L’IA per accelerare la scienza dei materiali

Francisco Martin-Martinez

Senior Lecturer, King’s College London

martinmartinezlab.com

SCUOLA ESTIVA VIRTUALE CCAI 2026 · LEZIONE SETTORIALE · 13 AGOSTO

CHI SONO?

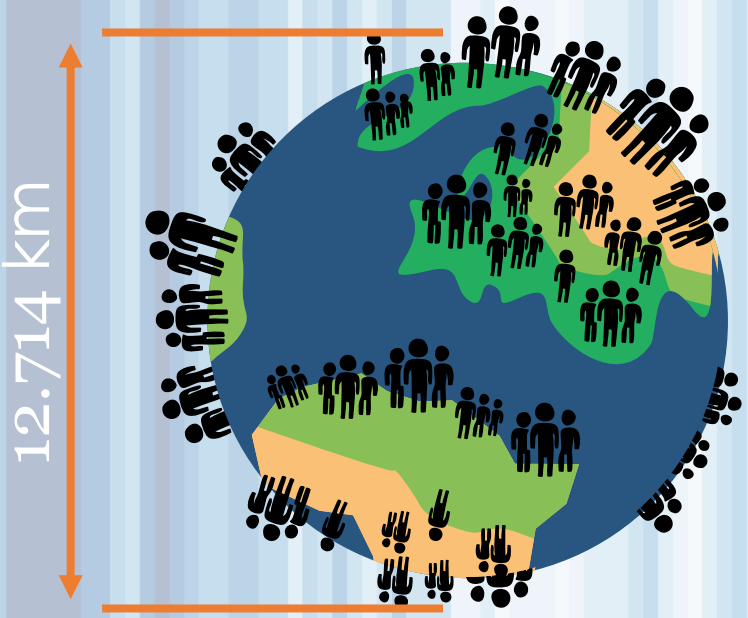
Chi sono e perché questo tema mi sta a cuore



Uno scienziato che usa l'IA come strumento o un'IA che fa scienza in modo autonomo: dov'è il confine?

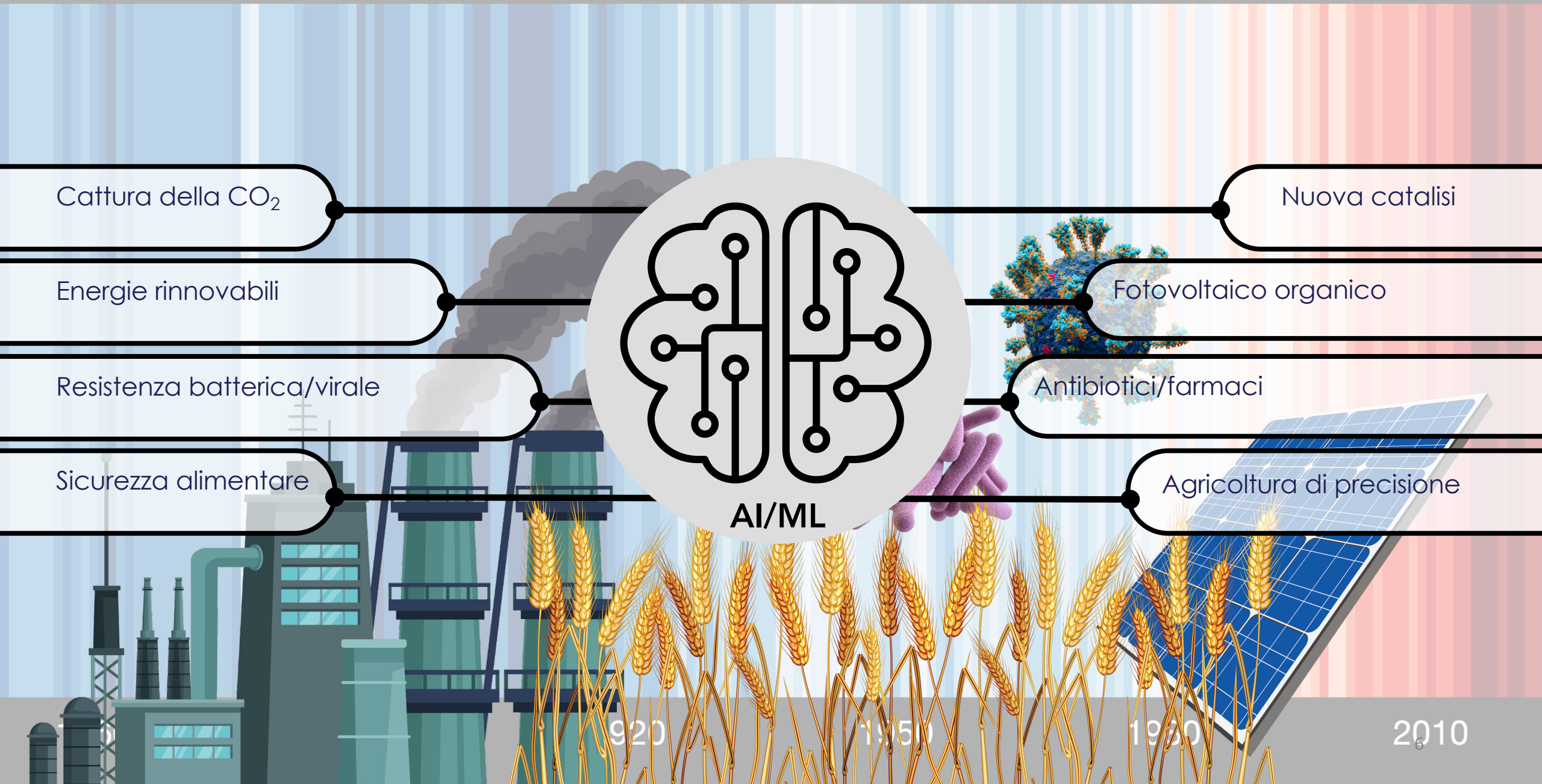
PERCHÉ MI STA A CUORE?

Global temperature change (1850-2022)



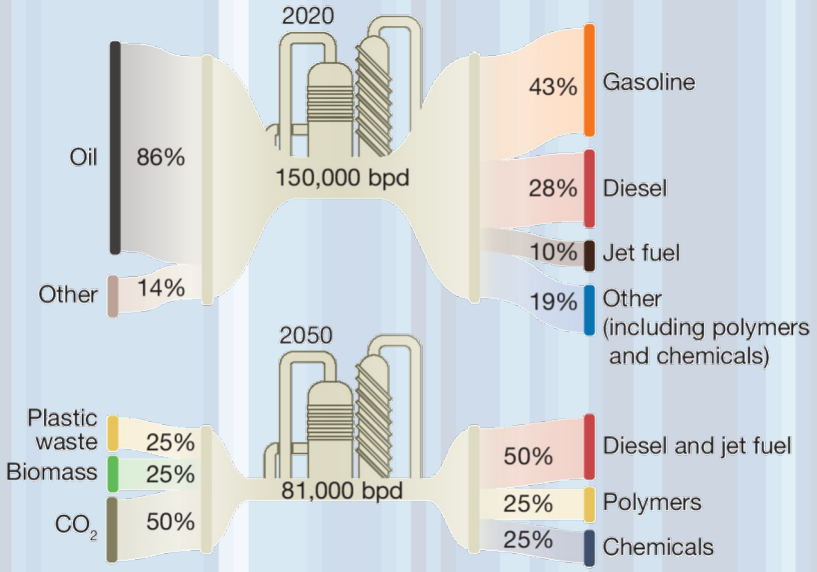
1860 1890 1920 1950 1980 2010

Global temperature change (1850-2022)



