Karlsruhe Institute of Technology

Institute for Applied Materials – Microstructure Modelling and Simulation (IAM-MMS)

KadiAI and CIDS: Data-integrated Artificial Intelligence

A. Koeppe[1,2], M. Selzer[2,3], Y. Zhao[1,2], G. Tosato[1,2], J. Grolig[1,2], L. Griem[1,2], D. Rajagopal[1,2], A. Cierpka[1,2], and M. Kocak[1,2]

[1]Institute for Applied Materials – Microstructure Modelling and Simulation (IAM-MMS), Karlsruhe Institute of Technology (KIT), Straße am Forum 7, 76131 Karlsruhe, Germany
[2]Institute of Nanotechnology – MicroStructure Simulation (INT-MSS), Karlsruhe Institute of Technology (KIT), Straße am Forum 7, 76131 Karlsruhe, Germany
[3]Institute for Digital Materials Science (IDM), Hochschule Karlsruhe - University of Applied Sciences (HKA), Moltkestrasse 30, 76133 Karlsruhe, Germany

Introduction

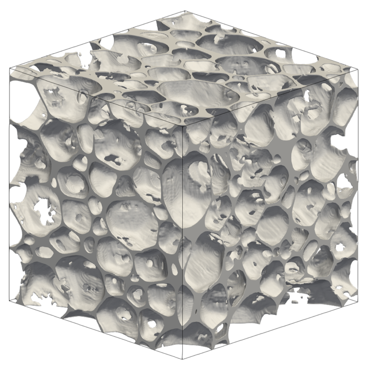
Enabling data-driven modeling

All research generates data, extracts knowledge, and develops models within scientific workflows. Manual knowledge extraction and execution are often implicitly used for conventional static datasets but becomes unfeasible for vast, dynamically changing datasets. However, the data-driven modeling paradigm necessitates efficient interfaces between data and models through research data management.

Aims for Artificial Intelligence (AI) in materials science

Effective research data management, efficient interfaces and AI enable to
- Automatically extract knowledge from data (unsupervised learning),
- Train generic models to predict directly from data (supervised learning), and
- Control scientific workflows based on data (active and reinforcement learning).

