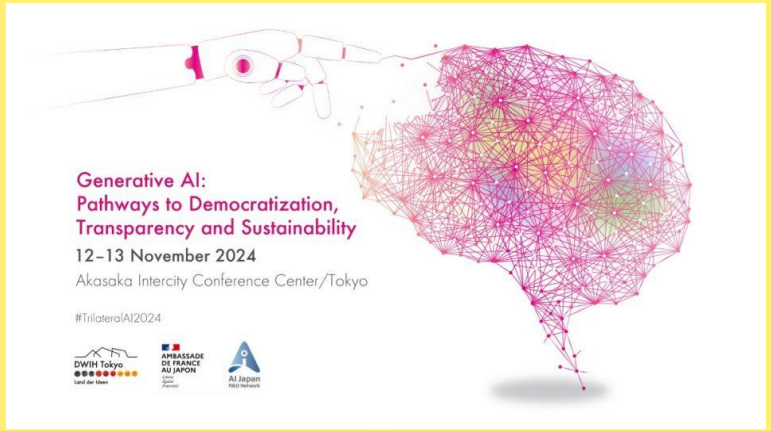


Parallel Workshop: GenAI and the Future of Research
12 -13 November 2024, Tokyo

Generative Approaches in Material Sciences: Accelerating Discovery and Innovation

Jean-Claude Crivello

http://link.cnrs.fr

Link IRL 3629

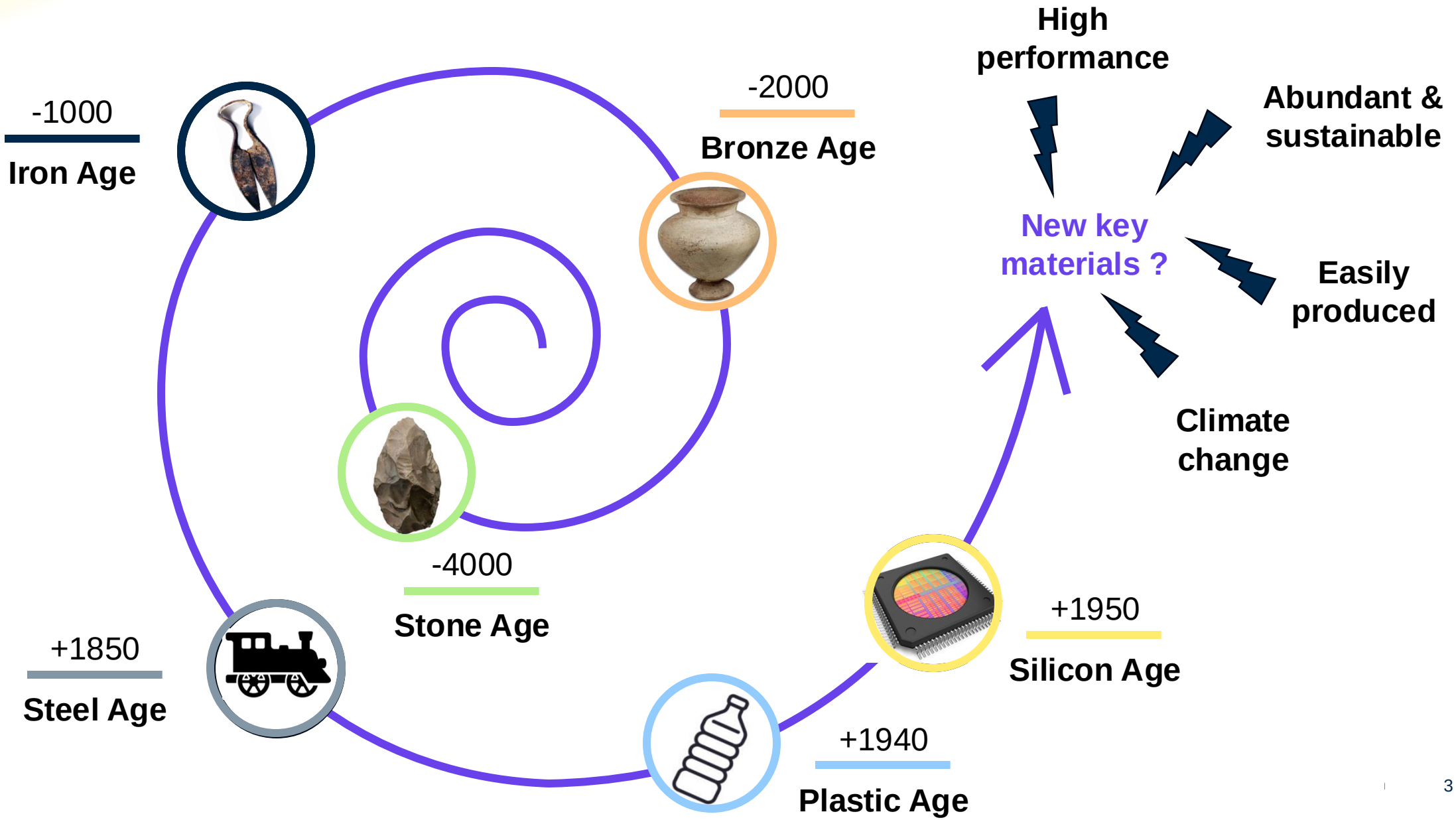
1. Introduction

2. How the Story Begins

3. Gen-AI in Materials Science

4. Limitations of AI approaches

Materials as Milestones in Humankind



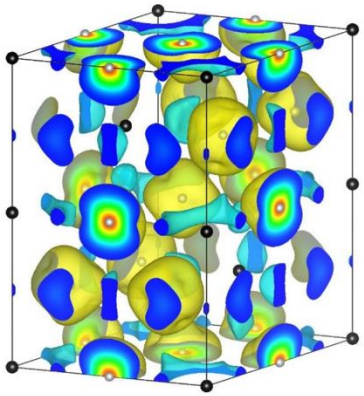
A new Paradigm in Materials Science

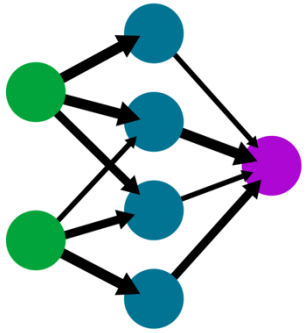
Heat capacity: $C_V \equiv \left(\frac{\partial U}{\partial T}\right)_V$ $C_P \equiv$

$$dQ + dW = CvdT$$
$$dW = -PdV$$
$$dQ = CvdT + PdV$$

Let V=RT/P :

$$dQ = C_V dT + RT\frac{dV}{V}$$
$$dQ = C_P dT - RT\frac{dP}{P}$$
$$\rightarrow \boxed{dQ = \frac{C_V}{R}VdP + \frac{C_P}{R}PdV}$$





1600 **1950** **2000**

1st paradigm — Empirical science

2nd paradigm — Model-based theoretical science

3rd paradigm — Computational science (simulations)

4th paradigm — Data driven science

J. Gray (2007)
4th paradigm is Science

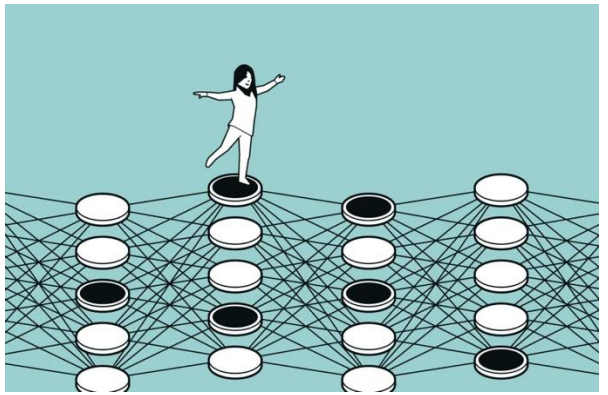
Agrawal *et al*. APL Materials (2016)
Perspective: Materials informatics and big data: realization of the « fourth paradigm » of science in materials science

Recognition by Nobel prizes 2024

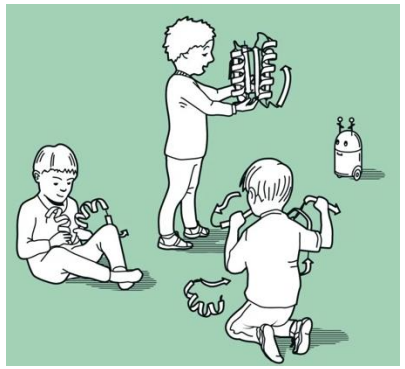
Physics: for foundational discoveries and inventions that enable machine learning with artificial neural networks

John Hopfield. Ill. Niklas Elmehed © Nobel Prize

Geoffrey Hinton. Ill. Niklas Elmehed © Nobel Prize



Chemistry: for computational protein design & for protein structure prediction



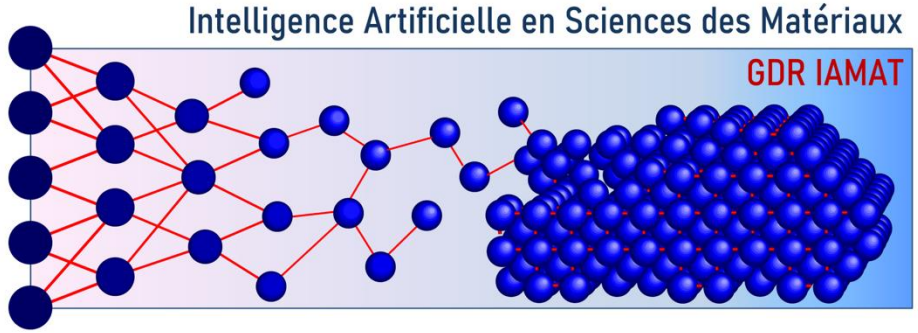
David Baker. Ill. Niklas Elmehed © Nobel Prize