Nuova estrazione di materiali dai rifiuti

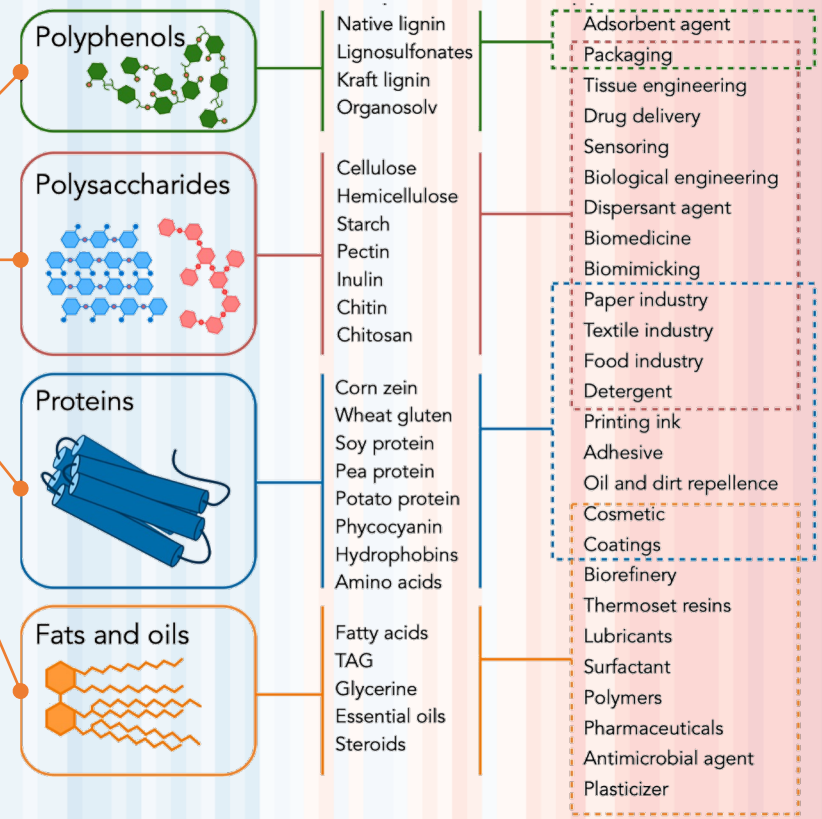
(Bio)raffineria 2050



Nature 629, 295, **2024**

Rifiuti di biomassa
140 Gt/anno

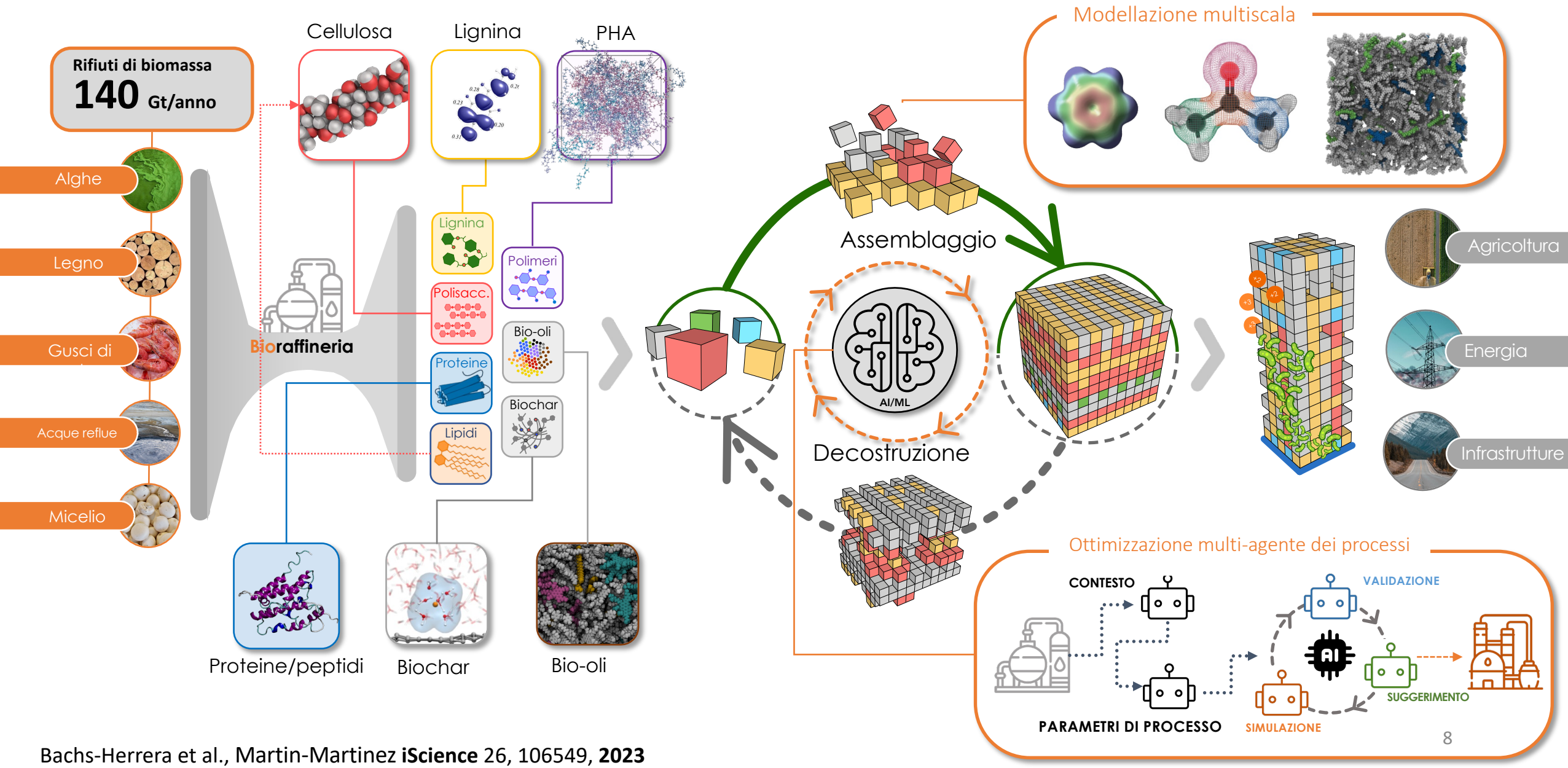
AI/ML



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

Valorizzazione dei rifiuti in molecole e materiali a base biologica

Materiali circolari biobased ➡ Verso una bioraffineria quantistica guidata dall'IA



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

DI COSA PARLEREMO OGGI?

DI COSA PARLEREMO

01 Perché i materiali sono un collo di bottiglia climatico

02 Il paradigma tradizionale (ancora utile) della modellazione e simulazione dei materiali

03 Un nuovo paradigma nella scoperta di materiali

04 La cassetta degli attrezzi dell'IA

05 Parlare di chimica ai computer

04 IA responsabile

05 IA spiegabile in chimica

PARTECIPAZIONE

Sondaggio dal vivo – Usate la chat

Quale sfida legata ai materiali è più urgente per il clima?
Non ci sono risposte sbagliate: serve a rompere il ghiaccio

DOMANDE?

IL COLLO DI BOTTIGLIA DEI MATERIALI

COLLO DI BOTTIGLIA

Dove interviene l’IA nella filiera dei materiali

1

Rappresentare

Codificare molecole/cristalli per l’AA (fingerprint, grafi OHE, testo)

2

Prevedere

Previsione di proprietà, modelli surrogati, **potenziali interatomici di AA**

3

Generare

Progettazione inversa/generativa di nuovi candidati

4

Sintetizzare

Laboratori autonomi e a guida autonoma che chiudono il ciclo

5

Caratterizzare

Analisi assistita dall’IA di spettri, immagini e strutture

COLLO DI BOTTIGLIA

I materiali, un collo di bottiglia climatico

Fotovoltaico

Efficienza e stabilità limitate dalla chimica dei materiali

Batterie

Densità energetica, durata, dipendenza dai minerali critici

Combustibili alternativi

Catalizzatori e vettori per H_2 verde ed e-fuel

Sorbenti di CO_2

Materiali di cattura economici, durevoli e scalabili

Altro: infrastrutture sostenibili, sicurezza alimentare, scoperta di farmaci, antibiotici…

Il problema: il ciclo di scoperta è lento e costoso.

