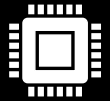
Scaling deep learning for material discovery

Jannis Steiger

Technische Universität München

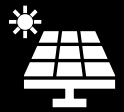
Seminar Deep Learning in Natural Sciences

Munich, 10th December 2024



computer chips

climate change

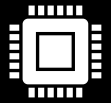


solar power

challenges in 2024

chemistry

battery

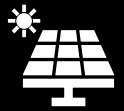


computer chips

chemistry

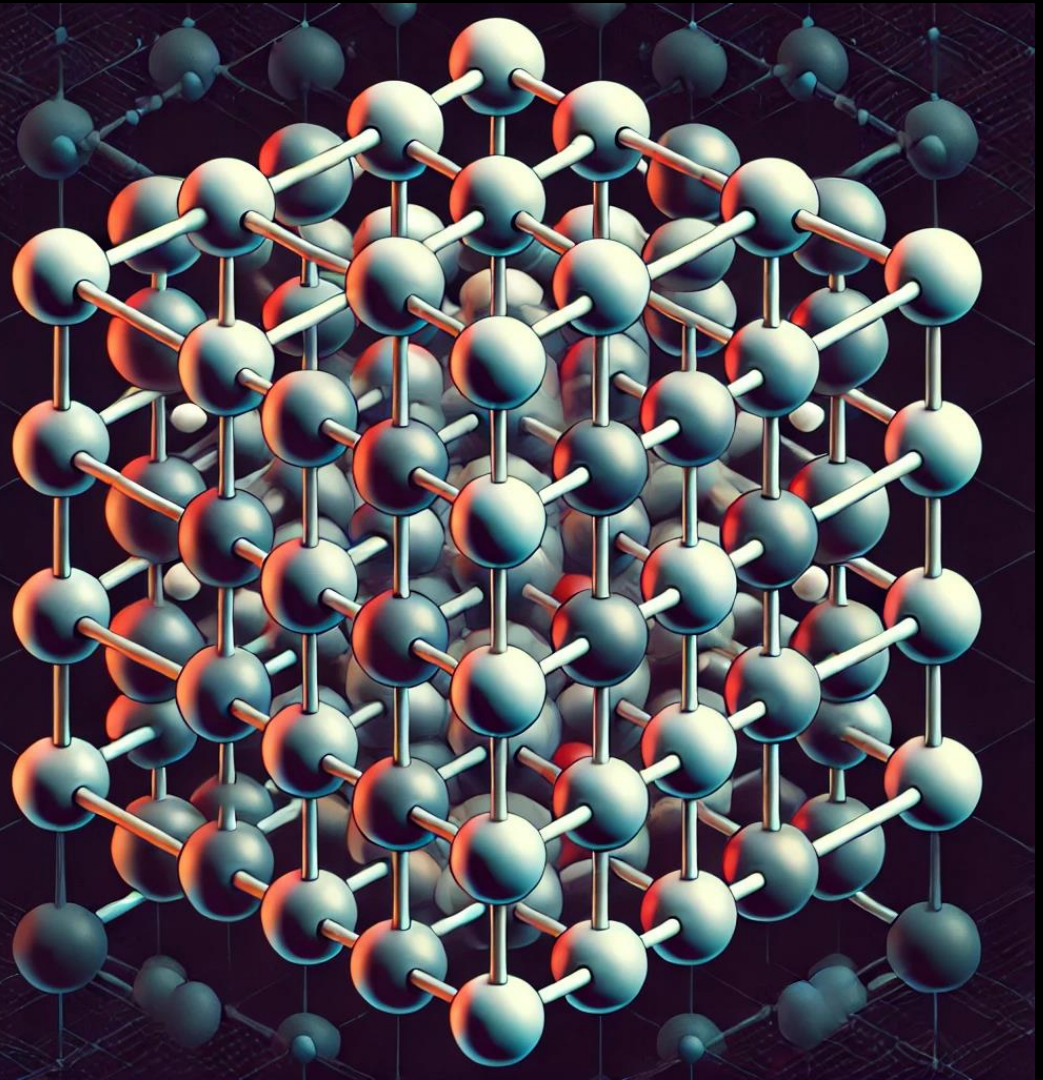
climate change

battery



solar power

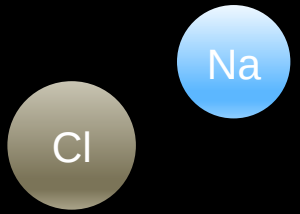
all of these technologies rely on crystals



Agenda

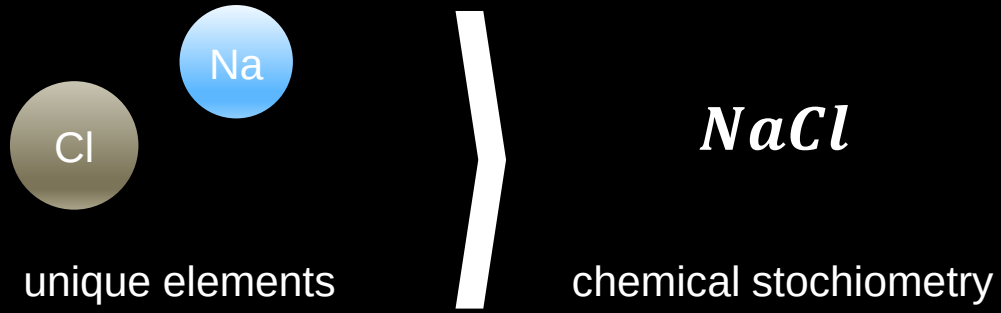
1. Motivation
2. Material discovery before GNoME
3. ML basics for GNoME
4. Material discovery with GNoME
5. Evaluation

How to build a crystal

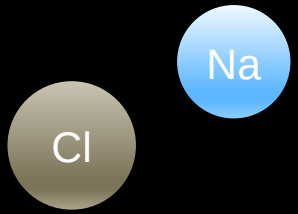


unique elements

How to build a crystal



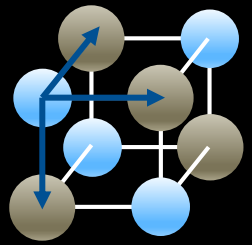
How to build a crystal



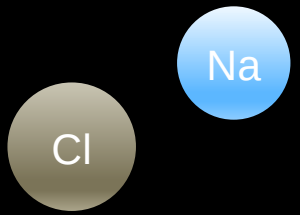
unique elements

$NaCl$

chemical stochiometry



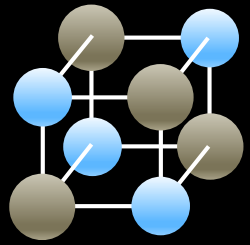
crystal structure



unique elements

$NaCl$

chemical stochiometry



crystal structure

stability

$E_{system} \overset{!}{<} 0$ → stable

DFT

Density Functional Theory (DFT)

