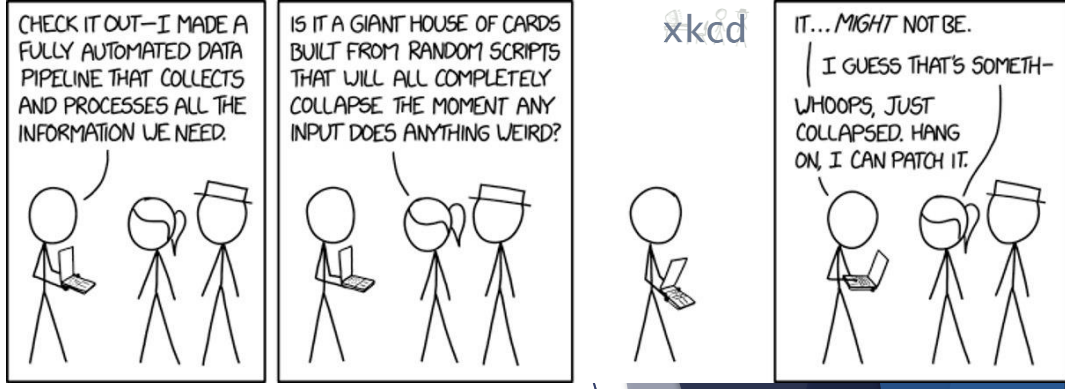
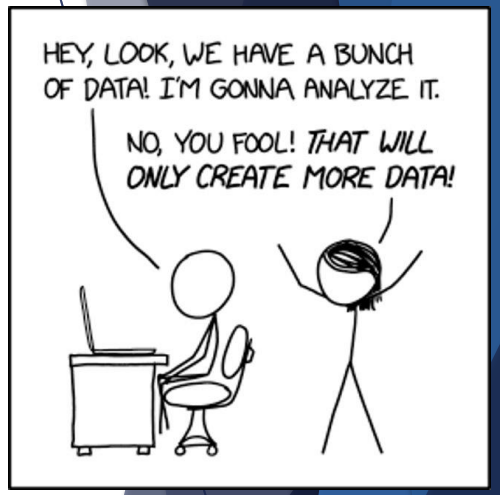
Science and Technology Facilities Council | Scientific Computing



Data to knowledge[to data]

Alin M Elena
Elliott Kasoar, Federica Zanca



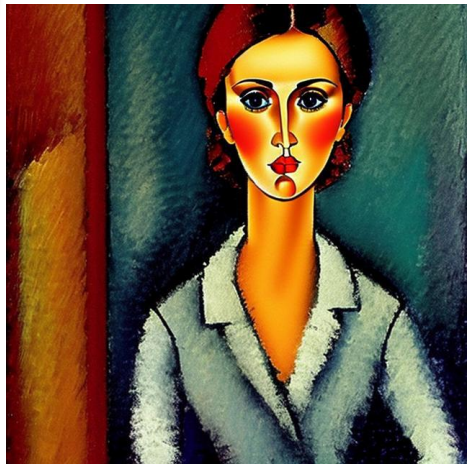
HEC MCC UK

CCP5 Collaborative Computational Project Computer Simulation of Condensed Phases

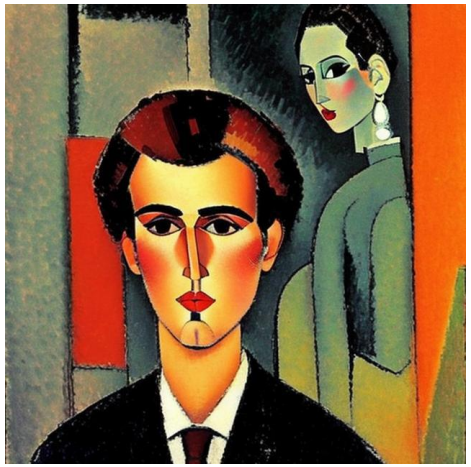
PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

https://www.psdi.ac.uk/

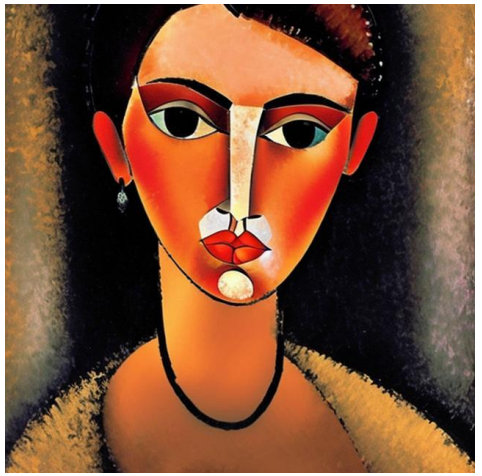
People



Alin Elena



Elliott Kasoar

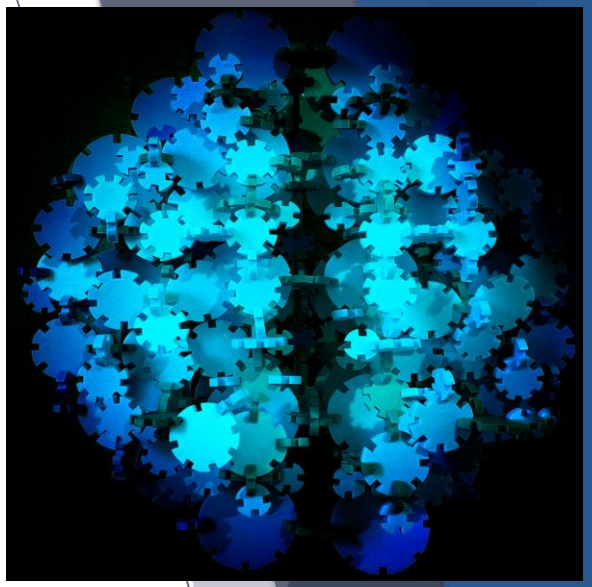


Federica Zanca

Openart.ai

SCD organogram

PSDI



Facilities, Institutes & Hubs

Examples:

- Catalysis Hub
- CCFE
- Central Laser Facility
- Diamond
- Future Manufacturing Hub
- ISIS
- Royce Institute
- ATI

National Research Facilities

Examples:

- HarwellXPS
- NXCT
- NCS
- PSDS
- SuperSTEM
- UK High Field Solid-State NMR
- XMaS