Progettare → sintetizzare → caratterizzare → testare può richiedere anni per ogni candidato.

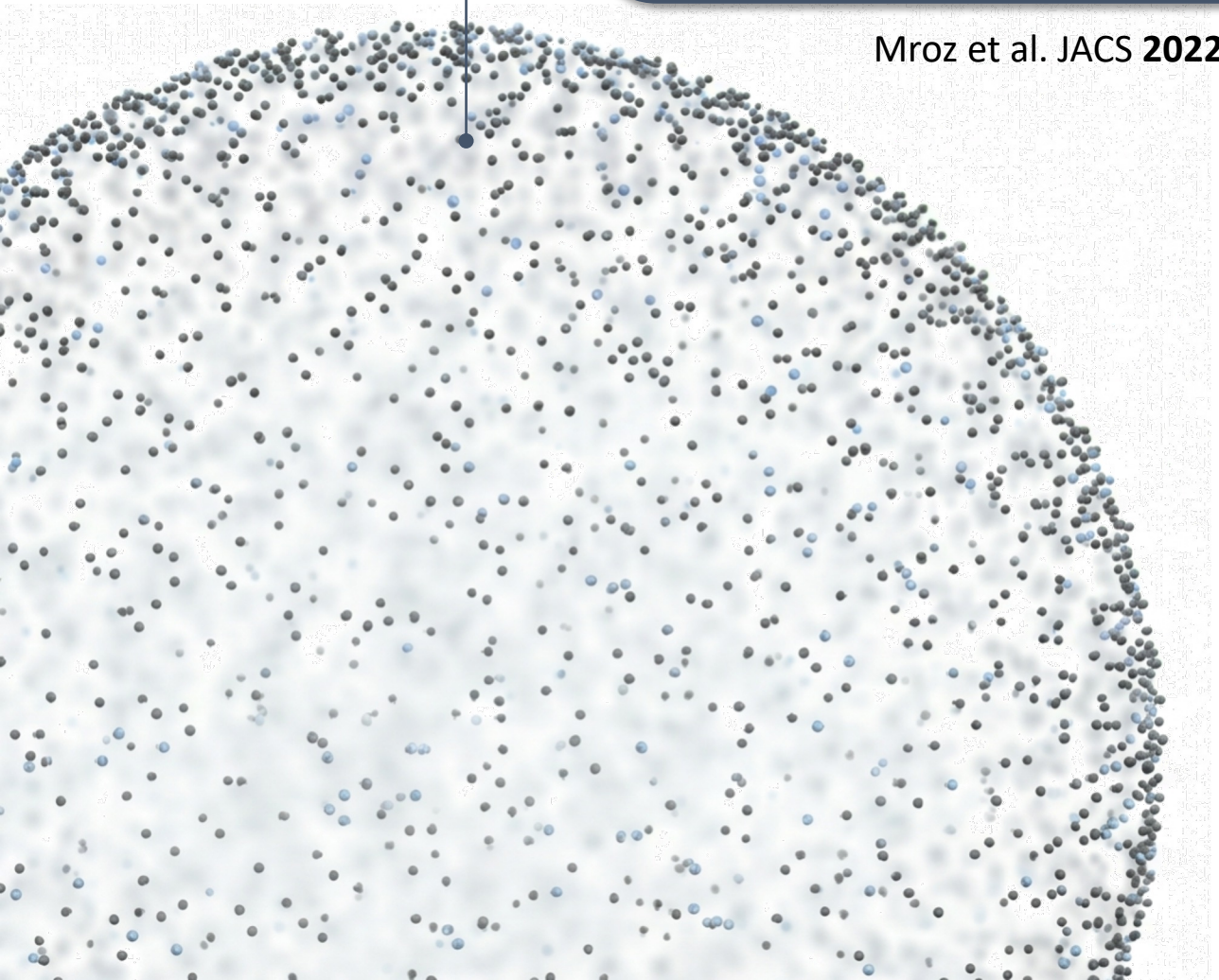
Sia la mitigazione SIA l'adattamento dipendono dall'ottenere i materiali giusti, più in fretta e a costi minori.

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

Il collo di bottiglia della scoperta

10^{60} Piccole molecole (simil-farmaco) per la progettazione di materiali

Mroz et al. JACS **2022**



Lo spazio chimico combinatorio è immenso

VS

La validazione sperimentale è lenta e costosa

LO SAPEVATE?

Esistono più piccole molecole organiche possibili che atomi nell'universo osservabile.

PARTECIPAZIONE

Quiz lampo – Vero/Falso

1. **Esistono più piccole molecole organiche possibili che atomi nell'universo osservabile.**

   Vero | Falso

2. **Un nuovo materiale arriva di solito sul mercato in poche settimane.**

   Vero | **Falso**

3. **L'IA interviene nella filiera nelle fasi rappresentare, prevedere, generare, sintetizzare e caratterizzare.**

   Vero | Falso

Messaggi chiave

PERCHÉ È IMPORTANTE PER IL CLIMA

1. I materiali sono un collo di bottiglia climatico: sia la mitigazione sia l'adattamento richiedono materiali migliori, più in fretta.

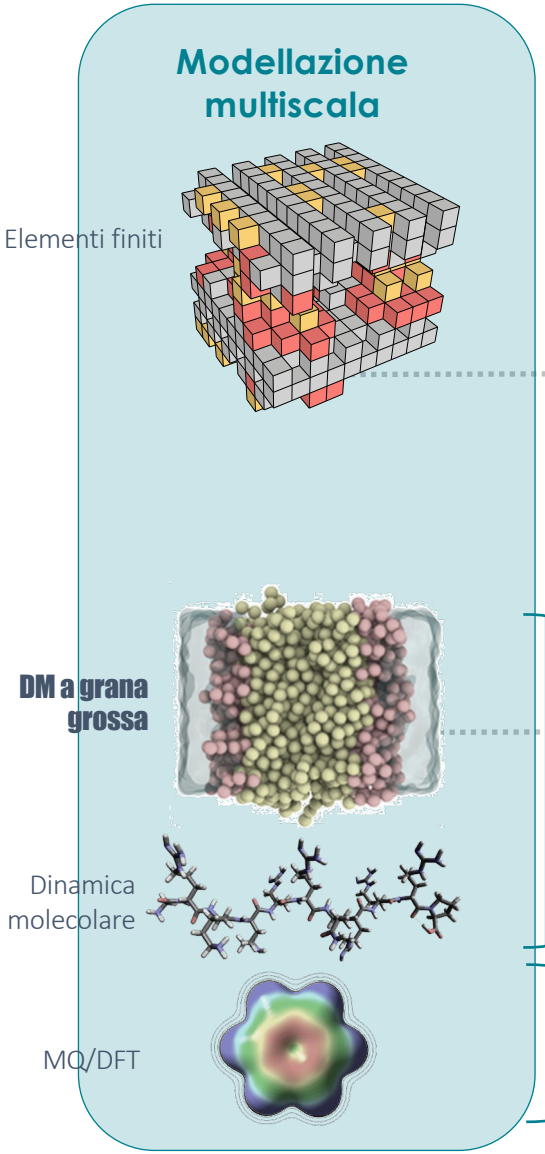
2. Lo spazio chimico non si può esplorare con la forza bruta. L'IA lo trasforma in una ricerca guidata.

3. Gli obiettivi climatici hanno scadenze fisse. Scoprire più in fretta significa diffondere più in fretta.

DOMANDE?