Assumptions:

1. Energy of system depends solely on electron density $\rho(r)$
2. The basic state of $\rho(r)$ minimizes the energy of the system $<=>$ $\boldsymbol{\rho_{basic}(r) = \underset{\rho}{argmin}\, E(\rho)}$

Energy is calculated as follows:

$$E(\rho) = T(\rho) + V_{ext}(\rho) + J(\rho) + \boldsymbol{E_{xc}(\rho)}$$

$\boldsymbol{E_{xc}(\rho)}$**:** exchange-correlation functional that describes quantum mechanical effects → **approximated**

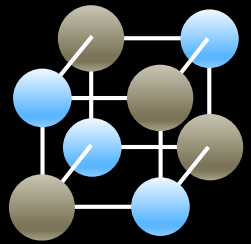
r²SCAN
PBE

+ DFT is an exact and efficient approach to calculate the stability of a crystal

- quality highly depends on good approximation of $\boldsymbol{E_{xc}(\rho)}$

Current Approaches

- currently mostly **trial-and-error** → depends on **human chemical intuition**
- experimental approaches catalogued **20,000 computationally stable crystals** in Inorganic Crystal Structure Database (ICSD)
- computational approaches lead to **48,000 computationally stable crystals**
- **ML** techniques tried but ineffective in estimating stability

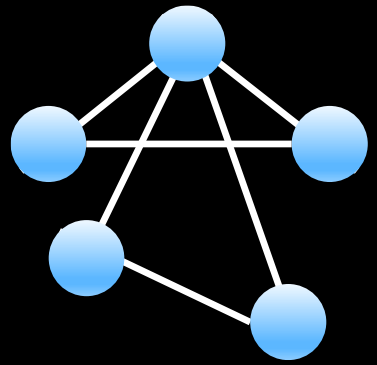


→ current approaches are either **inefficient** or **ineffective**

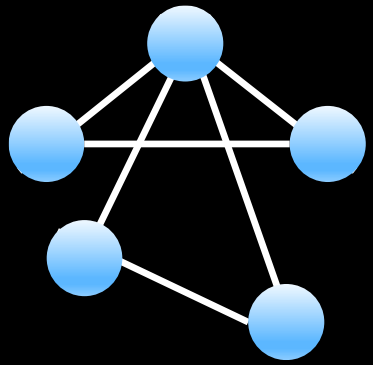
→ GNoME (graph network for material exploration) is the solution!

What is a graph neural network (GNN)?

- consists of graphs → just like crystals
- performs message passing
- has an attention mechanism



What is a graph neural network (GNN)?



- consists of graphs → just like crystals
- performs message passing
- has an attention mechanism

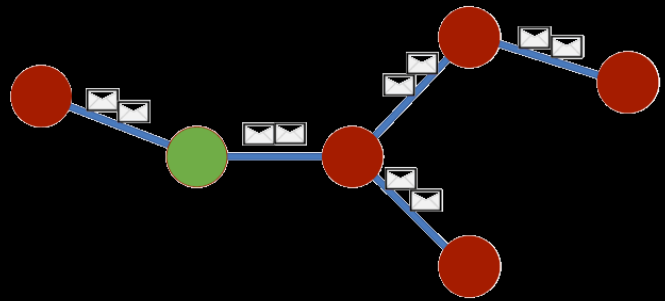
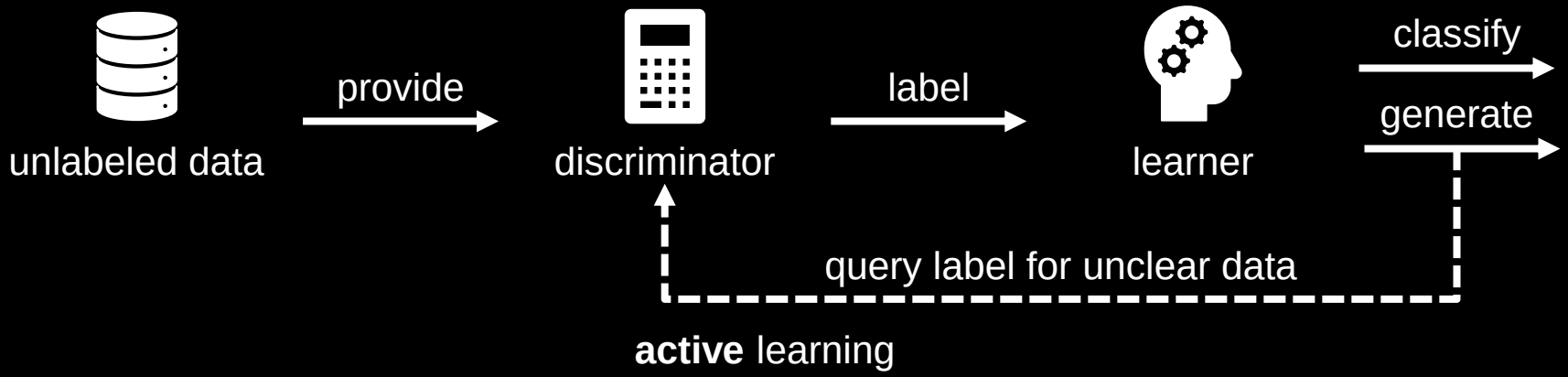


Figure: [machinelearningknowledge.ai]