Computational Initiatives

Examples:

- CCP5++
- CCP9
- CCP Biosim
- CCP EngSci
- CCPi
- CCP NC
- CCP NTH
- CCP QC
- CCP SAS
- CCP Turbulence
- CCP WSI
- SSI
- UK society of RSE
- HEC Biosim
- HEC Plasma
- MCC
- UKCTRF
- UKCP
- UKTC
- CoSeC
- EPSRC Tier2
- ExCALIBUR
- STFC Hartree Centre
- ARCHER

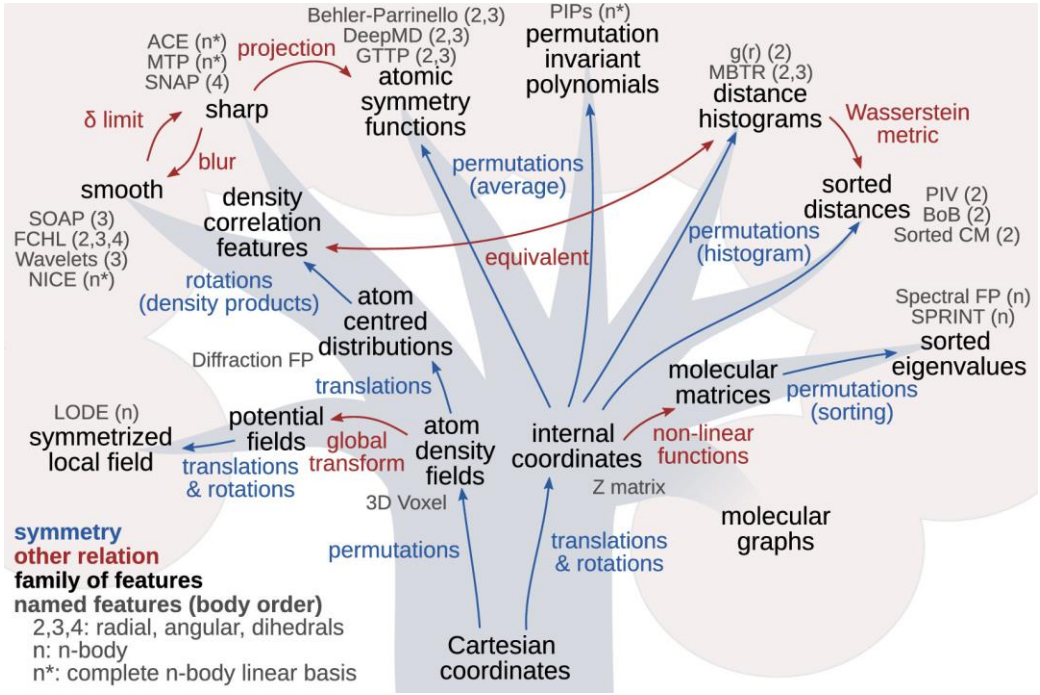
Research Institutions, Groups and Laboratories

Examples:

- Equipment Infrastructures
- Equipment Facilities
- University Labs
- ELNs
- Repositories
- Local Computing Resources

PHYSICAL SCIENCES DATA INFRASTRUCTURE

Aim

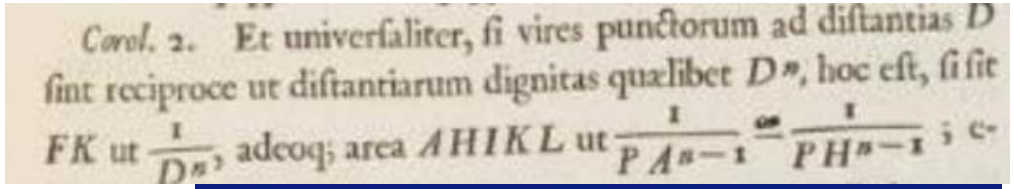


https://doi.org/10.1021/acs.chemrev.1c00021

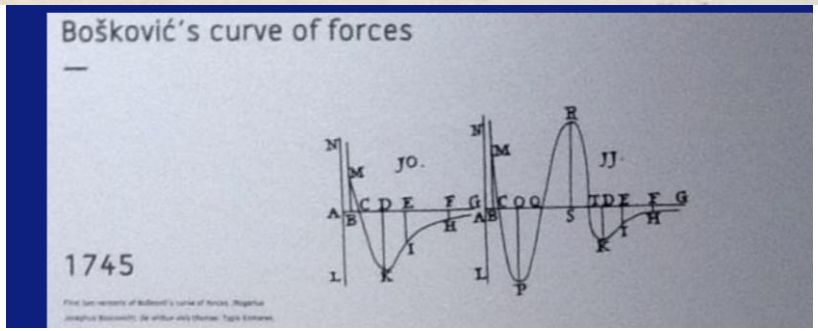
- **design and deploy the hardware infrastructure** to host both training data for machine learnt interatomic potentials – MLIP and potentials themselves
- **Features**: *Curate, store, distribute and interrogate* MLIPs and training data
- **Visualisation and training workflows**
- **Benchmarking and validation**
- **Ideally, HPC integration**

Science&history interlude

Newton (Pricipia... 1687 - 1/r^n)


Corol. 2. Et univerſaliter, ſi vires punctorum ad diſtantias D ſint reciproce ut diſtantiarum dignitas quælibet D^n, hoc eſt, ſi ſit FK ut $\frac{1}{D^n}$, adeoq; area $AHIKL$ ut $\frac{1}{PA^{n-1}} - \frac{1}{PH^{n-1}}$; &c.

