AI for accelerated materials science

Francisco Martin-Martinez

Senior lecturer, King's College London

martinmartinezlab.com

CCAI VIRTUAL SUMMER SCHOOL 2026 · SECTORAL LECTURE · 13th AUGUST 2026

WHO AM I?

Who am I, and why this matters to me



A scientist using AI as a tool vs. AI doing science autonomously, where is the line?

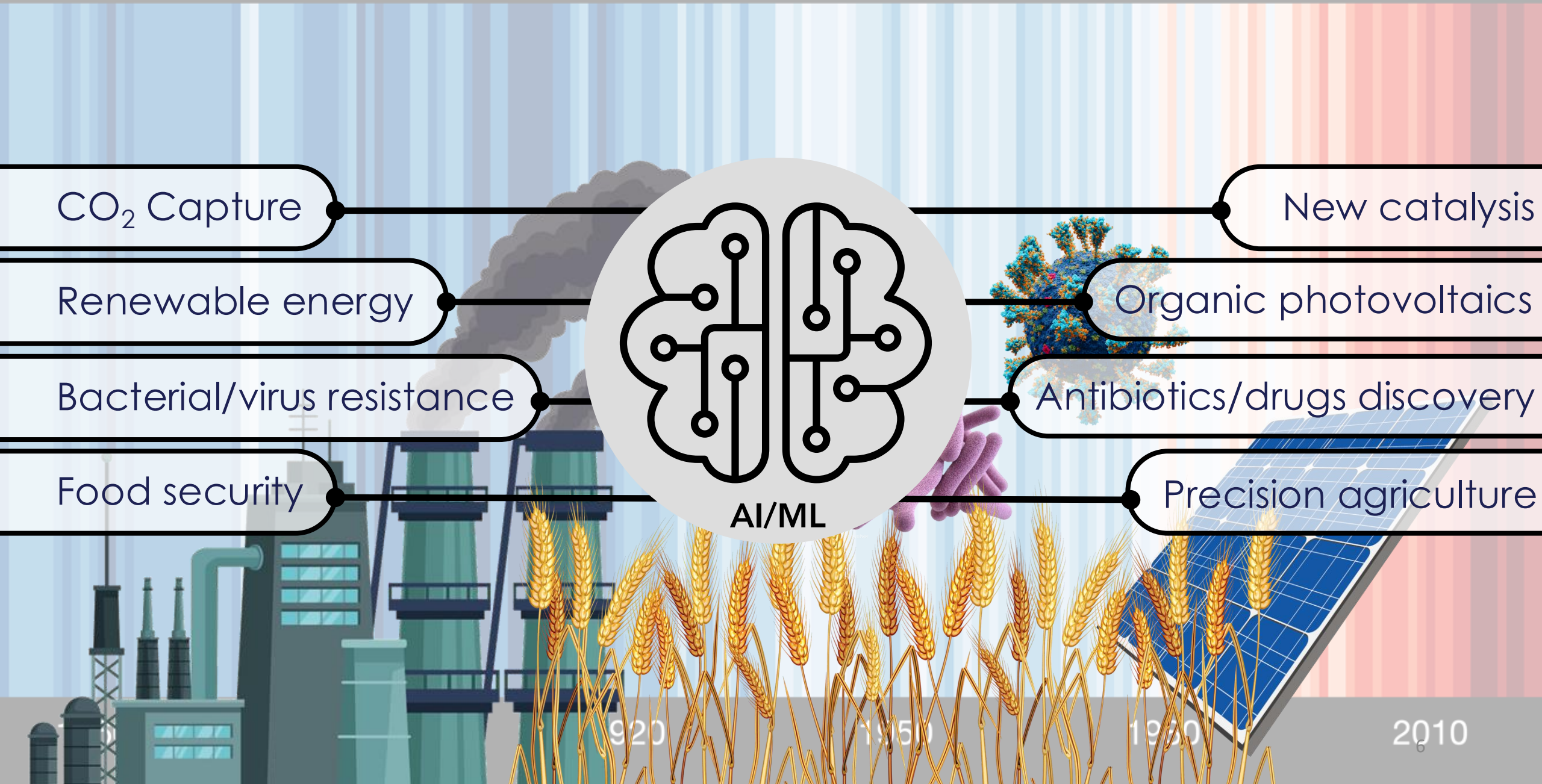
WHY THIS MATTERS TO ME?

Global temperature change (1850-2022)



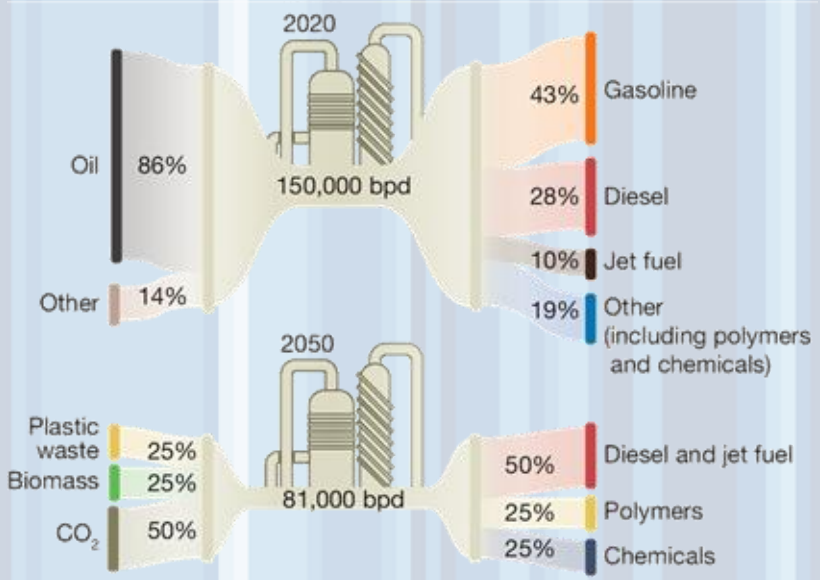
1860 1890 1920 1950 1980 2010

Global temperature change (1850-2022)



New mining of materials from waste

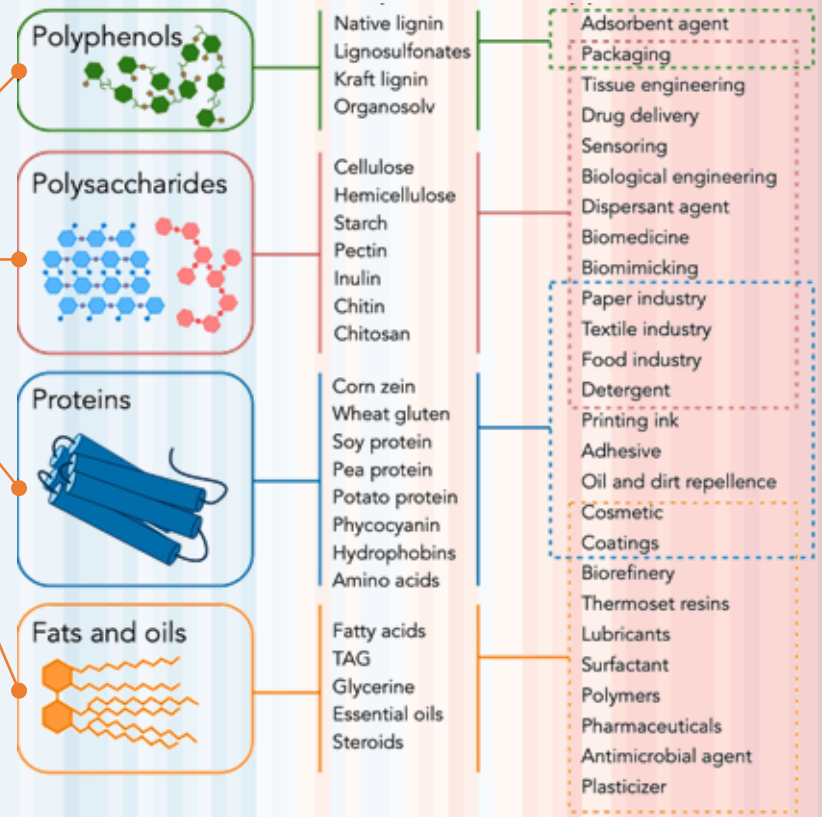
(Bio)refinery 2050



Nature 629, 295, **2024**

Biomass waste
140 Gt/yr

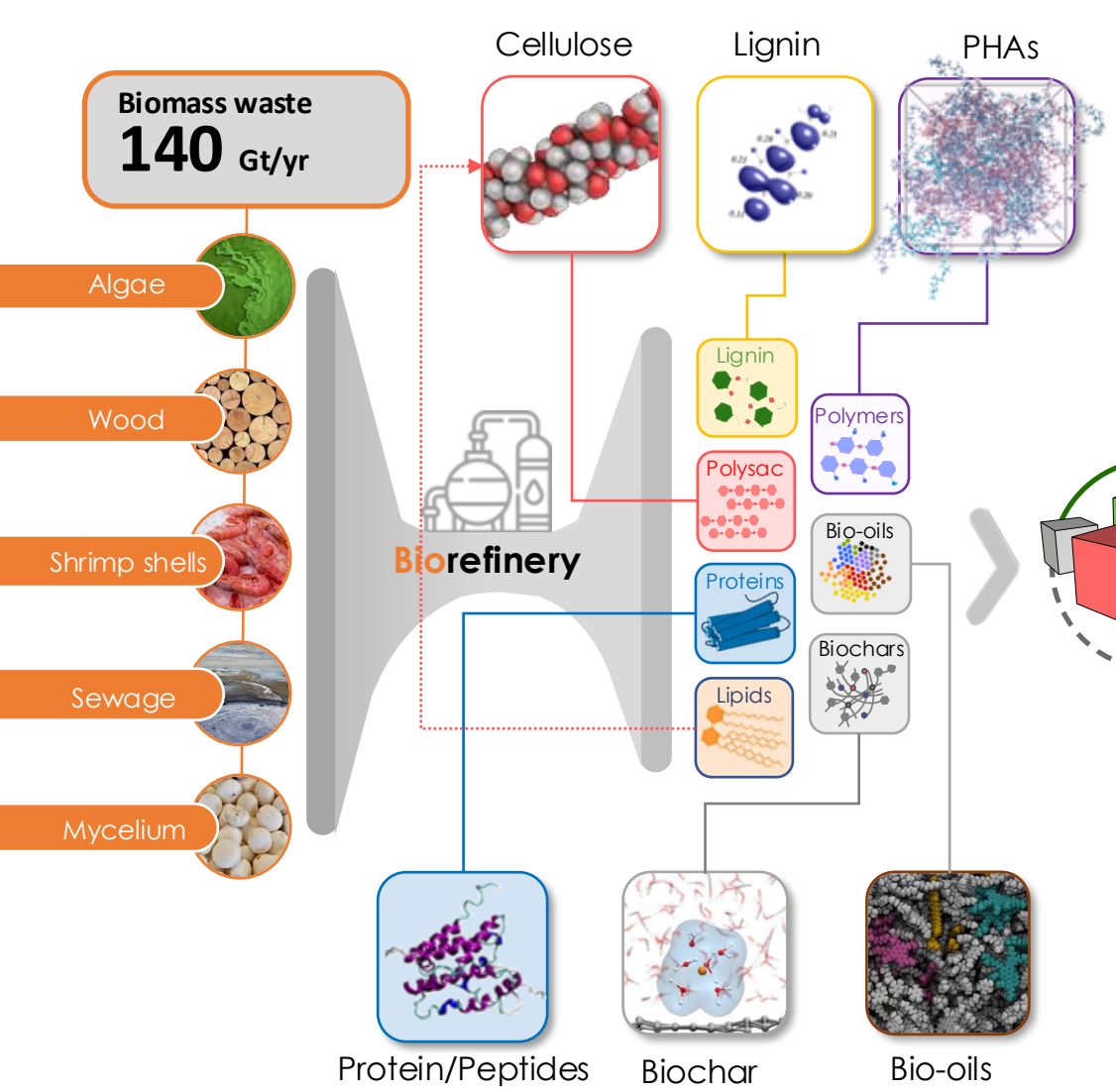
AI/ML

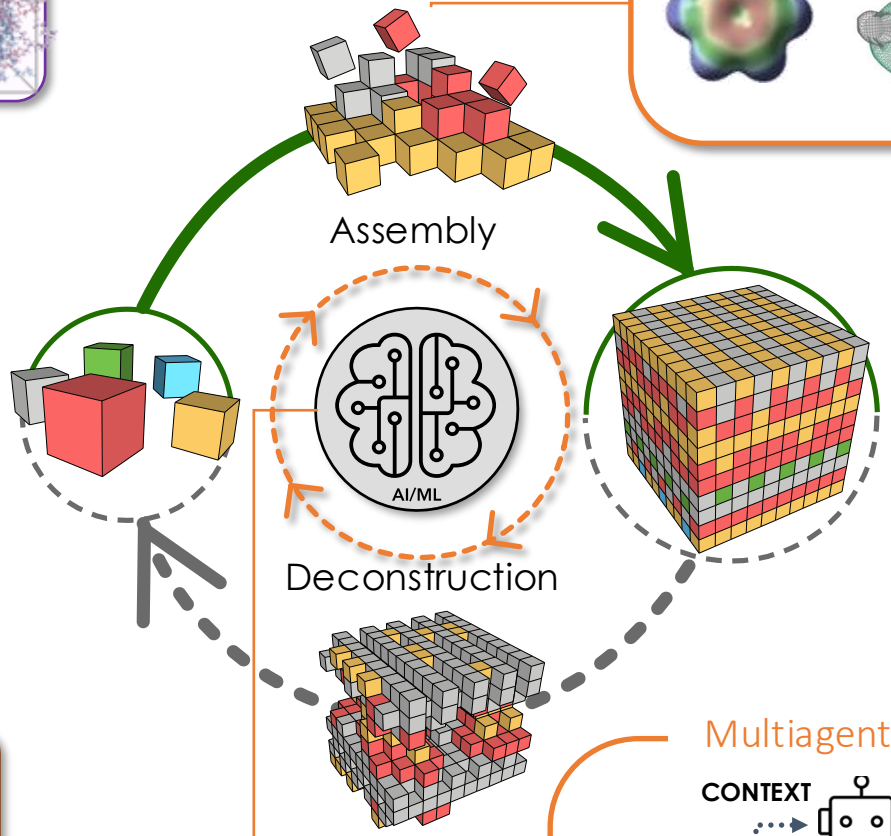


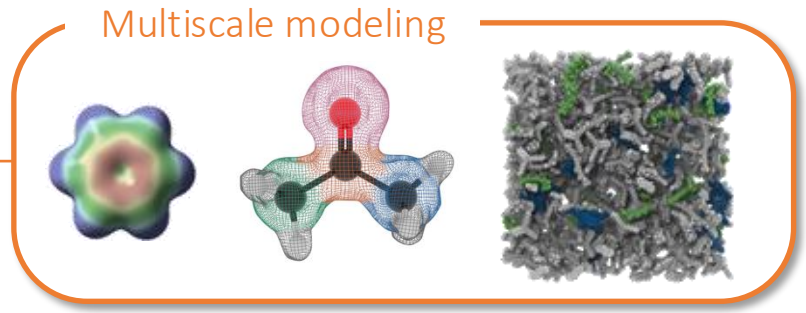
Bachs-Herrera, …, Martin-Martinez **iScience** 26, 106549, **2023**

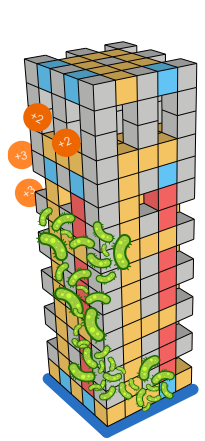
Waste valorization into biobased molecules and materials

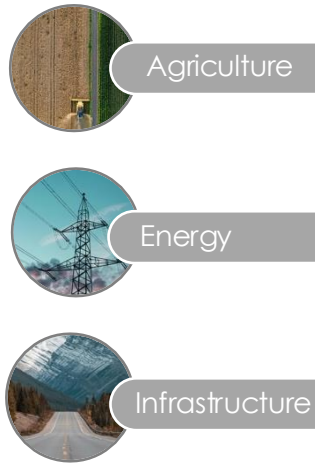
Biobased circular materials ➡ Towards an AI-driven quantum-informed biorefinery

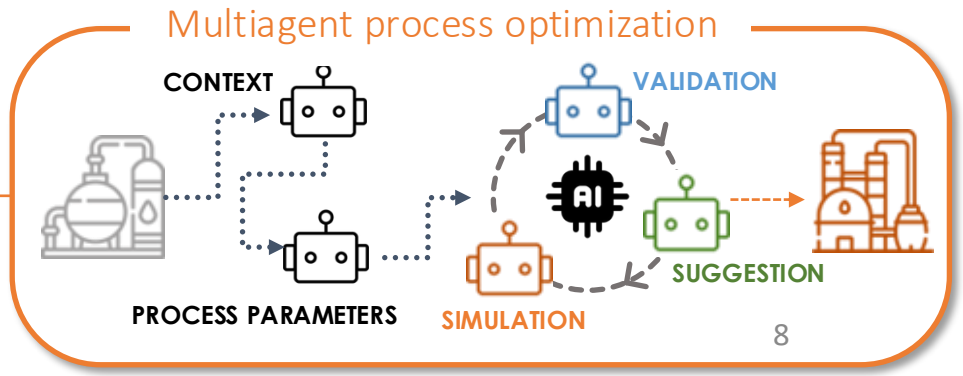












Bachs-Herrera et al., Martin-Martinez **iScience** 26, 106549, **2023**

WHAT WILL WE COVER TODAY?

WHAT WE'LL COVER

01 Why materials are a climate bottleneck

02 The traditional (still useful) paradigm of materials modeling and simulations

03 A new paradigm in materials discovery

04 The AI tool kit

05 Talking Chemistry to computers

04 Responsible AI

05 Explainable AI in Chemistry

ENGAGEMENT

Live poll – Use the chat

Which materials challenge do you think is most urgent for climate?

No wrong answer — this warms up the room

THE MATERIALS BOTTLENECK

Where AI plugs into the materials pipeline

1

Represent

Encode molecules/crystals for ML (fingerprints,OHE graphs, text)

2

Predict

Property prediction, surrogate models, **ML interatomic potentials**

3

Generate

Inverse / generative design of new candidates

4

Synthesize

Autonomous & self-driving labs close the loop

5

Characterize

AI-assisted analysis of spectra, images, structure

Materials are a climate bottleneck

Photovoltaics

Efficiency & stability limited by materials chemistry

Batteries

Energy density, lifetime, critical-mineral dependence

Alternative fuels

Catalysts & carriers for green H_2, e-fuels

CO_2 sorbents

Capture materials that are cheap, durable, scalable

Others: sustainable infrastructure, food security, drug discovery, antibiotics …

The problem: the discovery cycle is slow and expensive.

Design → synthesize → characterize → test can take years per candidate.

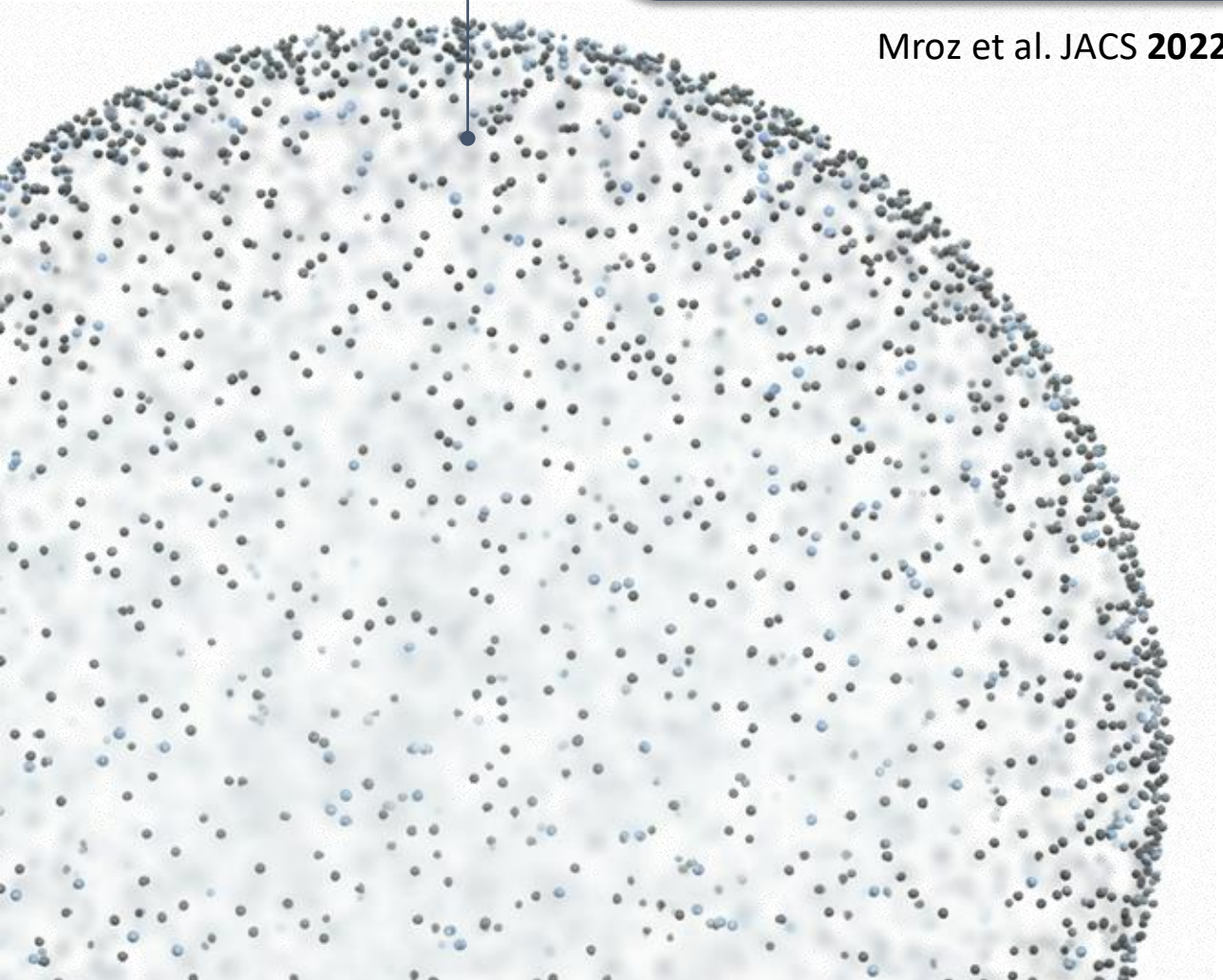
Mitigation AND adaptation both depend on getting the right materials, faster and more cheaply.