MODELLAZIONE E SIMULAZIONE DEI MATERIALI

Modellazione multiscala



Simulazioni di dinamica molecolare

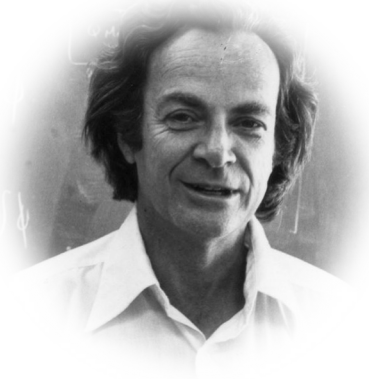


«Tutti i modelli sono sbagliati, ma alcuni sono utili».

– George E. P. Box



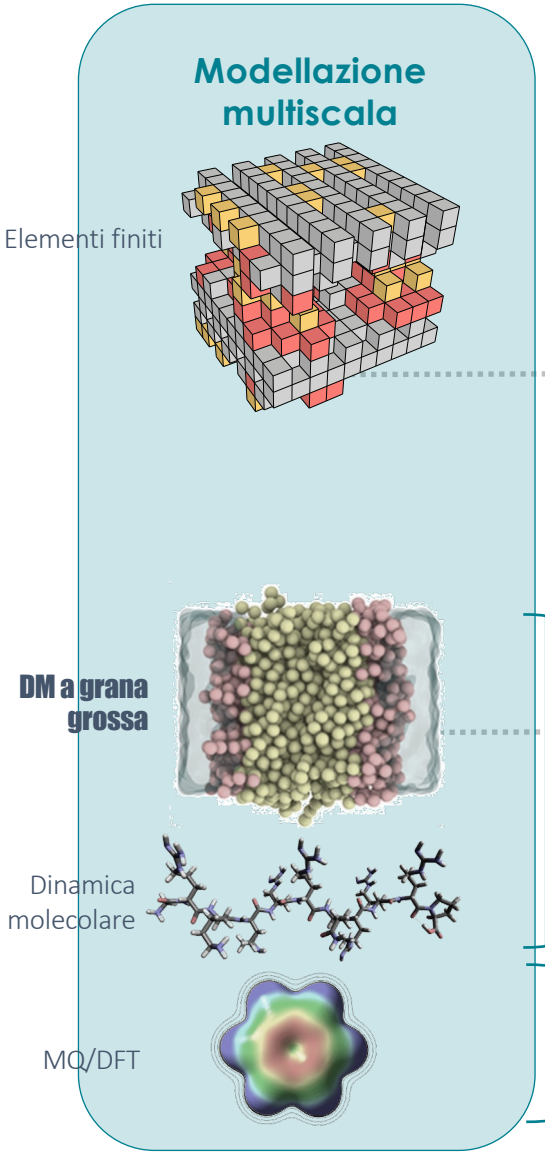
Teoria del funzionale del



«Penso di poter dire con sicurezza che nessuno capisce la meccanica quantistica».

– Richard Feynman

Modellazione multiscala



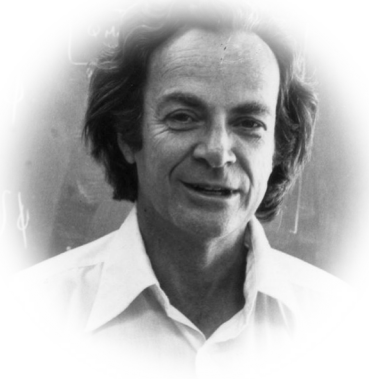
Simulazioni di dinamica molecolare



«Tutti i modelli sono sbagliati, ma alcuni sono utili».

– George E. P. Box

Teoria del funzionale della densità (DFT) – Meccanica quantistica

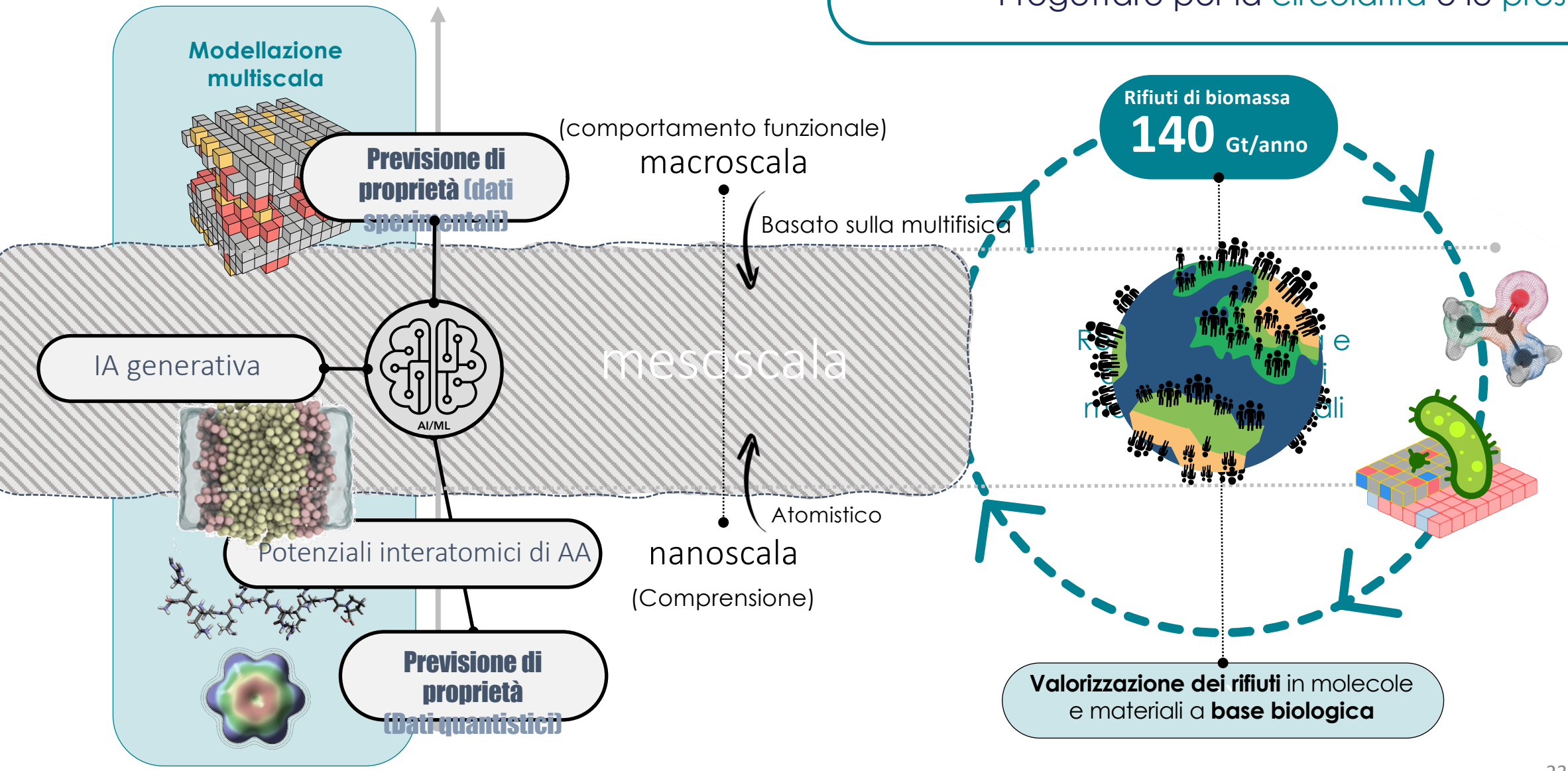


«Penso di poter dire con sicurezza che nessuno capisce la meccanica quantistica».