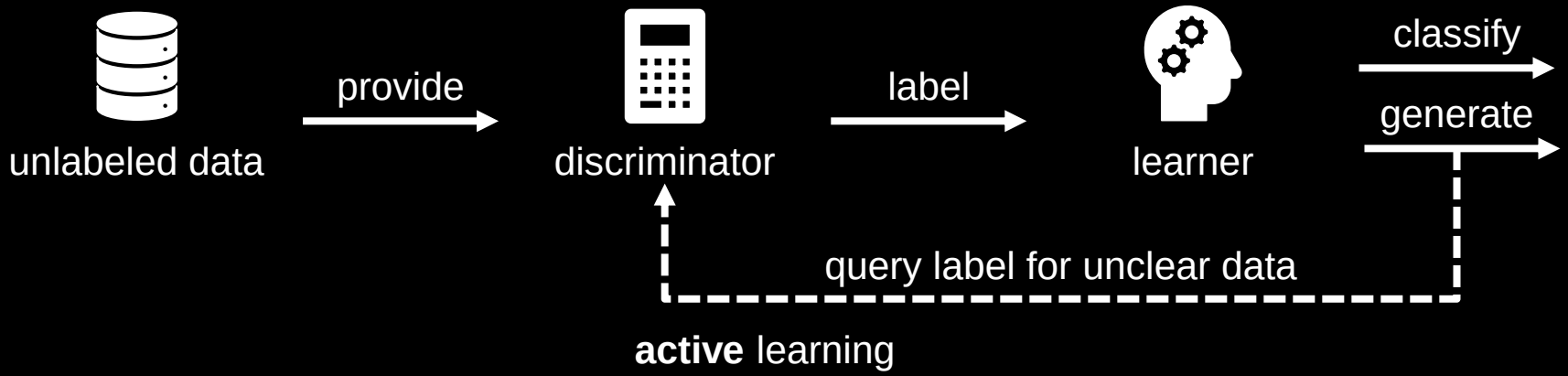
Active learning vs passive learning

passive learning

Active learning vs passive learning



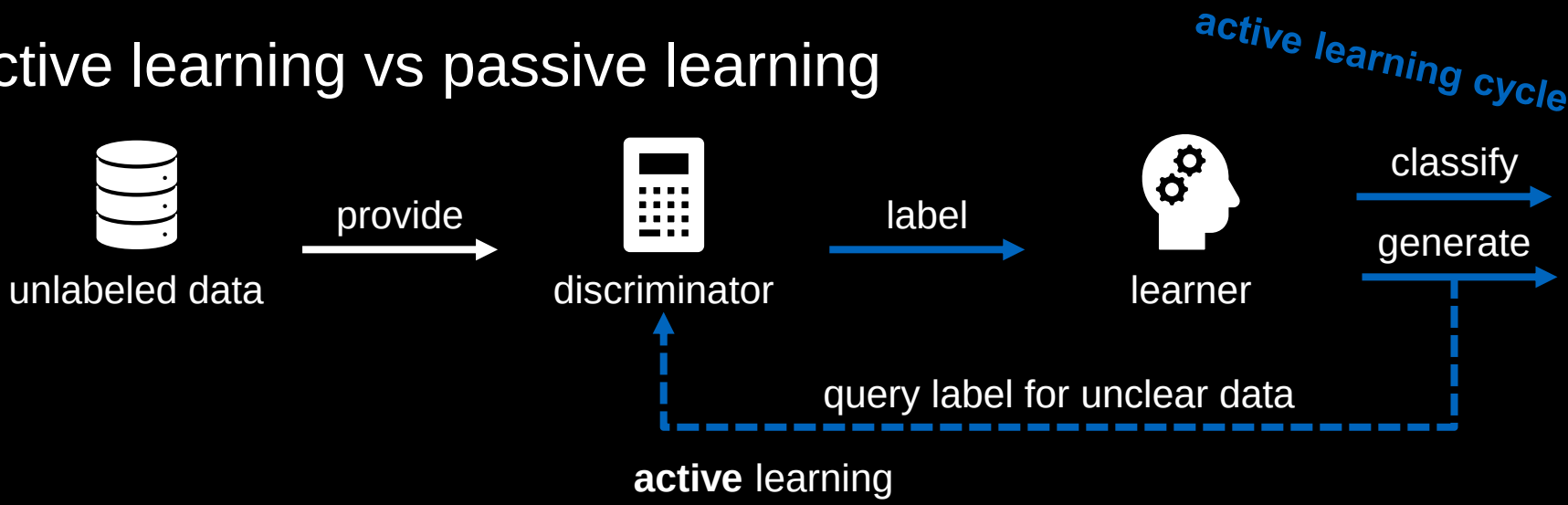
Active learning vs passive learning



- in active learning learner queries discriminator for new labels
  - ➢ amount of training data rises over time

→ **active** learning vs **passive** learning comparable to **lecture** vs **seminar**

Active learning vs passive learning



- in active learning learner queries discriminator for new labels
  - ➢ amount of training data rises over time

→ **active** learning vs **passive** learning comparable to **lecture** vs **seminar**