Yao et al., Nat. Rev. Mater. 2023 | Bachs-Herrera et al., iScience 2023

The materials-discovery bottleneck

10^{60} Small (drug-like) molecules for materials design

Mroz et al. JACS **2022**



Combinatorial chemical space is vast

VS

Experimental validation is slow and expensive

CURIOSITY

There are more possible small organic molecules than atoms in the observable universe.

ENGAGEMENT

Quick quiz – True/False

1. There are more possible small organic molecules than atoms in the observable universe.

   True | False

2. A new material usually reaches the market in a few weeks.

   True | **False**

3. AI plugs into the pipeline at represent, predict, generate, synthesize and characterize.

   True | False

Take-home messages

WHY IT MATTERS TO CLIMATE

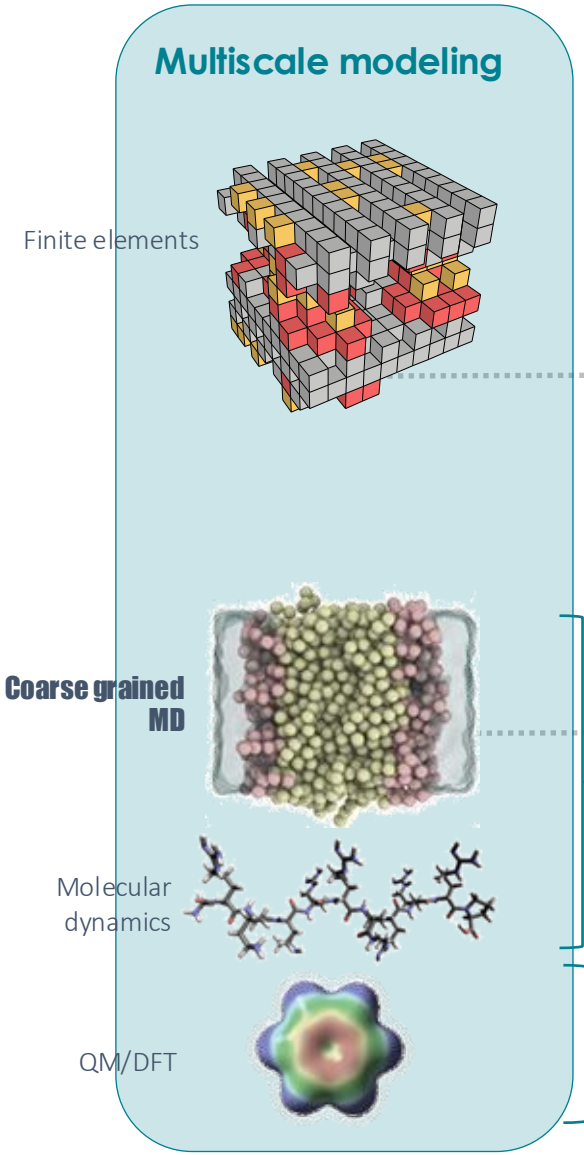
1. Materials are a climate bottleneck: mitigation and adaptation both need better materials, faster.

2. Chemical space is unsearchable by brute force. The use of AI reframes it as guided search.

3. Climate targets are on a fixed clock. Faster discovery means faster deployment.

MATERIALS MODELING AND SIMULATIONS

Modeling across scales



Molecular dynamics simulations



"All model are wrong, but some are useful."

– George E. P. Box



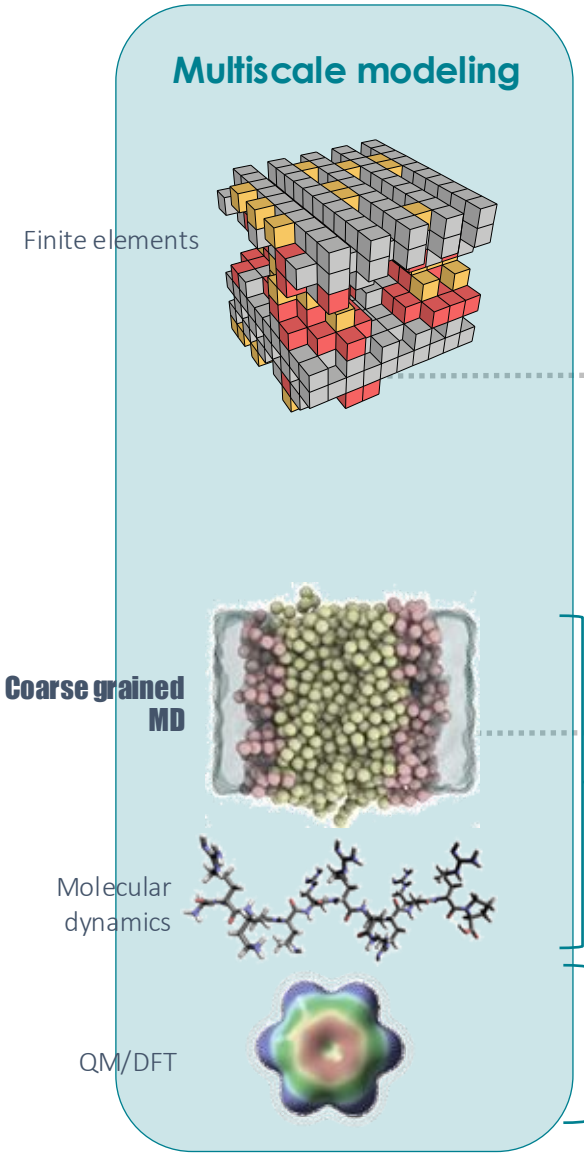
Density functional th



"I think I can safely say that nobody understands quantum mechanics."

– Richard Feynmann

Modeling across scales



Molecular dynamics simulations



"All model are wrong, but some are useful."

– George E. P. Box

Density functional theory (DFT) – Quantum mechanics

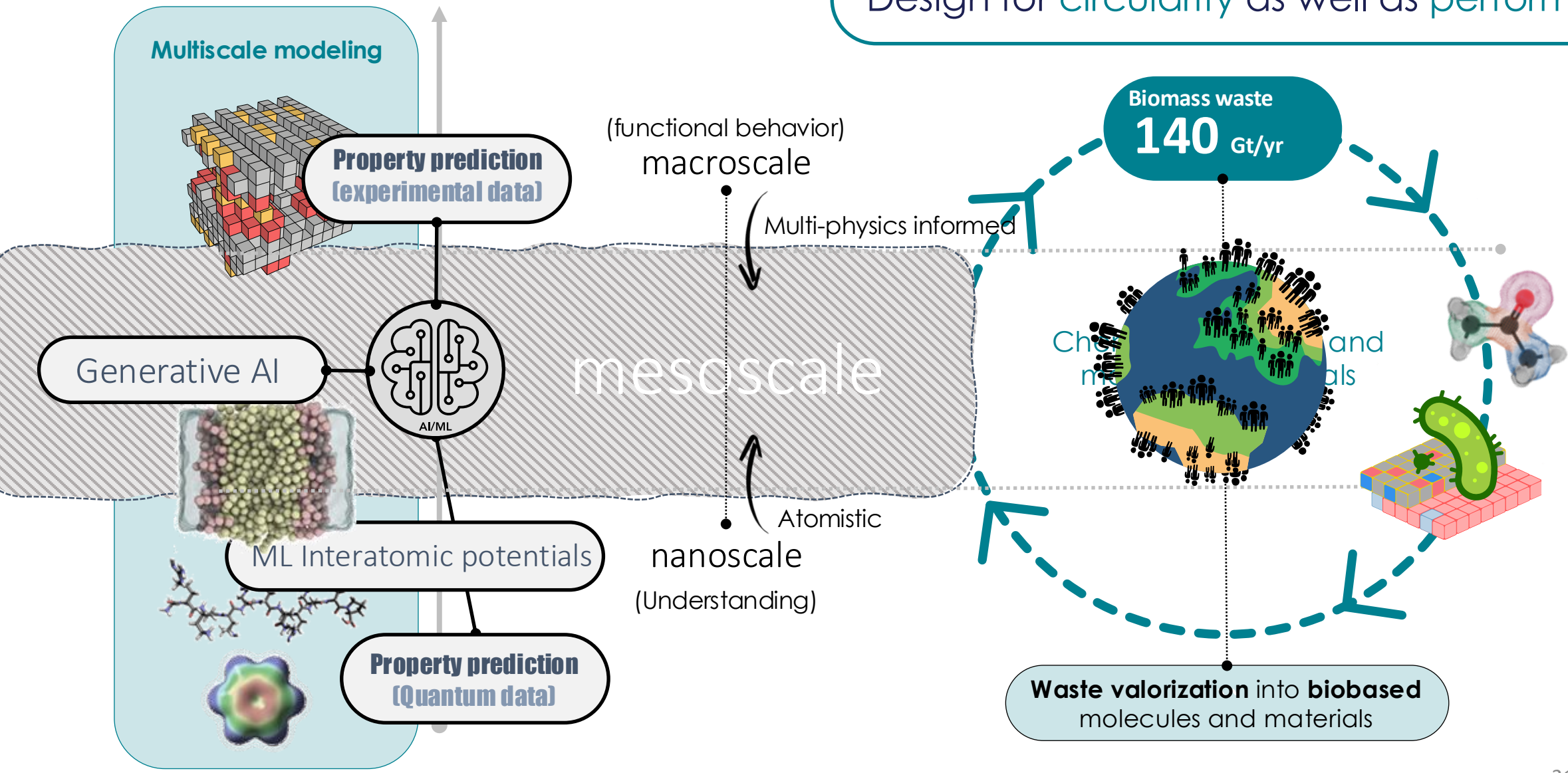


"I think I can safely say that nobody understands quantum mechanics."

– Richard Feynmann

Modeling across scales

Design for circularity as well as performance



Software for DFT, MD, and ML for Chemistry

Molecular DFT



ORCA

Solid state DFT



siesta™

MD (All atoms and coarse grained)

LAMMPS

GROMACS
FAST. FLEXIBLE. FREE.



NAMD
Scalable Molecular Dynamics



OpenMM





Integration and cheminformatics



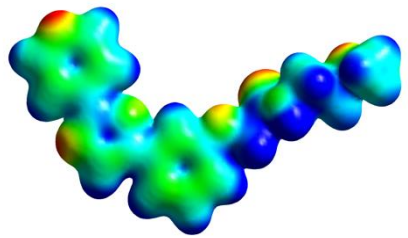


+ atomic simulation environment (ASE)

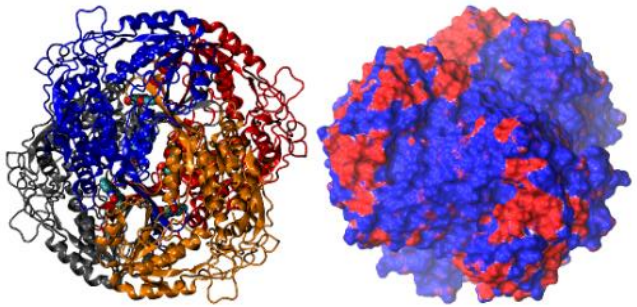
Databases to facilitate discovery in chemistry

Name	Description	Size (ca.)[a]	Availability
Chemical structures:			
ZINC[43,44]	commercially available compounds	35 M	open
ChemSpider[45]	structures and miscellaneous data	67 M	API
SureChEMBL[46]	structures from ChEMBL	1.9 M	open
Super Natural II[47]	natural product structures	325 k	open
SAVI[48]	enumerated synthetically accessible structures and their building blocks	283 M	open
Molecular properties:			
QM7/QM7b[84,85]	electronic properties (DFT)	7200	open
QM9[86]	electronic properties (DFT)	134 k	open
QM8[87]	spectra and excited-state properties	22 k	open
FreeSolv[88]	aqueous solvation energies	642	open
NIST Chemistry WebBook[89]	miscellaneous molecular properties	varies	open
Biological data:			
PubChem[36]	compounds and properties, emphasis on bioassay results	96 M	open
ChEMBL[58,59]	compounds and bioactivity measurements	1.9 M	open
ChEBI[60]	compounds and biological relevance	56 k	open
PDB[61]	biological macromolecular structures	150 k	open
PDBBind[62]	protein binding affinity	20 k	open
ProTherm[63]	thermodynamic data for proteins	10 k	open
LINCS[64]	cellular interactions and perturbations	varies	open
SKEMPI[65]	energetics of mutant protein interactions	7 k	open
xMoDEL[66]	MD trajectories of proteins	1700	open
GenBank[67]	species' nucleotide sequences	400 k	open
DrugBank[68]	drug compounds, associated chemical properties, and pharmacological information	13 k	open
BindingDB[69]	compounds and binding measurements	750 k; 1.7 M	open
CDD[70]	collaborative drug discovery database for neglected tropical diseases	>100 datasets	registration
ToxCast[71,72]	compounds and cellular responses	>4500	open
Tox21[73,74]	compounds and multiple bioassays	14 k	open

Density Functional Theory (DFT)



Molecular dynamics (MD)



Coley et al. Angew. Chem. Int. Ed. 2020, 59, 23414 – 23436

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) DFT (quantum)

D) Finite elements

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) **DFT (quantum)**

D) Finite elements

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) **DFT (quantum)**

D) Finite elements

Q2 **Why do we use multiscale modeling?**

A) One scale is enough

B) Different phenomena live at different scales

C) To save on citations

D) It is legally required

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) DFT (quantum)

D) Finite elements

Q2 **Why do we use multiscale modeling?**

A) One scale is enough

B) Different phenomena live at different scales

C) To save on citations

D) It is legally required

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) DFT (quantum)

D) Finite elements

Q2 **Why do we use multiscale modeling?**

A) One scale is enough

B) Different phenomena live at different scales

C) To save on citations

D) It is legally required

Q3 **Vs all-atom MD, coarse-grained MD reaches…**

A) Smaller systems

B) Larger systems & longer times

C) Higher quantum accuracy

D) No difference

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – multiple choice

Q1 **Which method is most accurate but most expensive?**

A) Classical MD

B) Coarse-grained MD

C) **DFT (quantum)**

D) Finite elements

Q2 **Why do we use multiscale modeling?**

A) One scale is enough

B) **Different phenomena live at different scales**

C) To save on citations

D) It is legally required

Q3 **Vs all-atom MD, coarse-grained MD reaches…**

A) Smaller systems

B) **Larger systems & longer times**

C) Higher quantum accuracy

D) No difference

MODELING & SIMULATION

Take-home messages - The cheapest place to fail is on a computer.

WHY IT MATTERS TO CLIMATE

1. No single scale suffices, and multiscale modeling is needed.

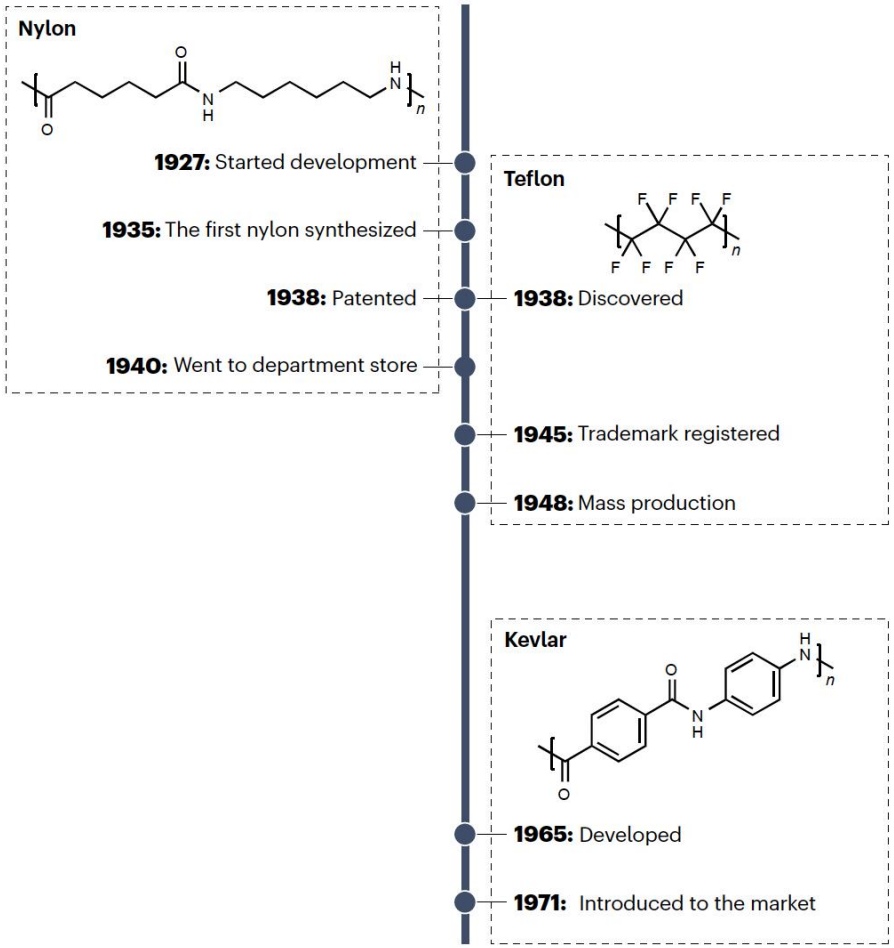
2. We must predict degradation and recyclability: circularity, not just performance.

3. Data is the substrate AI builds on. Good data makes good models.

A NEW PARADIGM IN MOLECULAR AND MATERIALS DISCOVERY

Slow materials, fast world

Polymer innovations over the past two centuries.



1833: Berzelius coins the term 'polymer'

1844: Goodyear patents the vulcanization of rubber

1846: Schonbein synthesizes nitrocellulose

1856: Parkes patents the first thermoplastic

1892: Cross, Bevan and Beadle patent 'viscose'

1907: Baekeland creates the first synthetic plastic

1912: Klatte invents the production process of PVC

1920: Staudinger publishes 'uber polymerization'

1933: Fawcett and Gibson discover LDPE

1937: Bayer discovers polyurethane

1938: Carothers patents nylon, i.e., Polyamide 6.6, and Plunkett discovers Teflonane

1944: McIntire discovers polystyrene

1954: Ziegler–Natta catalyst discovered for polypropylene production

1965: Kwolek creates Kevlar

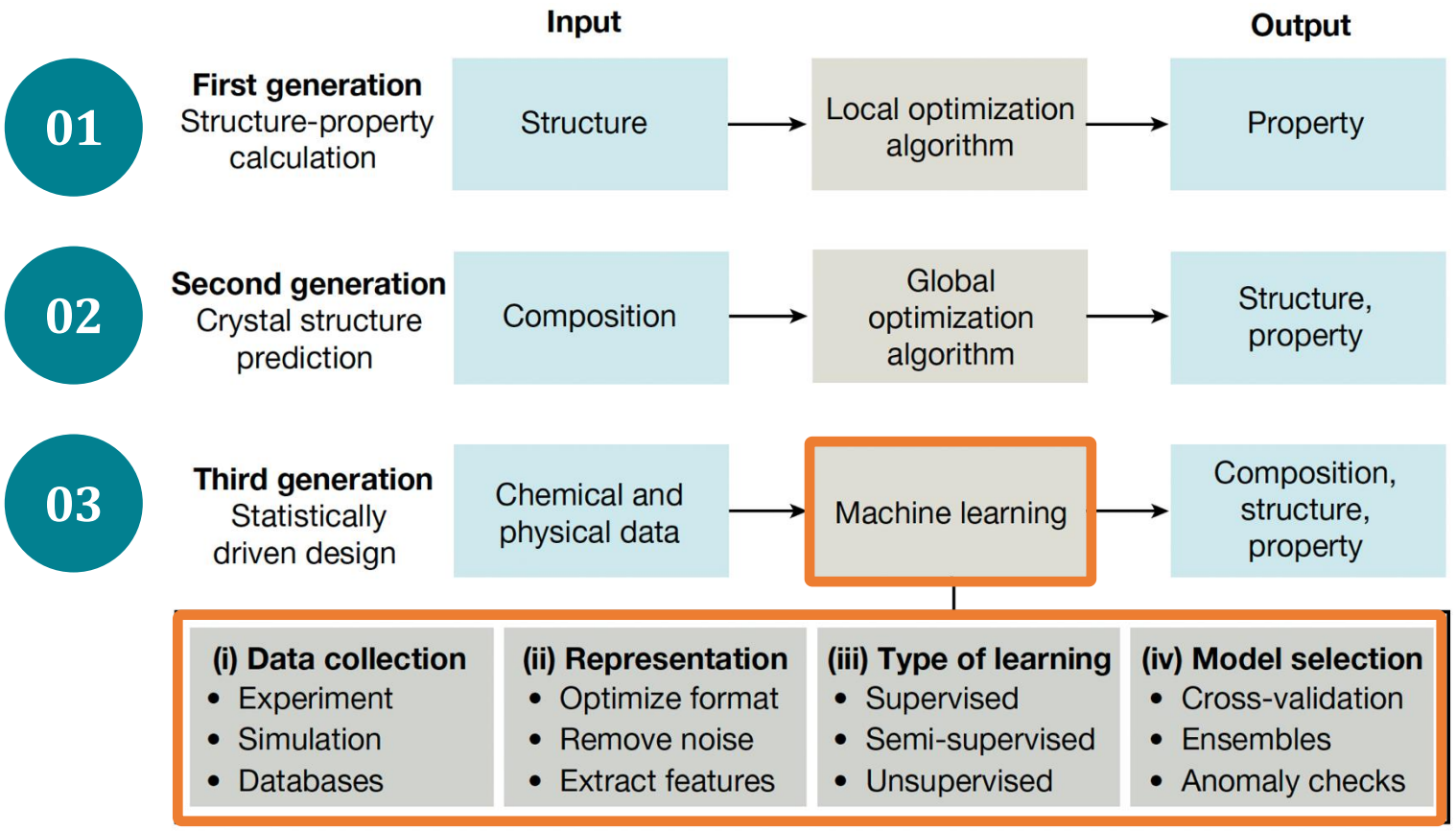
1977: Heeger, MacDiarmid and Shirakawa develop polyacetylene, the first conductive polymer