Roger Boscovich (De Virbus vivis, 1745)



Gustav Mie (1903)

$$\Phi_{12}(r) = \left(\frac{n}{n-m}\right)\left(\frac{n}{m}\right)^{m/(n-m)} \epsilon\left[\left(\frac{\sigma}{r}\right)^n - \left(\frac{\sigma}{r}\right)^m\right]$$

Edgar Grüneisen(1912)

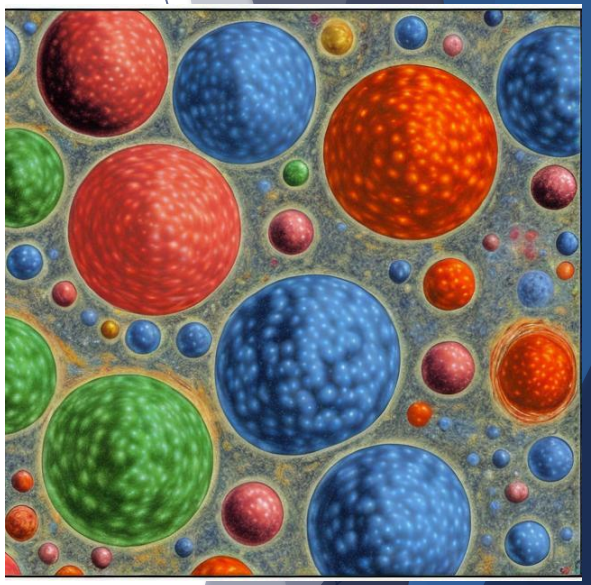
$$-\frac{\mathfrak{a}}{r^x} + \frac{\mathfrak{b}}{r^y}$$

Lennard-Jones (1924-1931)

$$f(r) = \frac{\lambda_n}{r^n} - \frac{\lambda_m}{r^m},$$

Fowler&Bernal (1933) n=12

$$V = -F(r) - C/r^6 + B/r^n,$$



199x-201x

Source: openart.ai prompt purgatory by hieronymus bosch

MACE

$$h_{i,k00}^{(0)} = \sum_z W_{kz} \delta_{zz_i}, \tag{1}$$

$$\bar{h}_{i,kl_2m_2}^{(s)} = \sum_{\tilde{k}} W_{k\tilde{k}l_2}^{(s)} h_{i,\tilde{k}l_2m_2}^{(s)}, \tag{2}$$

$$j_0^n\left(r_{\ddot{ij}}\right) = \sqrt{\frac{2}{r_{\text{cut}}}} \frac{\sin\left(n\pi \frac{r_{\ddot{ij}}}{r_{\text{cut}}}\right)}{r_{\ddot{ij}}} f_{\text{cut}}\left(r_{ij}\right), \tag{3}$$

$$R_{k\eta_1 l_1 l_2 l_3}^{(s)}\left(r_{\ddot{ij}}\right) = \text{MLP}\left(\left\{j_0^n\left(r_{\ddot{ij}}\right)\right\}_n\right), \tag{4}$$

$$\begin{aligned}\phi_{\ddot{ij},k\eta_1 l_3 m_3}^{(s)} &= \sum_{l_1 l_2 m_1 m_2} C_{\eta_1,l_1 m_1 l_2 m_2}^{l_3 m_3} R_{k\eta_1 l_1 l_2 l_3}^{(s)}\left(r_{\ddot{ij}}\right) \\ &\times Y_{l_1}^{m_1}\left(\hat{\boldsymbol{r}}_{\ddot{ij}}\right) \bar{h}_{j,kl_2m_2}^{(s)},\end{aligned} \tag{5}$$

$$A_{i,kl_3m_3}^{(s)} = \sum_{\tilde{k},\eta_1} W_{k\tilde{k}\eta_1 l_3}^{(s)} \sum_{j \in \mathcal{N}(i)} \phi_{\ddot{ij},\tilde{k}\eta_1 l_3 m_3}^{(s)}, \tag{6}$$

$$\boldsymbol{A}_{i,k\boldsymbol{lm}}^{(s),\nu} = \prod_{\xi=1}^{\nu} A_{i,kl_\xi m_\xi}^{(s)}, \tag{7}$$

$$\boldsymbol{B}_{i,\eta_\nu kLM}^{(s),\nu} = \sum_{\boldsymbol{lm}} \mathscr{C}_{\eta_\nu \boldsymbol{lm}}^{LM} \boldsymbol{A}_{i,k\boldsymbol{lm}}^{(s),\nu}, \tag{8}$$

$$m_{i,kLM}^{(s)} = \sum_\nu \sum_{\eta_\nu} W_{z_i \eta_\nu kL}^{(s),\nu} \boldsymbol{B}_{i,\eta_\nu kLM}^{(s),\nu}, \tag{9}$$

$$h_{i,kLM}^{(s+1)} = \sum_{\tilde{k}} W_{kL,\tilde{k}}^{(s)} m_{i,\tilde{k}LM}^{(s)} + \sum_{\tilde{k}} W_{kz_i L,\tilde{k}}^{(s)} h_{i,\tilde{k}LM}^{(s)}. \tag{10}$$