– Richard Feynman

Modellazione multiscala

Progettare per la circolarità e le prestazioni



Software di DFT, DM e AA per la chimica

DFT molecolare



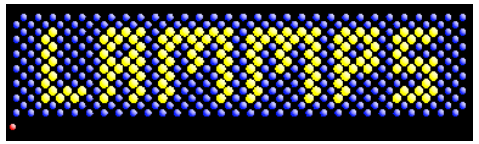
ORCA

DFT dello stato solido





DM (tutti gli atomi e grana grossa)





NAMD
Scalable Molecular Dynamics







Integrazione e chemioinformatica



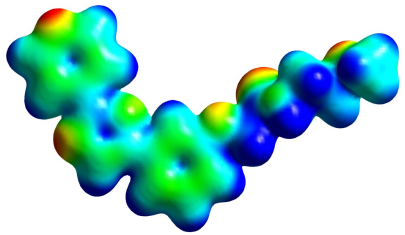


+ Atomic Simulation Environment (ASE)

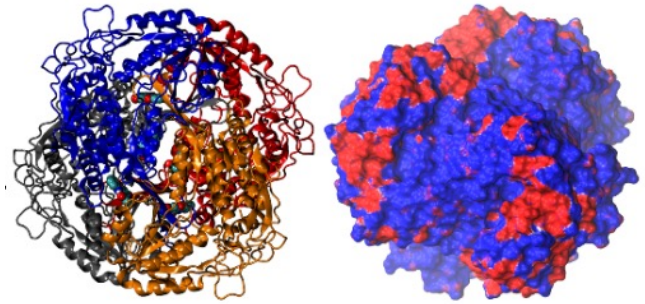
Banche dati per la scoperta in chimica

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

Teoria del funzionale della densità (DFT)



Dinamica molecolare (DM)



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

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) DFT (quantistica)

D) Elementi finiti

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) **DFT (quantistica)**

D) Elementi finiti

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) DFT (quantistica)

D) Elementi finiti

Q2 **Perché usiamo la modellazione multiscala?**

A) Basta una sola scala

B) I fenomeni variano con la scala

C) Per risparmiare citazioni

D) È richiesto dalla legge

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) **DFT (quantistica)**

D) Elementi finiti

Q2 **Perché usiamo la modellazione multiscala?**

A) Basta una sola scala

B) **I fenomeni variano con la scala**

C) Per risparmiare citazioni

D) È richiesto dalla legge

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) **DFT (quantistica)**

D) Elementi finiti

Q2 **Perché usiamo la modellazione multiscala?**

A) Basta una sola scala

B) **I fenomeni variano con la scala**

C) Per risparmiare citazioni

D) È richiesto dalla legge

Q3 **Rispetto alla DM a tutti gli atomi, la DM a grana grossa raggiunge…**

A) Sistemi più piccoli

B) Sistemi più grandi e tempi più lunghi

C) Maggiore accuratezza quantistica

D) Nessuna differenza

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – scelta multipla

Q1 **Quale metodo è il più accurato, ma anche il più costoso?**

A) DM classica

B) DM a grana grossa

C) **DFT (quantistica)**

D) Elementi finiti

Q2 **Perché usiamo la modellazione multiscala?**

A) Basta una sola scala

B) **I fenomeni variano con la scala**

C) Per risparmiare citazioni

D) È richiesto dalla legge

Q3 **Rispetto alla DM a tutti gli atomi, la DM a grana grossa raggiunge…**

A) Sistemi più piccoli

B) **Sistemi più grandi e tempi più lunghi**

C) Maggiore accuratezza quantistica

D) Nessuna differenza

MODELLAZIONE E SIMULAZIONE

Messaggi chiave – Il posto più economico per sbagliare è il computer.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. Nessuna scala basta da sola: serve la modellazione multiscala.

2. Dobbiamo prevedere degradazione e riciclabilità: circolarità, non solo prestazioni.

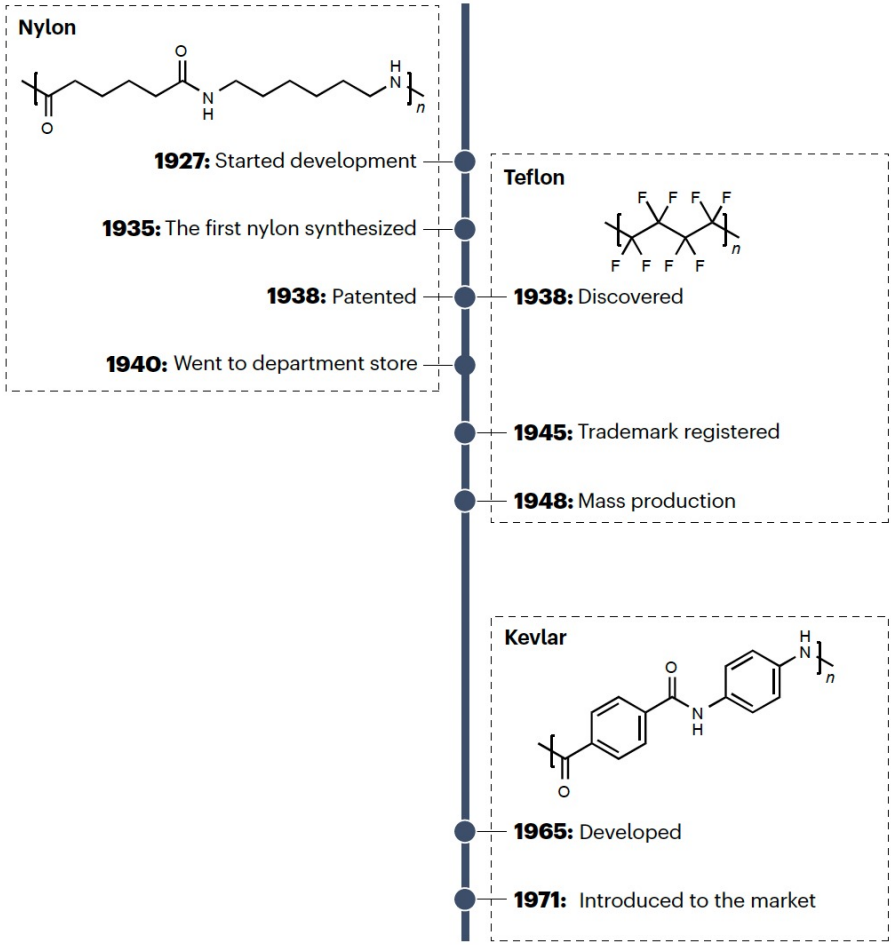
3. I dati sono la base su cui si costruisce l'IA. Buoni dati fanno buoni modelli.

DOMANDE?

UN NUOVO PARADIGMA NELLA SCOPERTA DI MOLECOLE E MATERIALI

Materiali lenti, mondo veloce

Innovazioni nei polimeri negli ultimi due secoli.



