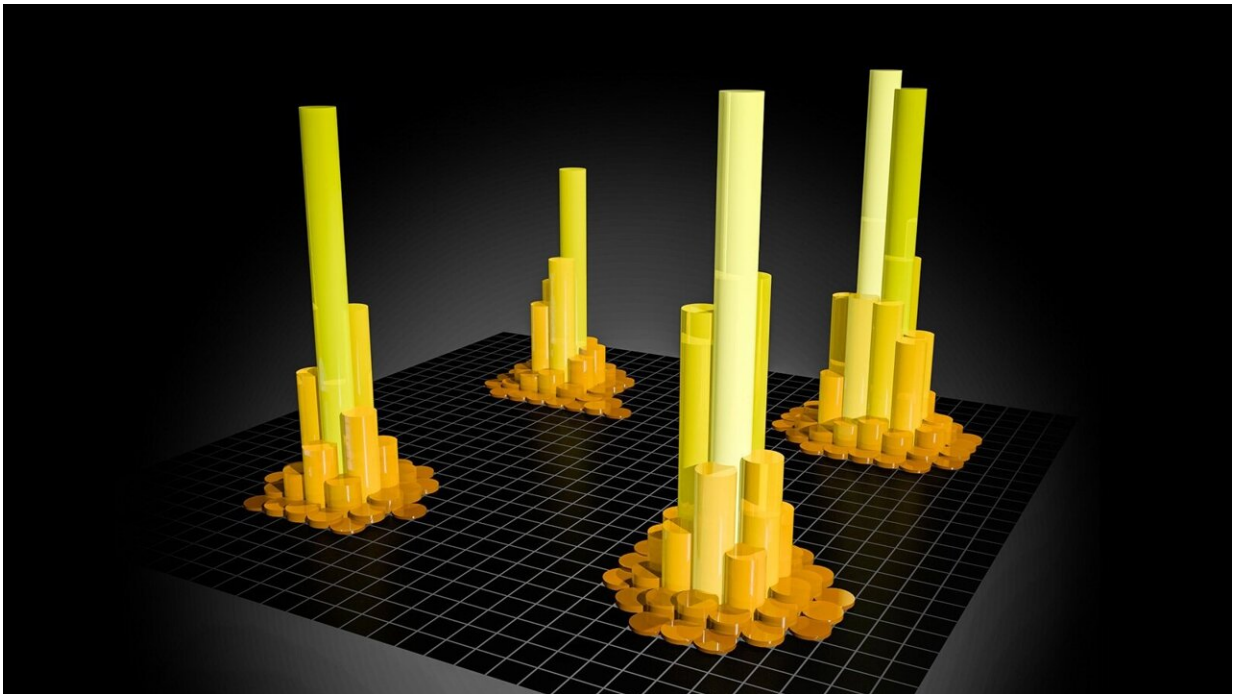


# Scientists articulate new data standards for AI models

November 10 2022, by Jared Sagoff

---



Scientific visualization of Bragg diffraction peaks in a 15x15 pixel patch of an undeformed bi-crystal gold sample. The height denotes photon counts. This data was acquired at the Advanced Photon Source and processed at the ThetaGPU supercomputer. Credit: Argonne National Laboratory/Eliu Huerta.

Aspiring bakers are frequently called upon to adapt award-winning recipes based on differing kitchen setups. Someone might use an eggbeater instead of a stand mixer to make prize-winning chocolate chip

cookies, for instance.

Being able to reproduce a recipe in different situations and with varying setups is critical for both talented chefs and [computational scientists](#), the latter of whom are faced with a similar problem of adapting and reproducing their own "recipes" when trying to validate and work with new AI models. These models have applications in [scientific fields](#) ranging from climate analysis to brain research.

"When we talk about data, we have a practical understanding of the digital assets we deal with," said Eliu Huerta, scientist and lead for Translational AI at the U.S. Department of Energy's (DOE) Argonne National Laboratory. "With an AI model, it's a little less clear; are we talking about data structured in a smart way, or is it computing, or software, or a mix?"

In a new study, Huerta and his colleagues have articulated a new set of standards for managing AI models. Adapted from recent research on automated [data management](#), these standards are called FAIR, which stands for findable, accessible, interoperable and reusable.

"By making AI models FAIR, we no longer have to build each system from the ground up each time," said Argonne computational scientist Ben Blaiszik. "It becomes easier to reuse concepts from different groups, helping to create cross-pollination across teams."

According to Huerta, the fact that many AI models are currently not FAIR poses a challenge to [scientific discovery](#). "For many studies that have been done to date, it is difficult to gain access to and reproduce the AI models that are referenced in the literature," he said. "By creating and sharing FAIR AI models, we can reduce the amount of duplication of effort and share best practices for how to use these models to enable great science."

To meet the needs of a diverse community of users, Huerta and his colleagues combined a unique suite of data management and high performance computing platforms to establish a FAIR protocol and quantify the "FAIR-ness" of AI models. The researchers paired FAIR data published at an online repository called the Materials Data Facility, with FAIR AI models published at another online repository called the Data and Learning Hub for Science, as well as with AI and supercomputing resources at the Argonne Leadership Computing Facility (ALCF).

In this way, the researchers were able to create a computational framework that could help bridge various hardware and software, creating AI models that could be run similarly across platforms and that would yield reproducible results. The ALCF is a DOE Office of Science user facility.

Two keys to creating this framework are platforms called funcX and Globus, which allow researchers to access high performance computing resources straight from their laptops. "FuncX and Globus can help transcend the differences in hardware architectures," said co-author Ian Foster, director of Argonne's Data Science and Learning division. "If someone is using one computing architecture and someone else is using another, we now have a way of speaking a common AI language. It's a big part of making AI more interoperable."

In the study, the researchers used an example dataset of an AI model that used diffraction data from Argonne's Advanced Photon Source, also a DOE Office of Science user facility. To perform the computations, the team used the ALCF AI Testbed's SambaNova system and the Theta supercomputer's NVIDIA GPUs (graphics processing units).

"We're excited to see the FAIR productivity benefits from model and data sharing to provide more researchers with access to high

performance computing resources," said Marc Hamilton, NVIDIA vice president for Solutions Architecture and Engineering. "Together we're supporting the expanding universe of [high performance](#) computing that's combining experimental data and instrument operation at the edge with AI to increase the pace of scientific discovery."

"SambaNova is excited to partner with researchers at Argonne National Laboratory to pursue innovation at the interface of AI and emergent hardware architectures," added Jennifer Glore, vice president for Customer Engineering at SambaNova Systems. "AI will have a significant role in the future of scientific computing, and the development of FAIR principles for AI models along with novel tools will empower researchers to enable autonomous discovery at scale. We're looking forward to continued collaboration and development at the ALCF AI Testbed."

A paper based on the study, "FAIR principles for AI models, with a practical application for accelerated high energy diffraction microscopy," appeared in *Scientific Data* on Nov. 10, 2022.

**More information:** Nikil Ravi et al, FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy, *Scientific Data* (2022). [DOI: 10.1038/s41597-022-01712-9](https://doi.org/10.1038/s41597-022-01712-9)

Provided by Argonne National Laboratory

Citation: Scientists articulate new data standards for AI models (2022, November 10) retrieved 19 April 2024 from <https://techxplore.com/news/2022-11-scientists-articulate-standards-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is

provided for information purposes only.