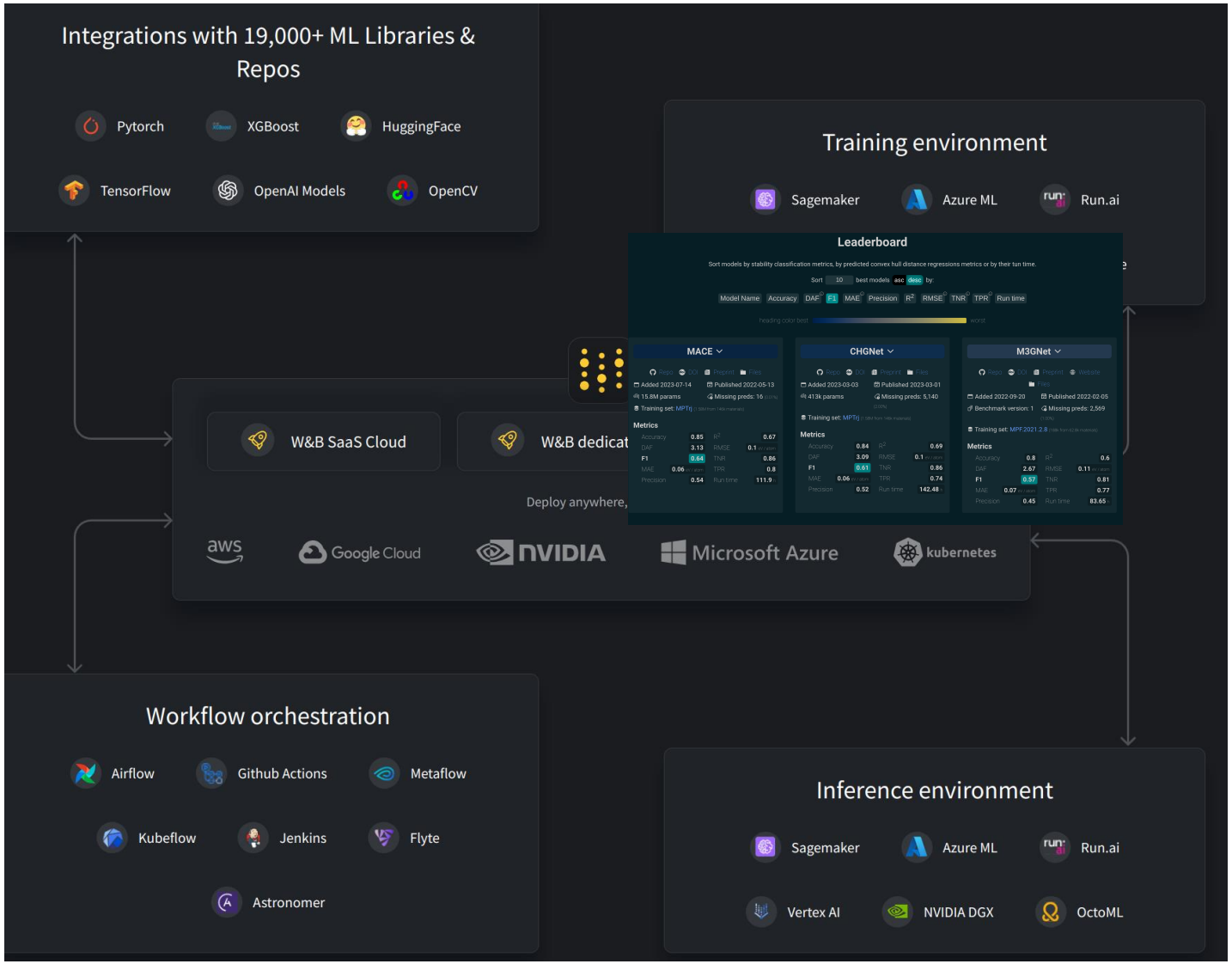
$$E_i = \sum_{s=1}^{S} E_i^{(s)} = \sum_{s=1}^{S} \mathscr{R}^{(s)}\left(\boldsymbol{h}_i^{(s)}\right),$$

$$\mathscr{R}^{(s)}\left(\boldsymbol{h}_i^{(s)}\right) = \begin{cases} \sum_k W_k^{(s)} h_{i,k00}^{(s)} & \text{if } s < S, \\ \text{MLP}\left(\left\{h_{i,k00}^{(s)}\right\}_k\right) & \text{if } s = S. \end{cases}$$

$$\boldsymbol{F} = -\nabla \sum_i E_i.$$

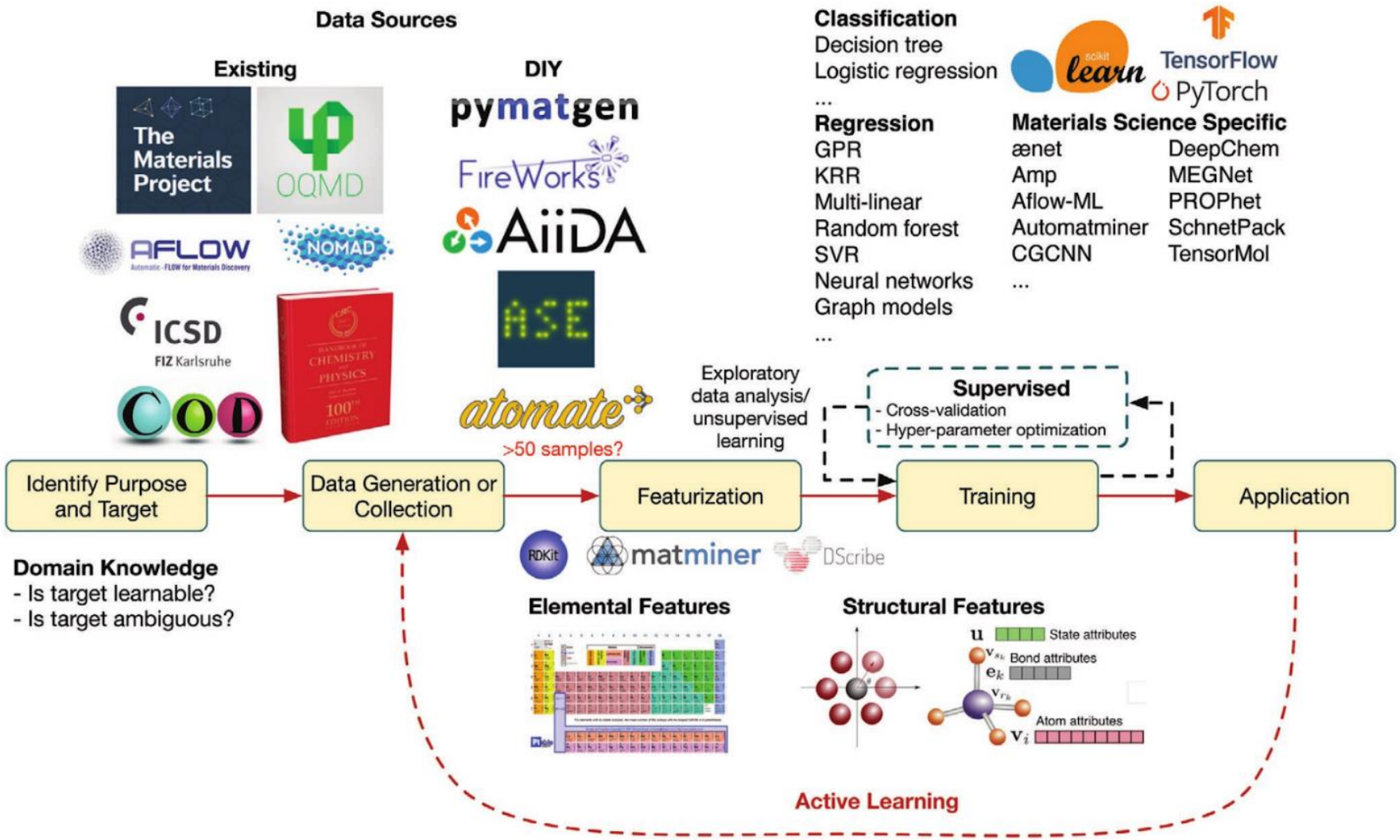
Ilyes Batatia *et al.* - Advances in Neural Information Processing Systems, 2022