2011: Launch of Materials Genome Initiative

2020: 100-year anniversary of modern polymer science

Tran, Gurnani, Kim, Pilania, Kwon, Lively, Ramprasad, "Design of functional and sustainable polymers assisted by AI," Nat. Rev. Mater. 9, 866–886 (2024). DOI: 10.1038/s41578-024-00708-8

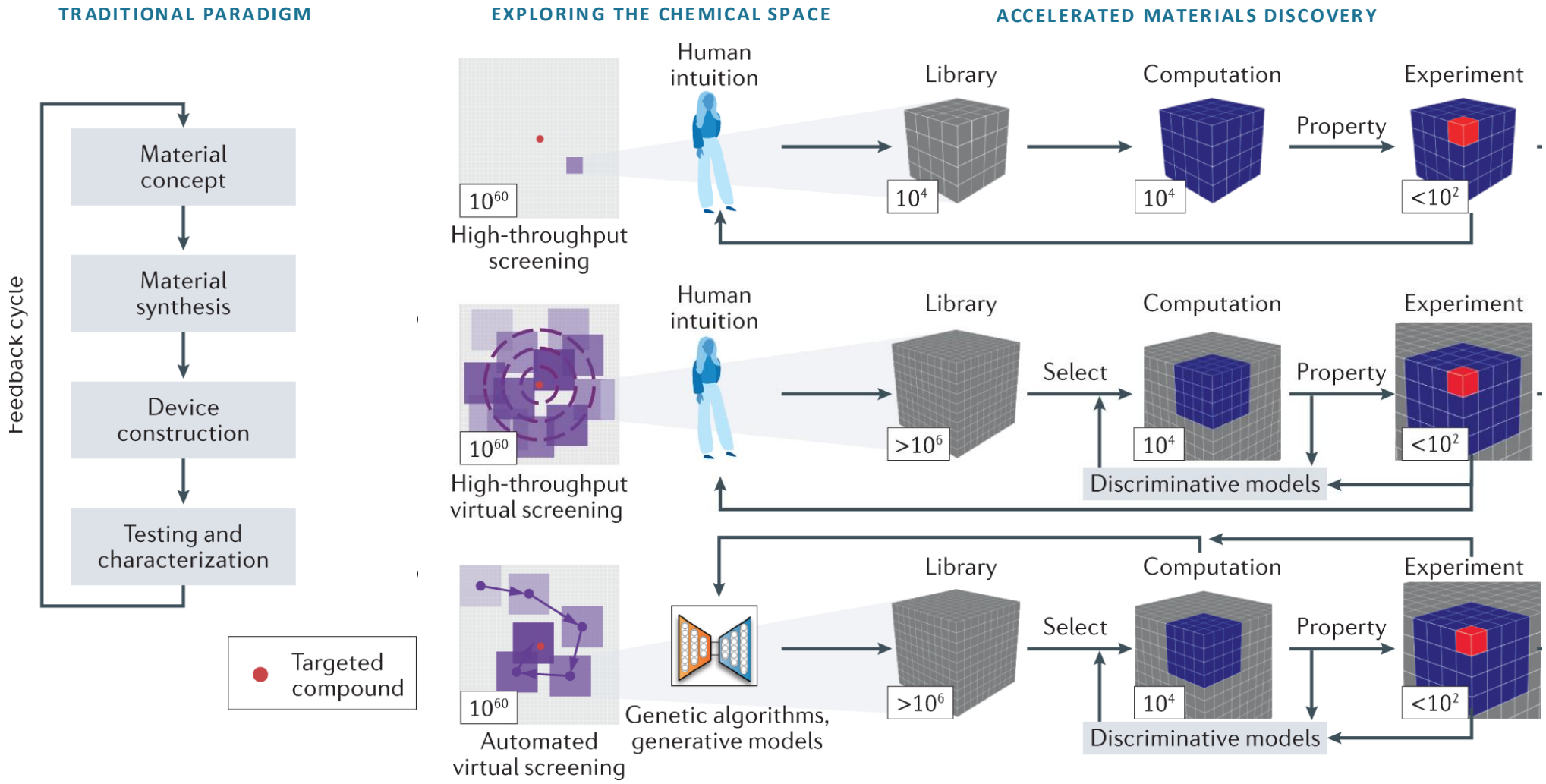
From trial-and-error to data-driven design:
evolution of the research workflow in computational chemistry



Butler, Davies, Cartwright, Isayev, Walsh, "Machine learning for molecular and materials science," Nature 559, 547–555 (2018). DOI: 10.1038/s41586-018-0337-2

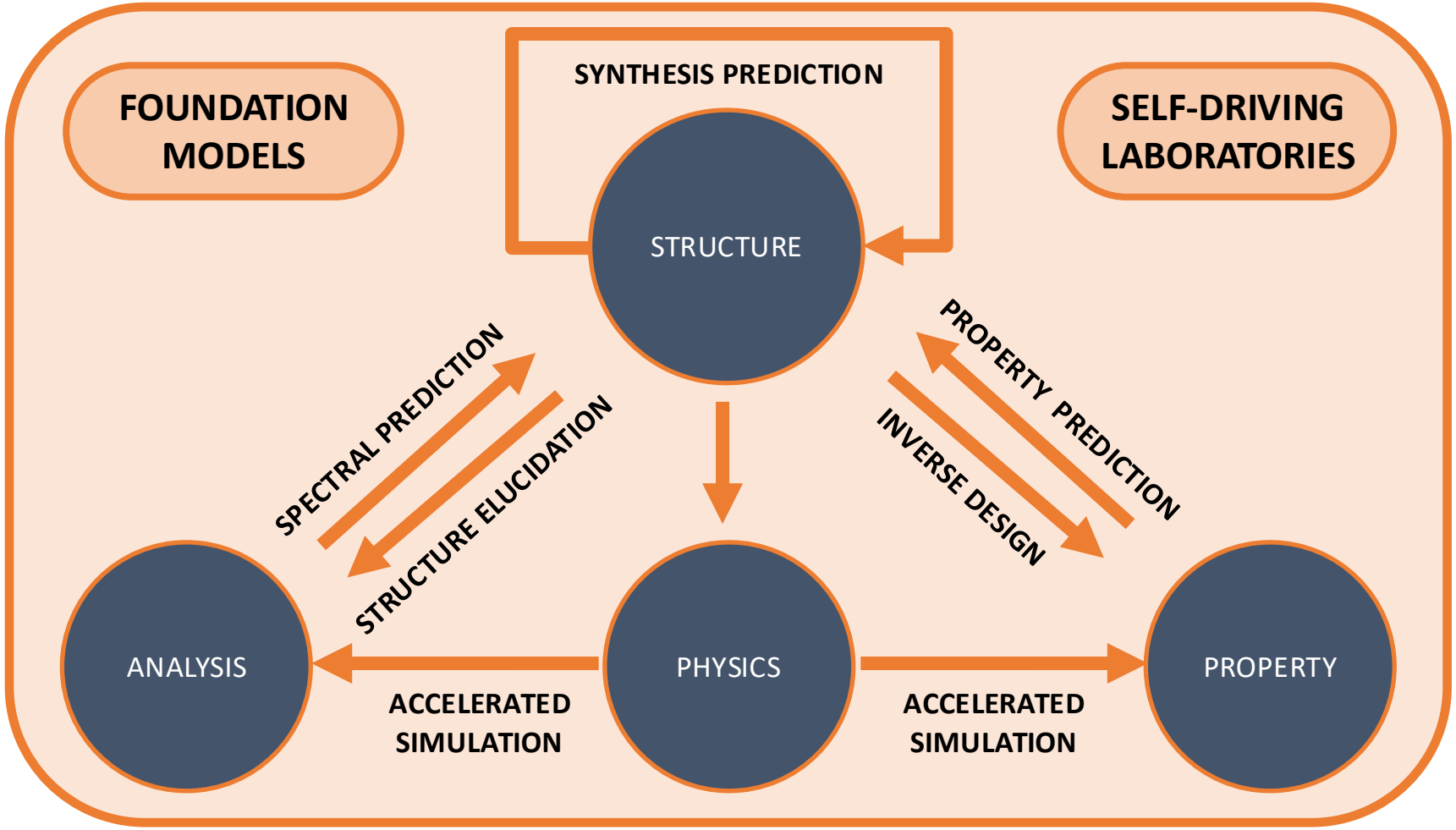
New paradigm in materials discovery

A NEW PARADIGM



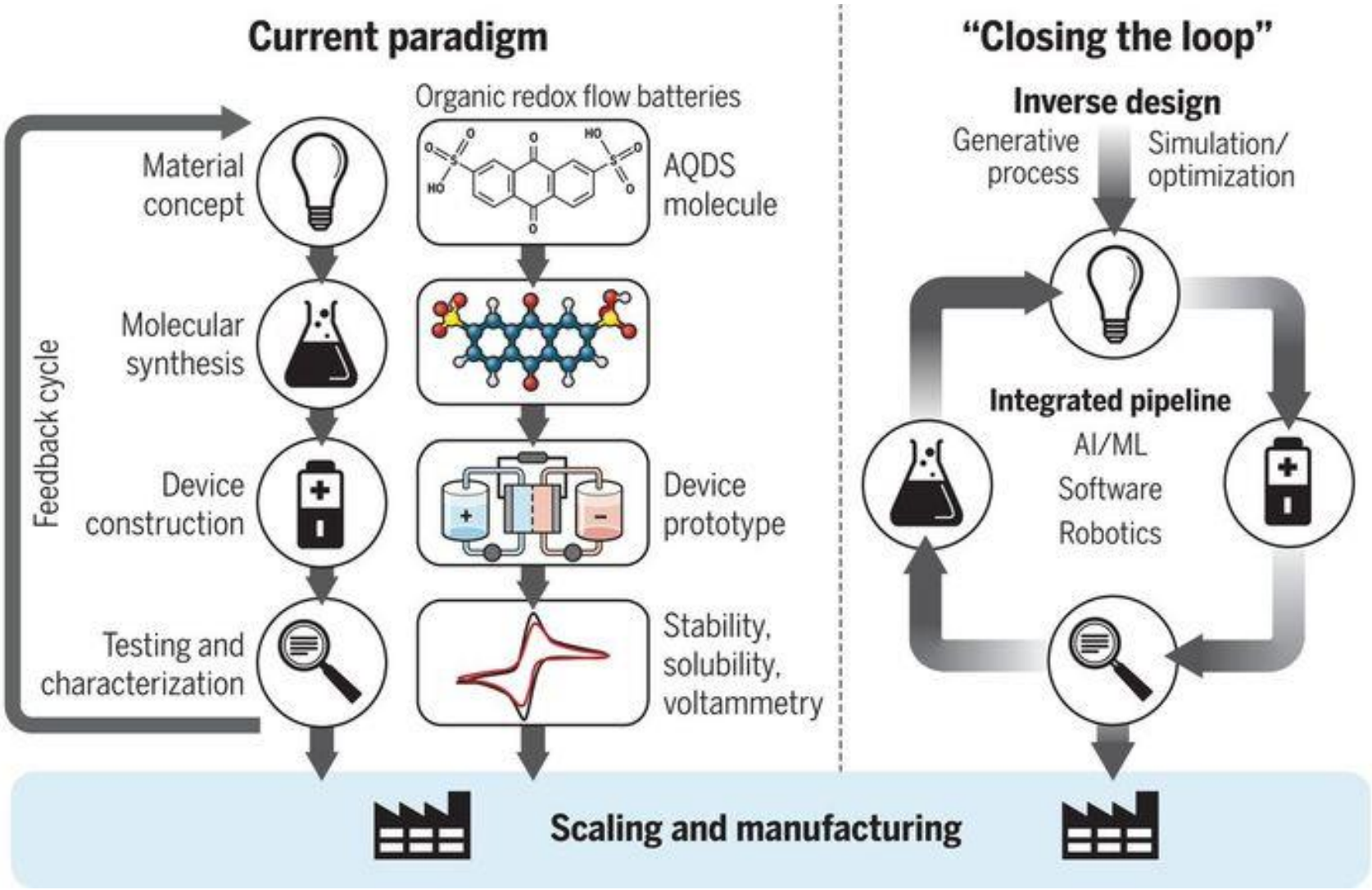
Yao, Lum, Johnston, Mejia-Mendoza, Zhou, Wen, Aspuru-Guzik, Sargent, Seh, "Machine learning for a sustainable energy future," Nat. Rev. Mater. 8, 202–215 (2023).

Machine learning for chemistry and materials



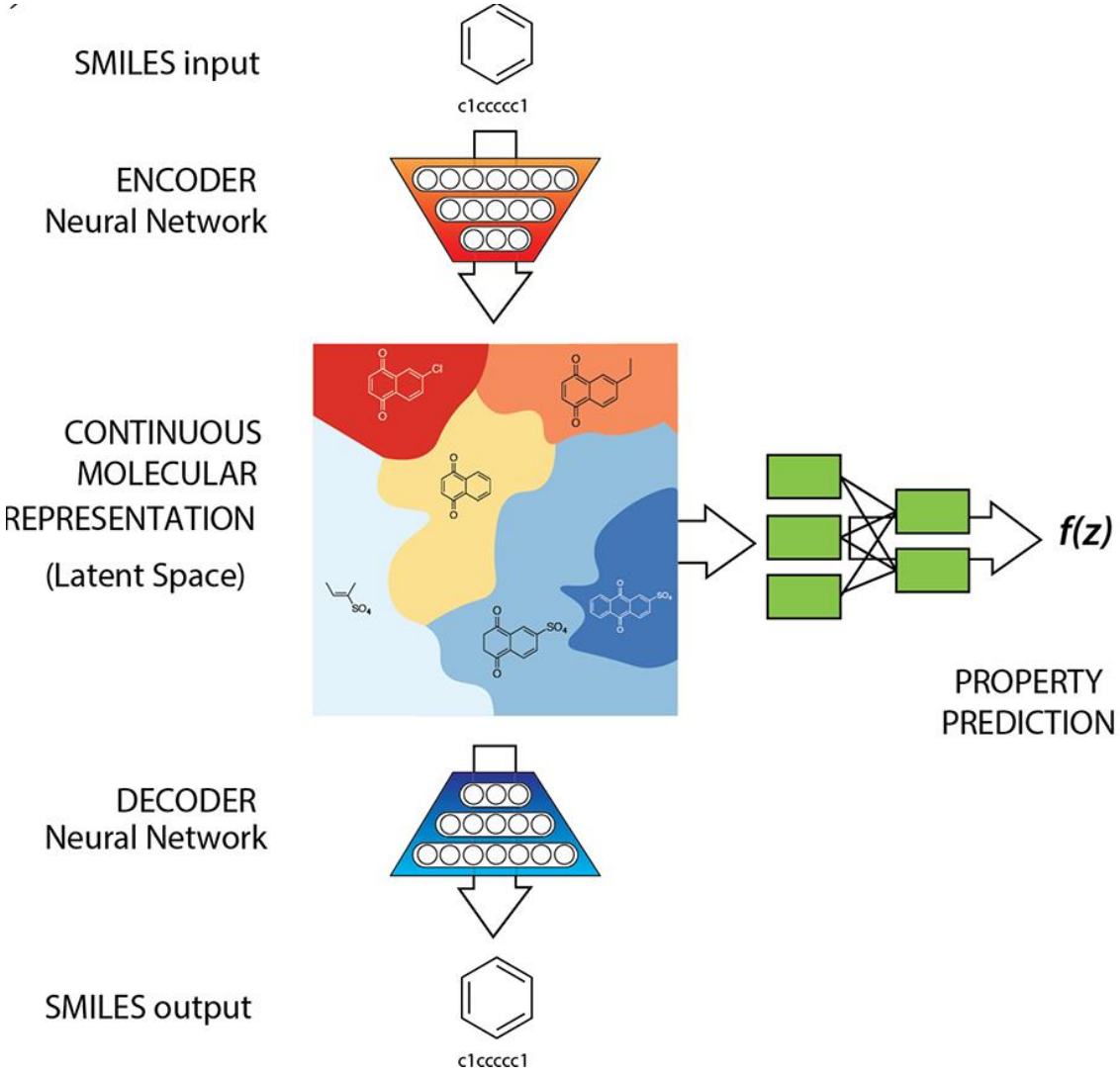
Cheng, Ser, Skreta, et al., Aspuru-Guzik, "Spiers Memorial Lecture: How to do impactful research in AI for chemistry and materials science," Faraday Discuss. 256, 10–60 (2025).

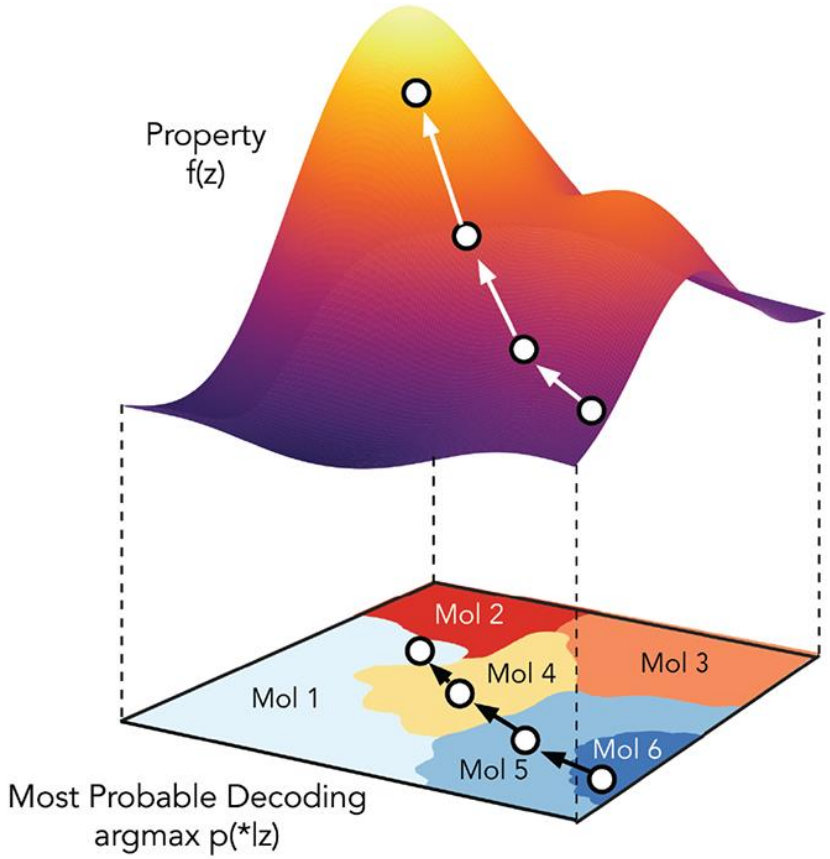
Direct vs Inverse design: example of redox flow batteries



Sanchez-Lengeling & Aspuru-Guzik, "Inverse molecular design using machine learning," Science 361, 360–365 (2018).

Chemical VAE: turn molecules into a map tooptimise





Gómez-Bombarelli et al. ACS Central Science 4(2), 268–276 (2018).

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – True/False

1. **In inverse design you specify the property you want and the model proposes a structure.**

   True | False

2. **A VAE gives a continuous latent space you can optimize in.**

   True | False

3. **The “third generation” of materials design is manual trial-and-error.**

   True | **False**

A NEW PARADIGM

Take-home messages - Design to a specification instead of stumbling onto it.

WHY IT MATTERS TO CLIMATE

1. We are in a ML-guided generation of materials design.

2. Inverse design = optimize in a learned latent space (continuous space of design).

3. There are two cross-cutting enablers: Foundation Models and Self-Driving Labs.

THE AI TOOLKIT

An AI toolkit

Six things ML actually does across the materials pipeline

1 Property prediction

PREDICT

Supervised models predict a property y (e.g. bandgap) from a structure x, aiding or replacing costly simulations.

2 Generative design

GENERATE

Unsupervised / generative models output new candidates x′, conditioned on target properties: inverse-design route.

3 Self-driving labs

SYNTHESIZE

Autonomous labs plan & run experiments: retrosynthesis, reaction prediction, active-learning optimization.

4 Aiding characterization

CHARACTERIZE

ML interprets measurements: crystal structure from TEM, coordination from XAS, symmetry from electron diffraction.

5 System management

DEPLOY

ML manages devices & grids: predicting battery life, forecasting loads, optimizing smart-grid control.

6 Accelerating theory

SIMULATE

ML makes intractable simulations feasible: cheaper potentials & functionals for larger scales.

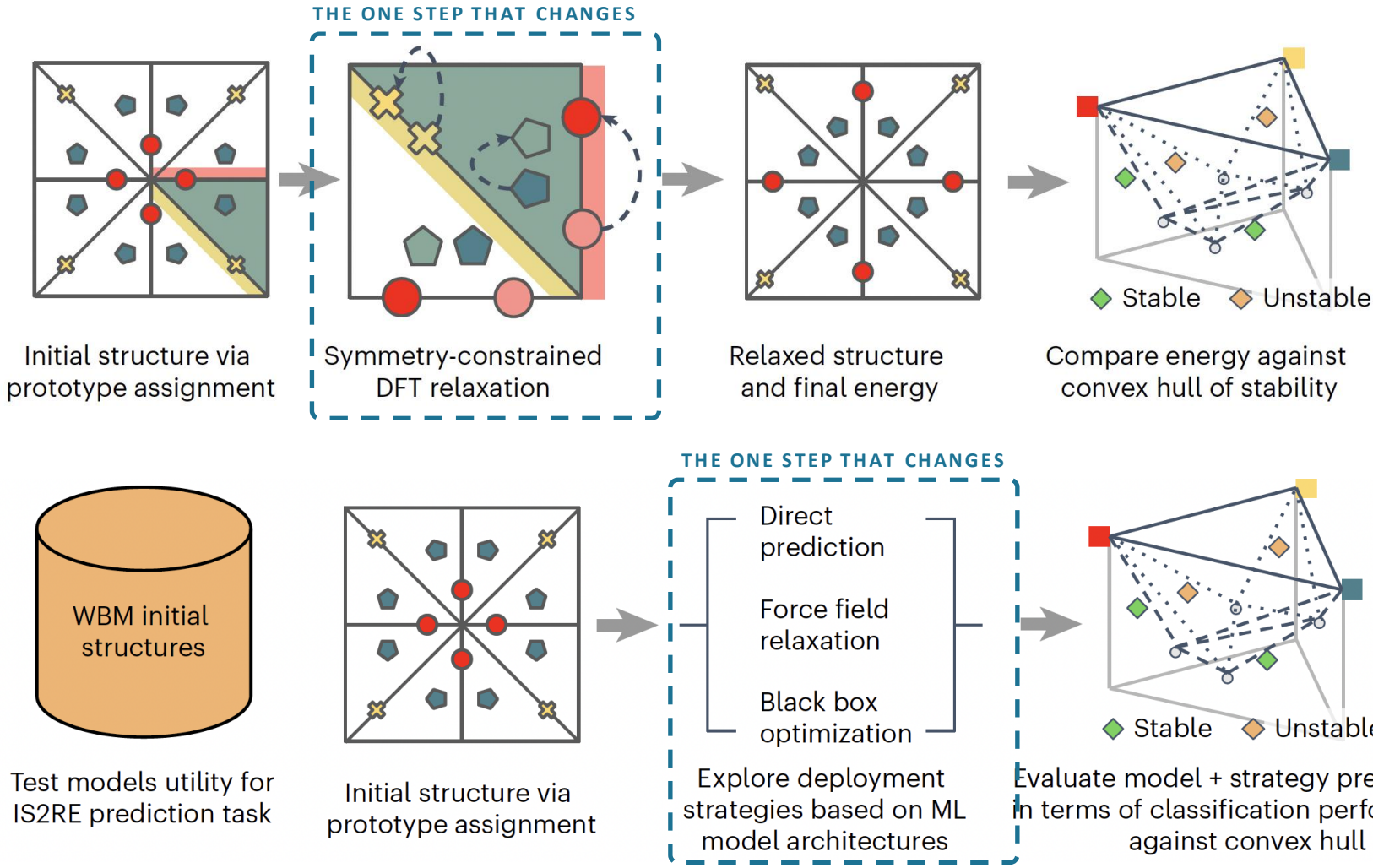
Adapted from Yao et al., “Machine learning for a sustainable energy future”, Nat. Rev. Mater. 2023, Box 1.