The GNoME framework

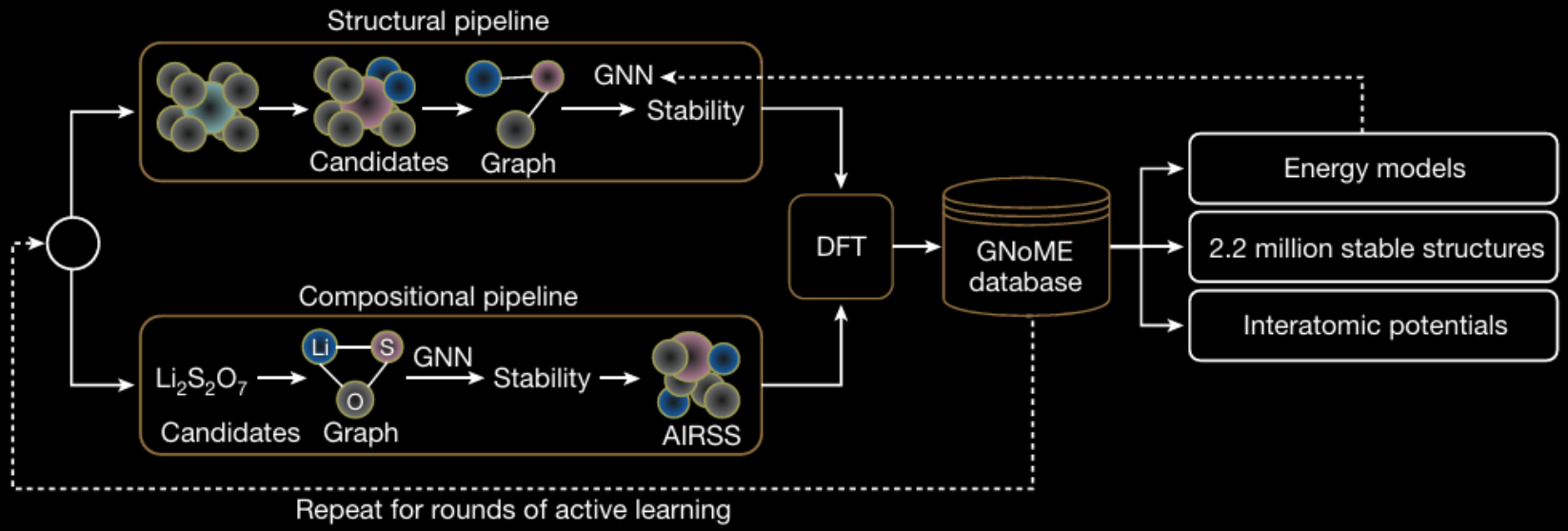


Figure: Merchant2023

The GNoME framework

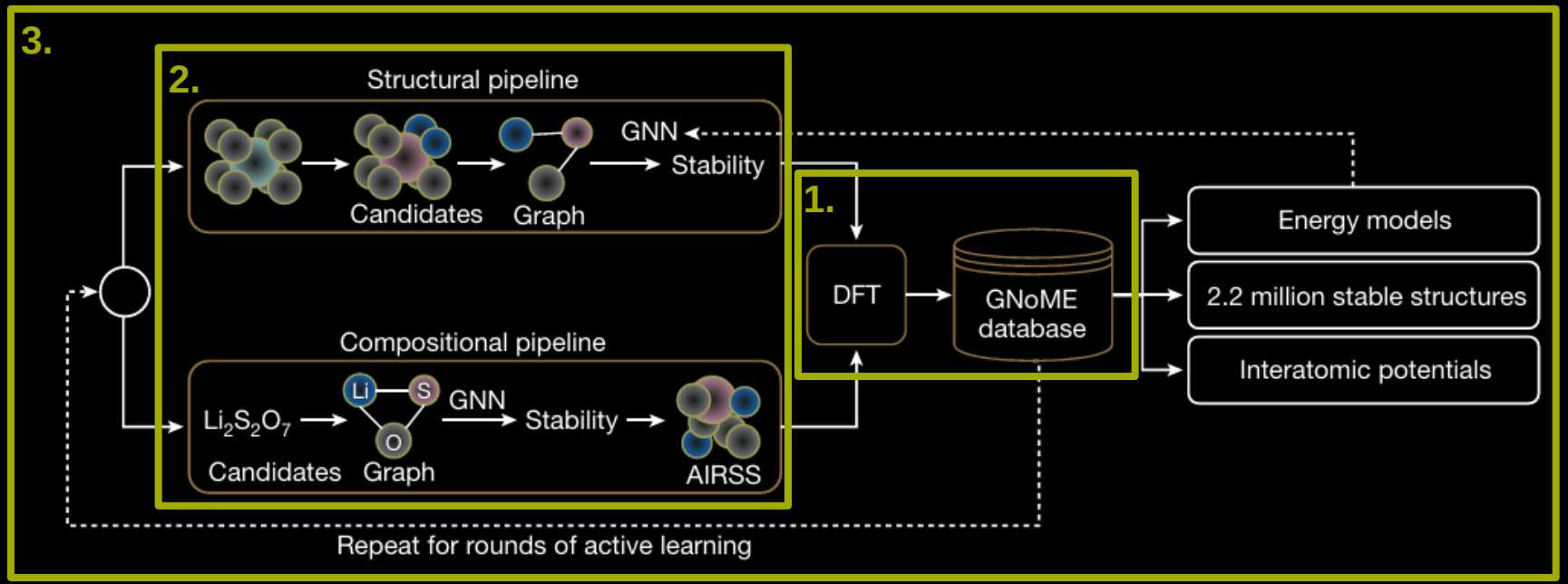


Figure: Merchant2023

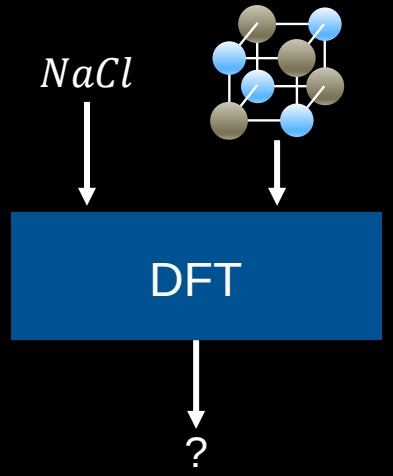
GNoME Database

Based on Experiments

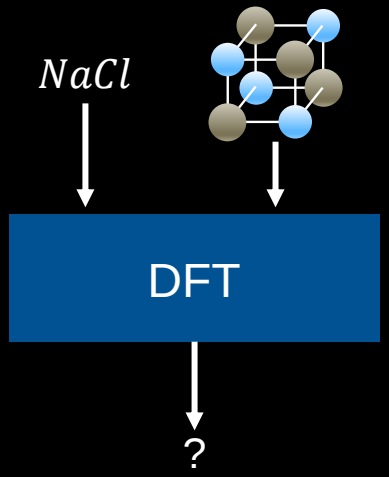
Inorganic Crystal Structure Database (**ICSD**)

- experimentally determined stable materials
- ~20.000 stable crystal structures

Based on Calculations



Based on Calculations



Materials Project (~100k materials)

electrical properties
thermal properties
mechanical properties
stability

OQMD (~1M materials)

stability
electrical properties
thermodynamical properties

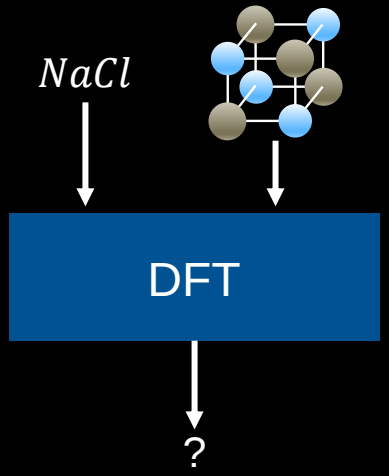
AFLOW (~3M materials)

mechanical properties
electrical properties
thermodynamical properties
(stability)

NOMAD (~100M materials)

combination of uploaded simulations (DFT, Molecular Dynamics, Monte Carlo)

Based on Calculations



Materials Project (~100k materials)

electrical properties

thermal properties

mechanical properties

stability

OQMD (~1M materials)

stability

thermodynamical properties

electrical properties

AFLOW (~3M materials)