Demis Hassabis. Ill. Niklas Elmehed © Nobel Prize

John Jumper. Ill. Niklas Elmehed © Nobel Prize

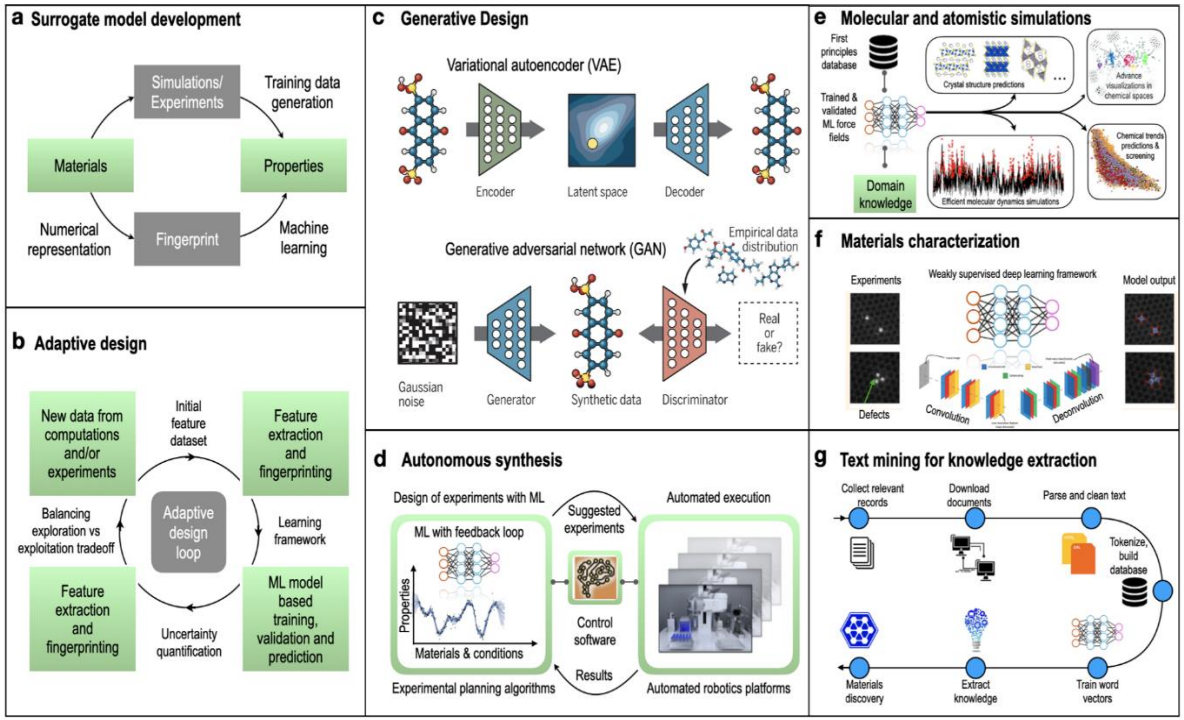
Literature: in 2042?

Structuration in France, Research Group GDR IA-MAT



Established in 2022 to unite AI and materials science research communities and foster interdisciplinary applications.

Classes of IA-MAT challenges	Artificial Intelligence Machine learning methods				
Machine learning in atomistic and multiscale modeling	Databases	Regression	Classification / Clustering	Feature identification	Optimisation
Artificial intelligence for high-throughput characterisation					
Materials design and structure-property relations					



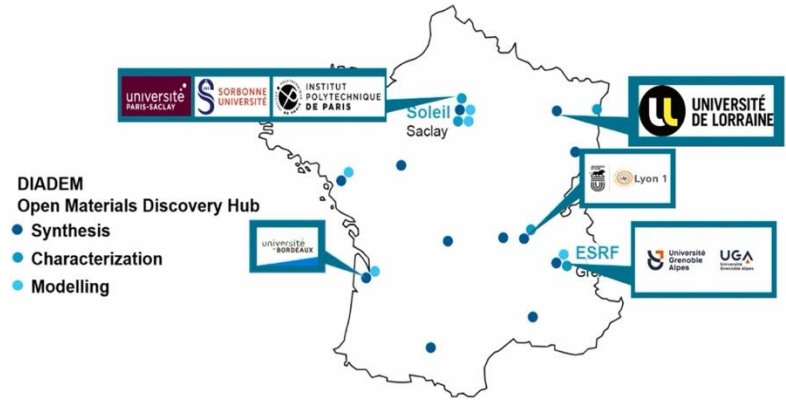
http://iamat.cnrs.fr

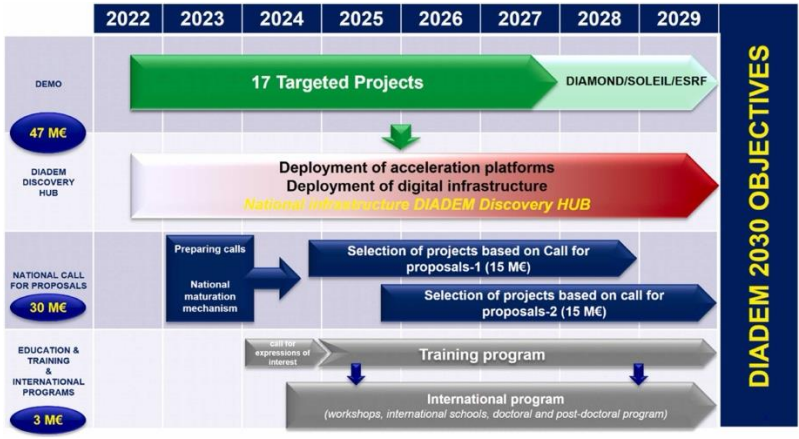
National Research Program, PEPR DIADEM

Led by

Financed by

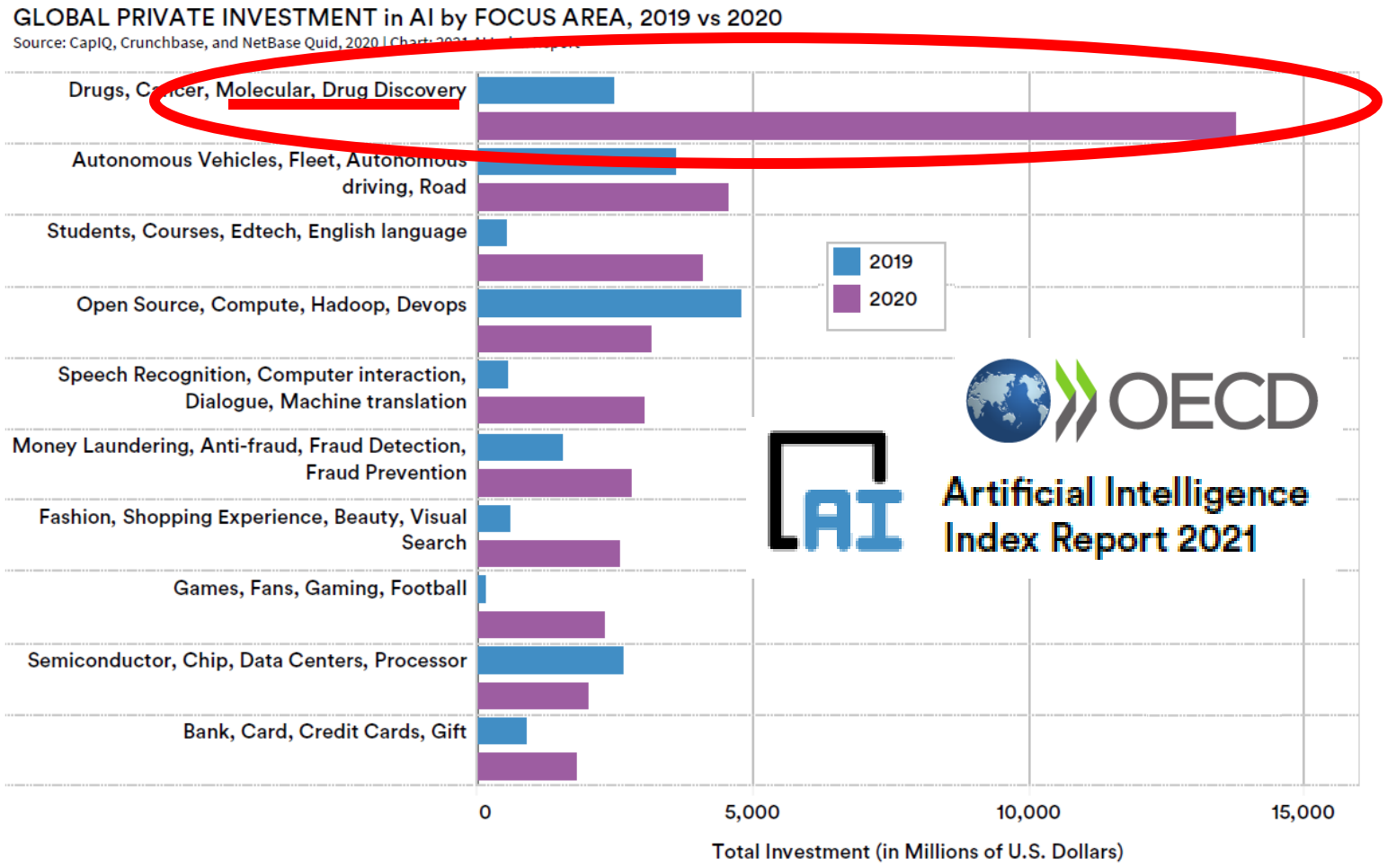
Operated by





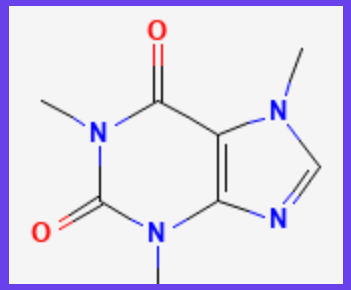
Lomello, Bard, Maglione, Schuster. Comput Struc Biotech J (2024)
PEPR DIADEM