Materials research
Domain problem example



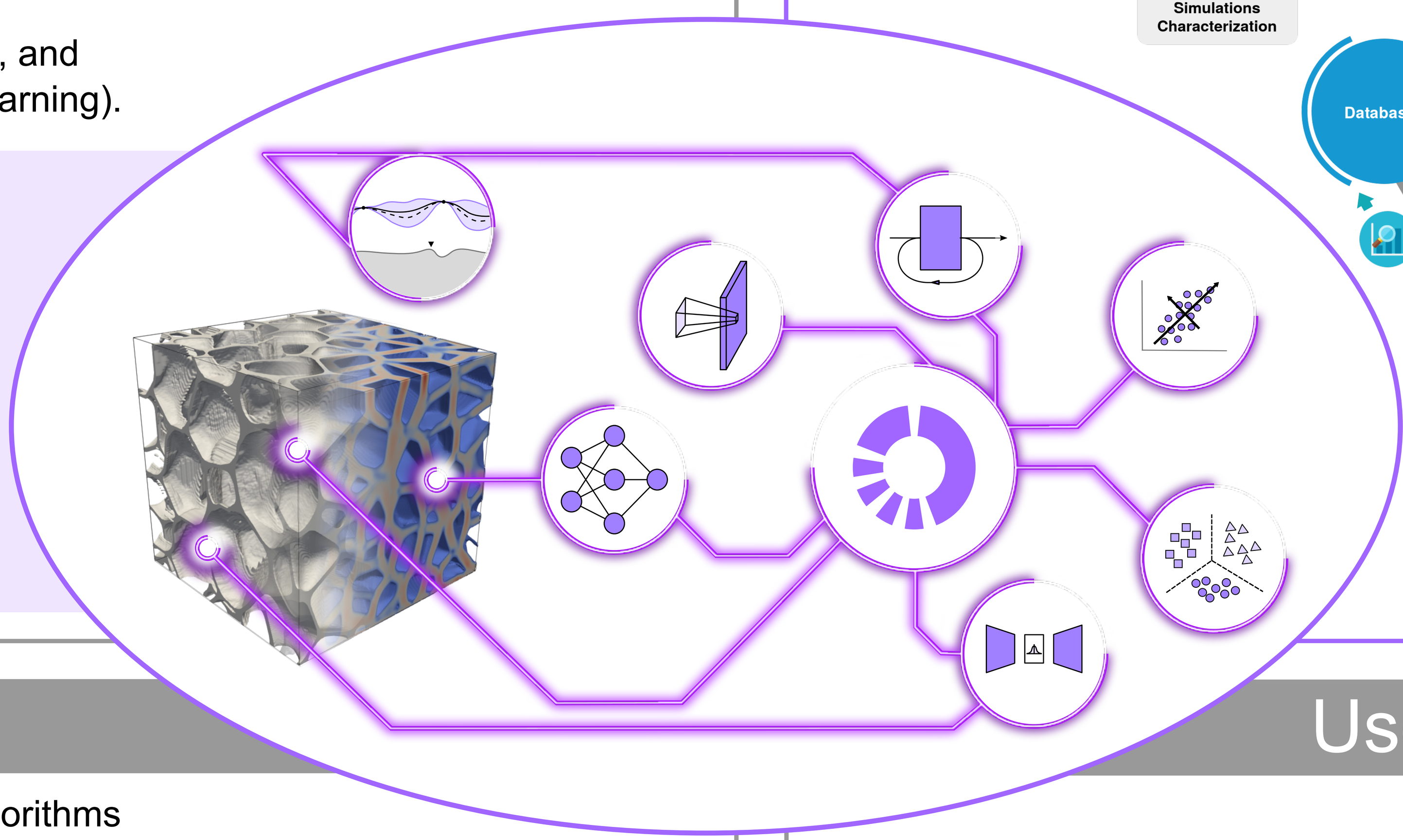
The Kadi Ecosystem for integrated AI

Kadi4Mat
Open-source platform for FAIR research data management [1]

KadiStudio
Electronic ab notebook and scientific workflow engines [2]

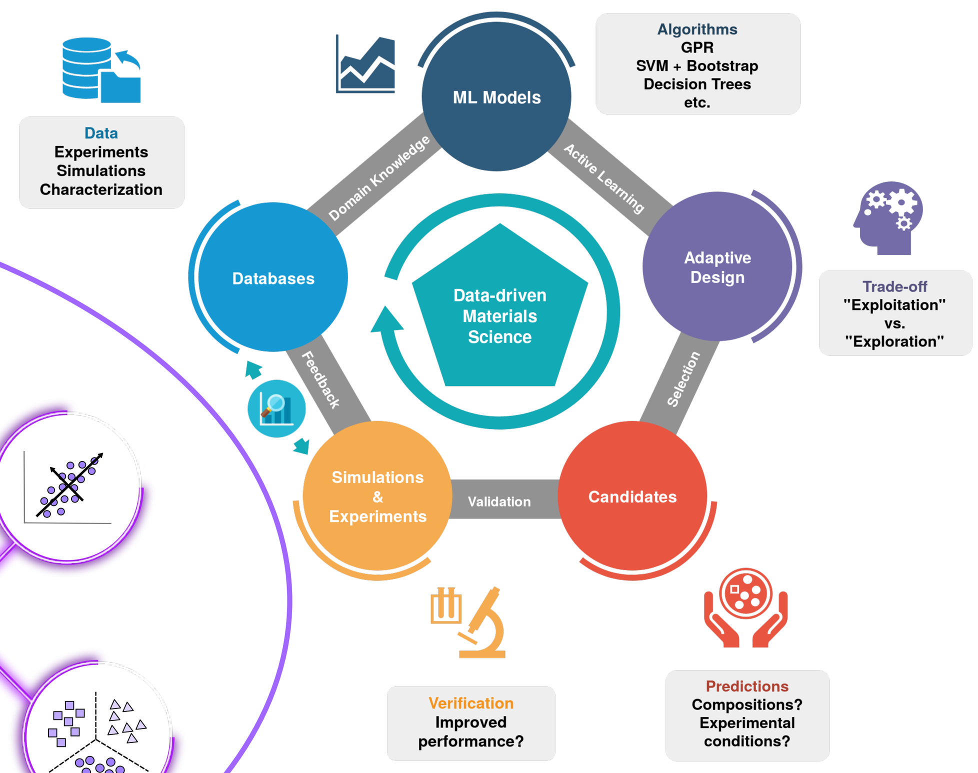
KadiAI
Interface between Kadi4Mat and machine learning tools [3]

cids tools
Computational Intelligence and Data Science framework [3]