mechanical properties

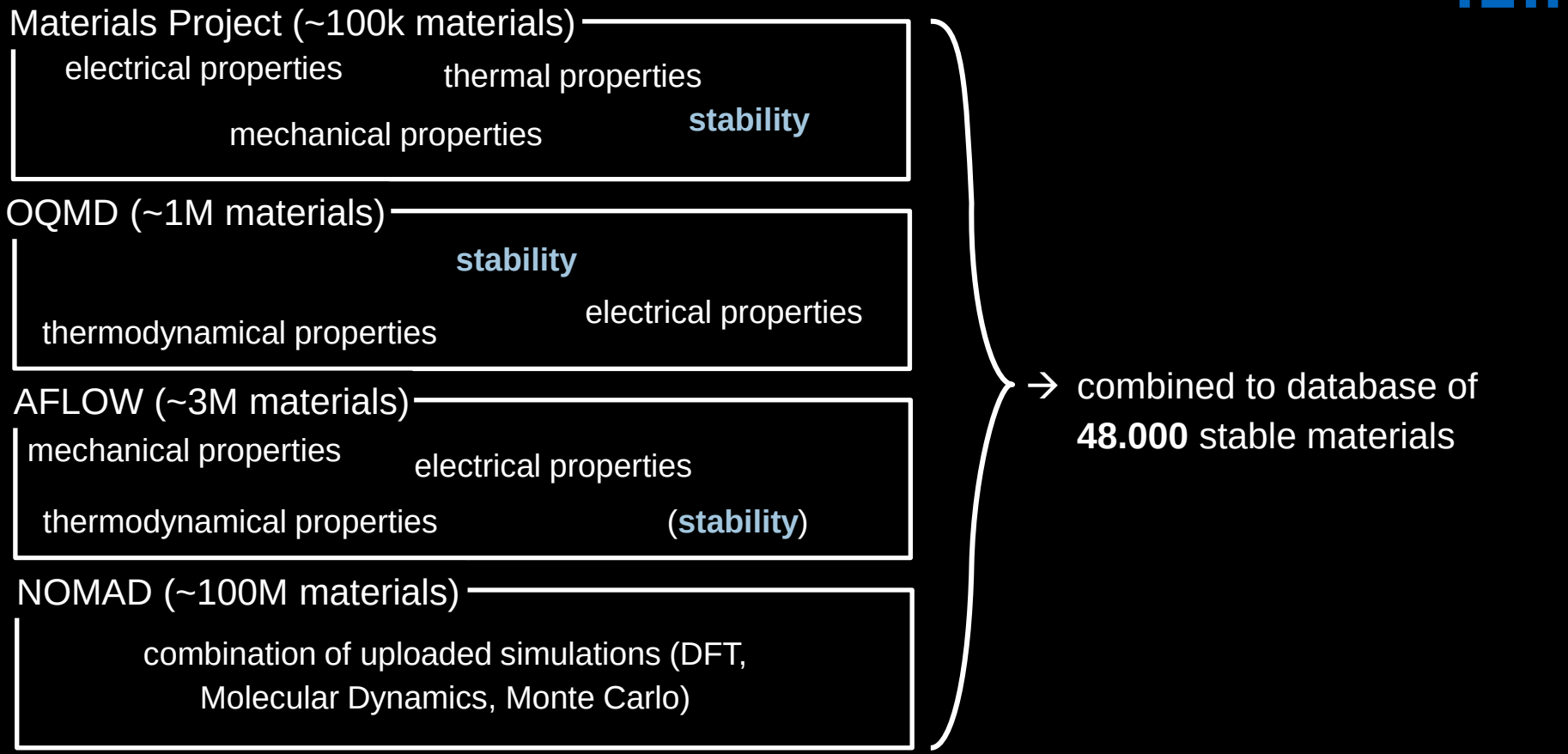
electrical properties

thermodynamical properties

(**stability**)

NOMAD (~100M materials)

combination of uploaded simulations (DFT, Molecular Dynamics, Monte Carlo)


Materials Project (~100k materials)
electrical properties
thermal properties
mechanical properties
stability
OQMD (~1M materials)
stability
thermodynamical properties
electrical properties
AFLOW (~3M materials)
mechanical properties
electrical properties
thermodynamical properties
(stability)
NOMAD (~100M materials)
combination of uploaded simulations (DFT, Molecular Dynamics, Monte Carlo)
→ combined to database of **48.000** stable materials

Density Functional Theory (DFT)

Assumptions:

1. Energy of system depends solely on electron density $\rho(r)$
2. The basic state of $\rho(r)$ minimizes the energy of the system $<=>$ $\boldsymbol{\rho_{basic}(r) = \underset{\rho}{argmin}\, E(\rho)}$

Energy is calculated as follows:

$$E(\rho) = T(\rho) + V_{ext}(\rho) + J(\rho) + \boldsymbol{E_{xc}(\rho)}$$

$\boldsymbol{E_{xc}(\rho)}$**:** exchange-correlation functional that describes quantum mechanical effects → **approximated**

r²SCAN

PBE

+ DFT is an exact and efficient approach to calculate the stability of a crystal

- quality highly depends on good approximation of $\boldsymbol{E_{xc}(\rho)}$

GNoME uses the Vienna Ab initio Simulation Package (**VASP)** to calculate the **DFT**

The GNoME framework

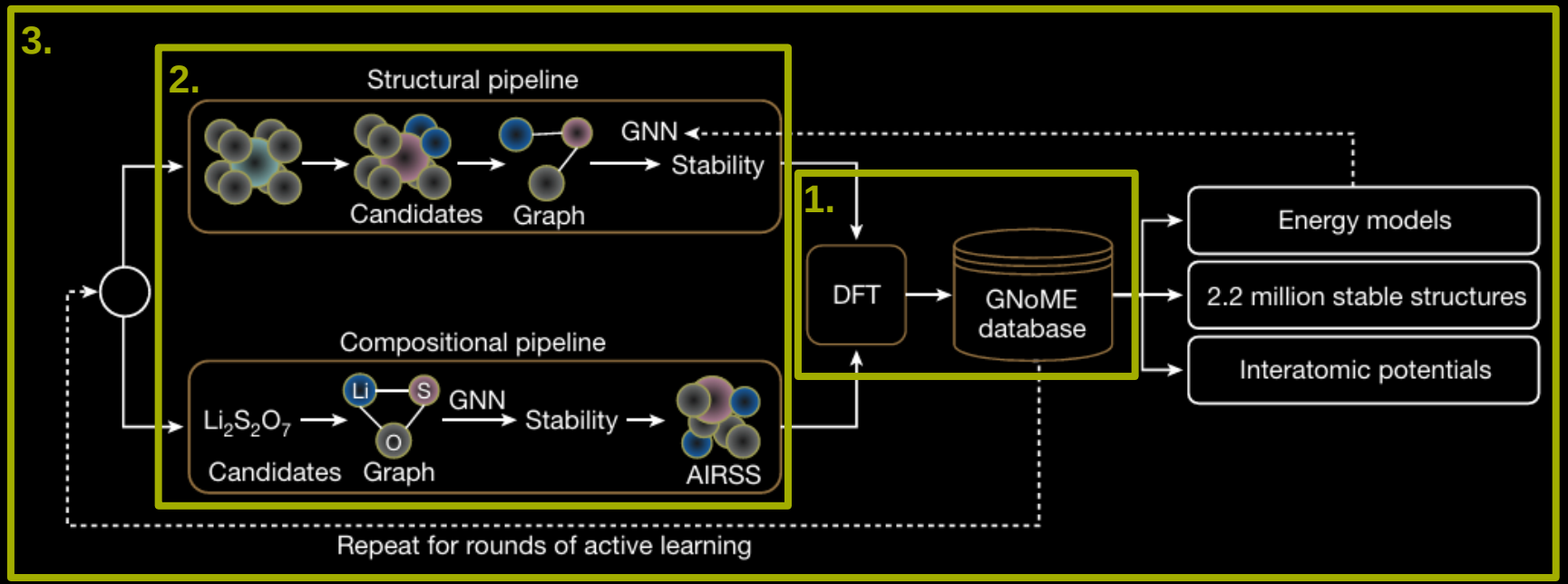
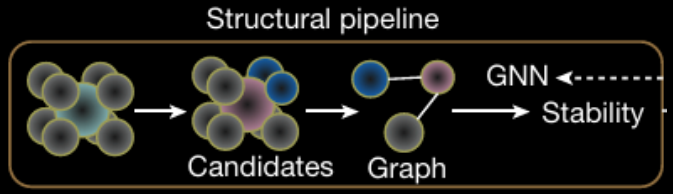
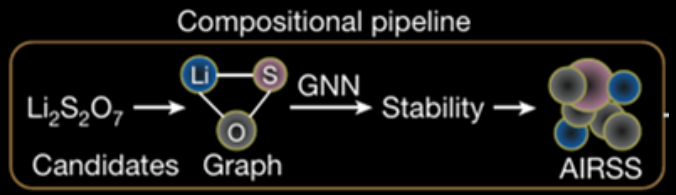