2. How the Story Begins



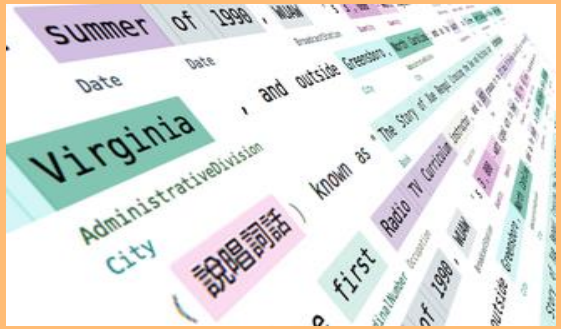
AI in Molecular Chemistry

SMILE Format

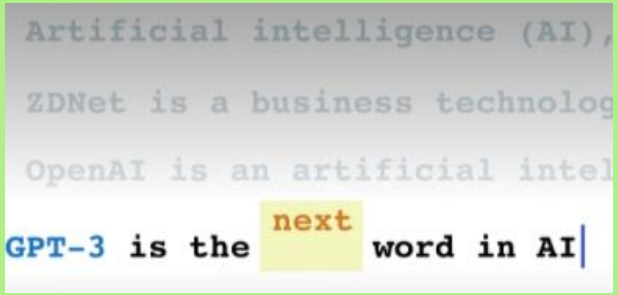


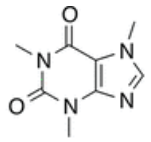
CN1C=NC2=C1C(=O)N(C(=O)N2C)C

Natural Language Processing



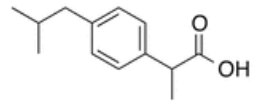
Generative Algorithms, *eg* Transformer





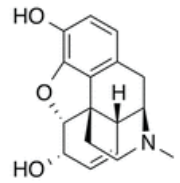
Caffeine

CN1c2ncn(C)c2C(=O)N(C)C1=O



Ibuprofen

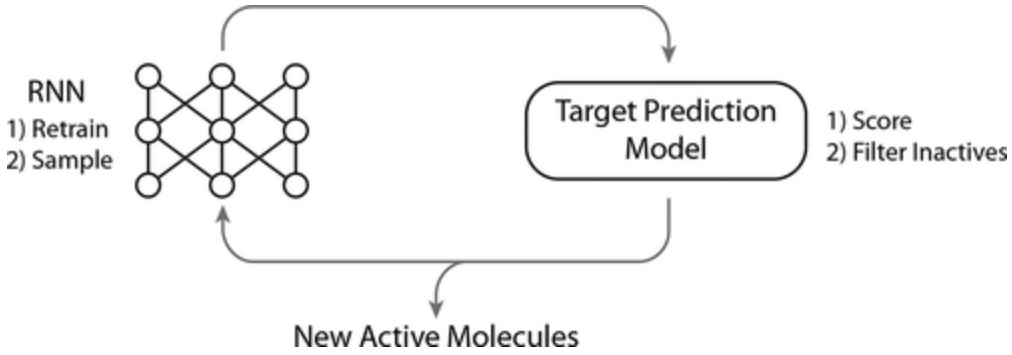
CC(C)Cc1ccc(cc1)C(C)C(O)=O



Morphine

[H][C@]12C=C[C@H](O)[C@@H]3Oc4c5c(C[C@H]1N(C)CC[C@@]235)ccc4O

Batch	Generated Example	valid
0	Oc.BK5i%ur+7oAFc7L3T=F8B5e=n)CS6RCTAR((OVCp1CApb)	no
1000	OF=CCC2OCCCC)C2)C1CNC2CCCCCCCCCCCCCCCCCCCCCCC	no
2000	O=C(N)C(=O)N(c1occc1OC)c2ccccc2OC	yes
3000	O=C1C=2N(c3cc(ccc3OC2CCC1)CCCc4cn(c5c(Cl)cccc54)C)C	yes



Segler *et al*. ACS Central Science (2018)
Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks

AI in Inorganic Materials

N. Nosenko. Nature (2016)
The material code

Table 3 | Publicly accessible structure and property databases for molecules and solids

Name	Description	URL
Computed structures and properties		
AFLOWLIB	Structure and property repository from high-throughput ab initio calculations of inorganic materials	http://aflowlib.org
Computational Materials Repository	Infrastructure to enable collection, storage, retrieval and analysis of data from electronic-structure codes	https://cmr.fysik.dtu.dk
GDB	Databases of hypothetical small organic molecules	http://gdb.unibe.ch/downloads
Harvard Clean Energy Project	Computed properties of candidate organic solar absorber materials	https://cepdb.molecularspace.org
Materials Project	Computed properties of known and hypothetical materials carried out using a standard calculation scheme	https://materialsproject.org
NOMAD	Input and output files from calculations using a wide variety of electronic-structure codes	https://nomad-repository.eu
Open Quantum Materials Database	Computed properties of mostly hypothetical structures carried out using a standard calculation scheme	http://oqmd.org
NREL Materials Database	Computed properties of materials for renewable-energy applications	https://materials.nrel.gov
TEDesignLab	Experimental and computed properties to aid the design of new thermo-electric materials	http://tedesignlab.org
ZINC	Commercially available organic molecules in 2D and 3D formats	https://zinc15.docking.org
Experimental structures and properties		
ChEMBL	Bioactive molecules with drug-like properties	https://www.ebi.ac.uk/chembl
ChemSpider	Royal Society of Chemistry's structure database, featuring calculated and experimental properties from a range of sources	https://chemspider.com
Citrination	Computed and experimental properties of materials	https://citrination.com
Crystallography Open Database	Structures of organic, inorganic, metal–organic compounds and minerals	http://crystallography.net
CSD	Repository for small-molecule organic and metal–organic crystal structures	https://www.ccdc.cam.ac.uk
ICSD	Inorganic Crystal Structure Database	https://icsd.fiz-karlsruhe.de
MatNavi	Multiple databases targeting properties such as superconductivity and thermal conductance	http://mits.nims.go.jp
MatWeb	Datasheets for various engineering materials, including thermoplastics, semi-conductors and fibres	http://matweb.com
NIST Chemistry WebBook	High-accuracy gas-phase thermochemistry and spectroscopic data	https://webbook.nist.gov/chemistry
NIST Materials Data Repository	Repository to upload materials data associated with specific publications	https://materialsdata.nist.gov
PubChem	Biological activities of small molecules	https://pubchem.ncbi.nlm.nih.gov

AFLOW Automatic-FLOW for Materials Discovery

M P

OPTIMADE Open Databases Integration for Materials Design

OQMD The Open Quantum Materials Database

NOMAD

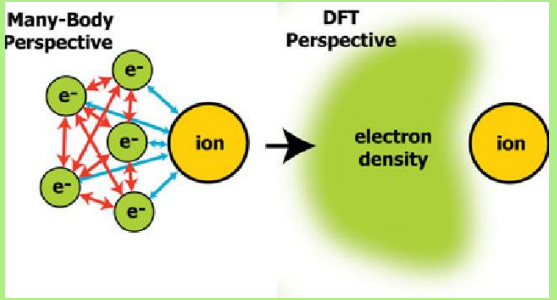
ICSD

MatNavi

DICE

AtomWork Adv.

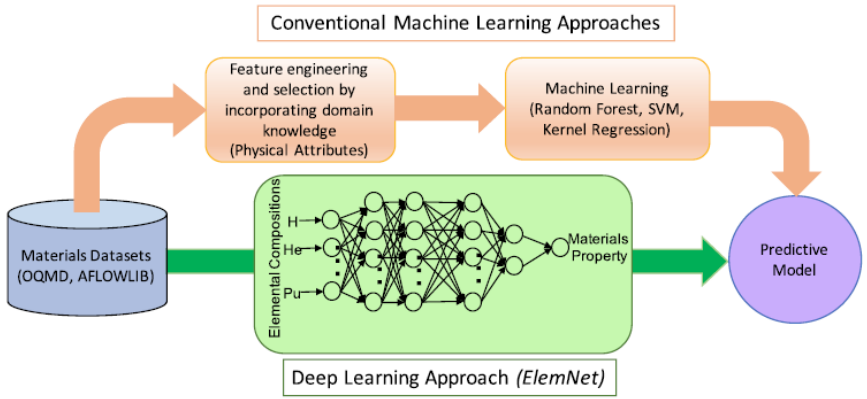
Quantum mechanics calculation



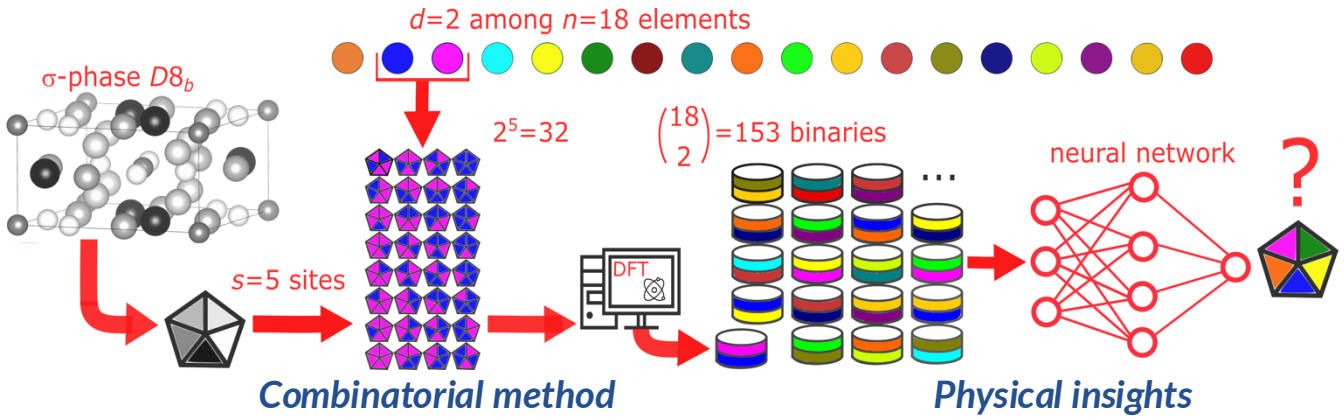
High Performance Computing

Big Data, *eg* Materials Genome Initiative

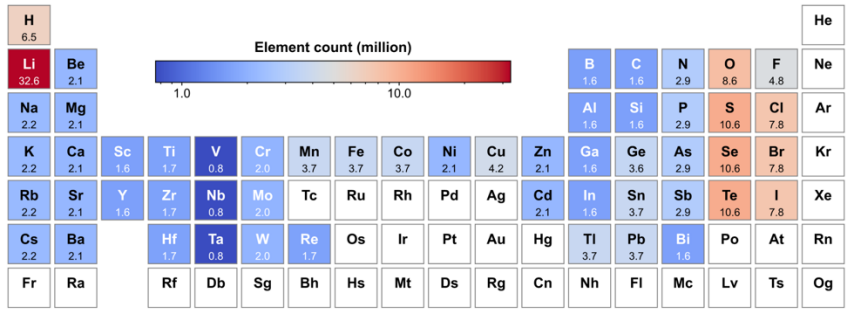
High-Throughput Screening & AI-accelerated Simulation