Data-integrated AI for Materials

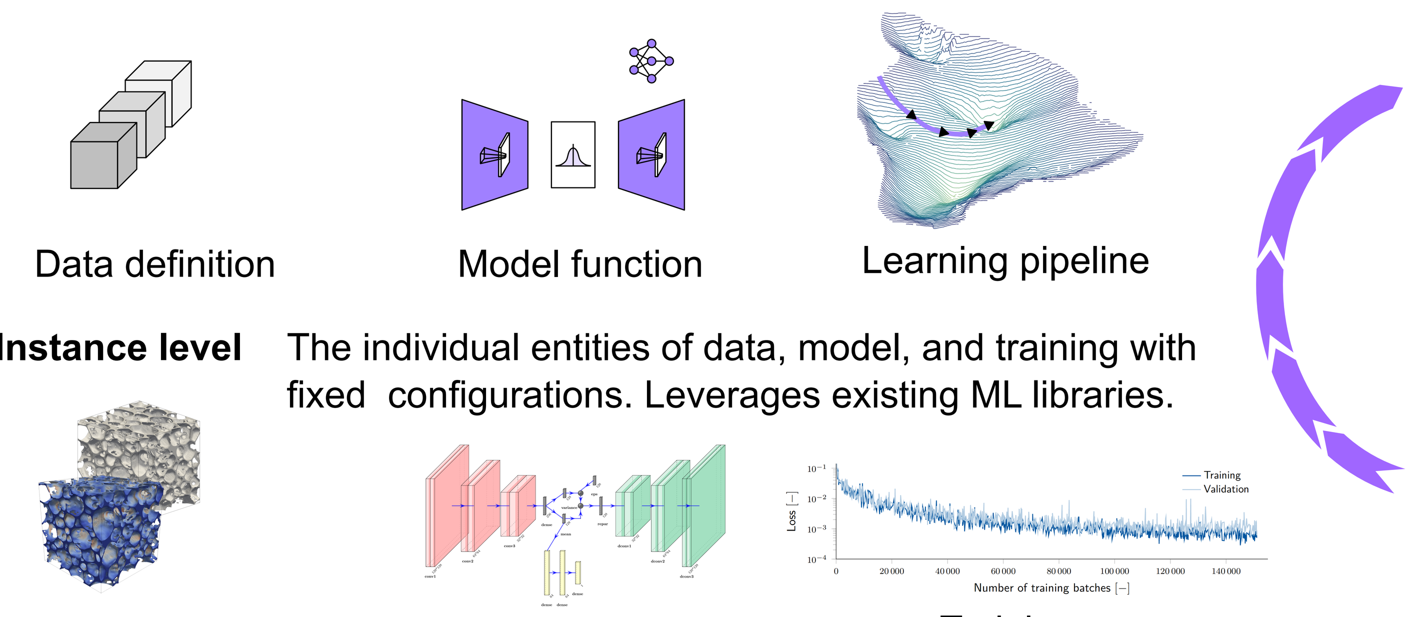
Efficient **interfaces between data and AI models** are necessary to enable data-driven modeling. Data-integrated AI models directly connect with the research data management solution to **extract knowledge, learn, and control scientific workflows** based on data.



AI Models and Learning

Concept level The meta information for data, models, and learning algorithms defines the boundaries in which an AI solution is investigated.