Workflows – AI community view

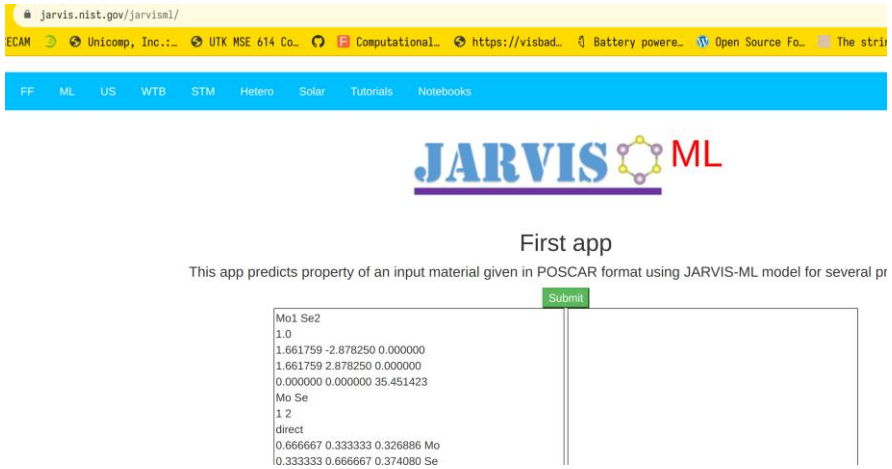


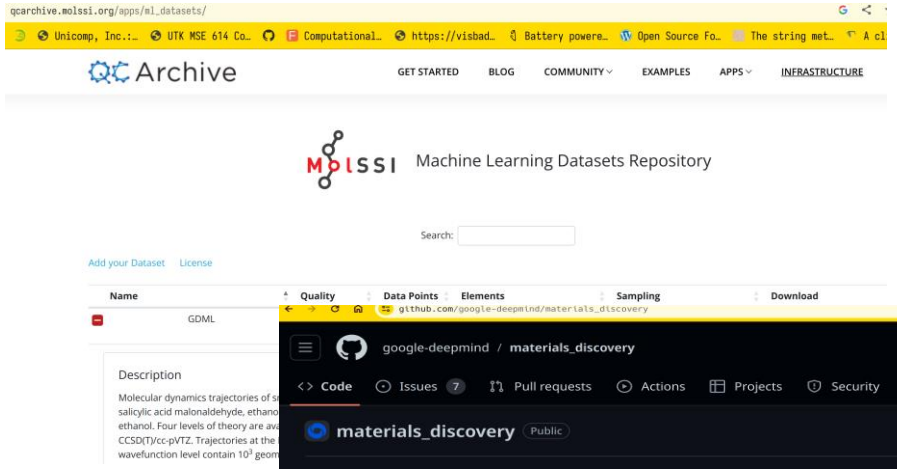
source: wandb

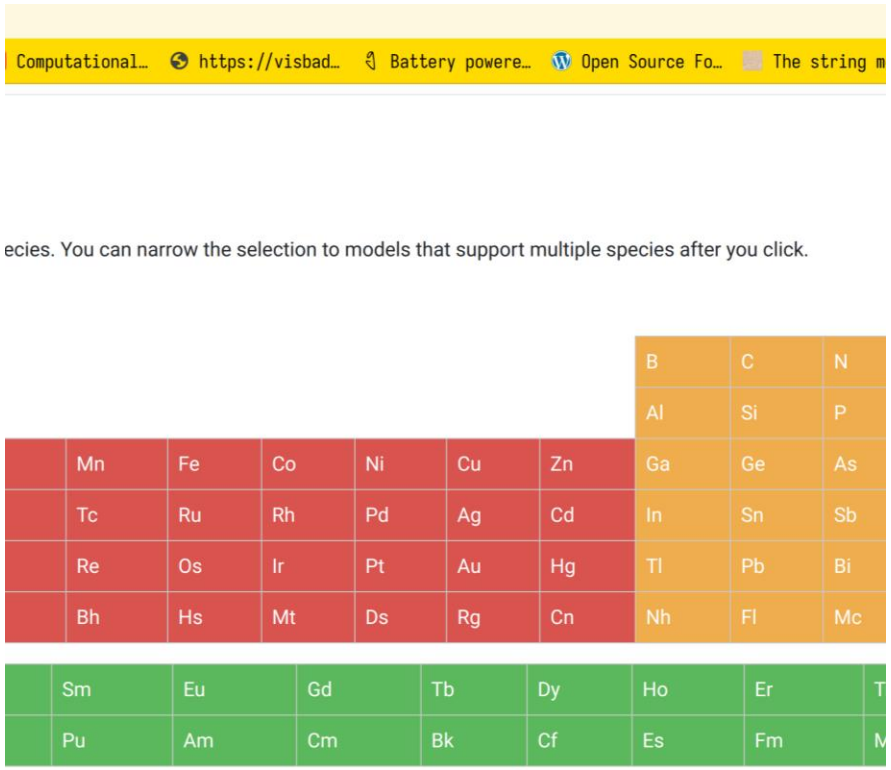
Workflows– science view

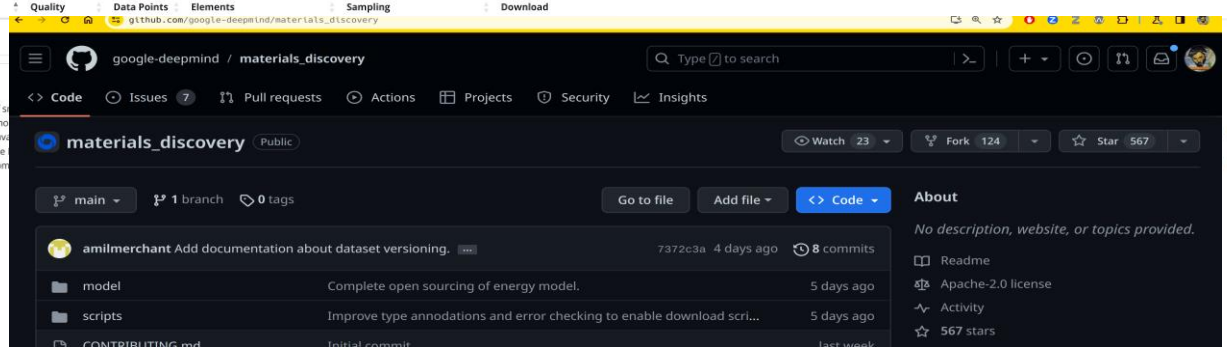


Current databases example









unique KIM potential is

Accessibility

Interatomic potentia directly with many