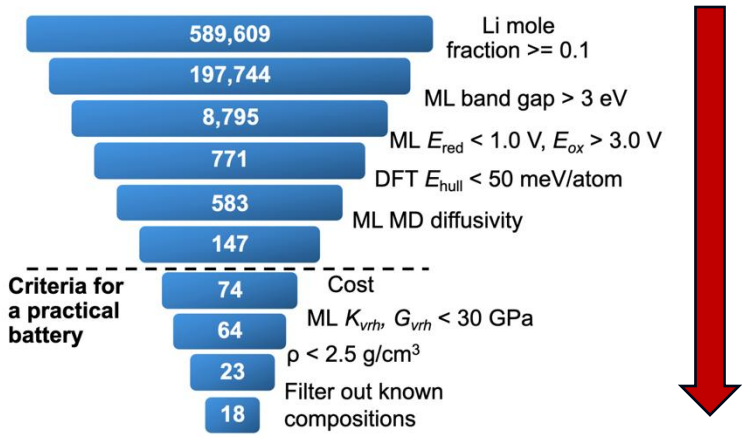
Jha, Wolverton, Agrawal *et al*. Scientific Report (2018)
ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition



Crivello *et al*. Computational Materials Science (2022)
Supervised deep learning prediction of the formation enthalpy of complex phases

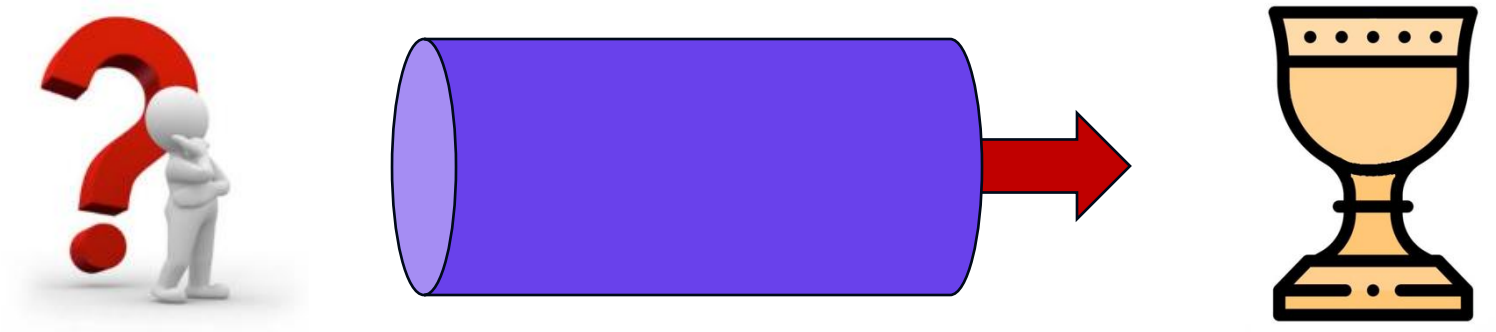


Solid electrolytes application



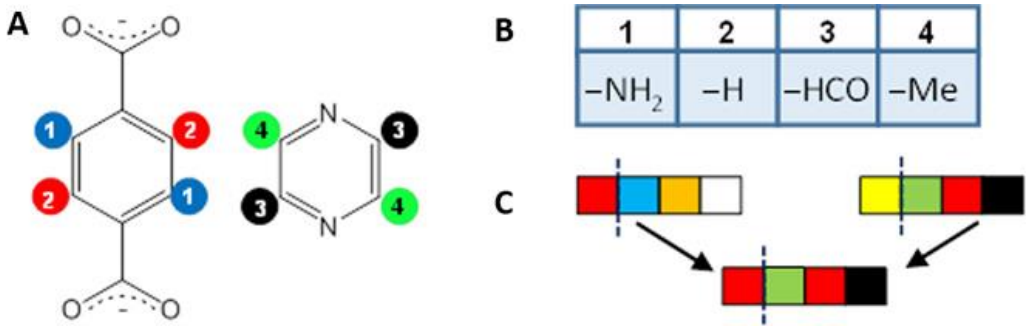
Chen *et al*.. ArXiv 2401.04070 (2024) *Accelerating computational materials discovery with artificial intelligence and cloud high-performance computing: from large-scale screening to experimental validation*