Figure: Merchant2023

The GNoME framework





Structural pipeline:

- known crystals as input
- augment known crystals
- predict stability

Compositional pipeline:

- predict stability by unstructured candidates
- ab initio random structure searching (AIRSS) creates structure candidates

Figure: Merchant2023

The GNoME framework

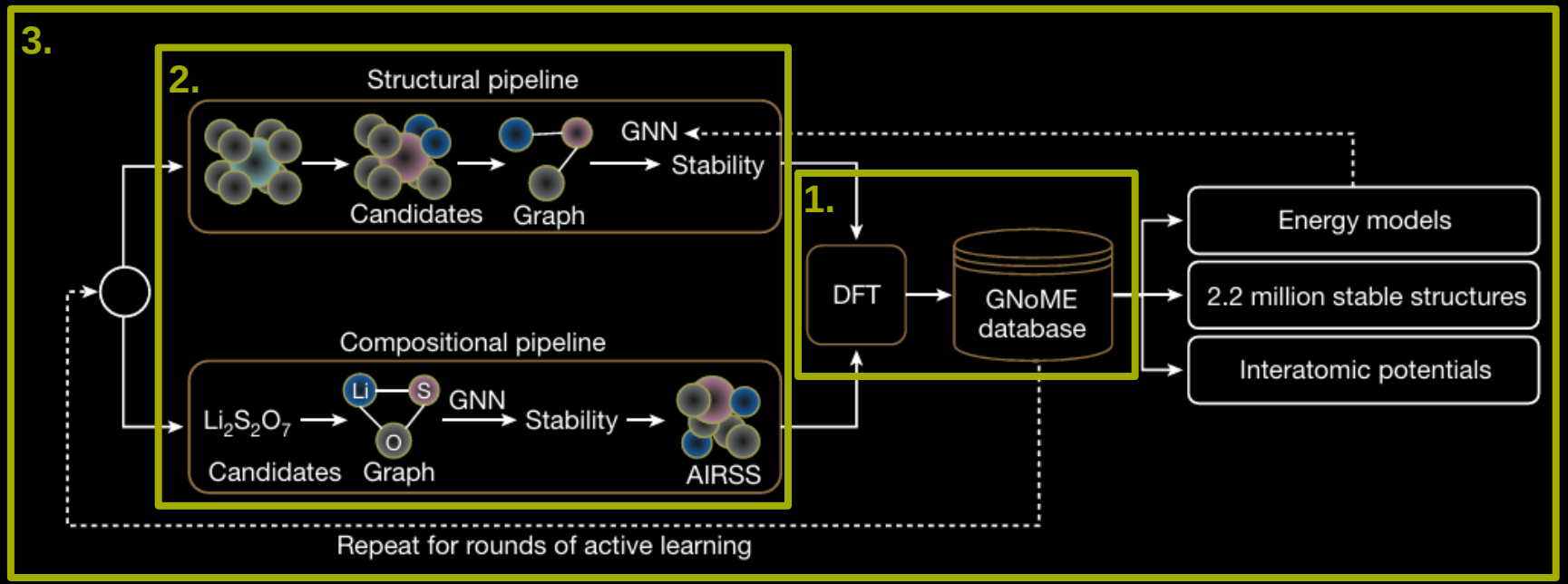


Figure: Merchant2023

The GNoME framework

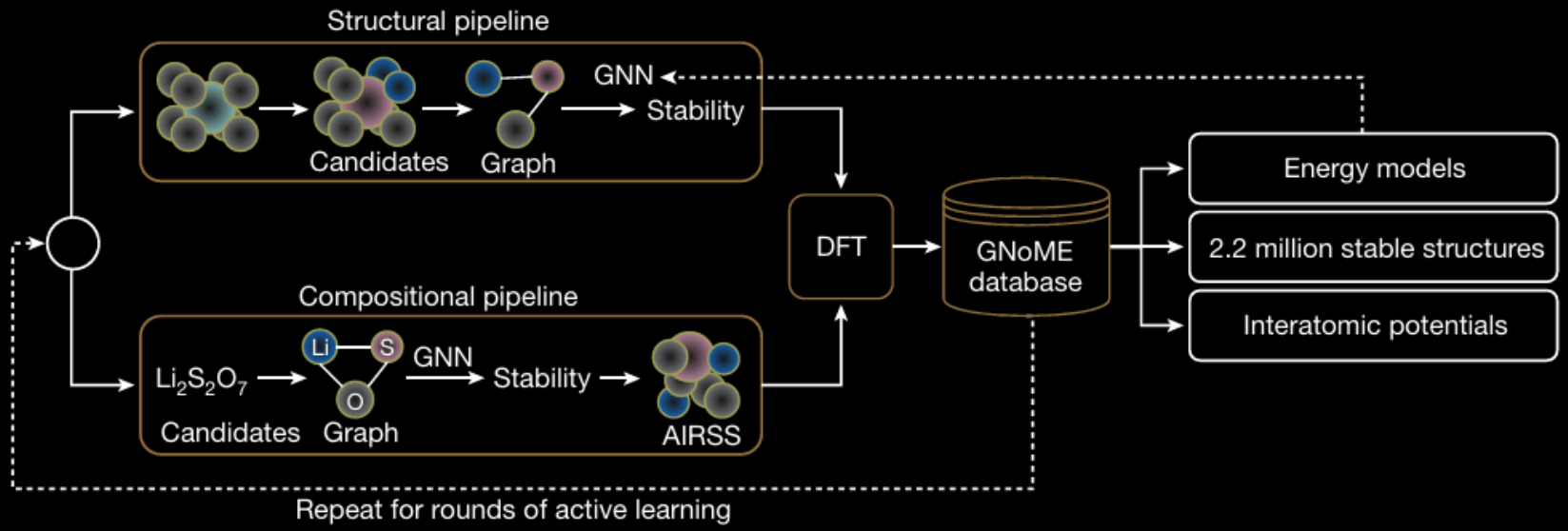


Figure: Merchant2023

The GNoME framework

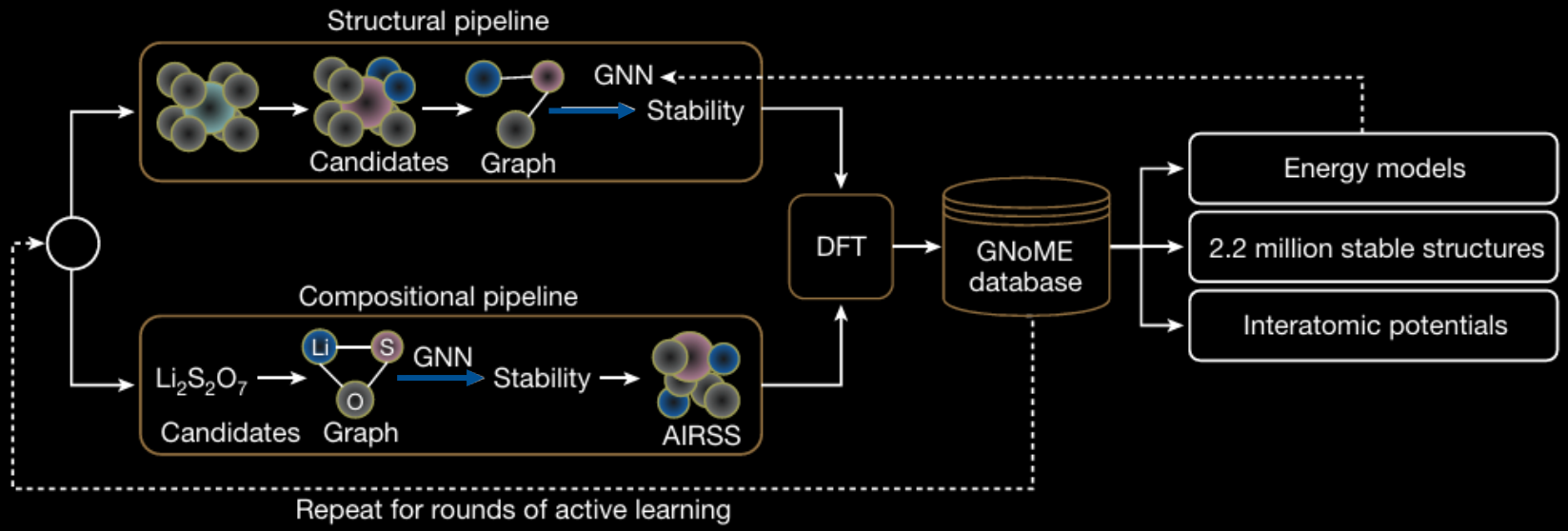


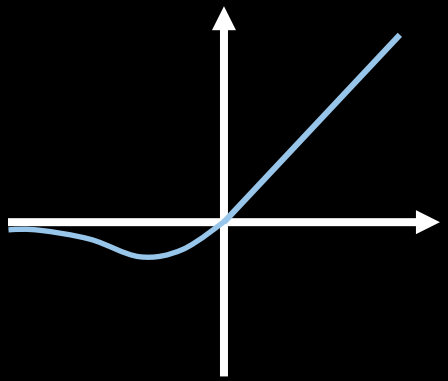