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

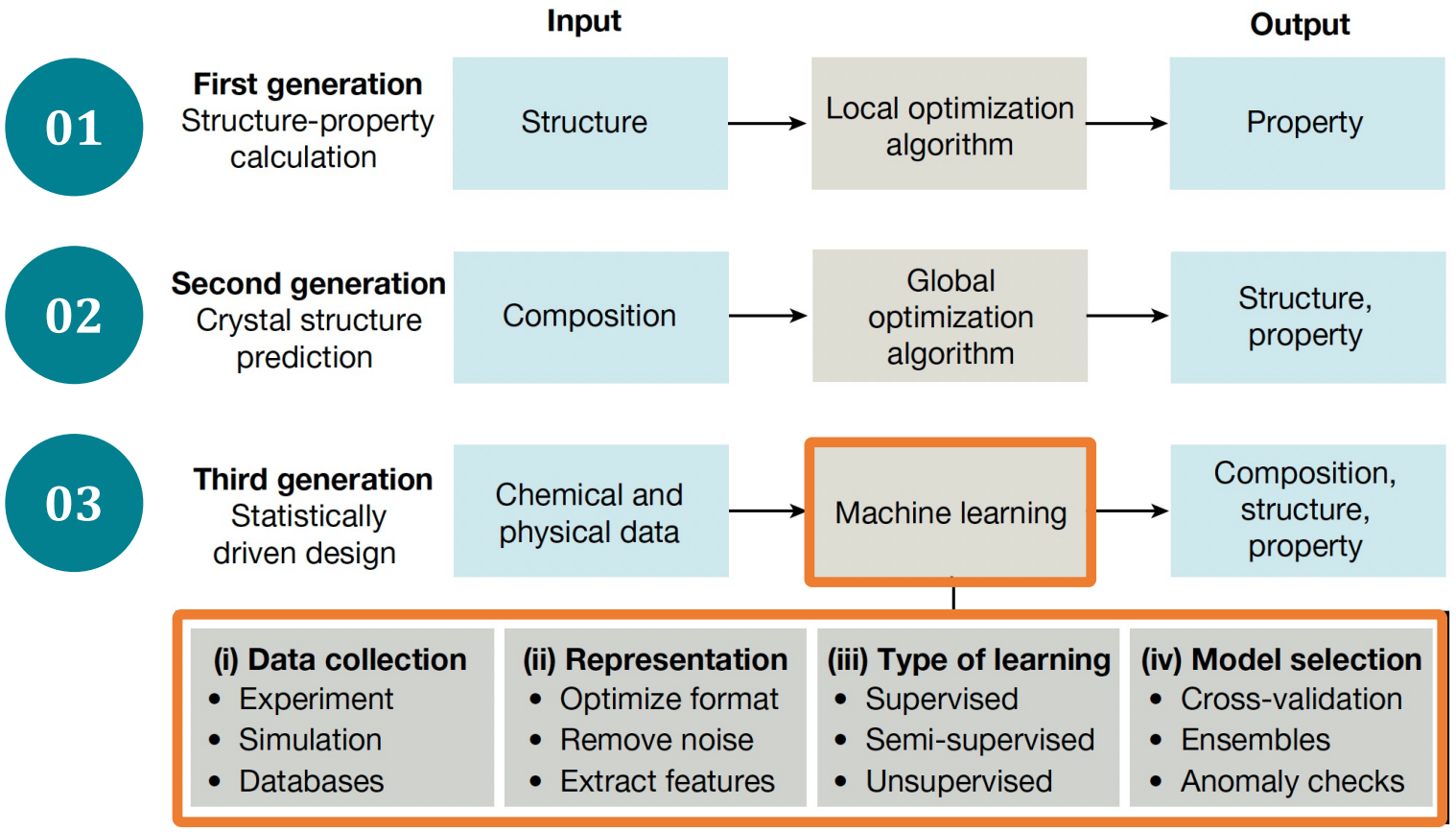
1977: Heeger, MacDiarmid and Shirakawa develop polyacetylene, the first conductive polymer

2011: Launch of Materials Genome Initiative

2020: 100-year anniversary of modern polymer science

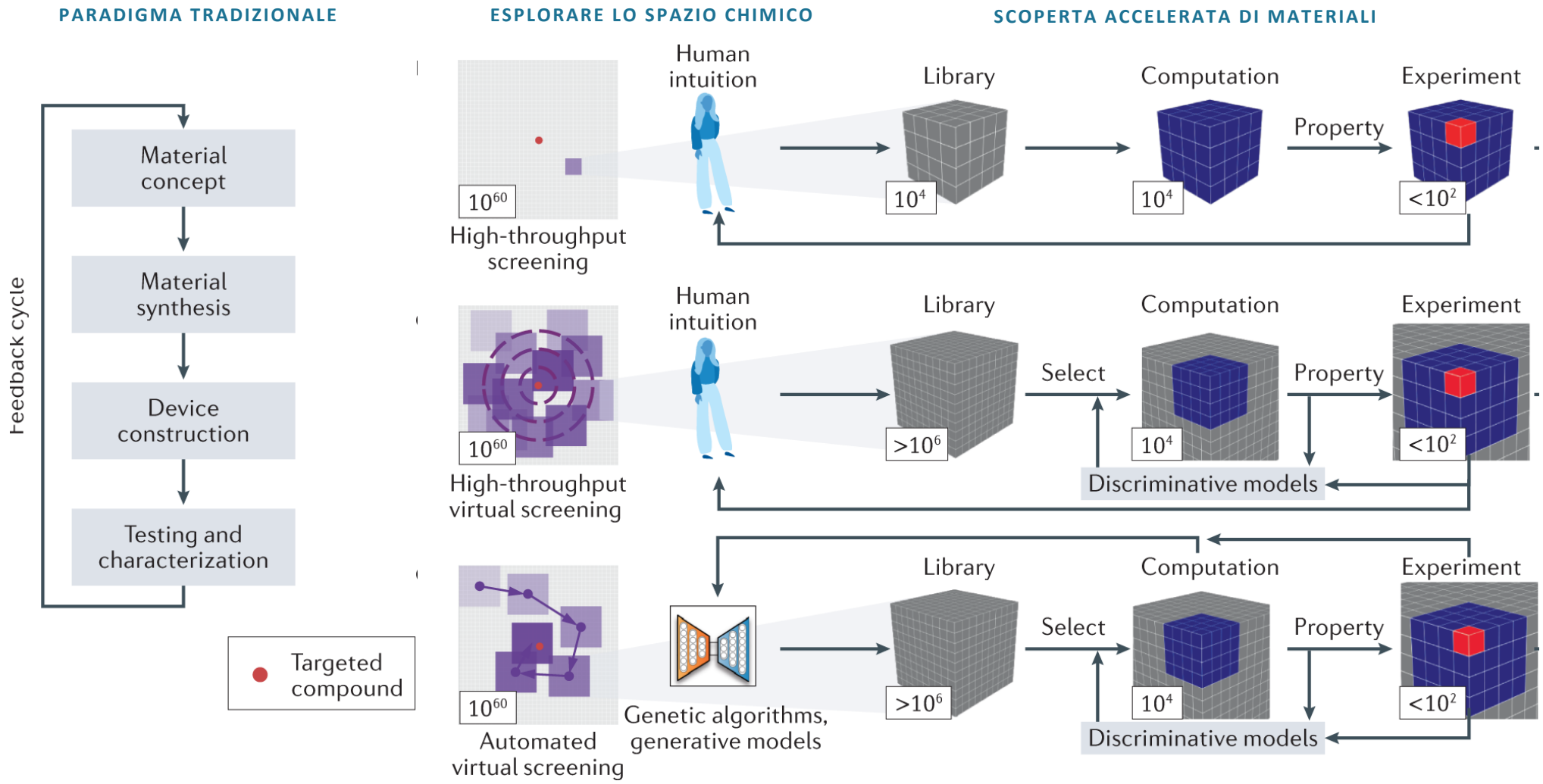
Tran, Gurnani, Kim, Pilania, Kwon, Lively, Ramprasad, Nat. Rev. Mater. 9, 866–886 (2024).