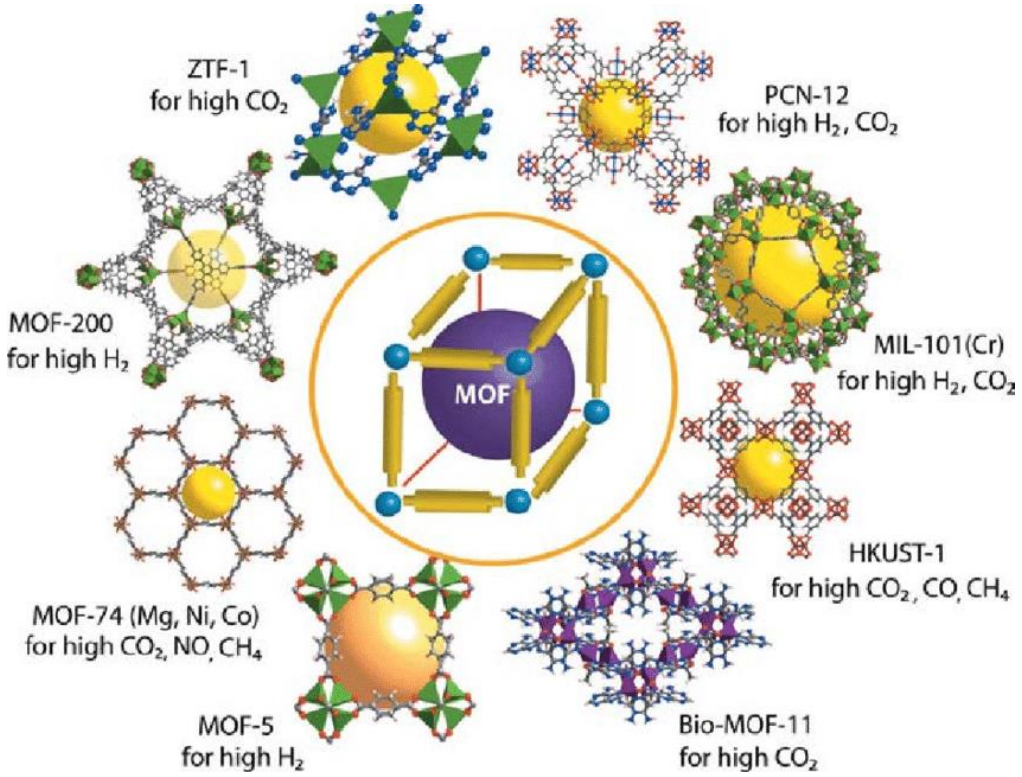
3. Generative AI in Materials Science



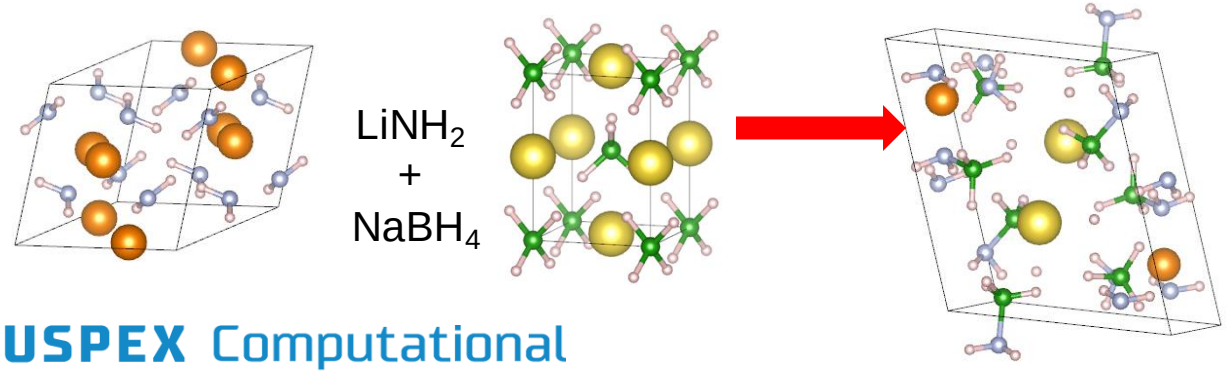
Evolutionary Algorithm for Gas Storage

Metal-organic Frameworks (MOF) for CO_2 capture





Metal-hydride (MH) for Hydrogen storage

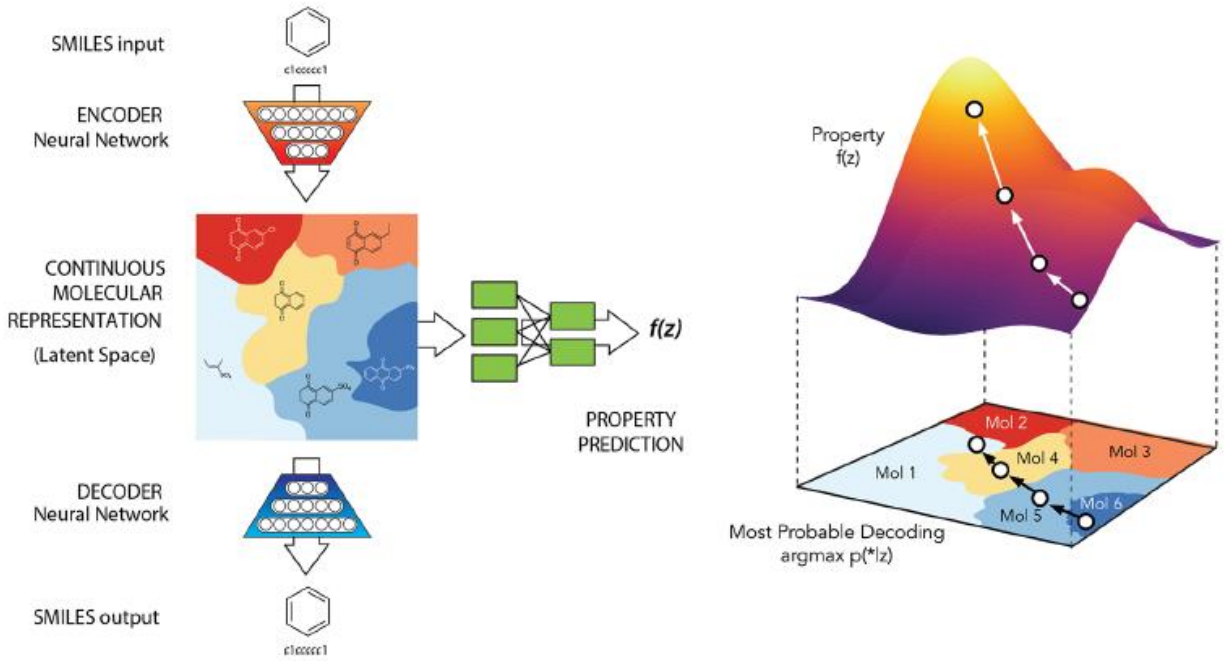


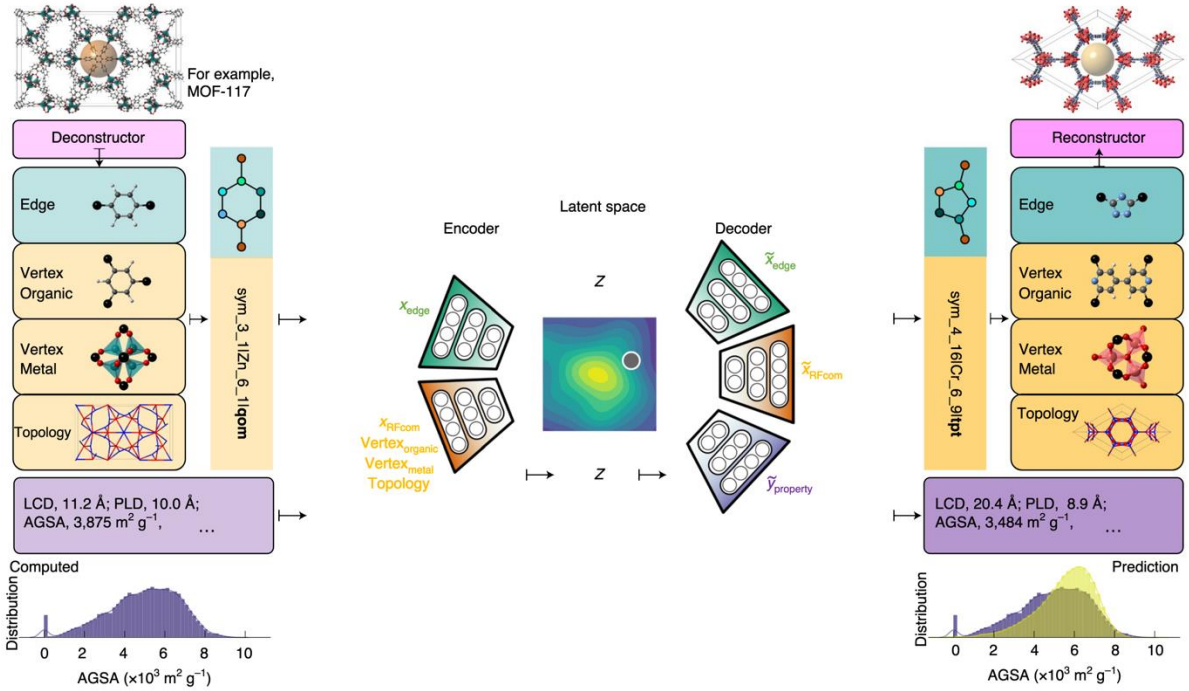
USPEX Computational Materials Discovery

Baturin *et al* ACS Applied Energy Materials (under review)
Novel Layered Calcium Borohydride Polymorph via an Enhanced Evolutionary Algorithm

Lee *et al*. Korean J Chem Eng (2013)
Collins *et al* Science Advances (2016)
Materials design by evolutionary optimization of functional groups in metal-organic frameworks

Inverse Design by Variational Auto-Encoders (VAE)



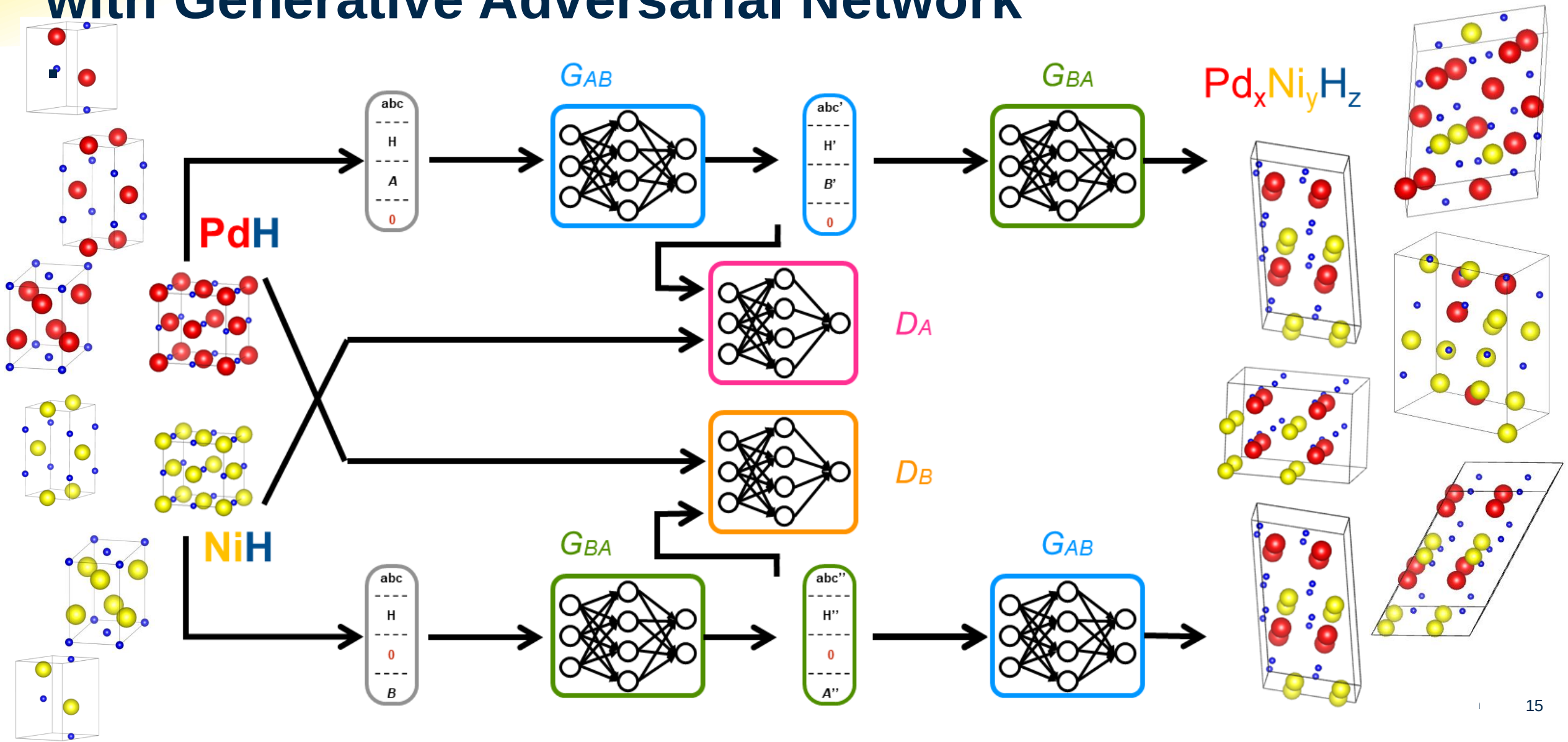


Gomez-Bombarelli *et al*. ACS central science (2018)
Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules

Yao *et al*. Nature Machine Intelligence (2021)
Inverse design of nanoporous crystalline reticular materials with deep generative models

Discovering new crystal structures with Generative Adversarial Network

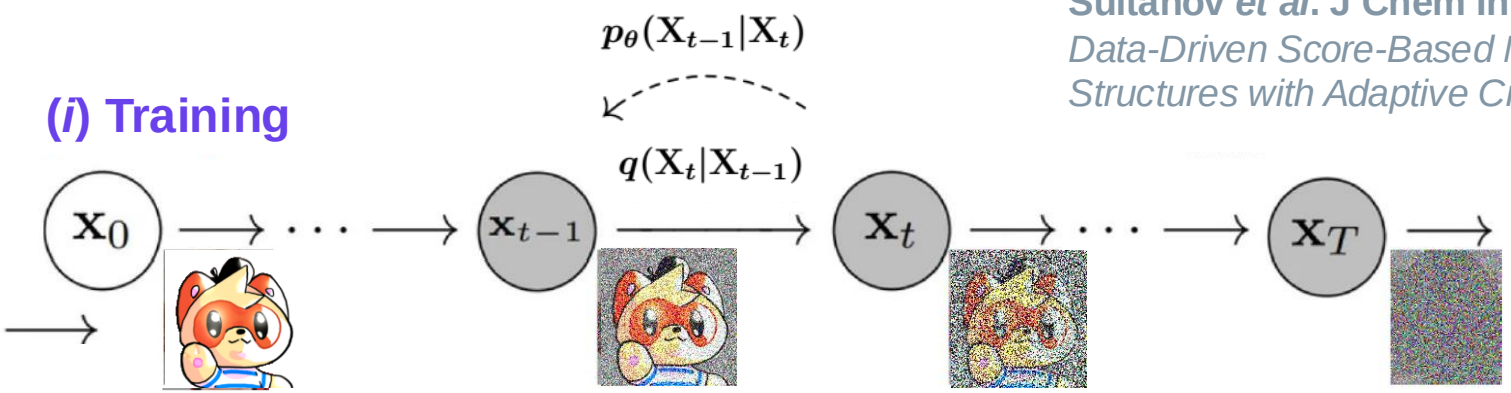
Nouira *et al*. AAAI-Make, Stanford (2019)
CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks