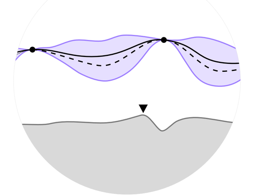
Instance level The individual entities of data, model, and training with fixed configurations. Leverages existing ML libraries.



cids tools

Tuning interface

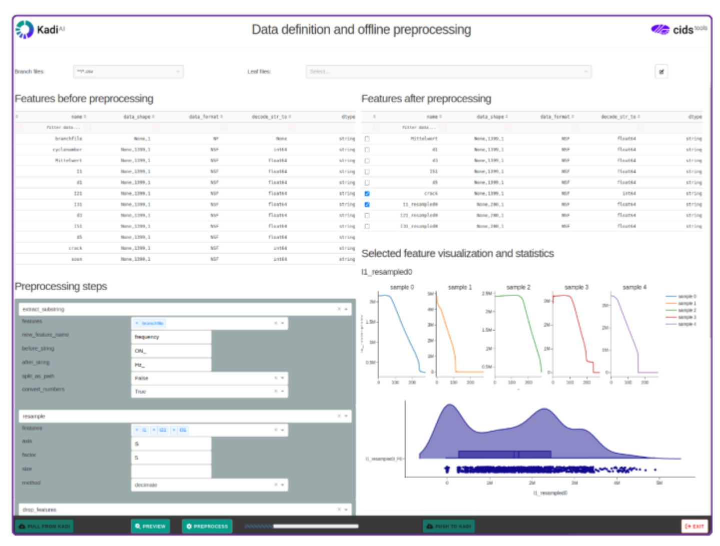
Search algorithms and active learning models explore and investigate the boundaries for the best AI configuration



User Interaction

Data and model definition through interactive dashboards

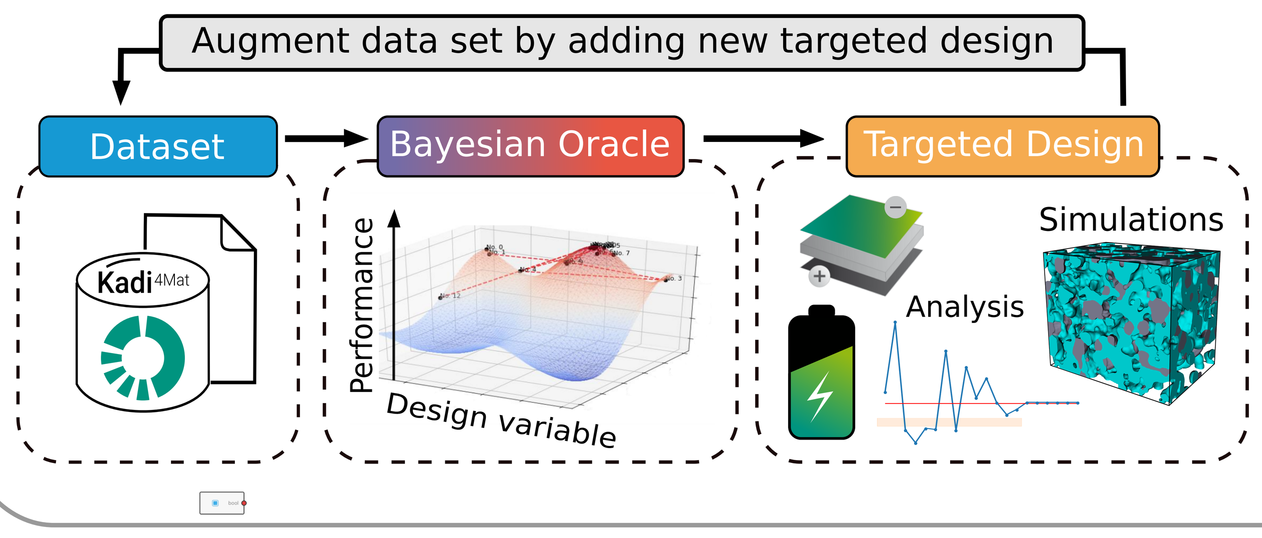
Interactive dashboards help domain experts aggregating, preprocessing, and visualizing their data from multiple files. Furthermore, the dashboards help the users select suitable AI modes and learning algorithms. Finally, the dashboards monitor and visualize the learning process.



