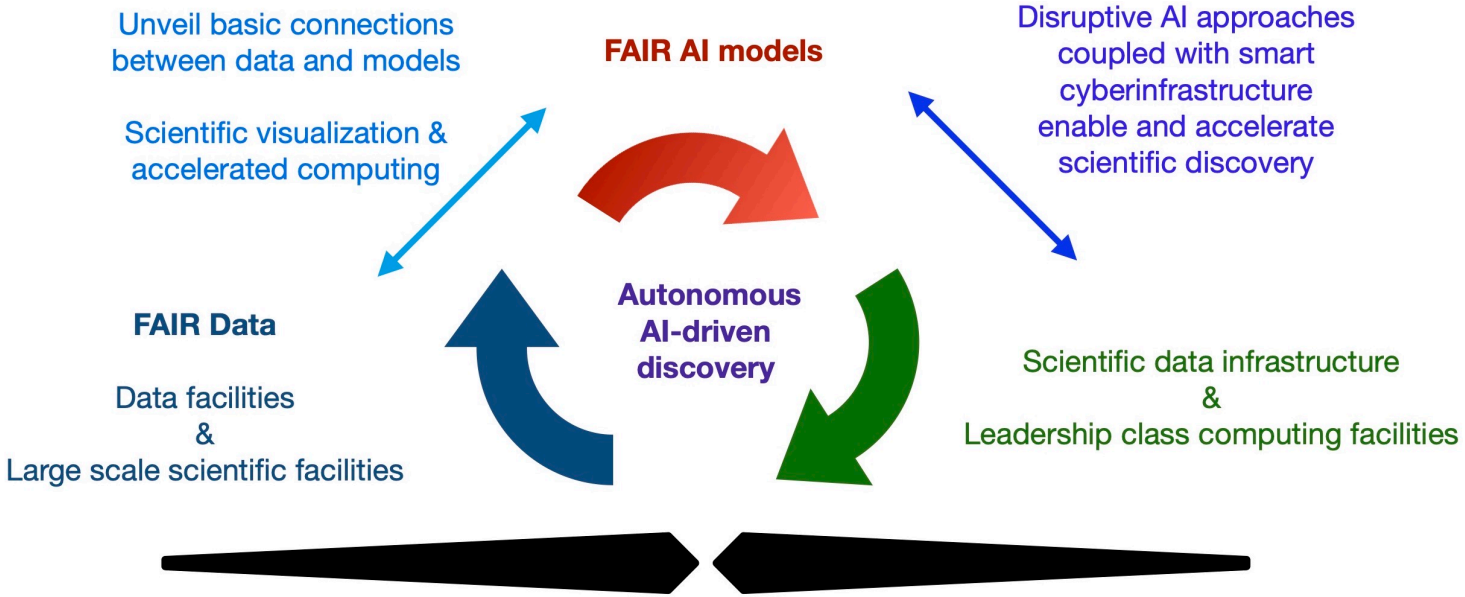
End user experience similar to running AI models directly from personal laptop

Identify high performing AI models

Accelerate state-of-the-art to state-of-practice cycle

Seamless combination of scientific data infrastructure and leadership class supercomputing

VISION



AI learns to describe natural phenomena bridging the gap between approximate models & simulations and experimental data

Workflows connect disparate data and computing resources to enable autonomous scientific discovery

ACKNOWLEDGEMENTS

This work was supported by the FAIR Data program and the Braid project of the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, under contract number DE-AC02-06CH11357. It used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

This work was performed under financial assistance award 70NANB14H012 from U.S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Material Design (CHiMaD), and by the National Science Foundation under award 1931306 “Collaborative Research: Framework: Machine Learning Materials Innovation Infrastructure.”

Development of DLHub has been supported by LDRD funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357. This work was also supported by the National Science Foundation under NSF Award Number: 2209892 ”Frameworks: Garden: A FAIR Framework for Publishing and Applying AI Models for Translational Research in Science, Engineering, Education, and Industry’’

We also thank SambaNova Systems, Inc., for engineering support to make our BraggNN AI models work efficiently on their system at the ALCF AI-Testbed.

We thank NVIDIA for their continued support.


Argonne
NATIONAL LABORATORY
75
1946–2021