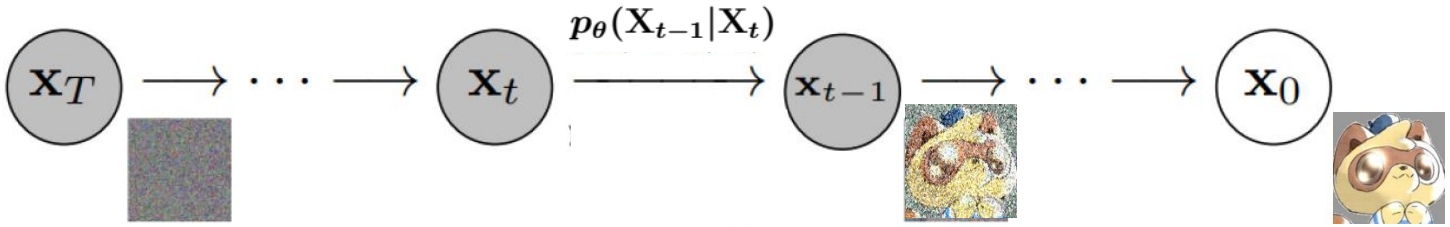
Diffusion Models Applied to Crystal Structures

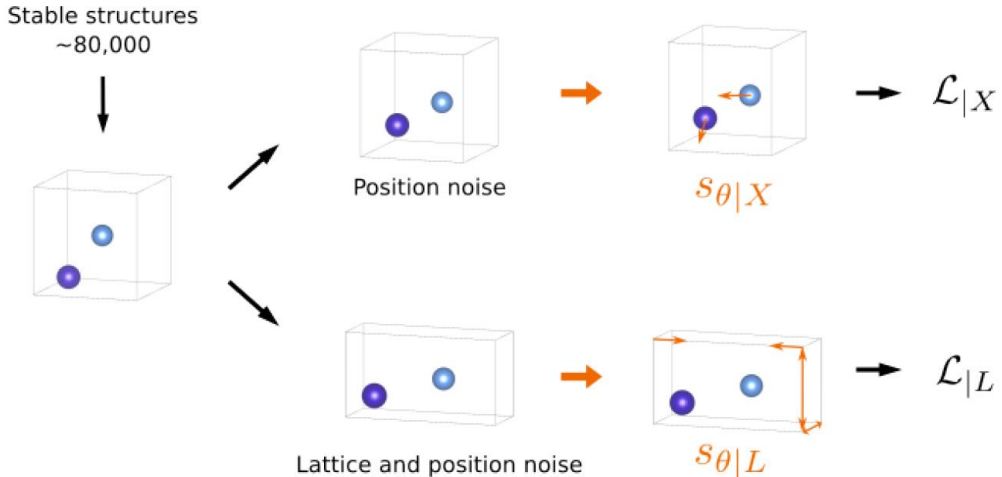
Sultanov *et al*. J Chem Inf Model (2023)
Data-Driven Score-Based Models for Generating Stable Structures with Adaptive Crystal Cells

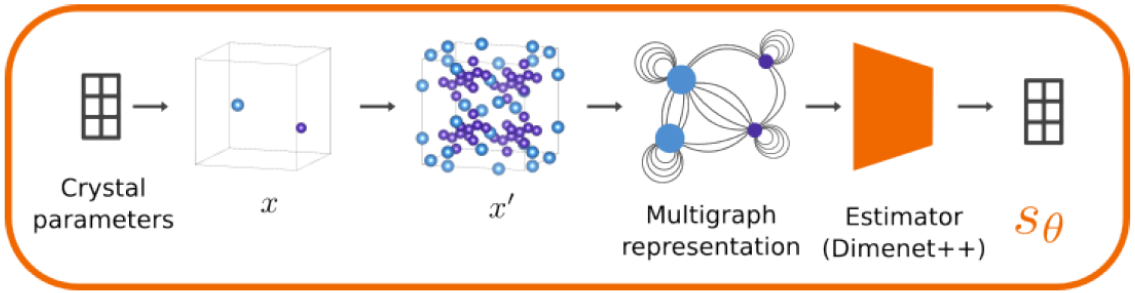
(*i*) Training



(*ii*) Sampling







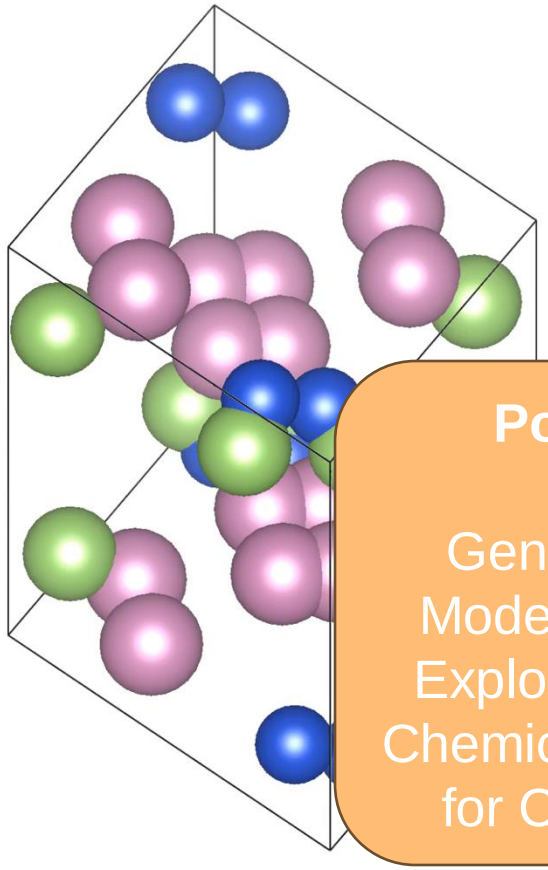
Diffusion Model to Find Stable Crystal Structures

Cu

In

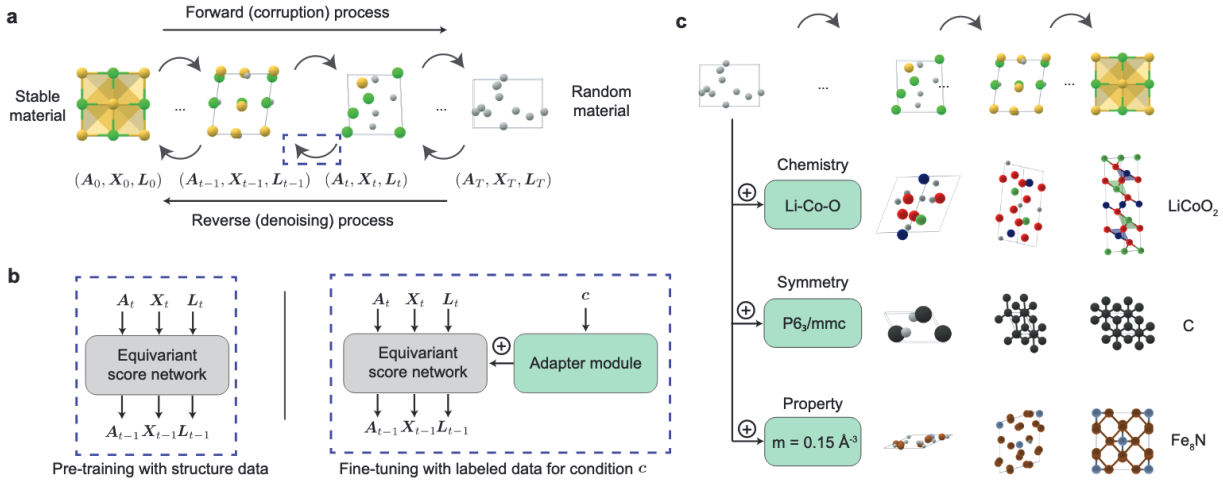
Ga

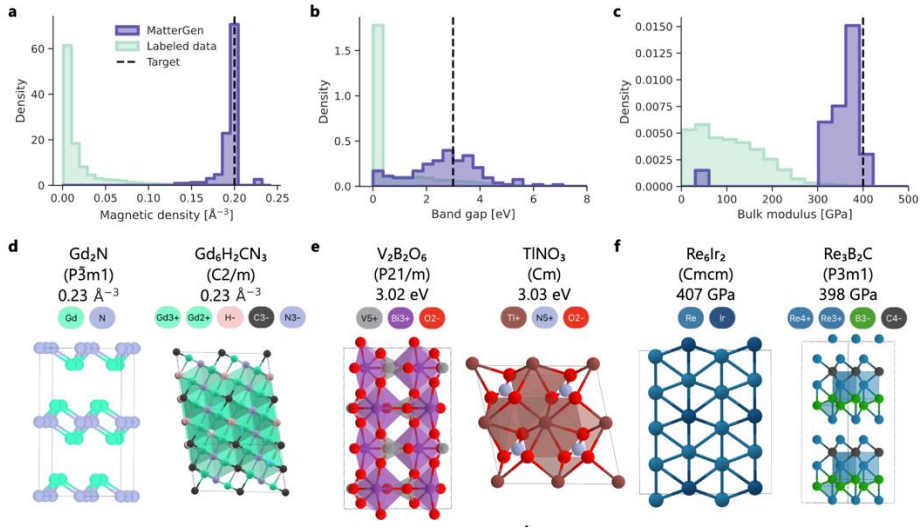
Space group 139 *I4/mmm* ?