Figure: Merchant2023

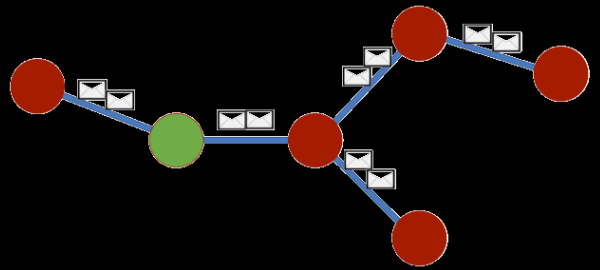
Message-passing in GNoME

- shallow multilayer perceptrons

  → only few message passing iterations

- messages normalized by average number of neighbors for an atom across the whole dataset
- swish nonlinearity





Training process in GNoME

initial model

trained on **69,000 materials** from Materials Project, **best architecture** kept
→ $MAE = 20meV/atom$ compared to $MAE = 28meV/atom$ benchmark

model training

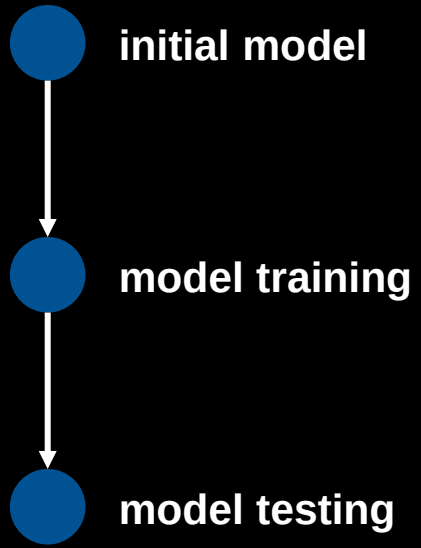
final training with whole data set → $MAE = 11meV/atom$