1

Property prediction

PREDICT

Predicting DFT properties with an ML surrogate



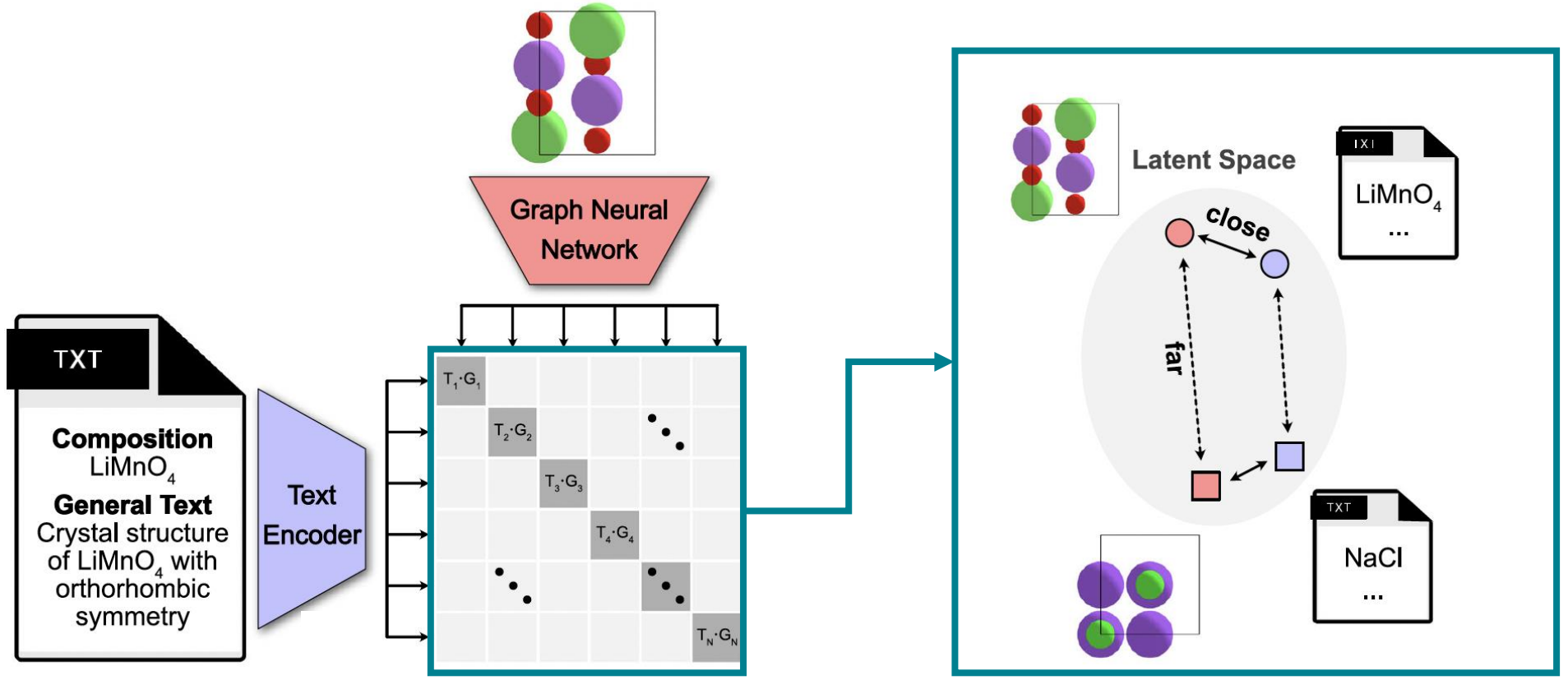
Adapted from Riebesell et al., “A framework to evaluate machine learning crystal stability predictions,” Nat. Mach. Intell. 7, 836–847 (2025).

2

Generative / inverse design

GENERATE

Chemeleon: turning a text prompt into a crystal, in three steps



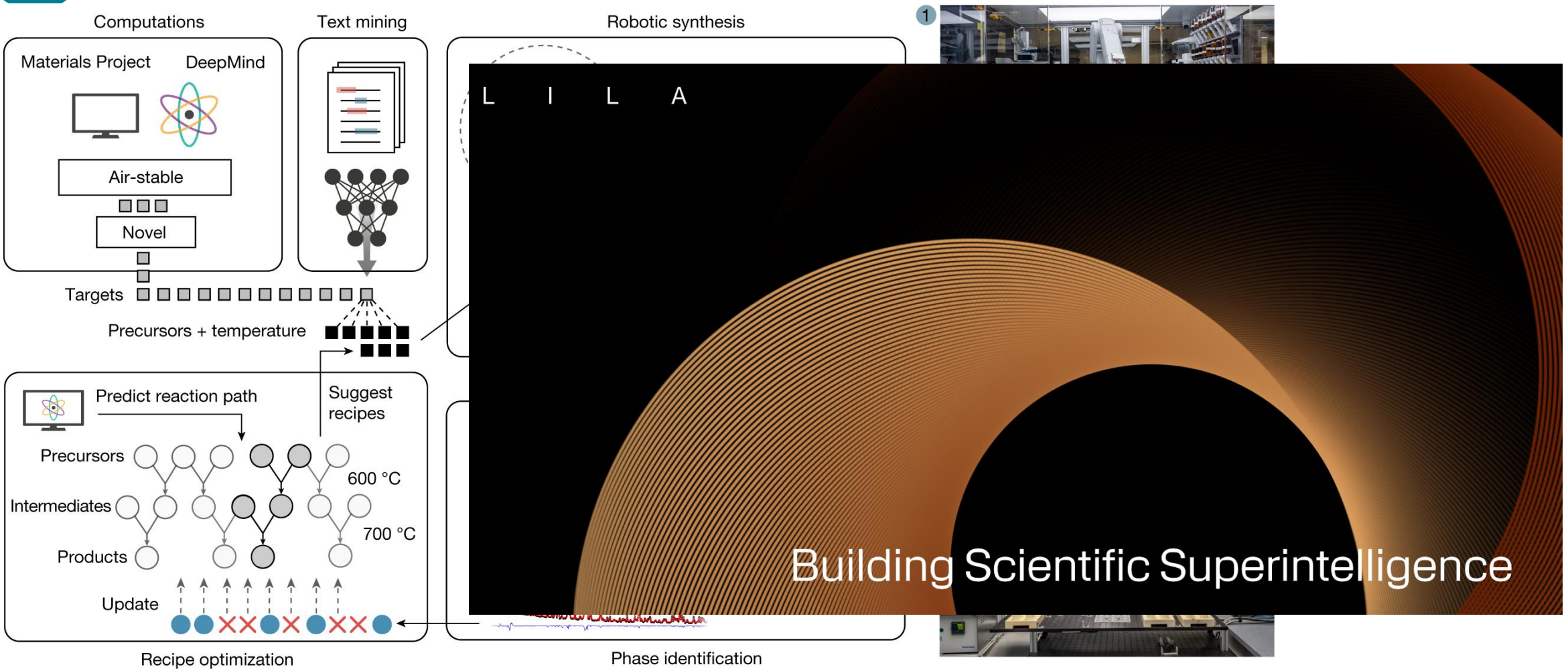
Adapted from Park, Onwuli & Walsh, Nat. Commun. 16 (2025)

3

Self-driving labs

SYNTHESIZE

An autonomous laboratory: the A-Lab



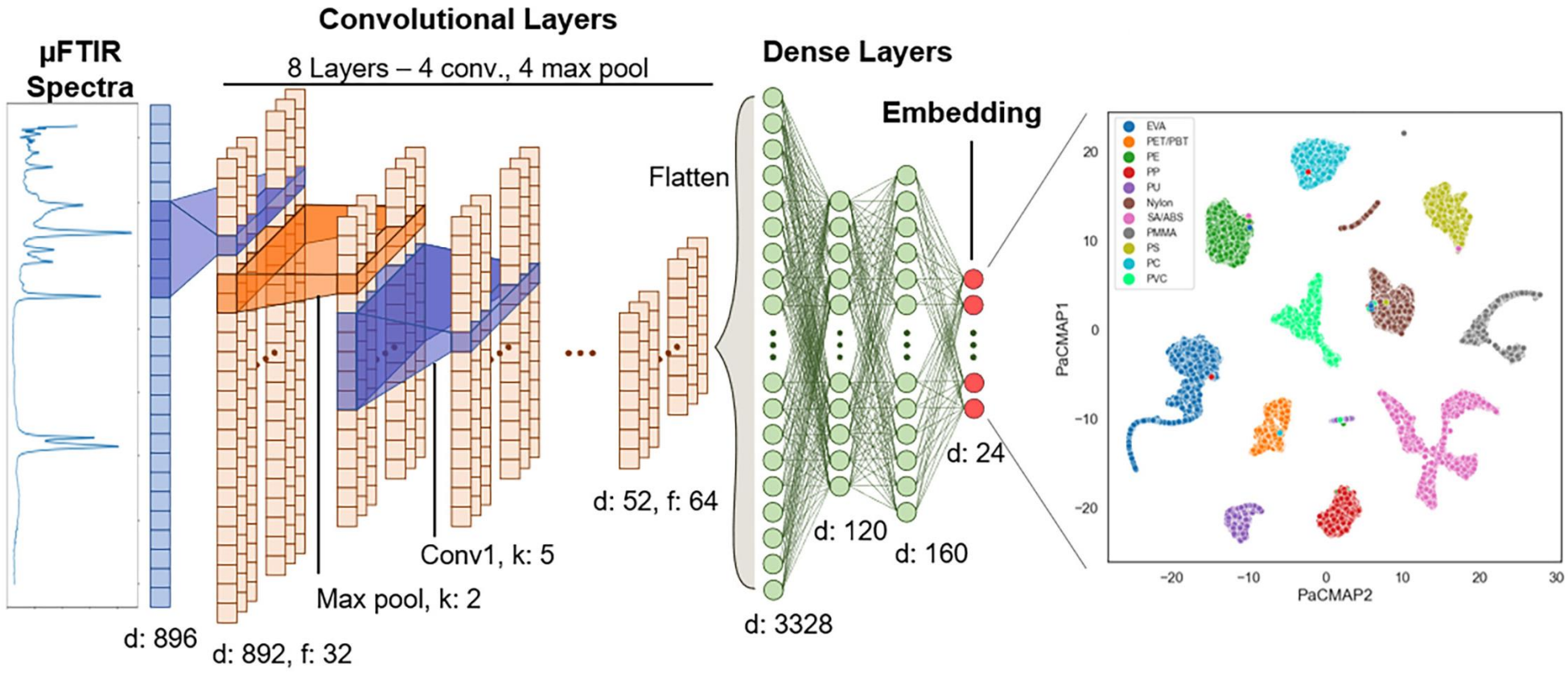
Adapted from Szymanski et al., Nature 624, 86 (2023)

4

Aiding characterization

CHARACTERIZATION

Similarity learning on μFTIR spectral imaging of microplastics



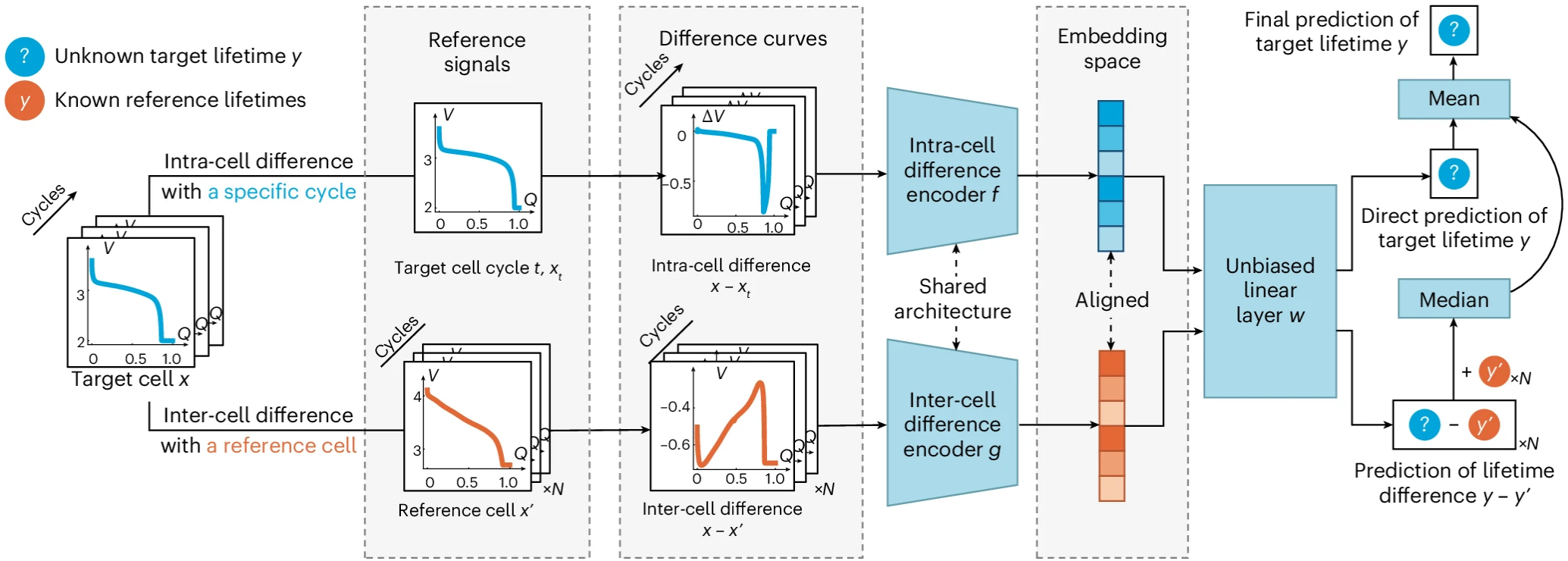
Smolen, Moore, Perez & Wooley, PNAS 122(42), e2509745122 (2025)

5

System management

DEPLOY

Predicting battery cycle life early

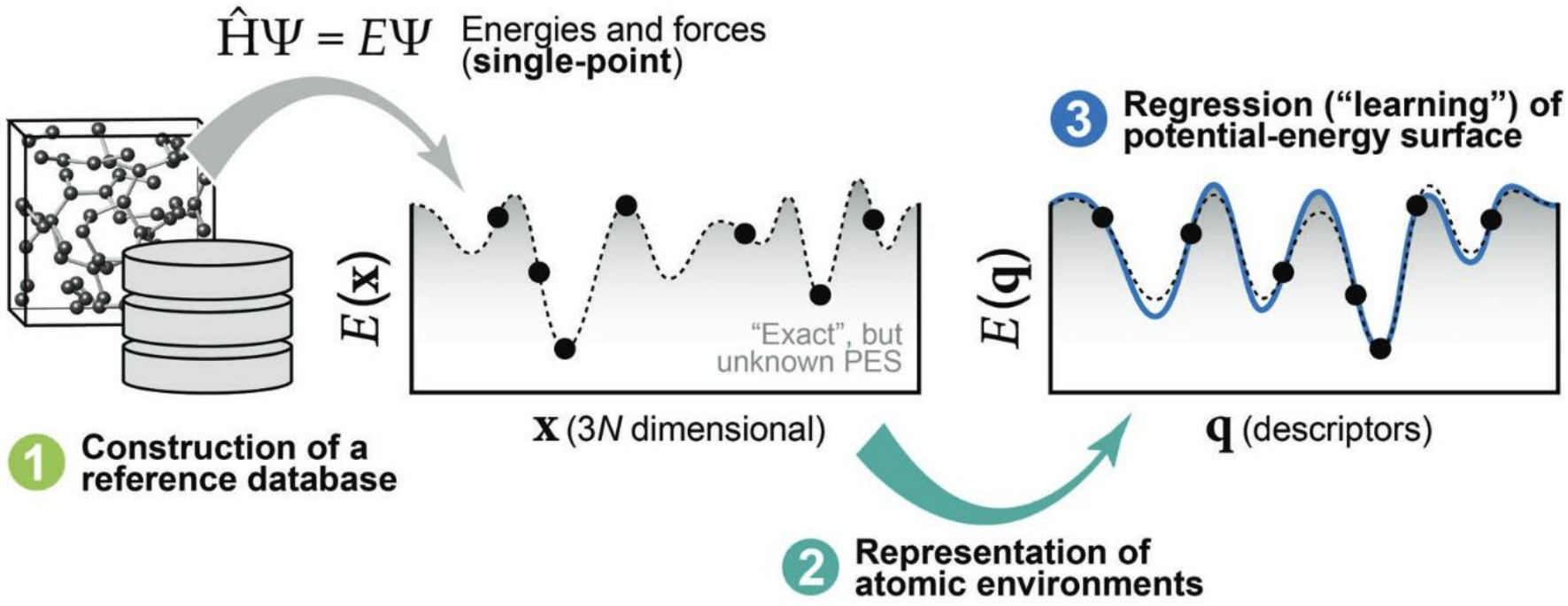


BatLiNet compares the target cell to a fully-cycled "reference" cell whose lifetime is already known

Zhang et al., "Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning" (BatLiNet), Nature Machine Intelligence 7 (2025)

How a machine-learning interatomic potential works

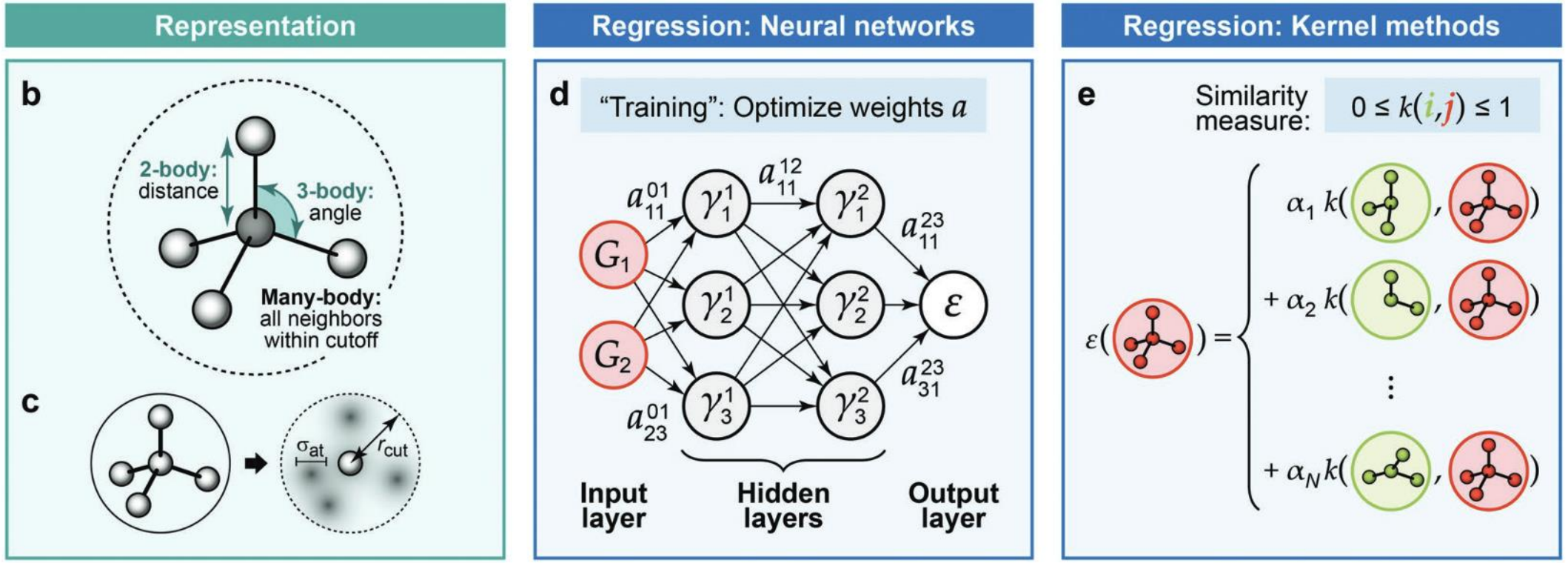
One recipe underlies them all: local atomic environment → energy, learned from DFT



Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. Adv. Mater. 31, 1902765 (2019). 10.1002/adma.201902765

How a machine-learning interatomic potential works

One recipe underlies them all: local atomic environment → energy, learned from DFT