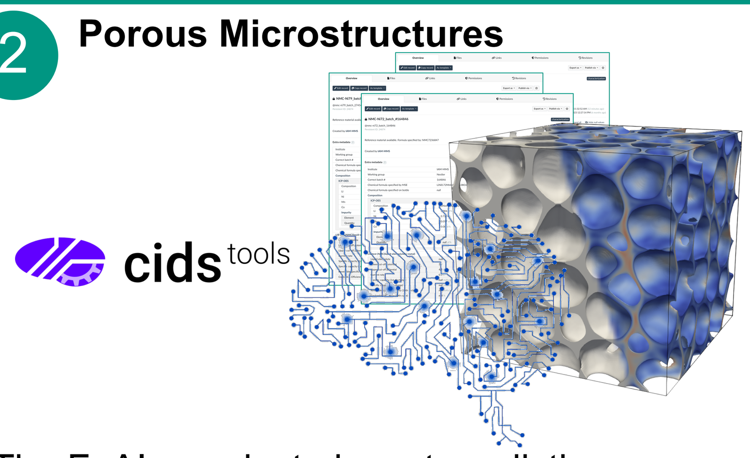
Use cases

1 Solid Oxide Fuel Cells

An active-learning framework uses Bayesian Optimization to automate the search of optimal parameters for phase-field simulation studies of aging Ni-GDC anodes in Solid Oxide Fuel Cells.
It provides an efficient exploration of complex, high-dimensional parameters space to create highly informative datasets.



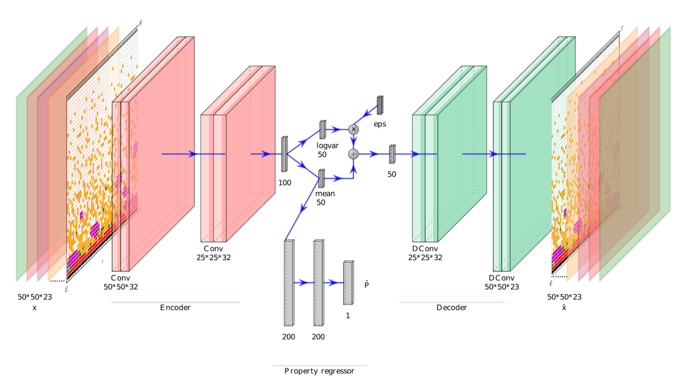
2 Porous Microstructures



The FoAIm project aims at predicting mechanical properties and designing foams. The complete ML pipeline from data conversion to model evaluation is implemented with CIDS. It combines experimental and machine-learning methods by linking foam microstructures to their mechanical properties and generating microstructures for given properties.

3 Characterization of Solid Electrolyte Interphase using Deep Generative Model

A deep generative model characterizes the solid electrolyte interphase in batteries. Using interactive input features and output feature definition, the model's architecture is controlled by specified hyperparameter ranges. The latent space is organized by an additional regressor based on physical properties. This method streamlines the process, complementing traditional characterization techniques.



Acknowledgement and references.

[1] N. Brandt et al., 'Kadi4Mat: A Research Data Infrastructure for Materials Science', Data Science Journal, vol. 20, no. 1, Art. no. 1, Feb. 2021, doi: 10.5334/dsj-2021-008.
[2] L. Griem et al., 'KadiStudio: FAIR Modelling of Scientific Research Processes', Data Science Journal, vol. 21, no. 1, Art. no. 1, Sep. 2022, doi: 10.5334/dsj-2022-016.
[3] A. Koeppe and The CIDS Team, 'cids: 3.1'. Zenodo, Jan. 11, 2023. doi: 10.5281/zenodo.7524476.

A special thanks to Jana Holland-Cunz for the design consultation.

This work is funded by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK-BW), in the project MoMaF--Science Data Center (grant number: 34-7547.222).
This work is funded by the BMBF and MWK-BW as part of the Excellence Strategy of the German Federal and State Governments in the project Kadi4X.
This work was [partly] carried out with the support of the Karlsruhe Nano Micro Facility (KNMFi, www.knmf.kit.edu), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT, www.kit.edu).

arnd.koeppe@kit.edu
www.iam.kit.edu/mms