model testing

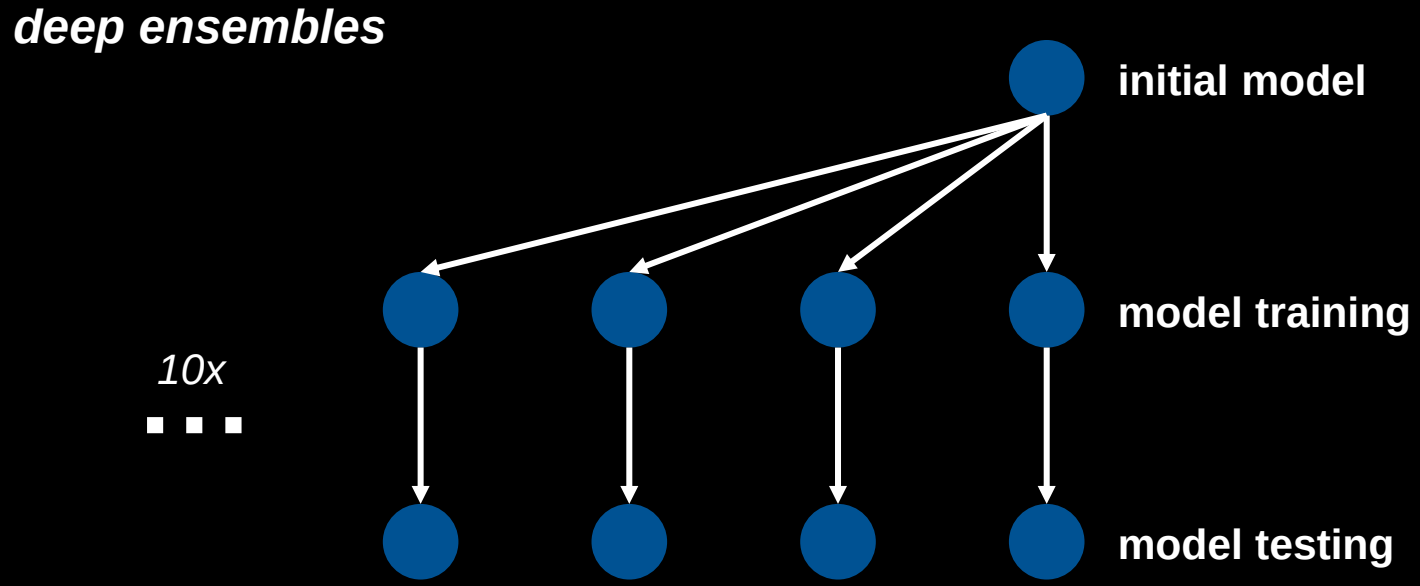
scaling of lattice vectors from 0.8 to 1.2 **augments test set**

Training process in GNoME

*Is there a simple way to improve **generalization**?*



Training process in GNoME



Did GNoME enhance material discovery or **revolutionize** it?

- improved data from 48k computationally stable materials to 2.2m computationally stable materials
  - ➢ factor $4.5 * 10^5$ (!)
- especially performs on crystals where chemical intuition fails (5 or more unique element)
- can lead to better understanding (e.g. interatomic potentials)

→ GNoME **revolutionized** material science

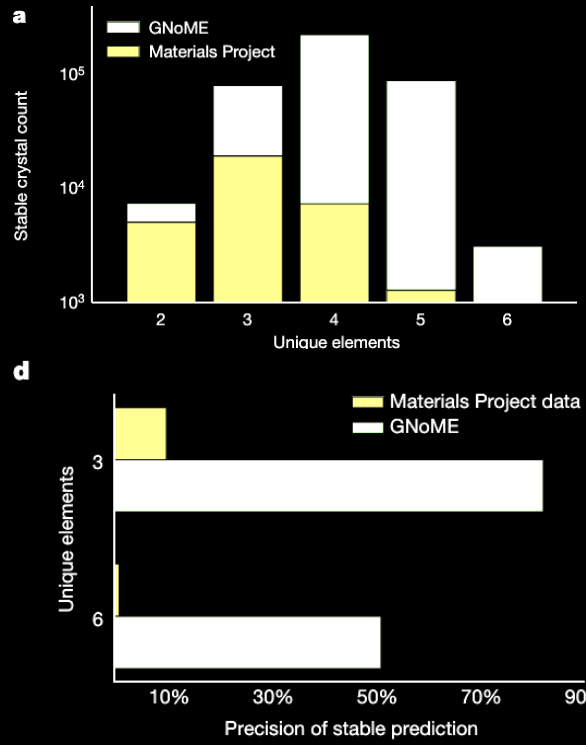


Figure: Merchant2023

Where did GNoME impact research already?

DeAngeles2024*

- identify potential candidates for batteries

↯

Cheetham2024**

- GNoME has not made practical contributions to experimental material science
- many discovered materials are slight variations of known materials
- industry application still decades away

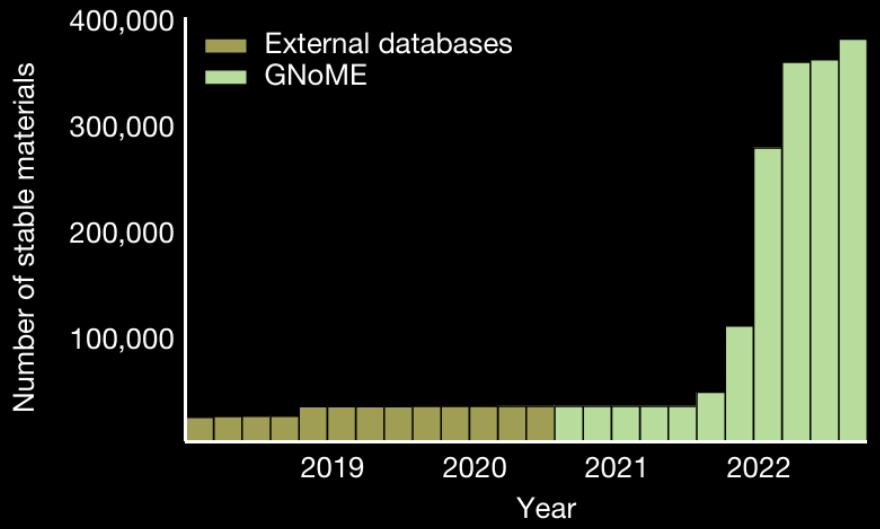


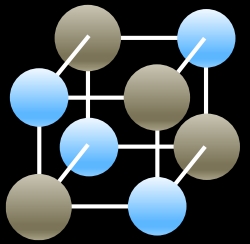
Figure: Merchant2023

*DE ANGELIS, Paolo, et al. Energy-GNoME: A Living Database of Selected Materials for Energy Applications. *arXiv preprint arXiv:2411.10125*, 2024

**CHEETHAM, Anthony K.; SESHADRI, Ram. Artificial intelligence driving materials discovery? perspective on the article: Scaling deep learning for materials discovery. *Chemistry of Materials*, 2024, 36. Jg., Nr. 8, S. 3490-3495

Main takeaways

1. GNoME exponentially **increased number of known stable crystals** but many materials are **slight variations** of already known crystals
2. GNoME opens **possibilities in many research areas** (battery, chemistry, solar, chips…) but actual **impact** is **to be evaluated**
3. Contributed massive impact on **understanding** of material science; especially for **complex crystals** and **interatomic potentials**



Any questions..?

References

Merchant2023: Scaling deep learning for materials discovery
DeAngeles2024: Energy-GNoME: A Living Database of Selected Materials for Energy Applications
Cheetham2024: Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery
Hellenbrand2004: The Inorganic Crystal Structure Database (ICSD)—present and future
Jain2013: The Materials Project: a materials genome approach to accelerating materials innovation
Saal2013: Materials design and discovery with high-throughput density functional theory: the Open Quantum Materials Database (OQMD)
Daxl2019: The NOMAD laboratory: from data sharing to artificial intelligence
Curtarolo2012: AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations
Graph Neural Networks (GNN) Explained for Beginners