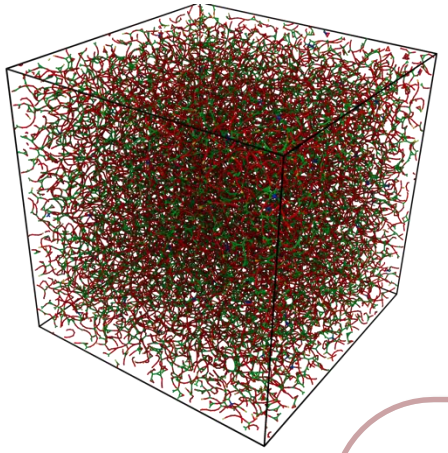


Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. Adv. Mater. 31, 1902765 (2019). 10.1002/adma.201902765

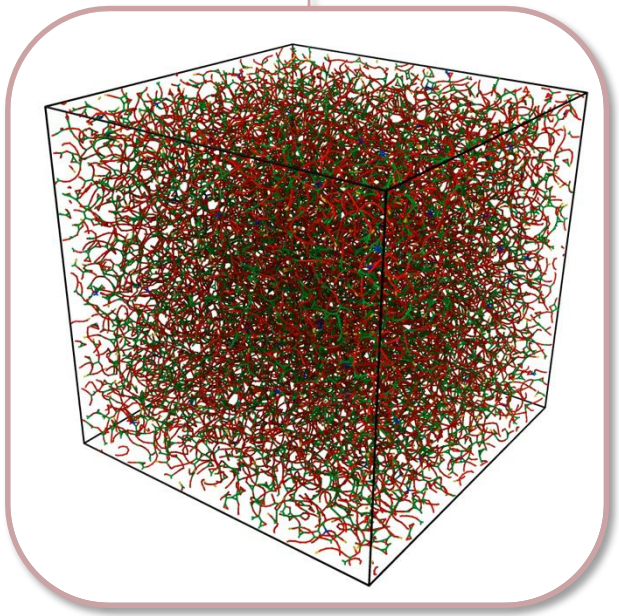
Modeling biochar with MLIPs (GAP20)

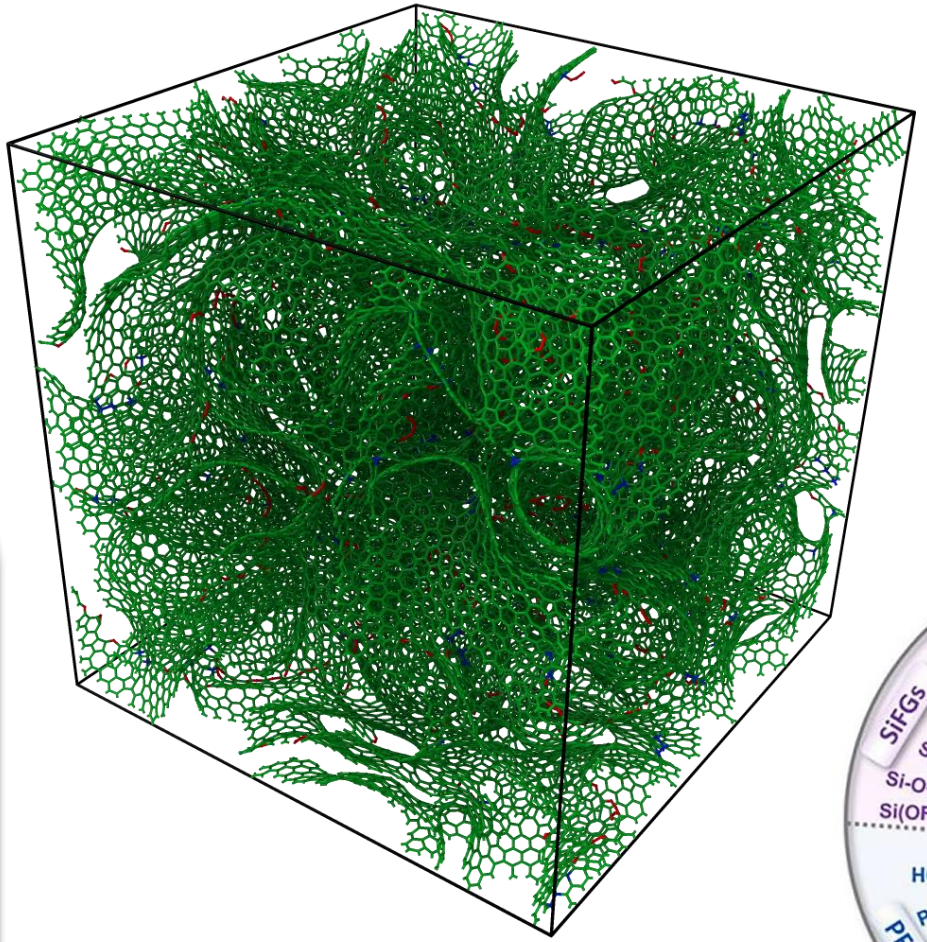
Atomic scale

Structure Characterization:
Porosity, Atomic Structure, Carbon Hybridization, Structure Order

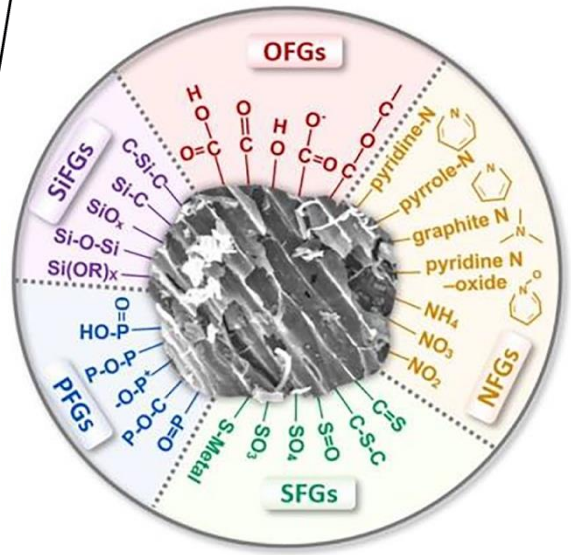


MD Simulation





Biochar Model



Yang et al., J. Hazard. Mater. Adv., 2022.

MLIP models at a glance

Representative models by family — key idea, typical use, and source

Model	Family	Key idea	Typical application	Reference
HDNNP	Descriptor + NN	symmetry functions → neural net	bulk solids, water	*Behler & Parrinello 2007*
GAP	Descriptor + kernel	SOAP + Gaussian process	Si, C, amorphous; data-efficient	*Bartók et al. 2010*
DeePMD	Learned descriptor + NN	end-to-end deep potential	large-scale MD (water, alloys)	*Zhang et al. 2018*
SNAP	Linear expansion	bispectrum components (linear)	metals (Ta, W)	*Thompson et al. 2015*
MTP	Linear expansion	moment tensors	alloys; active learning	*Shapeev 2016*
ACE	Body-ordered expansion	atomic cluster expansion	general; fast + accurate	*Drautz 2019*
SchNet	Invariant GNN	continuous-filter convolutions	molecules (MD17, QM9)	*Schütt et al. 2018*
NequIP	Equivariant GNN	E(3)-equivariant messages	data-efficient molecules + materials	*Batzner et al. 2022*
Allegro	Local equivariant GNN	strictly-local equivariant	very large-scale MD	*Musaelian et al. 2023*
MACE	Equivariant + high body-order	higher-order messages per layer	accurate & fast MD	*Batatia et al. 2022*
M3GNet	Universal GNN	MP-trained graph potential	crystal relaxation / screening	*Chen & Ong 2022*
CHGNet	Universal (charge-aware)	charge-informed GNN	Li-ion, oxides, magnetism	*Deng et al. 2023*
MACE-MP-0	Foundation (equivariant)	periodic-table foundation model	off-the-shelf MD across chemistries	*Batatia et al. 2024*

SOAP = smooth overlap of atomic positions · E(3) = 3-D rotations + translations + reflections · MP = Materials Project · GNN = graph neural network.

References

Machine-learning interatomic potentials, in roughly chronological order

1. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. Phys. Rev. Lett. 98, 146401 (2007). 10.1103/PhysRevLett.98.146401

2. Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials (GAP). Phys. Rev. Lett. 104, 136403 (2010). 10.1103/PhysRevLett.104.136403

3. Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral neighbor analysis method (SNAP). J. Comput. Phys. 285, 316–330 (2015). 10.1016/j.jcp.2014.12.018

4. Shapeev, A. V. Moment Tensor Potentials (MTP). Multiscale Model. Simul. 14, 1153–1173 (2016). 10.1137/15M1054183

5. Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics (DeePMD). Phys. Rev. Lett. 120, 143001 (2018). 10.1103/PhysRevLett.120.143001

6. Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet. J. Chem. Phys. 148, 241722 (2018). 10.1063/1.5019779

7. Drautz, R. Atomic cluster expansion (ACE). Phys. Rev. B 99, 014104 (2019). 10.1103/PhysRevB.99.014104

8. Batzner, S. et al. E(3)-equivariant GNNs for interatomic potentials (NequIP). Nat. Commun. 13, 2453 (2022). 10.1038/s41467-022-29939-5

9. Batatia, I.; Kovács, D. P.; Simm, G. N. C.; Ortner, C.; Csányi, G. MACE: Higher-Order Equivariant Message Passing. NeurIPS 35 (2022). arXiv:2206.07697

10. Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential (M3GNet). Nat. Comput. Sci. 2, 718–728 (2022). 10.1038/s43588-022-00349-3

11. Musaelian, A. et al. Learning local equivariant representations (Allegro). Nat. Commun. 14, 579 (2023). 10.1038/s41467-023-36329-y

12. Deng, B. et al. CHGNet: charge-informed universal neural network potential. Nat. Mach. Intell. 5, 1031–1041 (2023). 10.1038/s42256-023-00716-3

13. Merchant, A. et al. Scaling deep learning for materials discovery (GNoME). Nature 624, 80–85 (2023). 10.1038/s41586-023-06735-9

14. Batatia, I. et al. A foundation model for atomistic materials chemistry (MACE-MP-0). arXiv:2401.00096 (2024).

ENGAGEMENT

SECTION CLOSER · QUIZ

Quick quiz – True/False

1 **An ML interatomic potential aims for DFT-level accuracy at much lower cost.**

True | False

2 **A model with 100% training accuracy is always the most trustworthy on new data.**

True | **False**

3 **AI-generated candidate materials still need DFT / experimental validation.**

True | False

WHY IT MATTERS TO CLIMATE

One toolkit, every clean-energy material.

- ML potentials give near-DFT accuracy at a fraction of the energy and time: screen millions of candidates.
- Self-driving labs turn the best predictions into validated samples, fast.
- Validation keeps scarce lab effort off false positives.

Take-home messages

WHY IT MATTERS TO CLIMATE

1. Self-driving labs turn the best predictions into validated samples, fast.

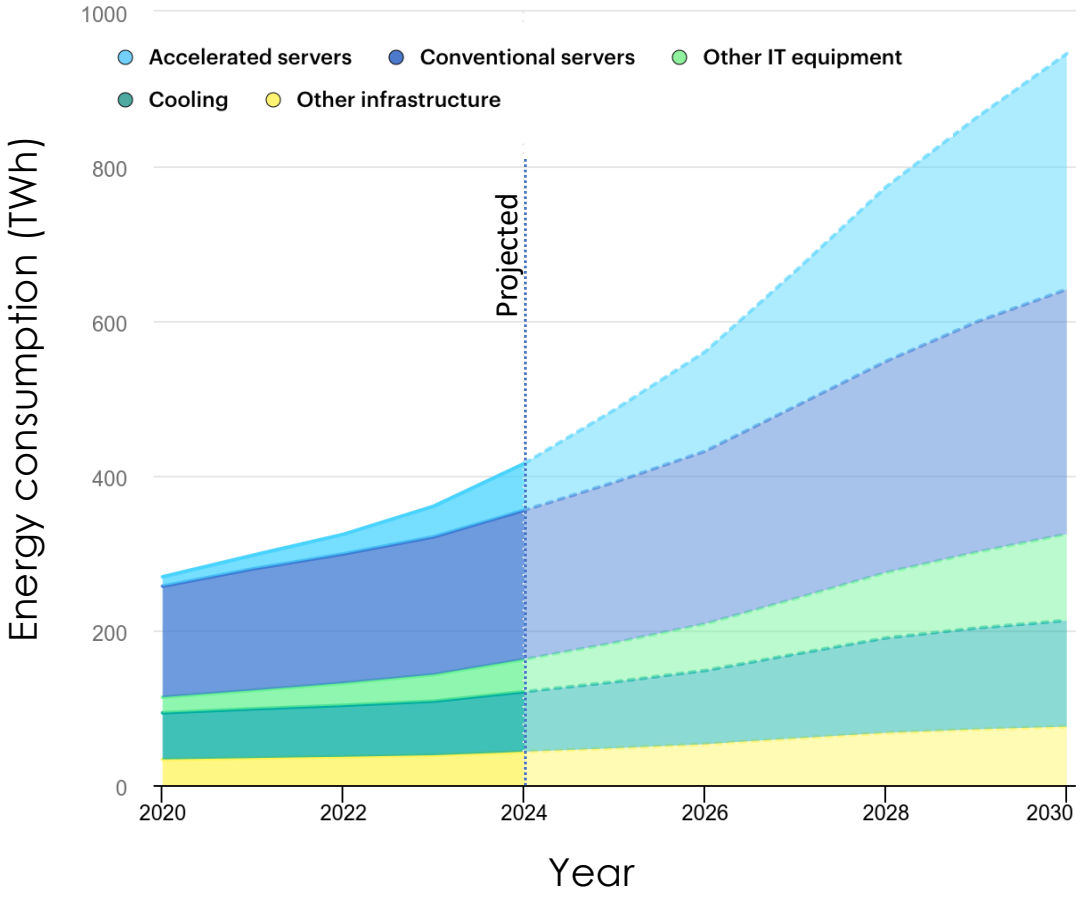
2. MLIPs deliver DFT accuracy at MD speed: the “accelerate simulation” box made real.

3. Generation is easy; validation is hard

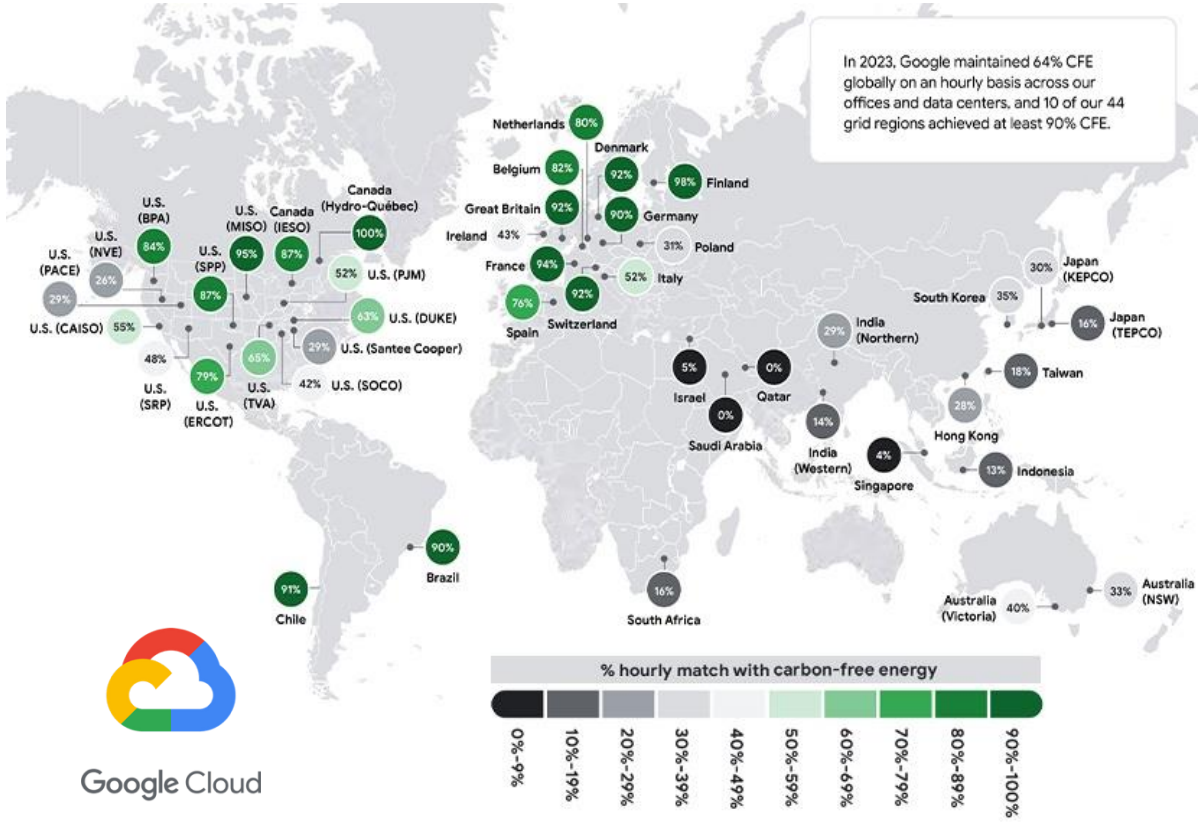
TALKING CHEMISTRY TO COMPUTERS

Increasing energy cost of AI and data centers

Need of greener and more democratized super computing

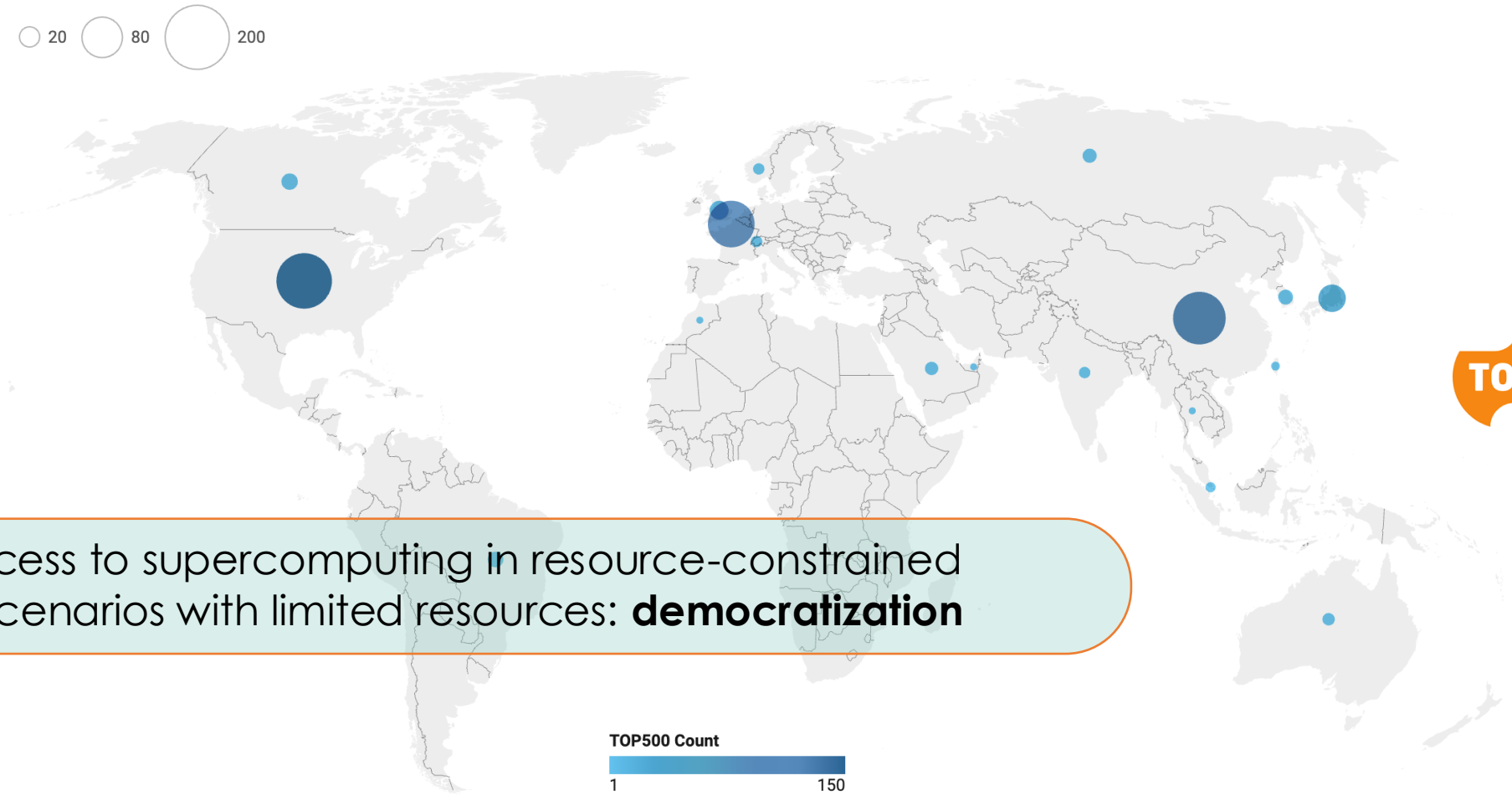


Source: International Energy Agency (IEA)



Source: Google Cloud

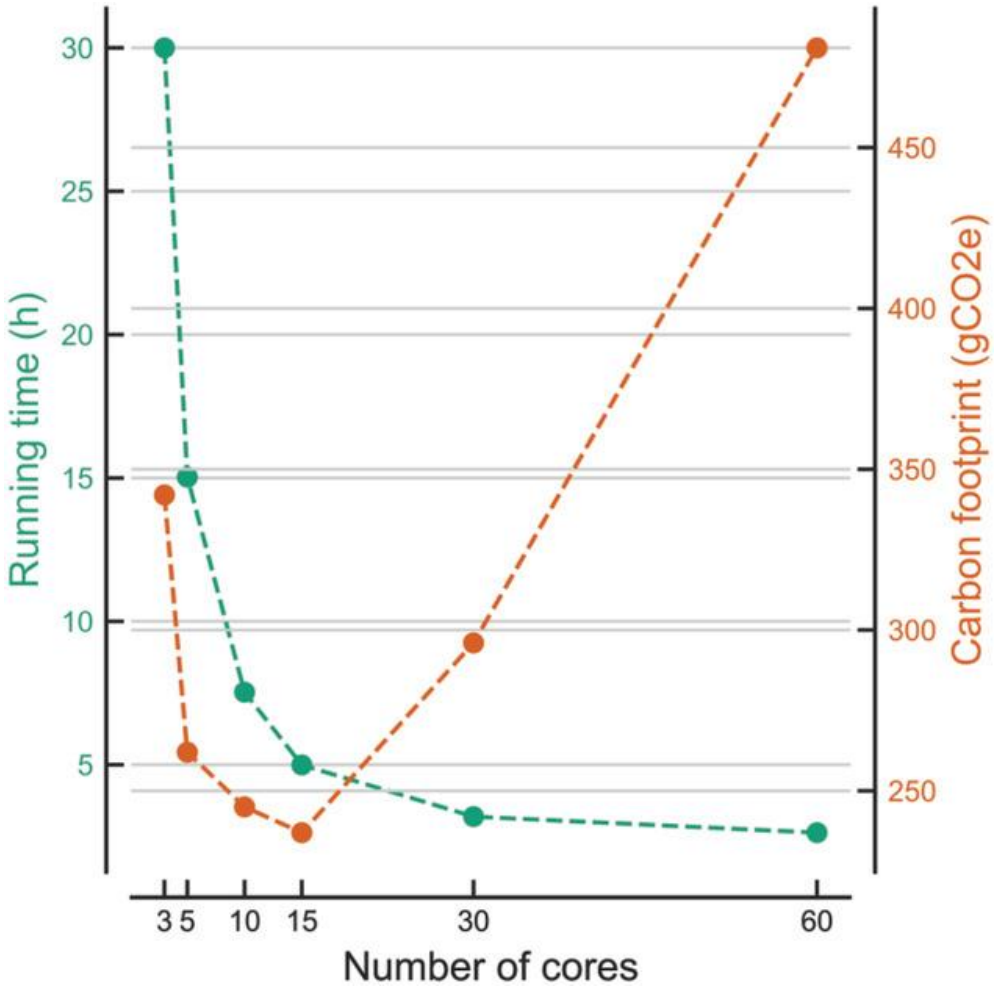
Where is supercomputing?