Dal tentativo ed errore alla progettazione guidata dai dati:
evoluzione del flusso di lavoro della ricerca in chimica computazionale



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

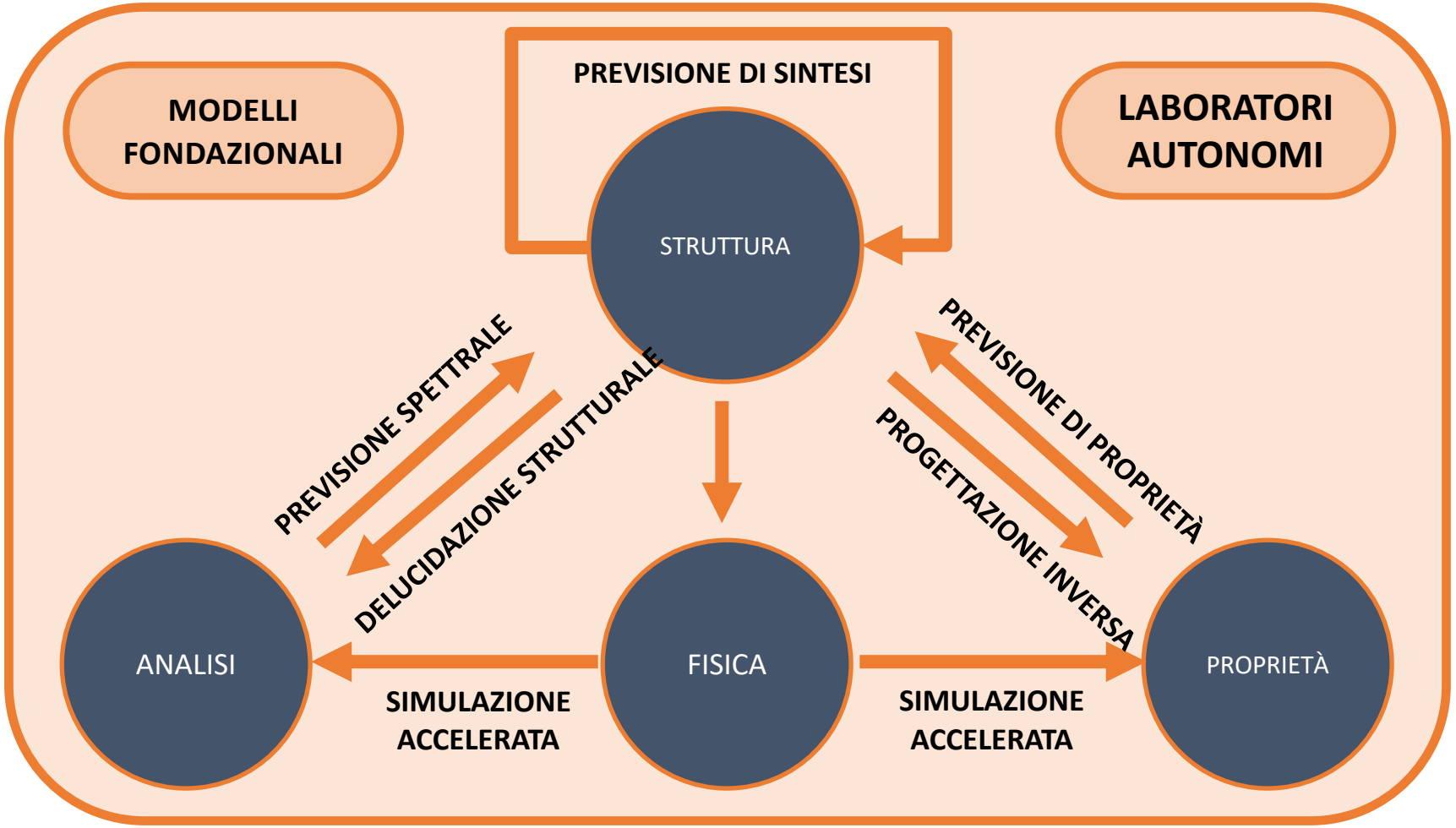
Nuovo paradigma nella scoperta di materiali



Yao, Lum, Johnston, Mejia-Mendoza, Zhou, Wen, Aspuru-Guzik, Sargent, Seh, Nat. Rev. Mater. 8, 202–215 (2023).

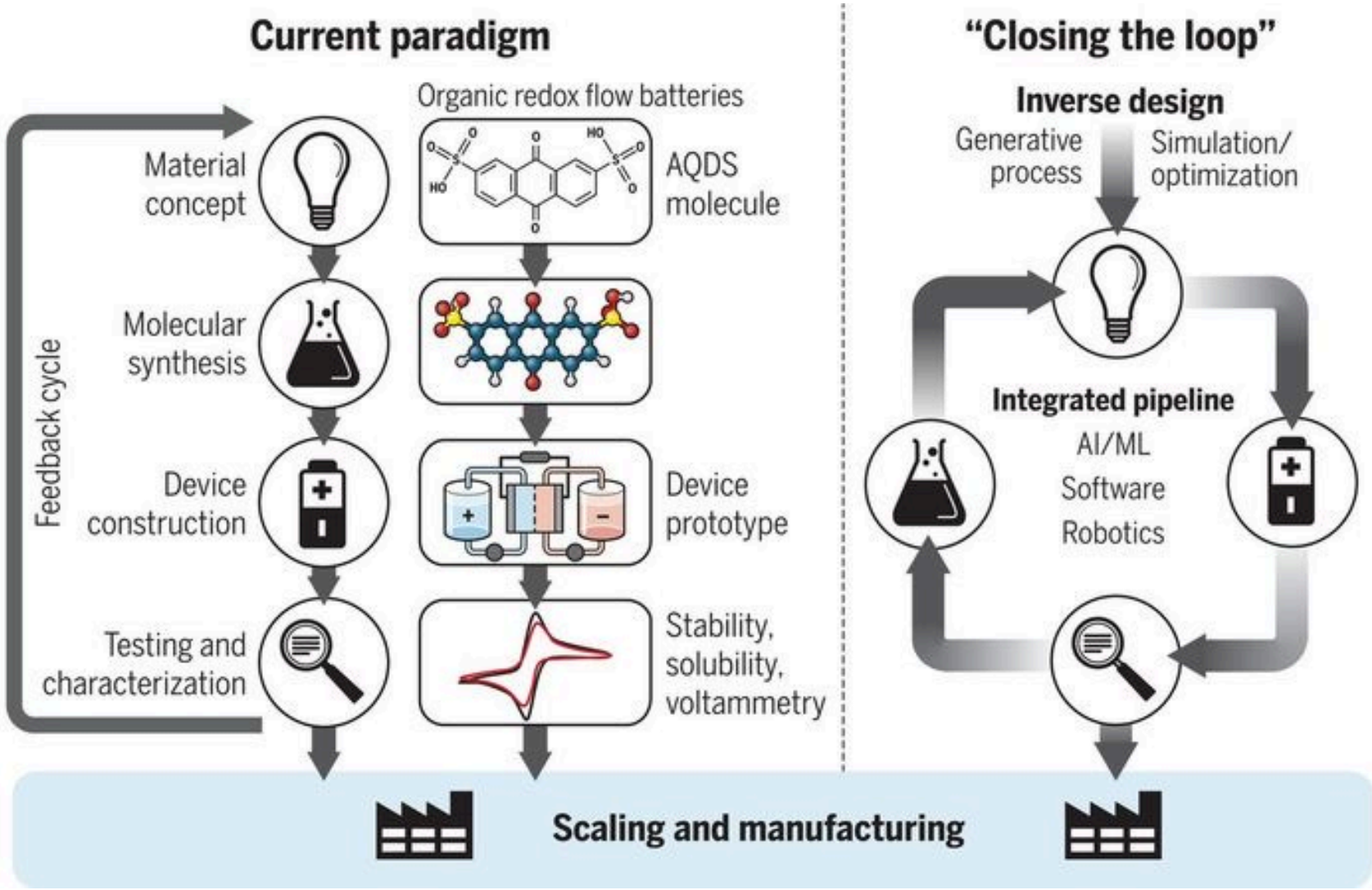
Apprendimento automatico per chimica e materiali