UiO66 HDNNP poter

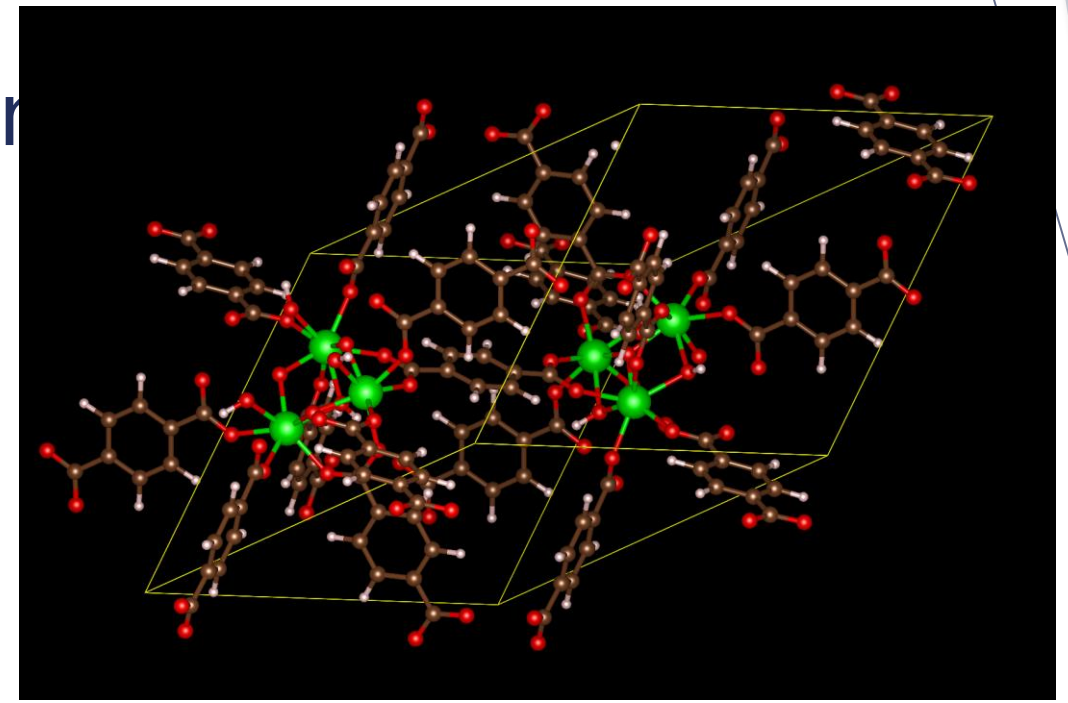
==========

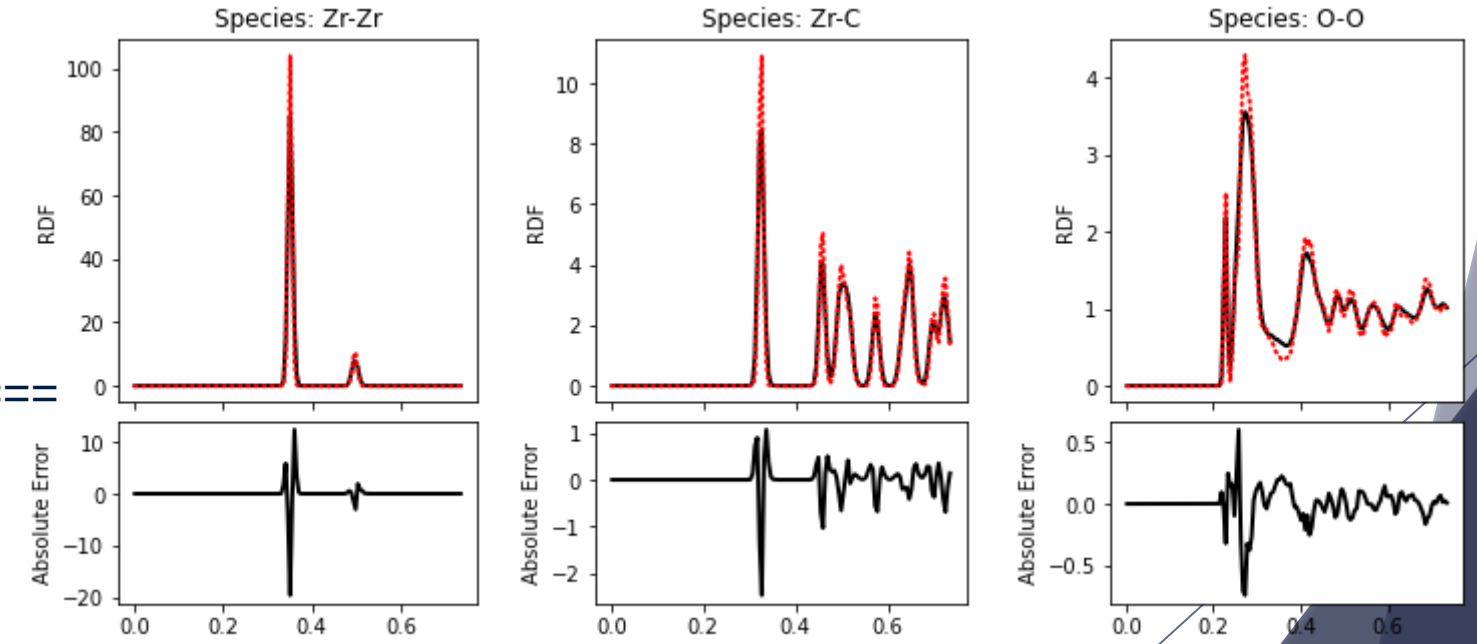
Score Summary

==========

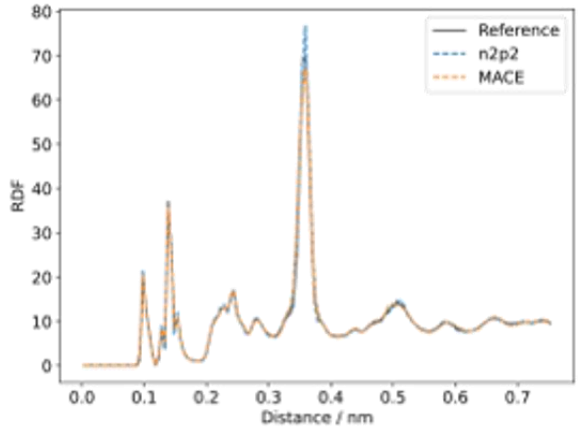
Label	Accuracy [%]
Zr-Zr	89.34
Zr-O	89.79
Zr-C	91.12
Zr-H	93.88
O-O	95.27
O-C	96.94
O-H	95.08
C-C	90.84
C-H	94.97
H-H	96.06
Mean	93.33