Poster

Generative Model-Based Exploration of Chemical Space for Crystals

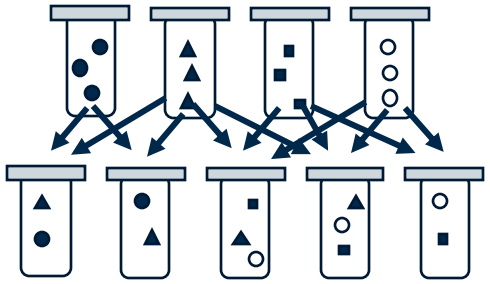




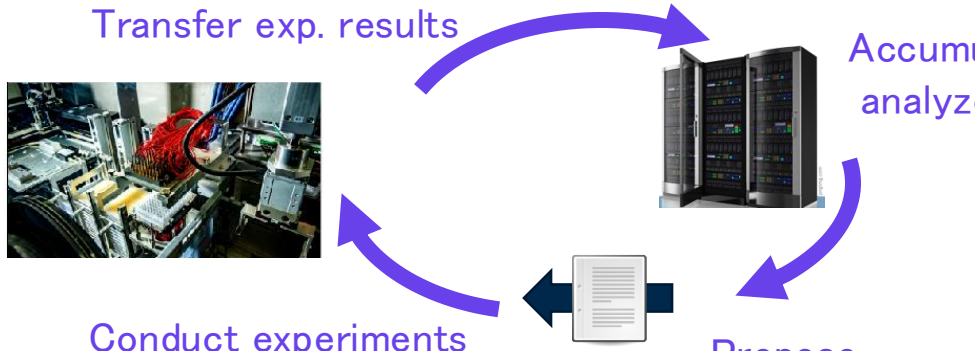
Sultanov *et al*. J Chem Inf Model (2023)
Data-Driven Score-Based Models for Generating Stable Structures with Adaptive Crystal Cells

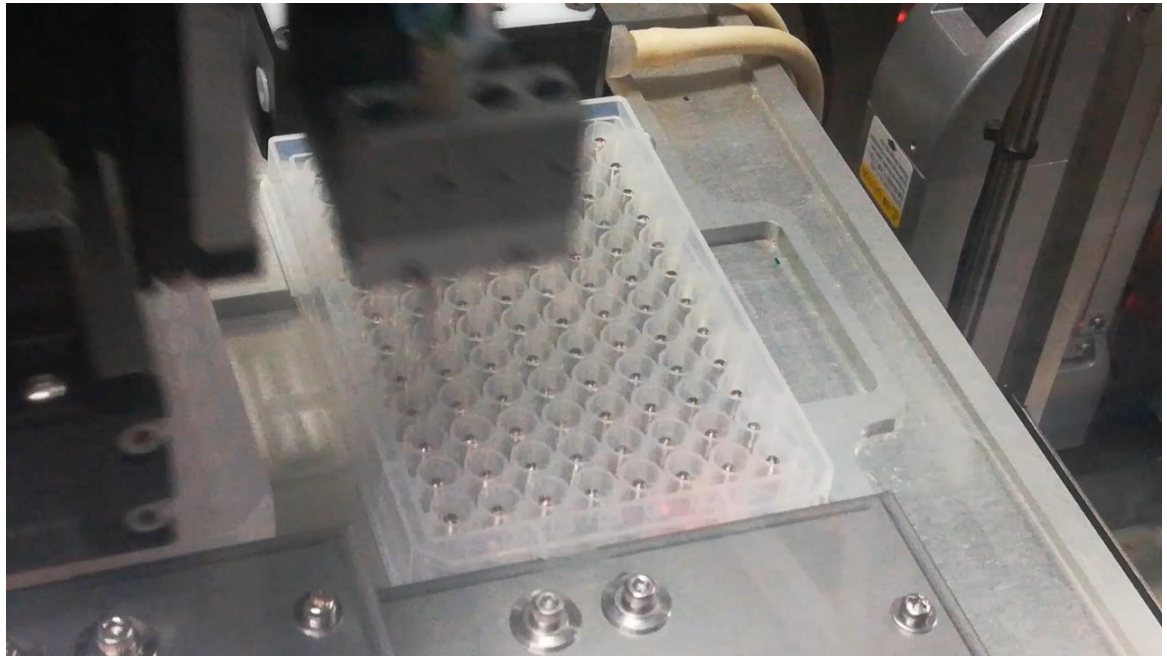
Zeni, Xie *et al*. Microsoft Research Forum (2024)
MatterGen: A Generative Model for Materials Design

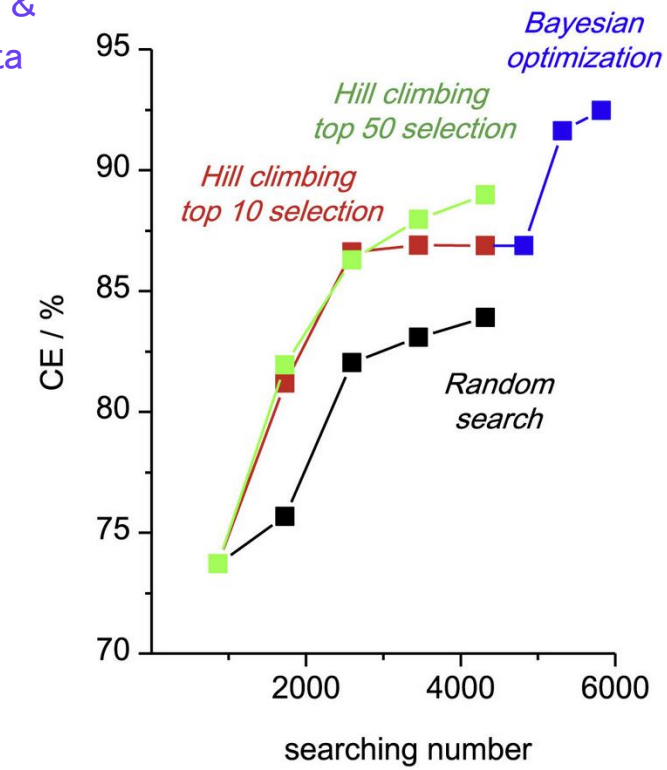
Autonomous experiment with high-throughput, automated robotic experiments and AI for electrochemistry



Combinatorial synthesis of electrolytes







S. Matsuda *et al*. Scientific Reports (2019)
Cell Reports Phys. Sci. (2022)
High-throughput combinatorial screening of multi-component electrolyte additives to improve the performance of Li metal secondary batteries

Large Language Models in Materials Science

Support for bibliography review

what are the key materials that require a real challenge to improve fuel cells?

Sources

[PDF] Grand Challenges in Fuel cell Technology towards Resource ...
resrecov · 1

The Technical and Industrial Challenges Faced by Fuel Cells—A ...
drpress · 2

Recent Advances on PEM Fuel Cells: From Key Materials to ...
link.springer · 3

View 5 more

Perplexity

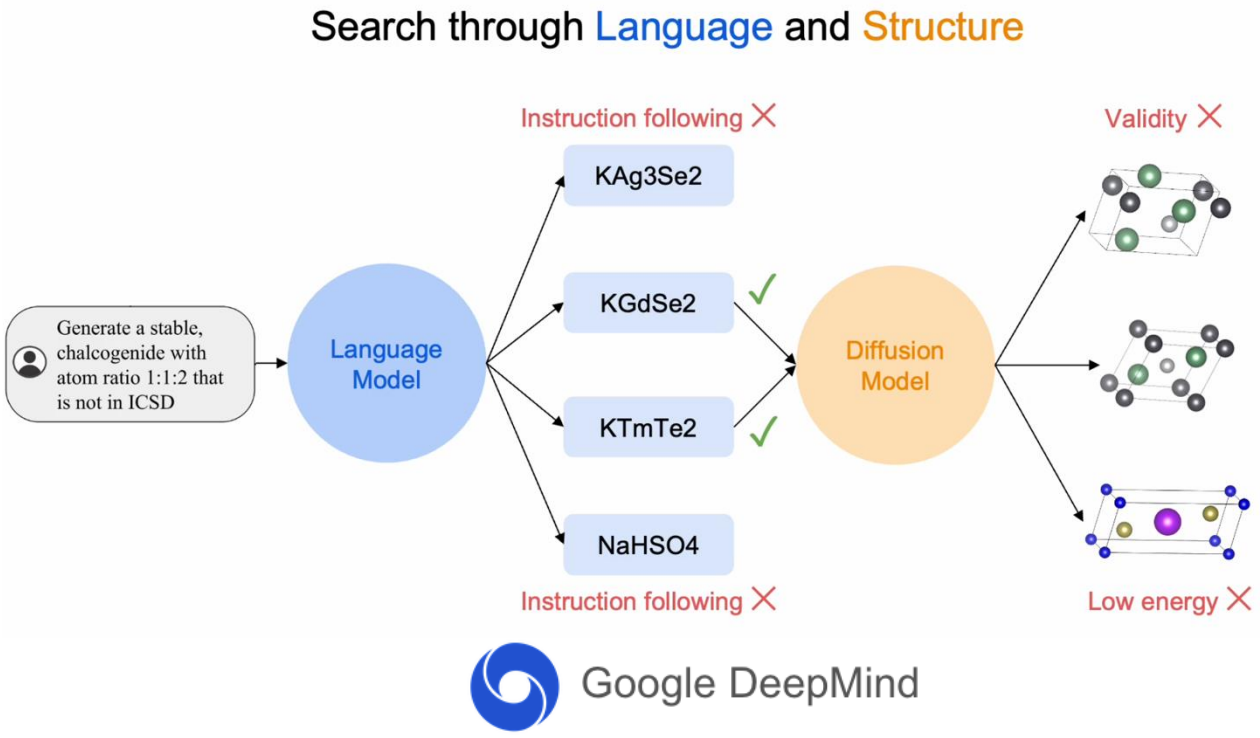
Based on the search results, several key materials present significant challenges for improving fuel cell performance and commercialization:

Proton Exchange Membranes (PEMs)

PEMs are critical components that face several challenges:

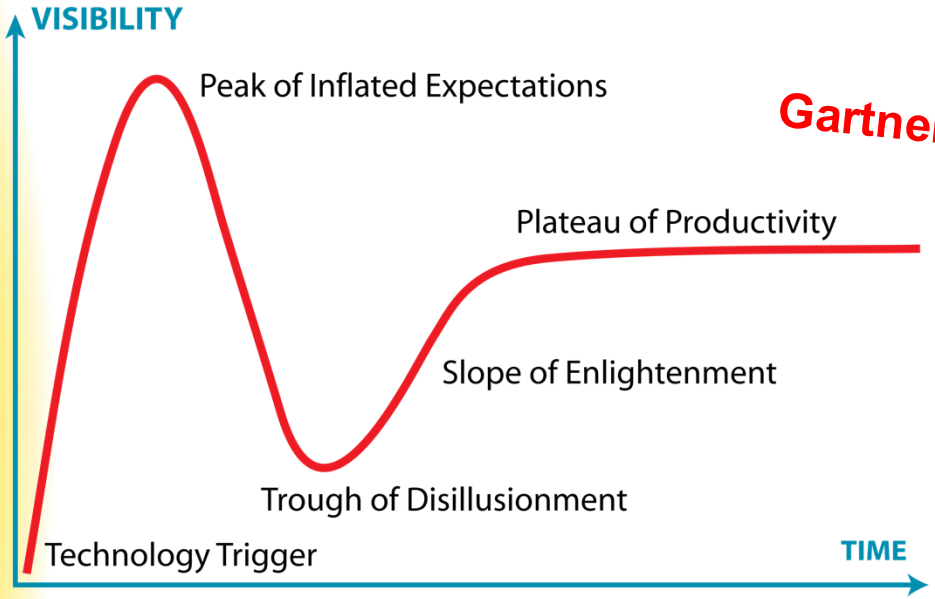
- Achieving high proton conductivity while maintaining good chemical and thermal stability [4].
- Improving mechanical performance to prevent morphological changes during wet-dry cycling [4].
- Reducing cost and environmental impact of production, as current PFSA membranes are expensive and polluting to manufacture [4].