Improve access to supercomputing in resource-constrained devices or scenarios with limited resources: **democratization**

Source: top500.org

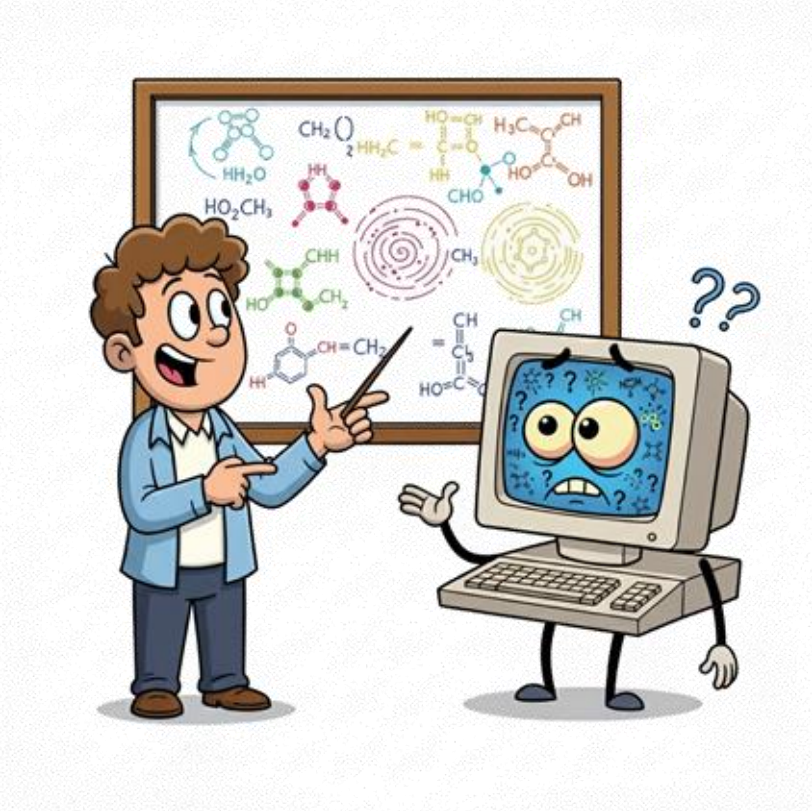
Green algorithms





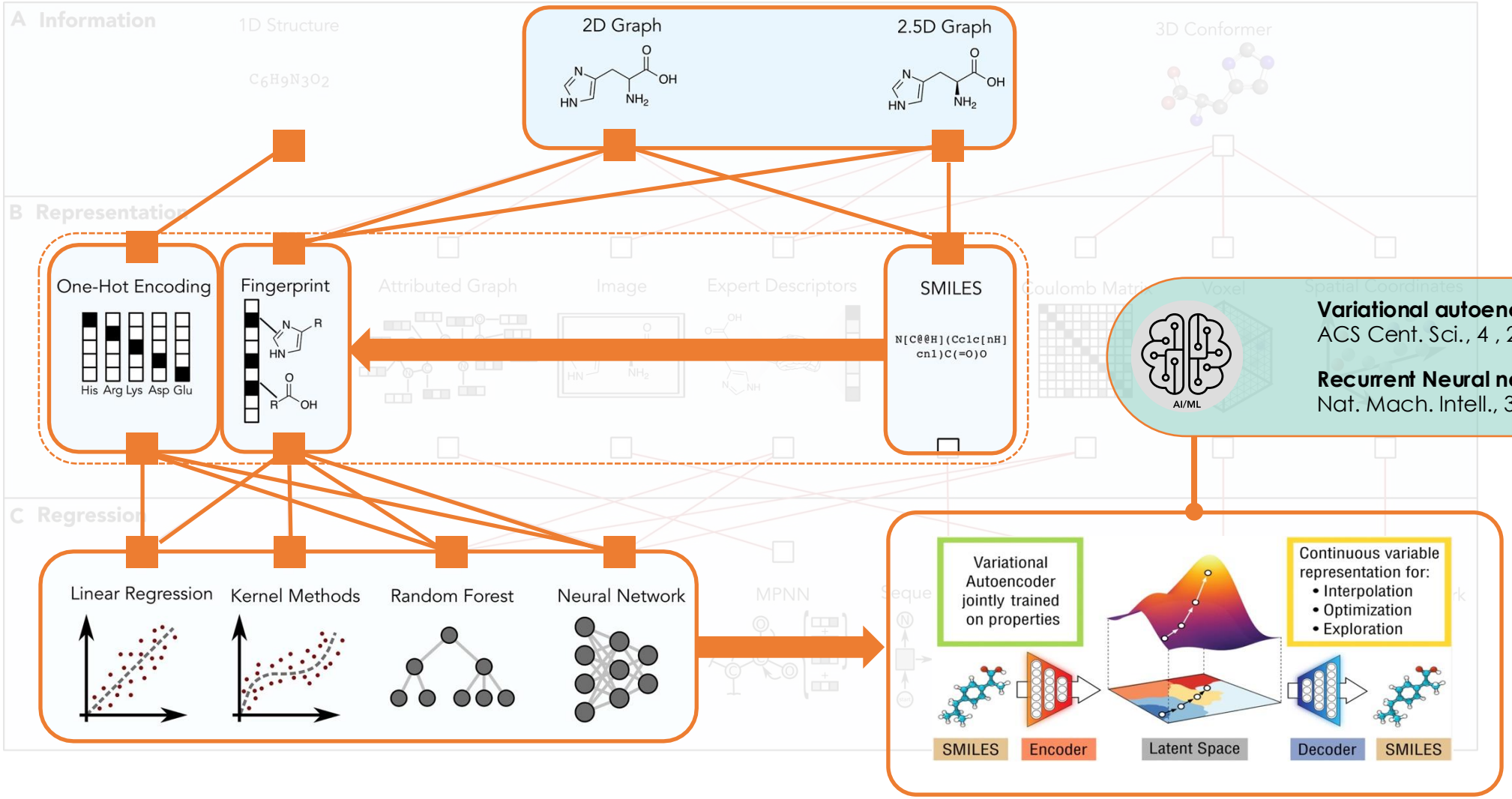
Lannelongue et al., Advanced Science, 8, 2021

One step back: How do we talk chemistry to computers?



Machine learning for chemistry requires **featurizing molecules**

ML for chemistry requires featurizing molecules



Variational autoencoders (VAE)
ACS Cent. Sci., 4 , 268, 2018`

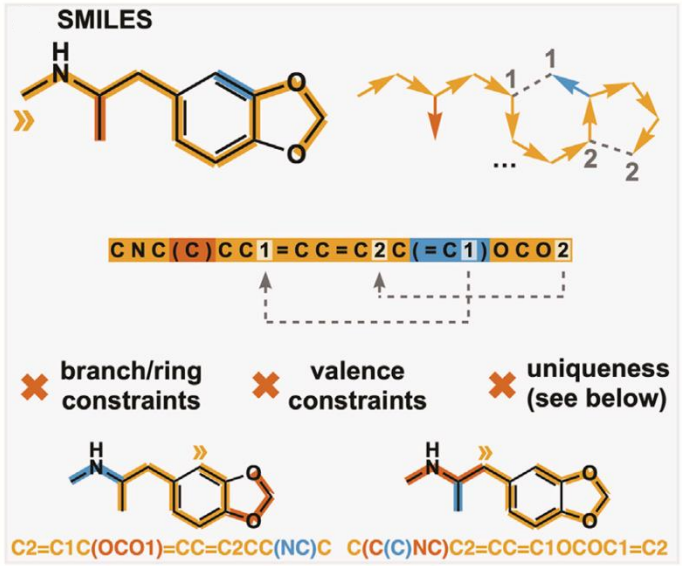
Recurrent Neural network (RNN)
Nat. Mach. Intell., 3 , 759, 2021

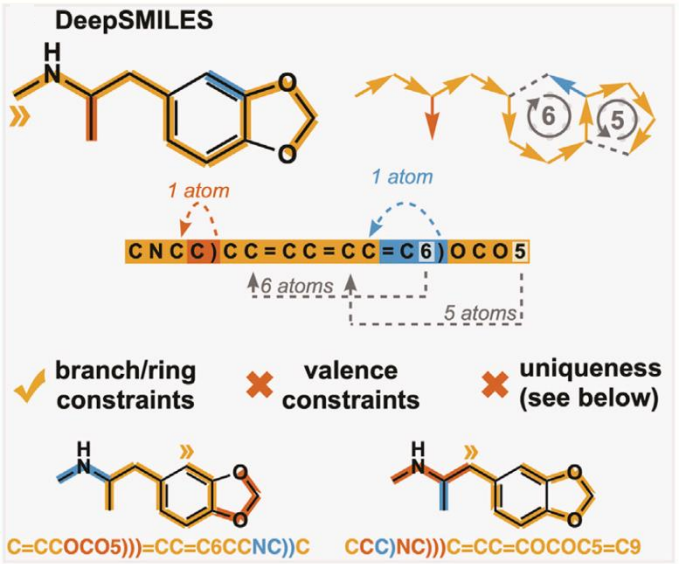
Pattanaik and Coley, **Chem 6, 1204-1207, 2020**

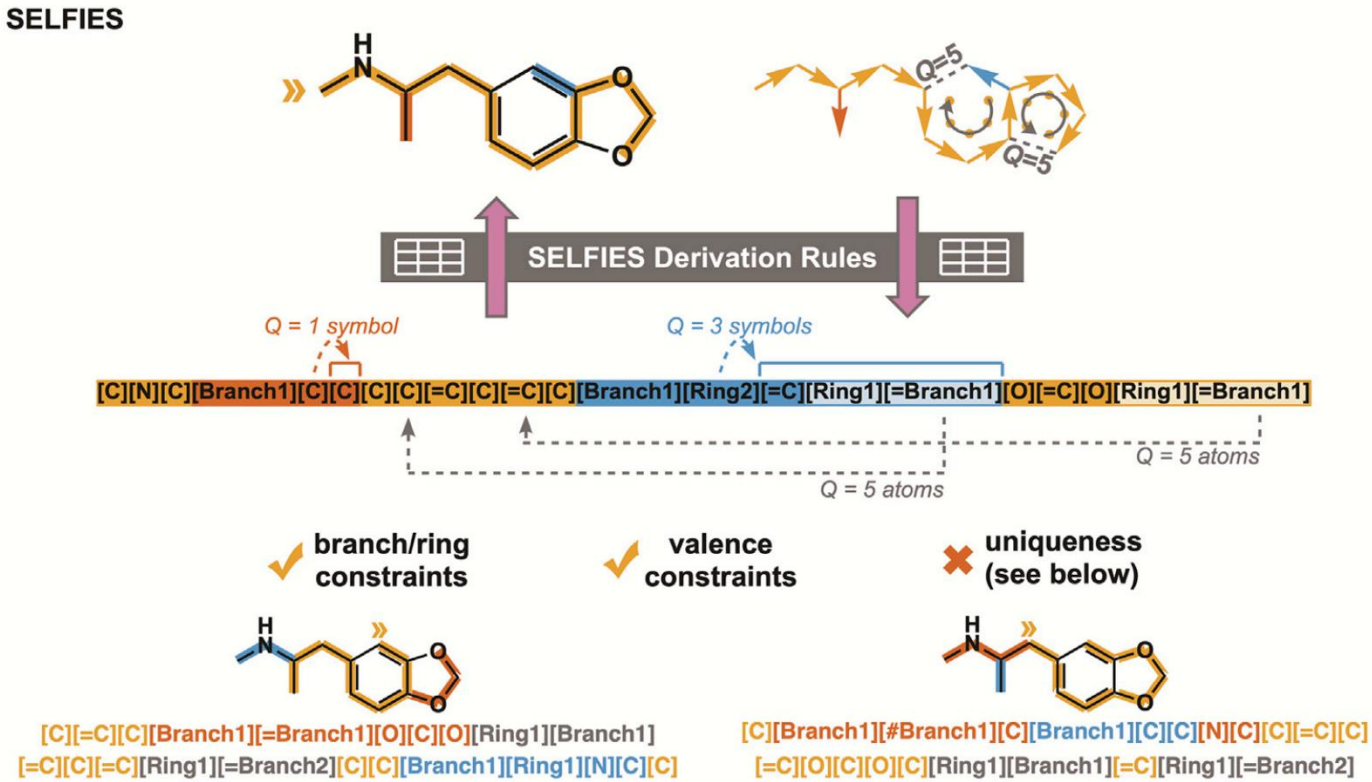
Skinnider et al. **Nat. Mach. Intell.**, 3 , 759, **2021**

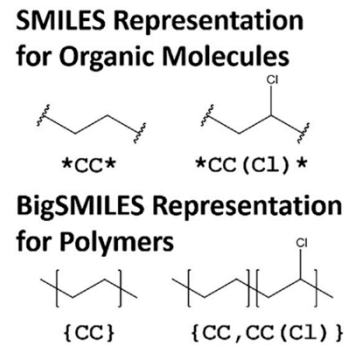
Gomez-Bombarelli et al. **ACS Cent Sci.** 4, 268–76, **2018**

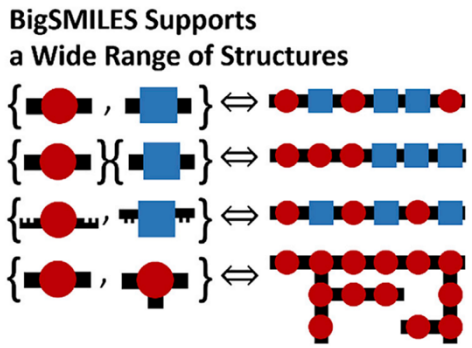
Machine-readable data encoding







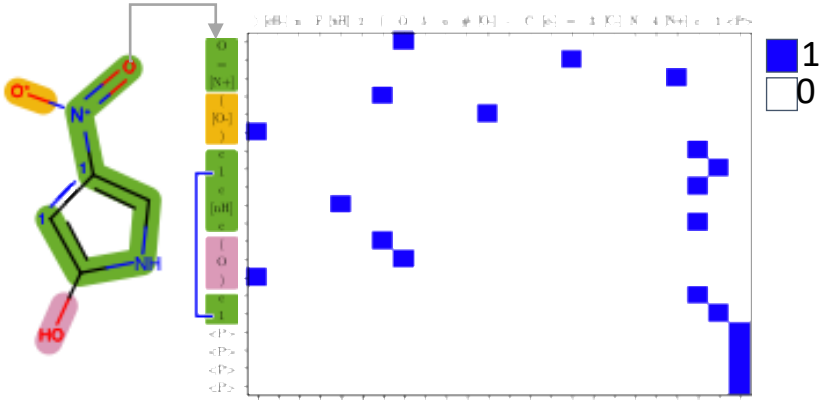




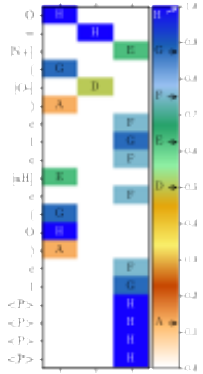
Krenn et al. Patterns 3, 10, 100588, 2022

embedded One Hot Encodings (eOHE)

One Hot Encoding (OHE)



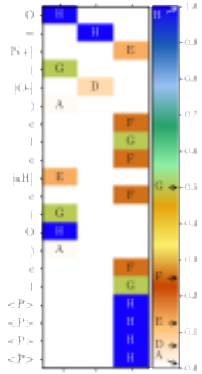
eOHE v1



Linear reduction
$q = 8$

$$v_1(r,q) = \frac{r+1}{q}$$

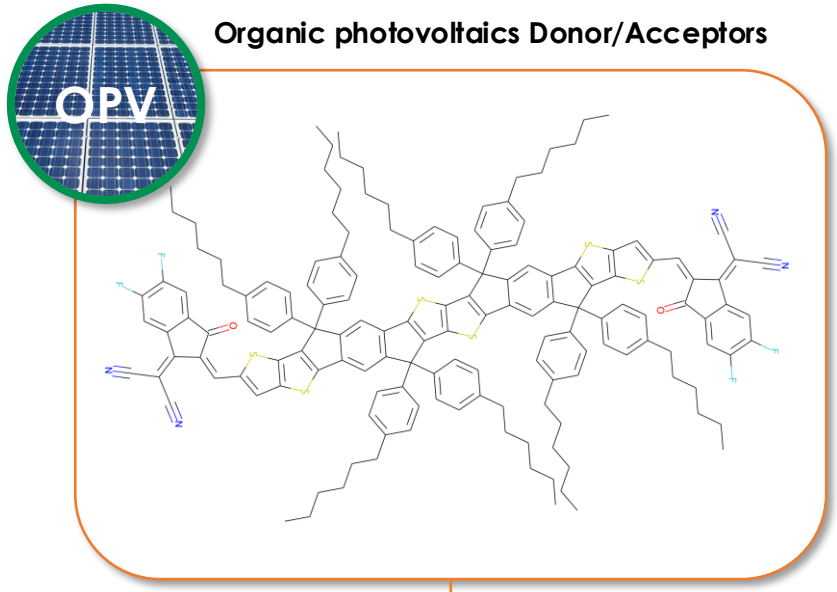
eOHE v2

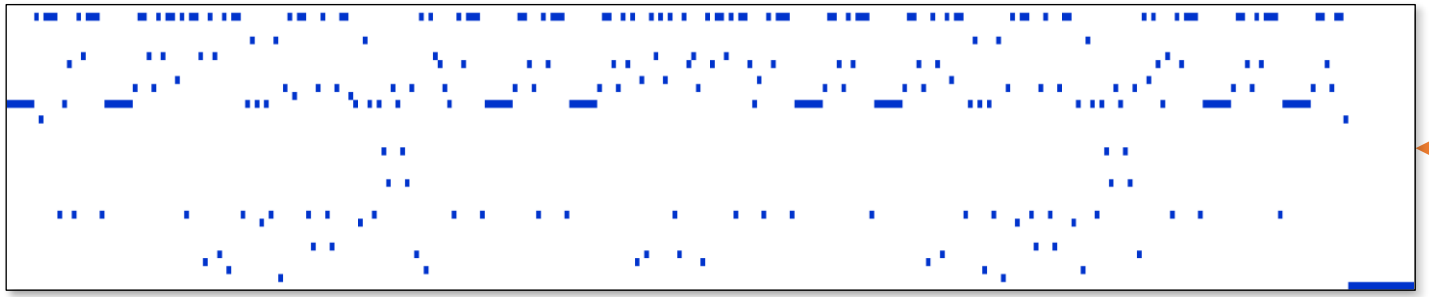


Power-of-2 reduction
$q = 8$

$$v_2(r,q) = \frac{2^r}{2^{q-1}}$$

Organic photovoltaics Donor/Acceptors

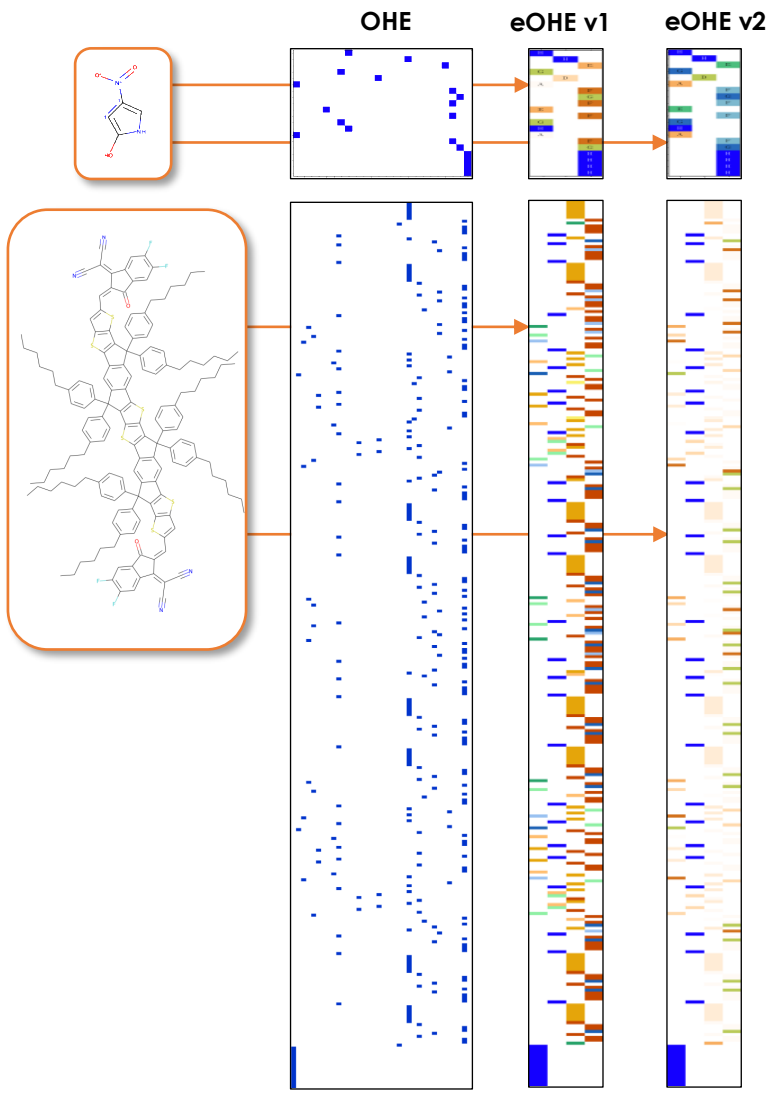


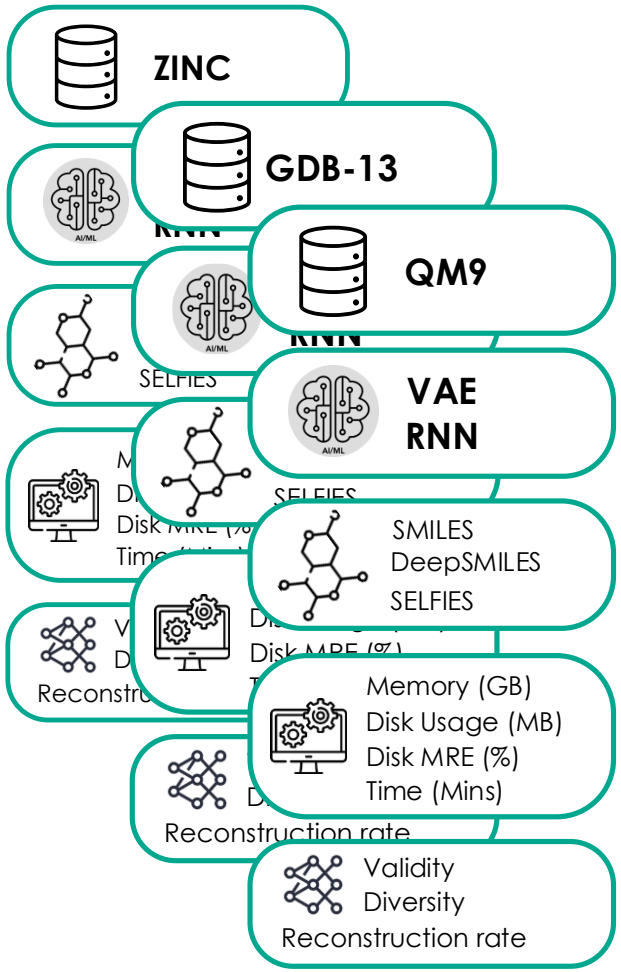


CCCCCCc1ccc(C2(c3ccc(CCCCCC)cc3)c3cc4c(cc3-c3sc5cc(C=C6C(=O)c7cc(F)c(F)cc7C6=C(C#N)C#N)sc5c32)C(c2ccc(CCCCCC)cc2)(c2ccc(CCCCCC)cc2)c2c-4sc3c4c(sc23)-c2cc3c(cc2C4(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)-c2sc4cc(C=C5C(=O)c6cc(F)c(F)cc6C5=C(C#N)C#N)sc4c2C3(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)cc1