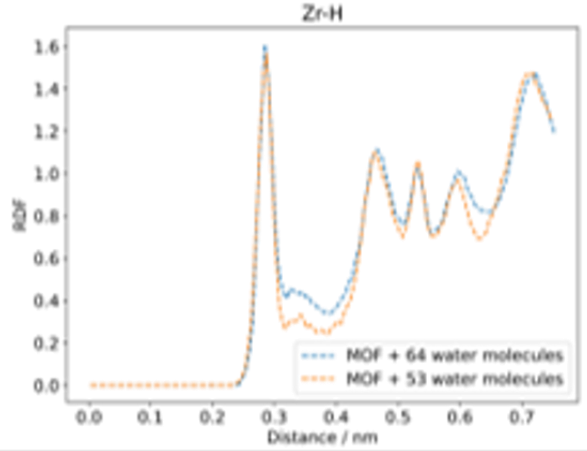
==========

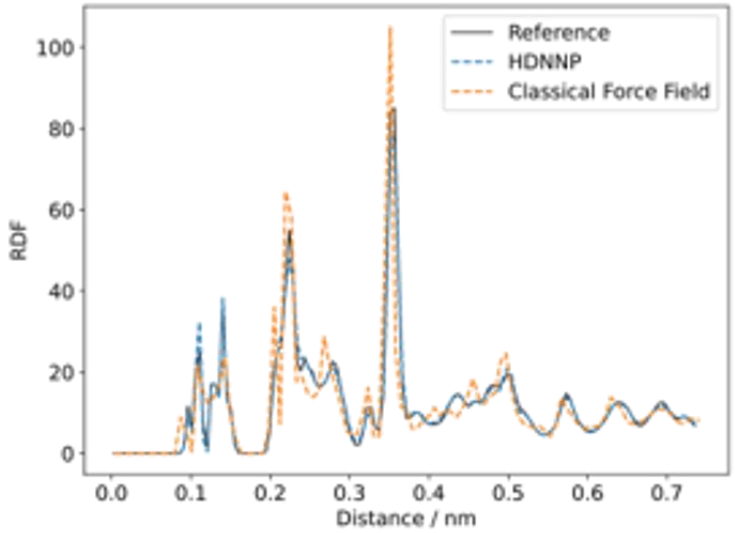


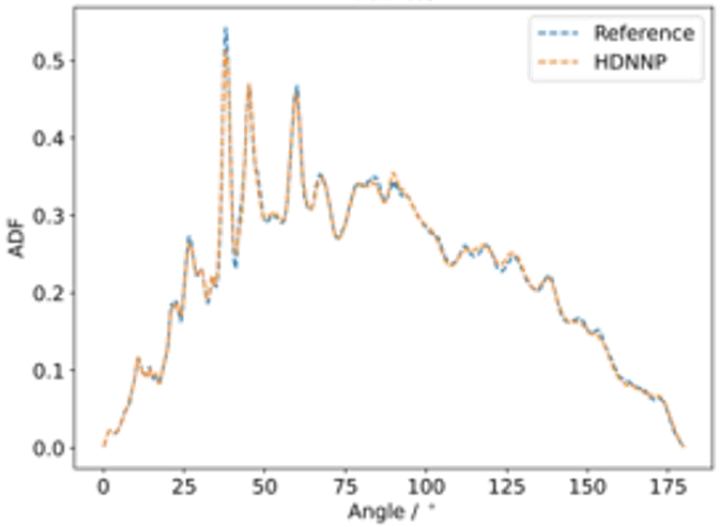


MACE

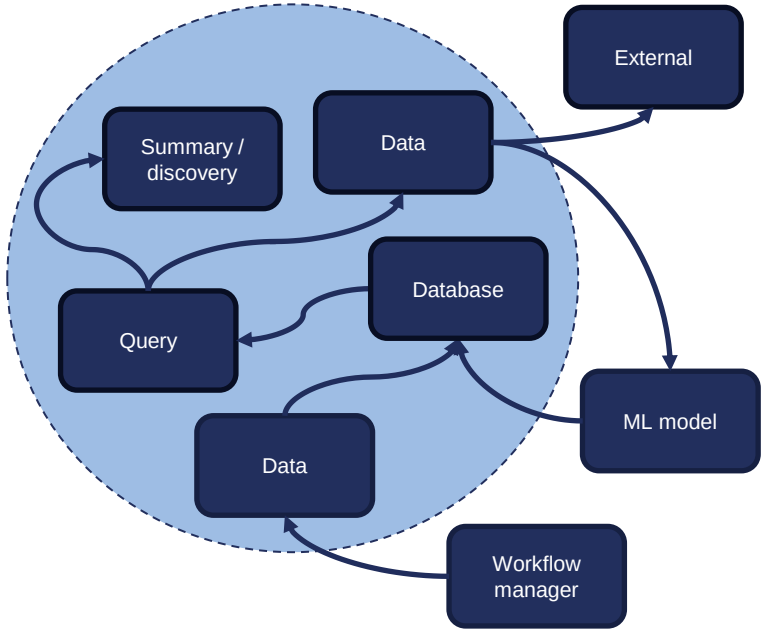

Reference
n2p2
MACE
RDF
Distance / nm


Zr-H
MOF + 64 water molecules
MOF + 53 water molecules
RDF
Distance / nm


Reference
HDNNP
Classical Force Field
RDF
Distance / nm


Reference
HDNNP
ADF
Angle / °

Database features

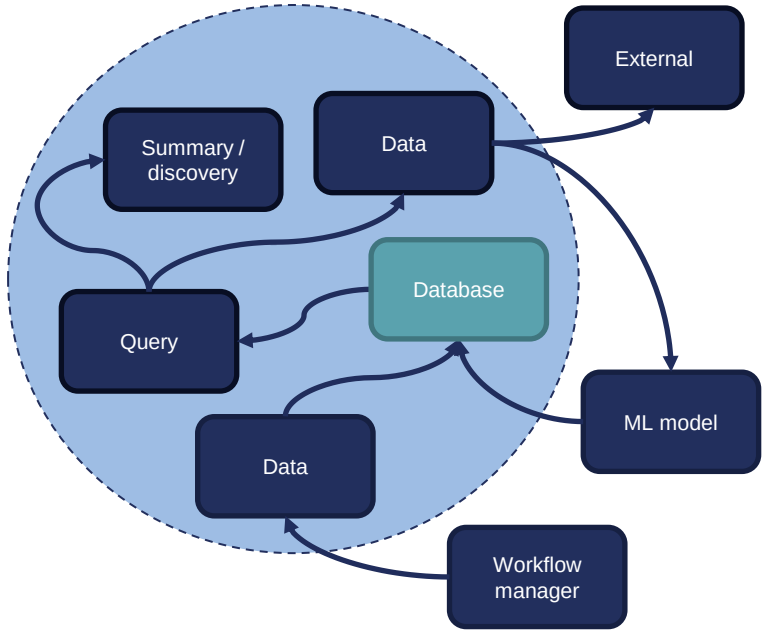


- Proof of concept database&app:
  - installable centrally within PSDI
  - local for user - blueprints
  - users can both interrogate, download and deposit data
  - reconfigurable database structure
  - API and web interface
- Advanced search features: opensearch techniques