GenMS: Generative Hierarchical Materials Search



Yang, Batzner *et al*. arXiv:2409.06762 (2024)
GenMS: Generative Hierarchical Materials Search

4. Limitations of AI approaches



Gartner hype cycle of Gen-IA?

- **Verification Required**
- **Loss of Expertise**
- **Ethics and Intellectual Property**
- **Open Data is Not Always Open**

Round Table Discussion

AI for science discoveries

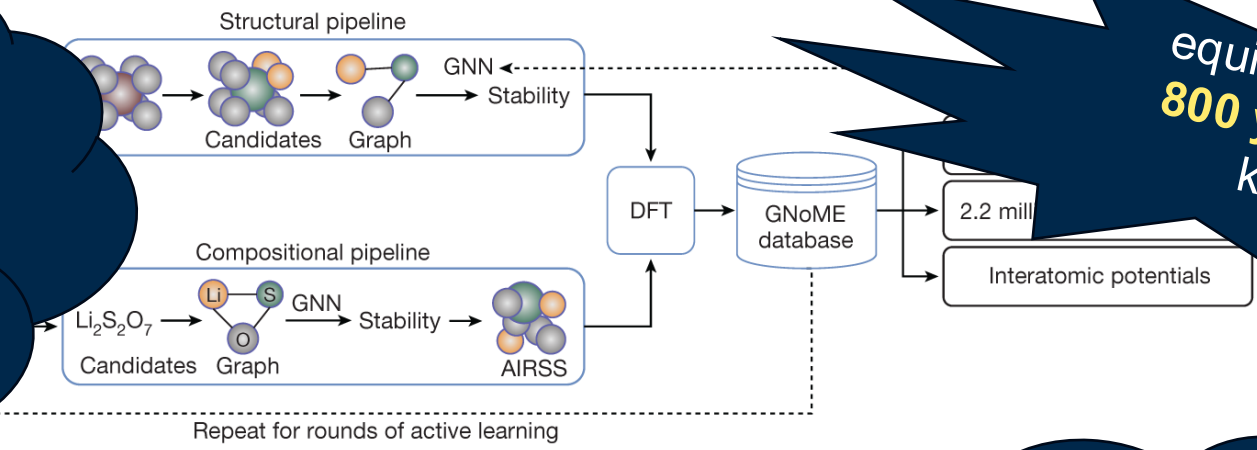
J.-C. Crivello & F. d'Alché-Buc

A New Era in Material Discovery?

Google DeepMind

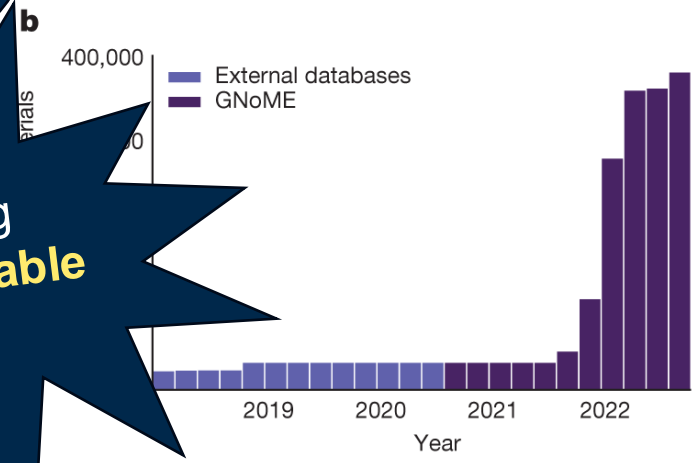
GNoME for *Graph Networks for Materials Exploration*

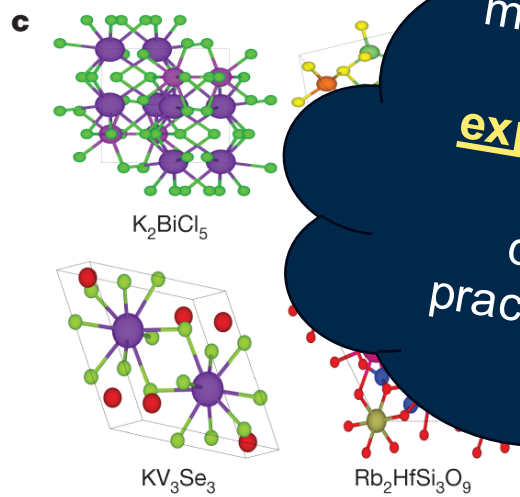
The team's breakthrough in predicting over **2.2 million new materials**



equivalent to nearly **800 years'** worth of knowledge

including **380,000 stable** ones





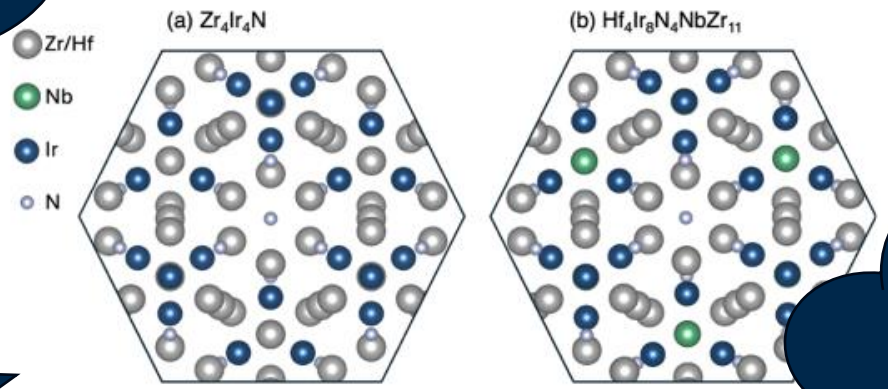
many of these materials have been **experimentally created and validated**, demonstrating the practical effectiveness of this method.

Merchant *et al*. Nature (2023)
Scaling deep learning for materials discovery

Artificial Intelligence Driving Materials Discovery?

"Not achieved by design or serediplity, but by exploring opportunities in the extensive repositiory of compounds that are already known"

Trivial adaptations of know compounds with higher symmetry



"Only periodic perfect compounds excluding all king of real materials: disordered phases, glass, polymers, composites, …"

including **380,000 stable**

"2000 compounds have been placed in the Gnome archive"

many of these materials have been **experimentally created and**

"Not checked by real synthetize but by quantum calculation (inclusion of radioactive elements) No functionality demonstrated"

Cheetham, Seshadri. Chem Matter (2024)
Perspective on the Article: Scaling Deep Learning for Ma Discovery

Conclusion remarks

Generative Approaches in Material Sciences:

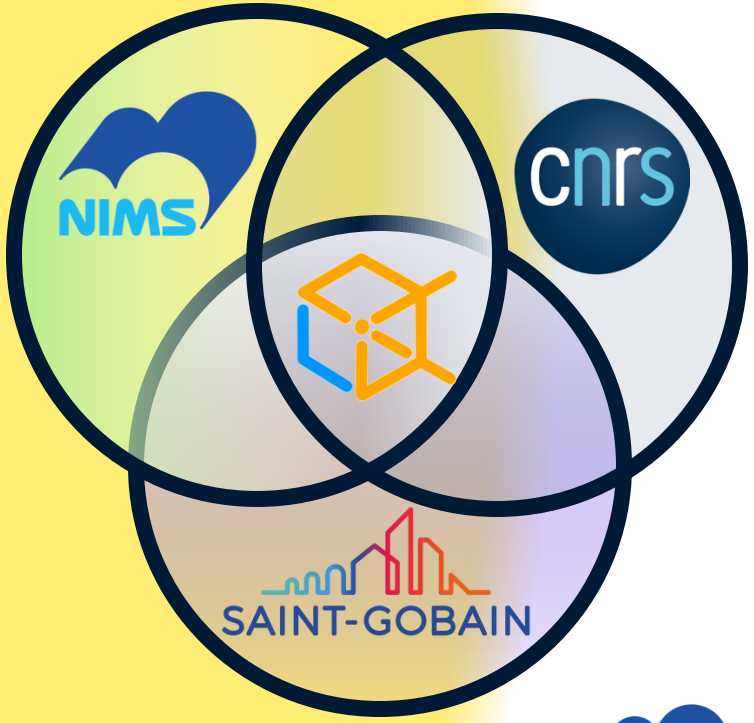
→ Accelerating Discovery and Innovation

Interdisciplinarity: Bridging Scientific Communities:

→ **Not Just Data Science:** Successful material discovery through AI requires the integration of domain expertise in crystallography, materials synthesis, and chemistry.

→ **Collaborative Approach:** AI experts, data scientists, and materials scientists must work together to interpret AI predictions, validate findings, and guide the iterative development of new materials.

Acknowledgements




Guillaume Lambard


Vladimir Baturin


J.-M. Joubert


A. Nouira


A. Sultanov


T. Rebafka


N. Sokolovska