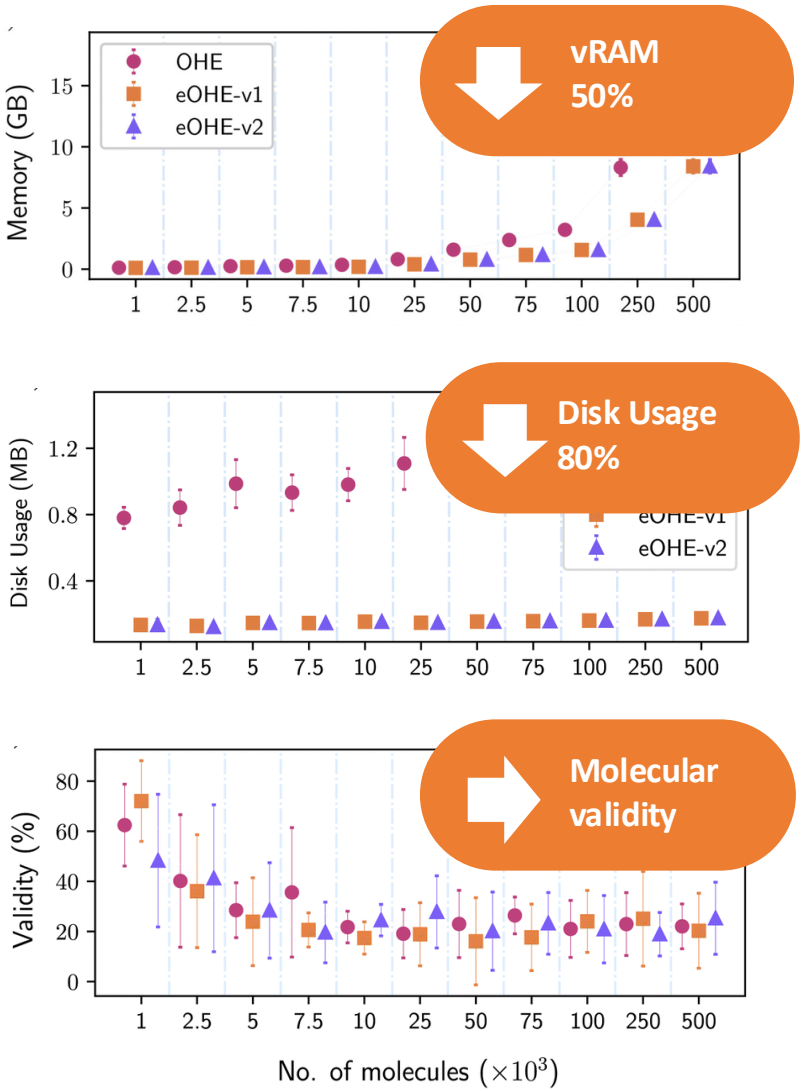
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez,. **Digital Discovery**, 4, 776-789, **2025**

More efficient machine-readable codification





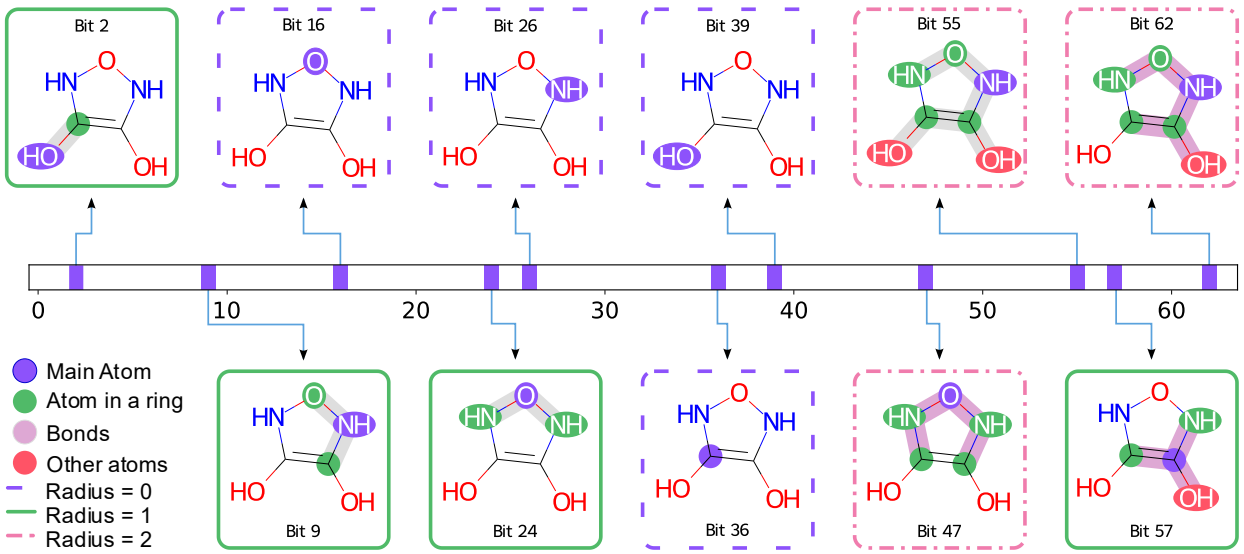
Performance in variational autoencoders



Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez,. **Digital Discovery**, 4, 776-789, **2025**

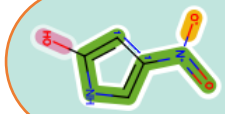
embedded Morgan Fingerprints (eMP)

A Morgan fingerprint is a fixed-size binary vector that encodes the presence (1's) or absence (0's) of substructures in a molecule.



QM9: 134000+ molecules
NFA: 50000 molecules
RedDB: 15000 molecules

Random Forest (RF),
Multi-layer Perceptron (MLP),
K-Neighbors Regressor (KNR),
Gradient Booster Regressor (GBR)
Deep Neural Network (DNN)

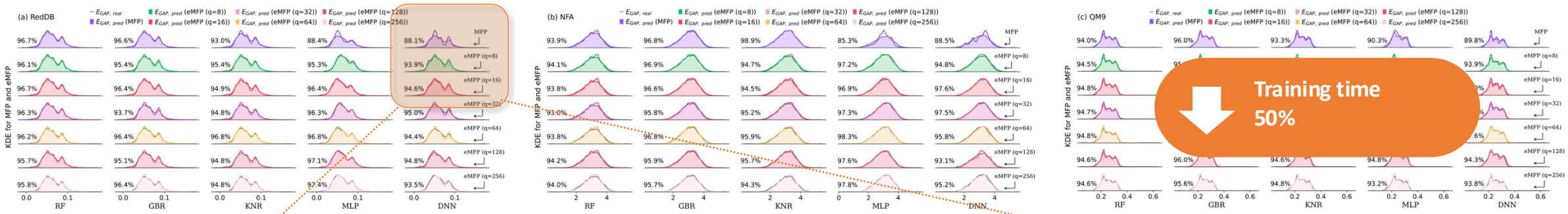


MFP
eMFP (q=8, 16, 32, 64, 128, 256)

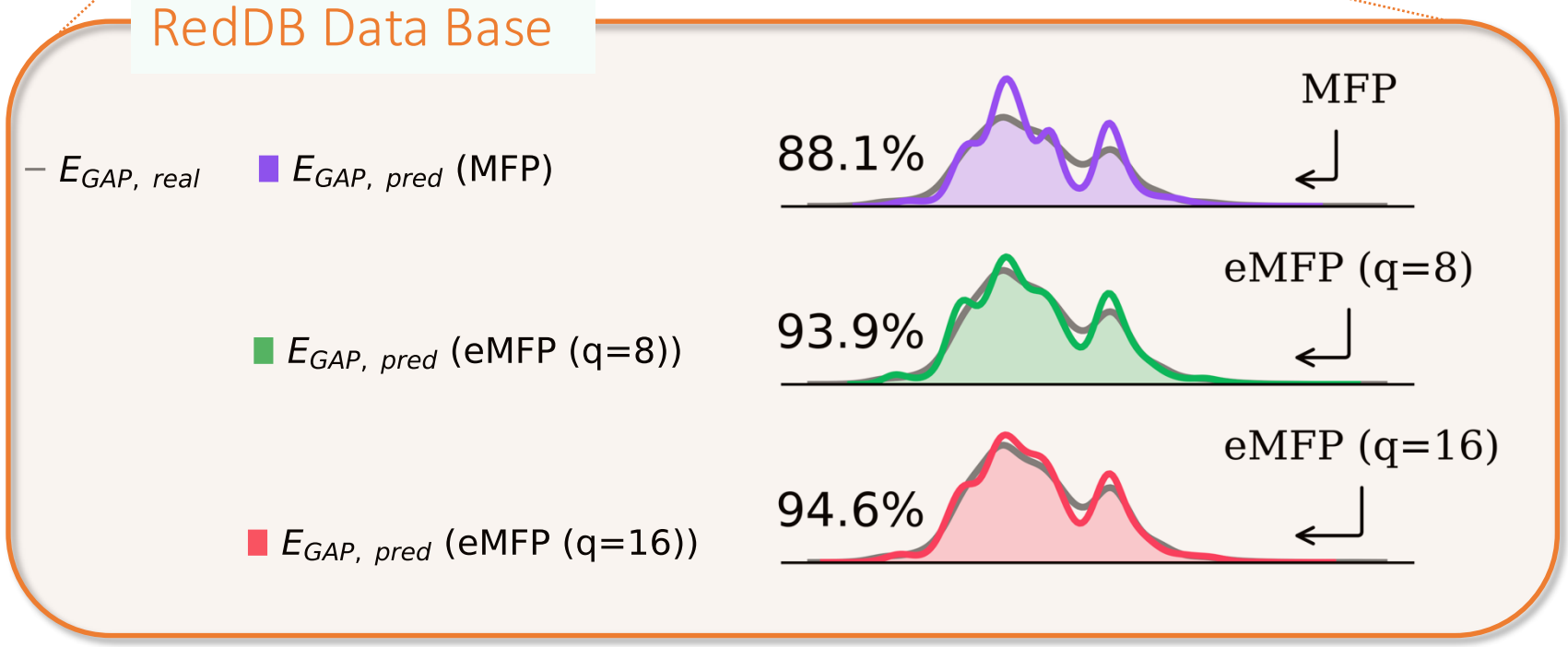
150 trials for15750 hyperparamiters optimisation

Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, *under review,* **2026**
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, **ChemRxiv. 2026**; doi:10.26434/chemrxiv-2025-6hfp8

Energy gap with embedded Morgan Fingerprints (eMP)



RedDB Data Base



Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, *under review,* **2026**
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, **ChemRxiv. 2026**; doi:10.26434/chemrxiv-2025-6hfp8

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES

B) SELFIES

C) InChI

D) None of these

TALKING CHEMISTRY

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES

B) SELFIES

C) InChI

D) None of these

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES

B) SELFIES

C) InChI

D) None of these

Q2 **eOHE / eMFP mainly improve…**

A) Accuracy only

B) Memory / disk / time at matched accuracy

C) Publication count

D) Nothing

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES

B) SELFIES

C) InChI

D) None of these

Q2 **eOHE / eMFP mainly improve…**

A) Accuracy only

B) Memory / disk / time at matched accuracy

C) Publication count

D) Nothing

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES

B) SELFIES

C) InChI

D) None of these

Q2 **eOHE / eMFP mainly improve…**

A) Accuracy only

B) Memory / disk / time at matched accuracy

C) Publication count

D) Nothing

Q3 **Why does the energy cost of AI matter?**

A) It doesn't

B) Sustainability & democratization

C) Compute is free

D) Only for gaming

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **Which representation guarantees a valid molecule for every string?**

A) SMILES
B) **SELFIES**
C) InChI
D) None of these

Q2 **eOHE / eMFP mainly improve…**

A) Accuracy only
B) **Memory / disk / time at matched accuracy**
C) Publication count
D) Nothing

Q3 **Why does the energy cost of AI matter?**

A) It doesn't
B) **Sustainability & democratization**
C) Compute is free
D) Only for gaming

TALKING CHEMISTRY

Take-home messages - Representation is a sustainability decision.

WHY IT MATTERS TO CLIMATE

1 Featurization is a first-class design decision, not a preprocessing afterthought.

2 eOHE / eMFP cut memory, disk and time at matched accuracy
efficiency = greener, more democratized compute.

3 Green, accessible computing is part of responsible acceleration.

LLM-GUIDED MATERIALS DEVELOPMENT

LLM-guided materials development: a new layer on the stack

Beyond predicting numbers, models that read, reason, plan and use tools

The shift: narrow ML maps one structure to one property. A large language model adds a natural-language interface, broad prior knowledge, step-by-step planning, and the ability to call external tools, **turning a predictor into a research collaborator**.

1

Language interface

Ask, adapt with a few examples, strong in the low-data limit where classical ML struggles.

2

World knowledge

Pretrained on the scientific literature, so it brings chemical priors and context, not just curve-fitting.

3

Planning & reasoning

Decompose a goal ("find a stable Na-ion cathode") into ordered, checkable steps.

4

Tool use

Call simulators, databases and robots, grounding answers in real computation instead of guessing.

…but be careful. LLMs are fluent yet unreliable. They hallucinate facts, mis-cite, and can be confidently wrong. Everything depends on **grounding in tools and validation/**

What LLMs actually do in materials development

Six capabilities — each grounded in a concrete, published example

1 Read the literature

Turn thousands of papers into structured, queryable databases (compositions, MOFs, dopants).

e.g. **Dagdelen et al. 2024**

2 Predict properties

Answer chemistry questions in natural language; competitive with ML, strong in the low-data limit.

e.g. **Jablonka et al. 2024**

3 Generate & inverse-design

Write molecules and crystals as text; property-directed generation.

e.g. **Gruver et al. 2024 · Chemeleon**

4 Plan synthesis

Propose synthetic routes and recipes; retrosynthesis and procedure design.

e.g. **ChemCrow 2024**

5 Reason with tools

Augment the model with expert tools (RDKit, calculators, search) so it stops guessing.

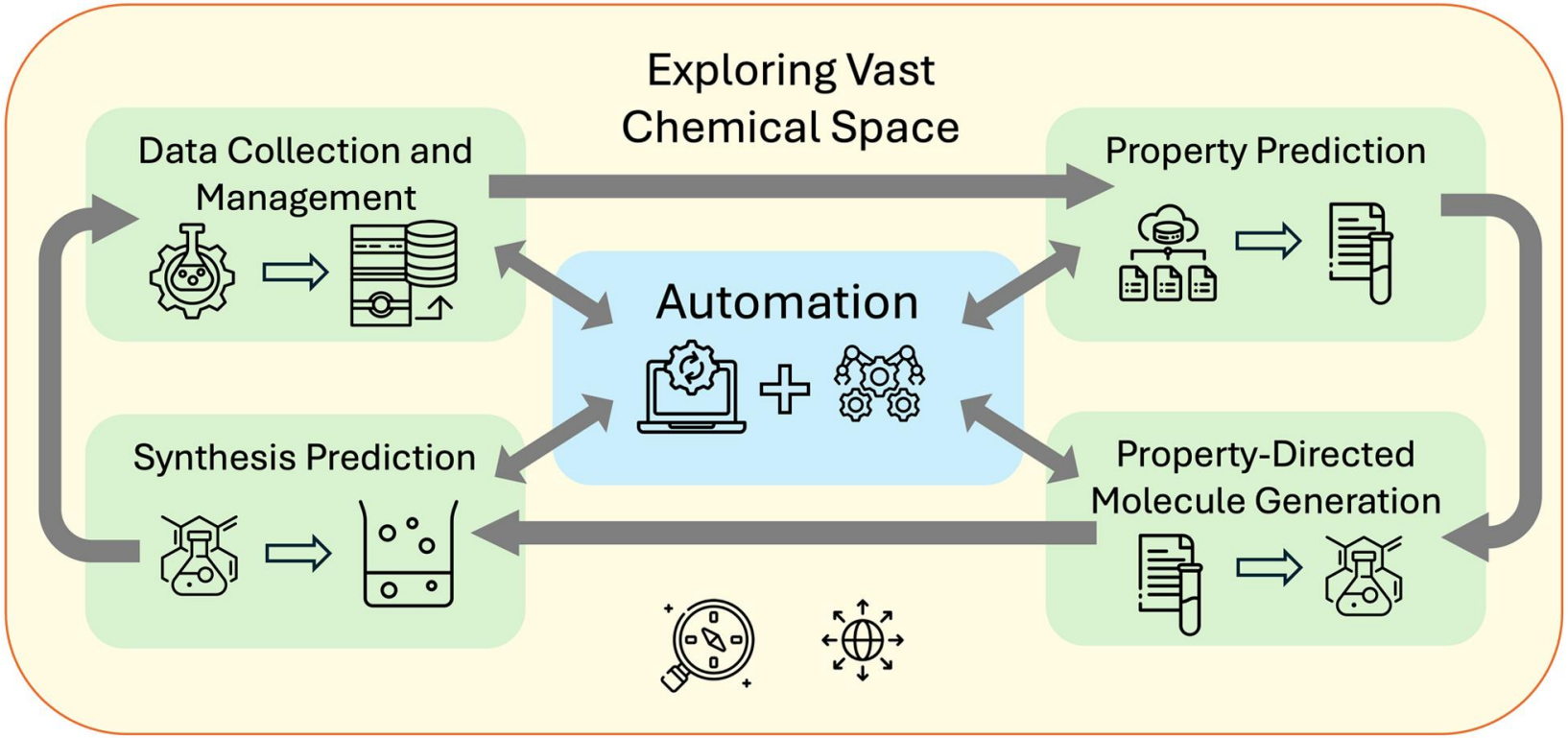
e.g. **ChemCrow — 18 tools**

6 Run experiments autonomously

Plan-act-observe loops that drive instruments and robots to test hypotheses.

e.g. **Coscientist 2023**

What can LLMs and agents actually do in chemistry?