- Integrate into ML workflows
- Reproduceable and Reusable

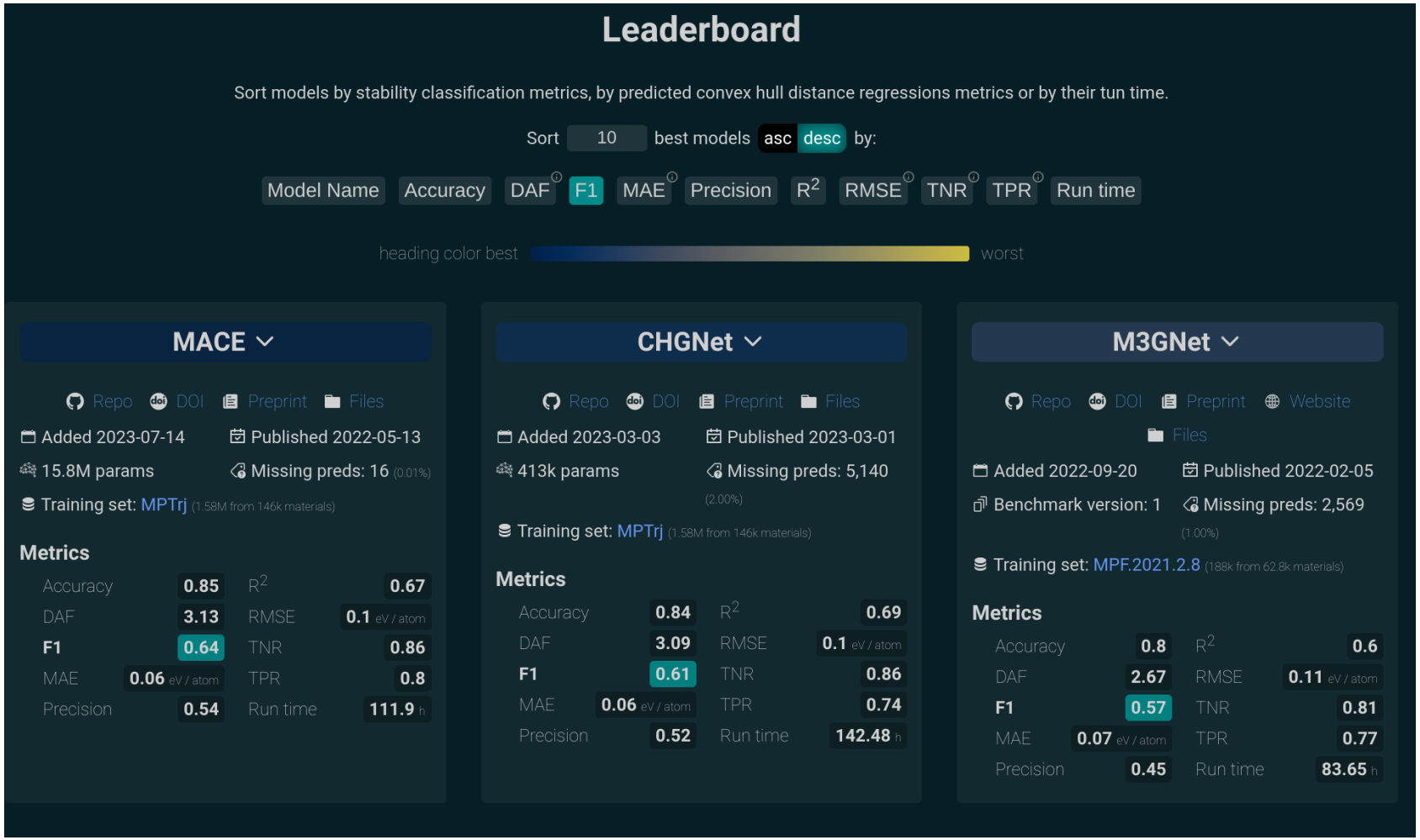
- Proof of concept in collaboration with Gábor Csányi@University of Cambridge

Database implementation

- JSON/dictionary data input
  - Atomistic data e.g. elements, positions, energies, forces
  - Metadata e.g. units, provenance
  - User-defined keys
- API access via Python client and CLI initially
- Complex, scalable queries via OpenSearch
- Data discovery due to unstructured data
  - List/frequency of keys
  - Visualisation
- Filtered data used to train ML models
  - Generate and store new structures and force fields in database



but



https://matbench-discovery.materialsproject.org/models

Q&AI