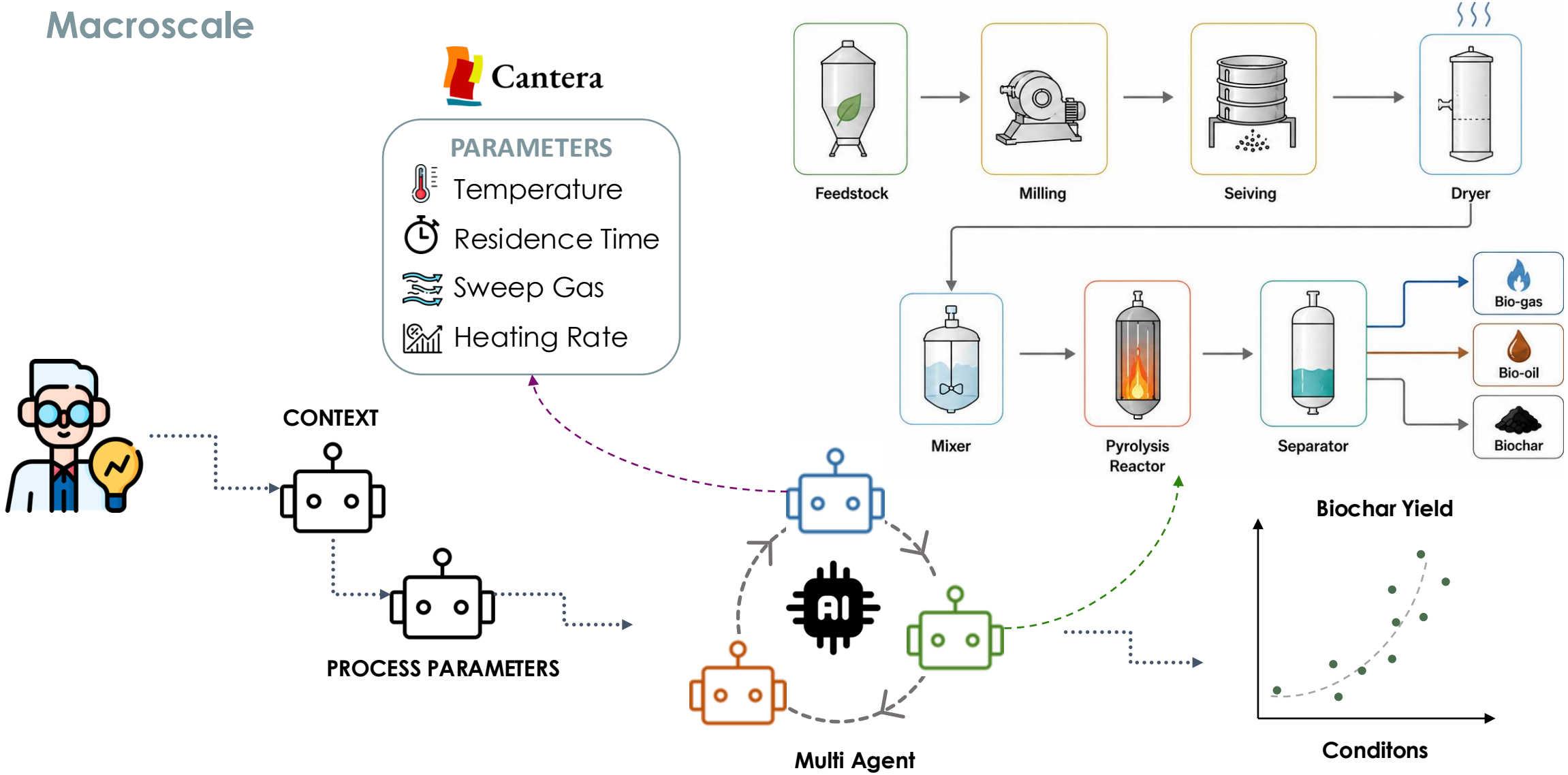


Butler, Davies, Cartwright, Isayev, Walsh, Nature 559, 547–555 (2018).

Modeling biochar – pyrolisis process engineering

LLM-GUIDED DEVELOPMENT

Macroscale



Landmark LLM systems for chemistry & materials

Representative systems. What they do, the approach, and the source

System	What it does	LLM / approach	Reference
Coscientist	designs, plans & runs experiments (e.g. Pd cross-couplings)	GPT-4 + web/docs/code + robotics	*Boiko et al., Nature 2023*
ChemCrow	organic synthesis, drug & materials design; ran real syntheses	GPT-4 + 18 expert tools	*Bran et al., Nat. Mach. Intell. 2024*
Predictive-chem LLM	property prediction & classification from text; low-data	fine-tuned GPT-3; in-context	*Jablonka et al., Nat. Mach. Intell. 2024*
Struct. extraction	reads papers → structured records (dopants, MOFs, compositions)	fine-tuned GPT-3 / Llama-2	*Dagdelen et al., Nat. Commun. 2024*
Crystal-text LLM	generates stable inorganic crystals written as text	fine-tuned LLaMA-2 70B	*Gruver et al., ICLR 2024*
Chemeleon	text-guided crystal-structure generation	contrastive text–crystal + diffusion	*Park et al., Nat. Commun. 2025*

References

LLMs and autonomous agents for chemistry & materials

1. Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous chemical research with large language models (Coscientist). Nature 624, 570–578 (2023). 10.1038/s41586-023-06792-0

2. M. Bran, A.; Cox, S.; Schilter, O.; Baldassari, C.; White, A. D.; Schwaller, P. Augmenting large language models with chemistry tools (ChemCrow). Nat. Mach. Intell. 6, 525–535 (2024). 10.1038/s42256-024-00832-8

3. Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging large language models for predictive chemistry. Nat. Mach. Intell. 6, 161–169 (2024). 10.1038/s42256-023-00788-1

4. Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; Jain, A. Structured information extraction from scientific text with large language models. Nat. Commun. 15, 1418 (2024). 10.1038/s41467-024-45563-x

5. Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-Tuned Language Models Generate Stable Inorganic Materials as Text. ICLR 2024. arXiv:2402.04379

6. Ramos, M. C.; Collison, C. J.; White, A. D. A review of large language models and autonomous agents in chemistry. Chem. Sci. 16, 2514–2572 (2025). 10.1039/D4SC03921A

7. Park, H.; Onwuli, A.; Walsh, A. Exploration of crystal chemical space using text-guided generative AI (Chemeleon). Nat. Commun. 16, 4379 (2025). 10.1038/s41467-025-59636-y

8. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 571, 95–98 (2019). 10.1038/s41586-019-1335-8

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions

B) Plans and acts via tools in a loop

C) Never uses tools

D) Runs only offline

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions

B) Plans and acts via tools in a loop

C) Never uses tools

D) Runs only offline

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions

B) Plans and acts via tools in a loop

C) Never uses tools

D) Runs only offline

Q2 **Adding tools (RDKit, search, simulators) to an LLM…**

A) Makes it worse

B) Grounds it & improves reliability

C) Removes all errors

D) Is impossible

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions

B) Plans and acts via tools in a loop

C) Never uses tools

D) Runs only offline

Q2 **Adding tools (RDKit, search, simulators) to an LLM…**

A) Makes it worse

B) Grounds it & improves reliability

C) Removes all errors

D) Is impossible

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions
B) Plans and acts via tools in a loop
C) Never uses tools
D) Runs only offline

Q2 **Adding tools (RDKit, search, simulators) to an LLM…**

A) Makes it worse
B) Grounds it & improves reliability
C) Removes all errors
D) Is impossible

Q3 **LLMs are especially useful when…**

A) Data are huge
B) Data are scarce (low-data limit)
C) Never
D) Only for images

ENGAGEMENT

SECTION CLOSER · QUIZ (multiple choice)

Multiple choice

Q1 **An LLM “agent” (vs assistant) is one that…**

A) Only answers questions
B) Plans and acts via tools in a loop
C) Never uses tools
D) Runs only offline

Q2 **Adding tools (RDKit, search, simulators) to an LLM…**

A) Makes it worse
B) Grounds it & improves reliability
C) Removes all errors
D) Is impossible

Q3 **LLMs are especially useful when…**

A) Data are huge
B) Data are scarce (low-data limit)
C) Never
D) Only for images

LLM-GUIDED DEVELOPMENT

Take-home messages - Unlock the knowledge trapped in text.

WHY IT MATTERS TO CLIMATE

1. LLMs add a language + reasoning + tool-use layer over narrow ML.

2. Mine decades of literature into usable data; plan syntheses; orchestrate tools.

3. Keep a human accountable and log every action. Provenance and oversight are what make autonomy trustworthy (and reproducible).

RESPONSIBLE AI

Responsible AI for materials & chemistry

Powerful tools bring responsibilities. The same acceleration cuts both ways

1 Integrity & reproducibility

Data leakage and benchmark-chasing inflate results.
Report enough that others can reproduce them.

2 Trustworthy predictions

Report uncertainty, respect the applicability domain, and validate downstream (DFT → experiment).

3 Data quality & bias

Datasets skew to well-studied, successful systems.
Garbage in, garbage out

4 Openness & FAIR

Share data, code and model weights
Findable, Accessible, Interoperable, Reusable (FAIR).

5 Safety & dual use

Powerful design tools can be inverted for harm.
Access control and human oversight.

6 Sustainability & compute

Training and inference cost energy.
Right-size models and reuse pretrained ones.

Data biases

Challenges

- Implicit beliefs, and biases from Western contexts may be normalized in AI, making it less globally responsible.
- A need to employ a global lens in core AI assumptions: training data, model performance, and **explainability** of systems
- Ask the difficult questions on positive and negative impacts early on, rather than introducing repairs and post-hoc fixes.

Actions

- Be Robust to Multilingualism and Diverse Literacies
- Personalize Experiences with the Shared Identity Model
- Explainability and Transparency Should Make Sense for New Technology Users
- Proactively Address Fairness in Data Skews

Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis

Xiwen Jia, Allyson Lynch, Yuheng Huang, Matthew Danielson, Immaculate Lang'at, Alexander Milder, Aaron E. Ruby, Hao Wang, Sorelle A. Friedler, Alexander J. Norquist & Joshua Schrier

Nature **573**, 251–255 (2019) | Cite this article

Machine-learning-assisted materials discovery using failed experiments

Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler, Joshua Schrier & Alexander J. Norquist

Nature **533**, 73–76 (2016) | Cite this article

ACM interactions, vol. 26, 2019, pp. 68-71

References

Responsible and trustworthy AI for science

1. Urbina, F.; Lentzos, F.; Invernizzi, C.; Ekins, S. Dual use of artificial-intelligence-powered drug discovery. Nat. Mach. Intell. 4, 189–191 (2022). 10.1038/s42256-022-00465-9

2. Kapoor, S.; Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns 4, 100804 (2023). 10.1016/j.patter.2023.100804

3. Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best practices in machine learning for chemistry. Nat. Chem. 13, 505–508 (2021). 10.1038/s41557-021-00716-z

4. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016). 10.1038/sdata.2016.18

5. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. Proc. ACL 2019, 3645–3650. arXiv:1906.02243

6. Miret, S.; Krishnan, N. M. A. Are LLMs Ready for Real-World Materials Discovery? arXiv:2402.05200 (2024).

RESPONSIBLE AI

ENGAGEMENT

SECTION CLOSER · QUIZ (true / false)

True or false?

1. **Published chemistry datasets are biased partly because failed experiments are rarely reported.**

   True | False

2. **A drug-design model can be inverted to design toxins.**

   True | False

3. **FAIR stands for Findable, Accessible, Interoperable, Reusable.**

   True | False

Take-home messages - Acceleration only helps if it is trustworthy and shared.

WHY IT MATTERS TO CLIMATE

1. Data are biased: publication bias, human bias, Western-context defaults.
2. Powerful generative/agentic tools are inherently dual-use.
3. Reproducibility, openness (FAIR) and compute cost are part of doing this well.

EXPLAINABLE AI

Explainable AI: from prediction to understanding

In science we usually want the "why", not just the number

The gap: a model can predict a property accurately and still teach you nothing. Explainability asks what the model learned, whether it is for the right reasons, and what new science it implies.

1 Understanding

Turn a fitted model into design rules and mechanisms. Recover the physics/chemistry behind the data.

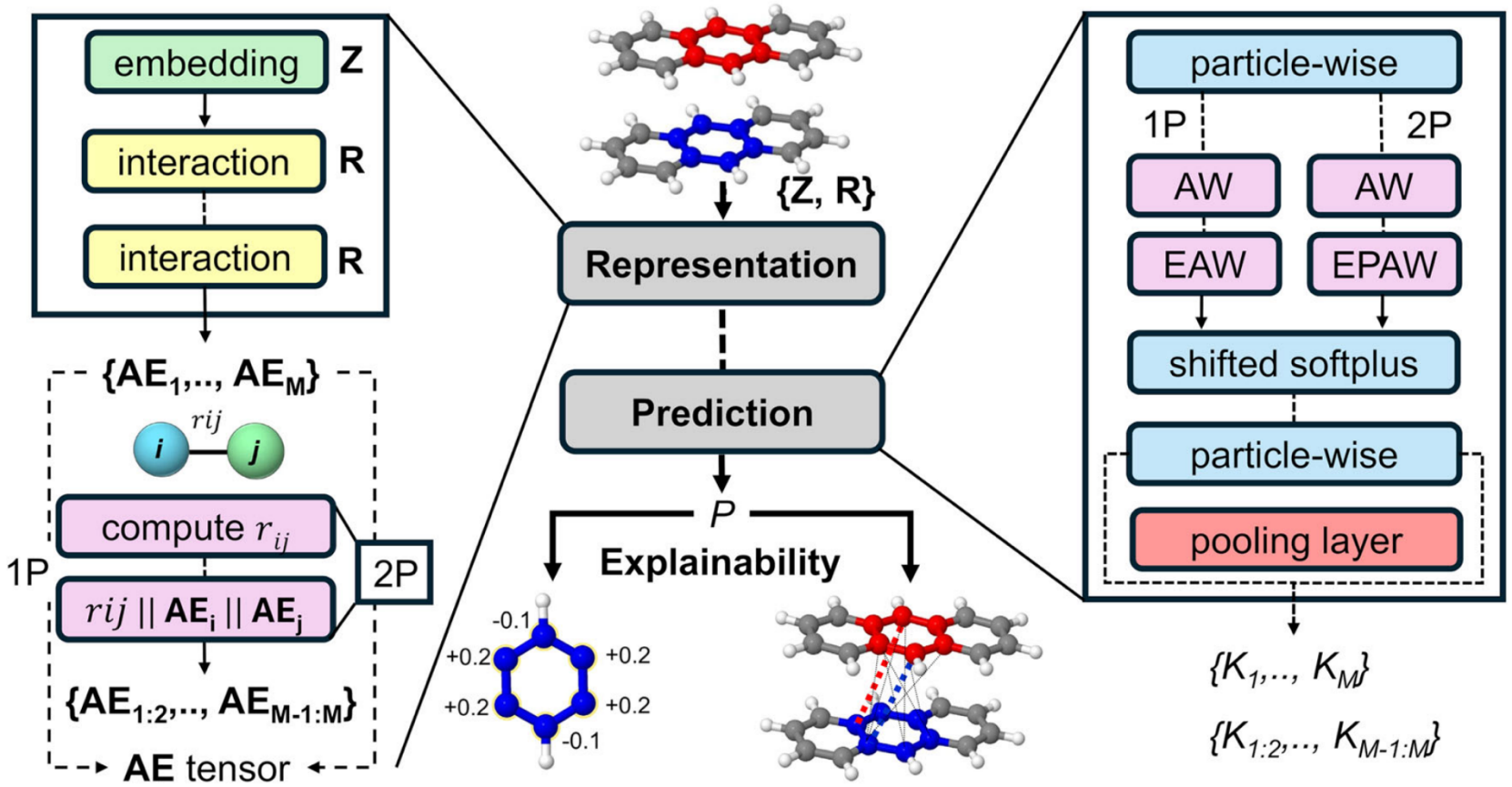
2 Trust & validation

Check the model predicts for the right reasons, not false correlations or dataset artefacts.

3 Discovery

Interpretation surfaces new structure–property relationships and testable hypotheses.

Explainable Chemical AI: SchNet4AIM



Gallegos, M.; Vassilev-Galindo, V.; Poltavsky, I.; Martín Pendás, Á.; Tkatchenko, A. "Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors." Nature Communications 15, 4326

References

Explainable & interpretable AI for science and materials

1. Ghiringhelli, L. M.; Vybíral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. Phys. Rev. Lett. 114, 105503 (2015). 10.1103/PhysRevLett.114.105503

2. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor. Phys. Rev. Materials 2, 083802 (2018). 10.1103/PhysRevMaterials.2.083802

3. Udrescu, S.-M.; Tegmark, M. AI Feynman: A physics-inspired method for symbolic regression. Sci. Adv. 6, eaay2631 (2020). 10.1126/sciadv.aay2631

4. Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier (LIME). KDD 2016. arXiv:1602.04938

5. Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions (SHAP). NeurIPS 2017. arXiv:1705.07874

6. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks (Integrated Gradients). ICML 2017. arXiv:1703.01365

7. Ying, R.; Bourgeois, D.; You, J.; Zitnik, M.; Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. NeurIPS 2019. arXiv:1903.03894

8. Oviedo, F.; Ferres, J. L.; Buonassisi, T.; Butler, K. T. Interpretable and Explainable Machine Learning for Materials Science and Chemistry. Acc. Mater. Res. 3, 597–607 (2022). 10.1021/accountsmr.1c00244

9. Roscher, R.; Bohn, B.; Duarte, M. F.; Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8, 42200–42216 (2020). 10.1109/ACCESS.2020.2976199

EXPLAINABLE AI

Take-home messages - Turn fast predictions into design rules we can trust.

WHY IT MATTERS TO CLIMATE

1. In science we want the why, not just the number.

2. Interpretable methods expose models that are right for the wrong reasons.

3. Treat every explanation as a hypothesis to verify against physics and experiment.

FINAL REFLECTIONS

EDITORIAL | 19 May 2026

nature

Why AI cannot do good science without humans

With the arrival of 'AI scientists', it's as well to remember that human wisdom, empathy and sheer messiness are as much part of progress as are process and efficiency.

Design for circularity
Design for performance
as well as performance



AI/ML

Online repositories — open-source toolkit for AI-accelerated materials science

A shareable starting kit. Our group's tools are highlighted; verify links are current before distributing.

Area	Tool	What it's for	Repository
Data & tools	**pymatgen**	Materials analysis, I/O & Materials Project API	*github.com/materialsproject/pymatgen*
Data & tools	**matminer**	Ready-made descriptors / featurization for ML	*github.com/hackingmaterials/matminer*
Data & tools	**RDKit**	Cheminformatics: molecules, fingerprints, descriptors	*github.com/rdkit/rdkit*
Data & tools	**DeepChem**	ML-for-chemistry toolkit: datasets, models, benchmarks	*github.com/deepchem/deepchem*
Data & tools	**ASE**	Atomistic Simulation Environment: build / run / analyze	*gitlab.com/ase/ase*
Representations	**SELFIES**	Always-valid molecular string representation	*github.com/aspuru-guzik-group/selfies*
Representations	**eOHE**	Embedded one-hot encodings (efficient inputs)	*github.com/MMLabCodes/eOHE*
Representations	**eMFP**	Embedded Morgan fingerprints (compact, faster)	*github.com/MMLabCodes/eMFP*
Generative	**chemical_vae**	Continuous latent-space molecular design	*github.com/aspuru-guzik-group/chemical_vae*
Generative	**MatterGen**	Diffusion generative model for inorganic crystals	*github.com/microsoft/mattergen*
ML potentials	**MACE**	ML interatomic potentials (+ MACE-MP-0 foundation)	*github.com/ACEsuit/mace*
ML potentials	**CHGNet**	Universal charge-informed neural network potential	*github.com/CederGroupHub/chgnet*
ML potentials	**MatGL (M3GNet)**	Universal graph potentials + property models	*github.com/materialsvirtuallab/matgl*
Autonomy	**Ax / BoTorch**	Bayesian optimization for self-driving labs	*github.com/facebook/Ax*
LLM agents	**ChemCrow**	LLM chemistry agent wired to expert tools	*github.com/ur-whitelab/chemcrow-public*
Explainability	**SHAP**	Feature-attribution explanations for any model	*github.com/shap/shap*
Explainability	**SISSO**	Interpretable descriptor / formula discovery	*github.com/rouyang2017/SISSO*
Green computing	**CodeCarbon**	Track the carbon cost of your compute	*github.com/mlco2/codecarbon*

Highlighted rows = our group's repositories (github.com/MMLabCodes). List is representative, not exhaustive.

Further reading

MATERIALS BOTTLENECK

Yao, Lum, Johnston, … Seh. Nat. Rev. Mater. 8, 202–215 (2023). 10.1038/s41578-022-00490-5

Mroz et al.. J. Am. Chem. Soc. (2022).

Bachs-Herrera, …, Martín-Martínez. iScience 26, 106549 (2023). 10.1016/j.isci.2023.106549

MODELING & SIMULATION

Coley, Eyke, Jensen. Angew. Chem. Int. Ed. 59, 23414–23436 (2020). 10.1002/anie.201909987

A NEW PARADIGM

Butler, Davies, Cartwright, Isayev, Walsh. Nature 559, 547–555 (2018). 10.1038/s41586-018-0337-2

Sanchez-Lengeling & Aspuru-Guzik. Science 361, 360–365 (2018). 10.1126/science.aat2663

Gómez-Bombarelli et al. ACS Cent. Sci. 4, 268–276 (2018). 10.1021/acscentsci.7b00572

Cheng, Ser, Skreta, … Aspuru-Guzik. Faraday Discuss. 256, 10–60 (2025). 10.1039/D4FD00153B

THE AI TOOLKIT

Riebesell et al.. Nat. Mach. Intell. 7, 836–847 (2025). 10.1038/s42256-025-01055-1

Park, Onwuli, Walsh. Nat. Commun. 16, 4379 (2025). 10.1038/s41467-025-59636-y

Szymanski et al.. Nature 624, 86–91 (2023). 10.1038/s41586-023-06734-w

TALKING CHEMISTRY

Krenn et al.. Patterns 3, 100588 (2022). 10.1016/j.patter.2022.100588

Gómez-Bombarelli et al. ACS Cent. Sci. 4, 268–276 (2018). 10.1021/acscentsci.7b00572

Lannelongue, Grealey, Inouye. Adv. Sci. 8, 2100707 (2021). 10.1002/advs.202100707

Nuñez-Andrade, Vidal-Daza, Gómez-Bombarelli, Ryan, Martín-Martínez. Digital Discovery 4, 776–789 (2025).

Nuñez-Andrade, … Martín-Martínez (eMFP). ChemRxiv 2025. doi:10.26434/chemrxiv-2025-6hfp8

RESPONSIBLE AI

Urbina, Lentzos, Invernizzi, Ekins. Nat. Mach. Intell. 4, 189–191 (2022). 10.1038/s42256-022-00465-9

Artrith et al. Best practices in ML for chemistry. Nat. Chem. 13, 505–508 (2021). 10.1038/s41557-021-00716-z

Wilkinson et al. (FAIR). Sci. Data 3, 160018 (2016). 10.1038/sdata.2016.18

Strubell, Ganesh, McCallum. Energy & policy considerations for deep learning. ACL 2019. arXiv:1906.02243

EXPLAINABLE AI

Ghiringhelli, Vybíral, Levchenko, Draxl, Scheffler. Phys. Rev. Lett. 114, 105503 (2015). 10.1103/PhysRevLett.114.105503

Oviedo, Ferres, Buonassisi, Butler. Acc. Mater. Res. 3, 597–607 (2022). 10.1021/accountsmr.1c00244

Thank you