NUOVO PARADIGMA



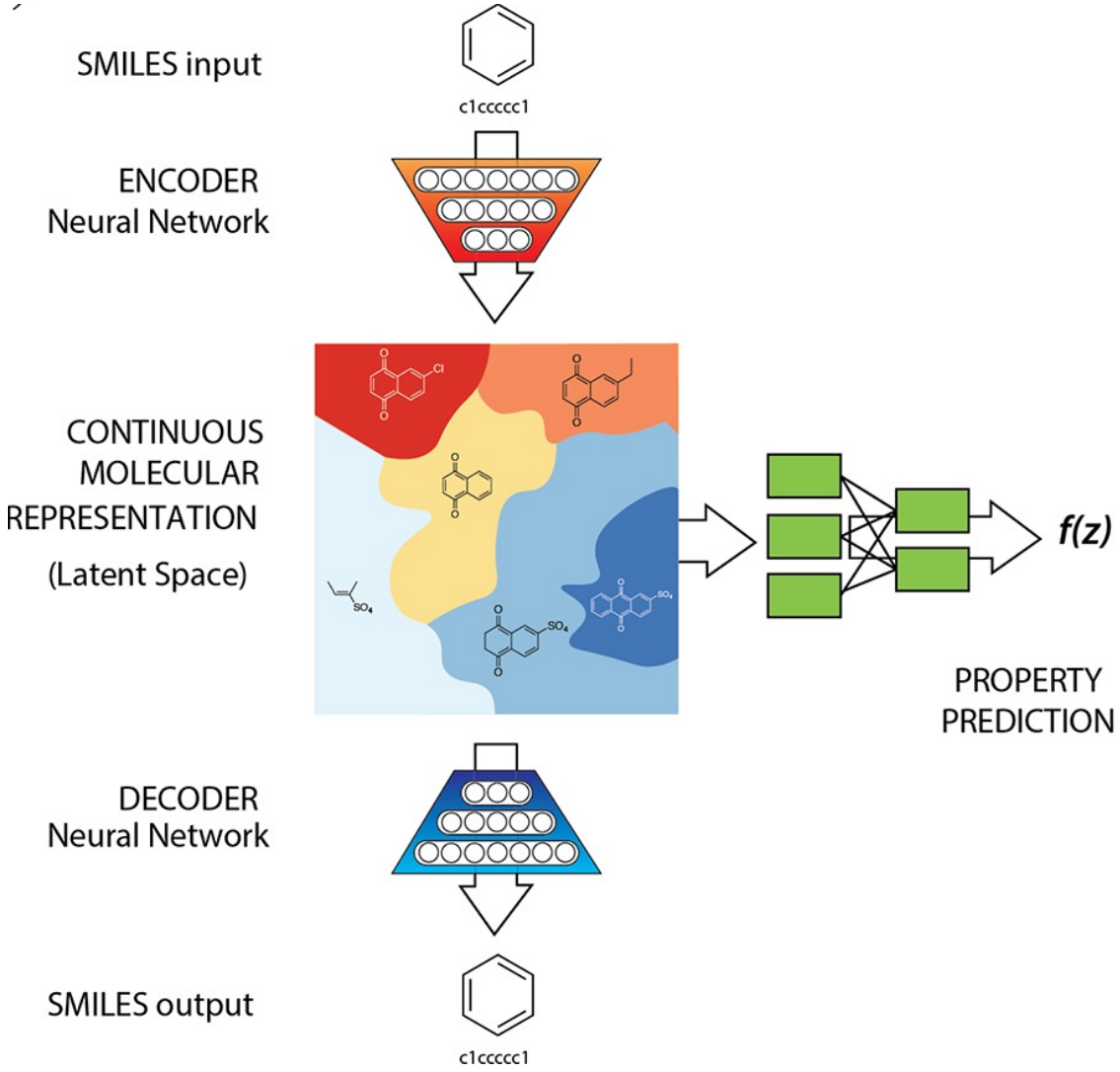
Cheng, Ser, Skreta, et al., Aspuru-Guzik, Faraday Discuss. 256, 10–60 (2025).

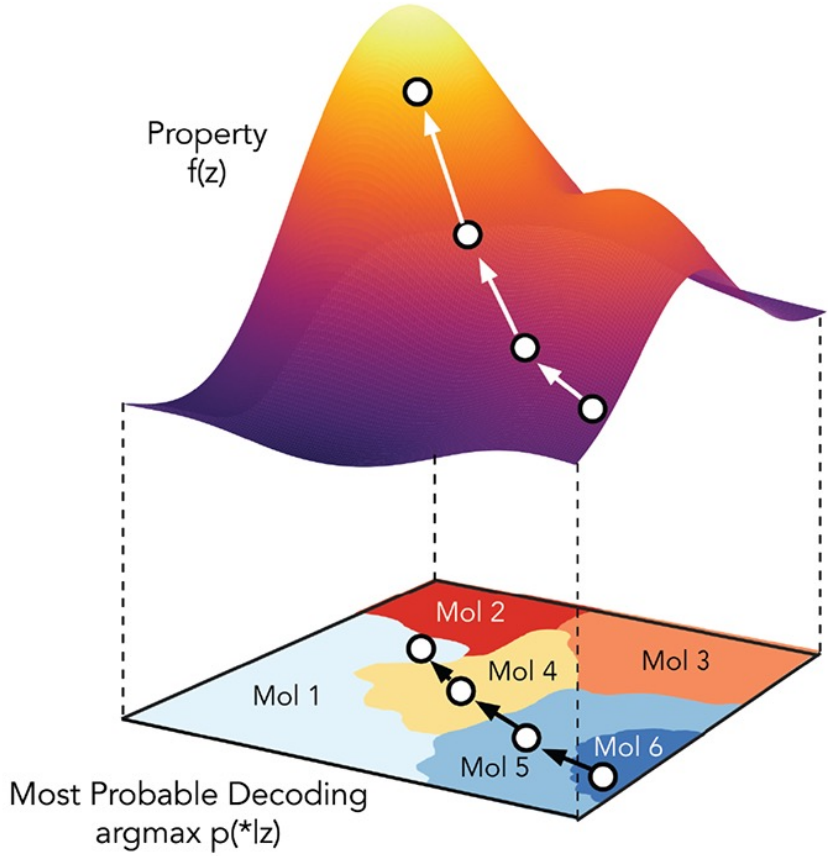
Progettazione diretta e inversa: l'esempio delle batterie a flusso redox



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

VAE chimico: trasformare le molecole in una mappa da ottimizzare





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

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – Vero/Falso

1. **Nella progettazione inversa si specifica la proprietà desiderata e il modello propone una struttura.**

   Vero | Falso

2. **Un VAE fornisce uno spazio latente continuo in cui è possibile ottimizzare.**

   Vero | Falso

3. **La «terza generazione» della progettazione di materiali è il tentativo ed errore manuale.**

   Vero | **Falso**

NUOVO PARADIGMA

Messaggi chiave – Progettare secondo una specifica invece di arrivarci per caso.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. Siamo in una generazione di progettazione di materiali guidata dall'AA.

2. Progettazione inversa = ottimizzare in uno spazio latente appreso (spazio di progettazione continuo).

3. Ci sono due fattori abilitanti trasversali: i modelli fondazionali e i laboratori autonomi.

DOMANDE?

LA CASSETTA DEGLI ATTREZZI DELL'IA

Una cassetta degli attrezzi dell'IA

Sei cose che l'AA fa davvero lungo la filiera dei materiali

1 **Previsione di proprietà**

PREVEDERE

I modelli supervisionati prevedono una proprietà y (ad es. il band gap) a partire da una struttura x, affiancando o sostituendo simulazioni costose.

2 **Progettazione generativa**

GENERARE

I modelli non supervisionati/generativi producono nuovi candidati x', condizionati a proprietà obiettivo: la via della progettazione inversa.

3 **Laboratori autonomi**

SINTETIZZARE

Laboratori autonomi pianificano ed eseguono esperimenti: retrosintesi, previsione di reazioni, ottimizzazione con apprendimento attivo.

4 **Supporto alla caratterizzazione**

CARATTERIZZARE

L'AA interpreta le misure: struttura cristallina da TEM, coordinazione da XAS, simmetria da diffrazione elettronica.

5 **Gestione dei sistemi**

IMPLEMENTARE

L'AA gestisce dispositivi e reti: previsione della durata delle batterie, previsione dei carichi, ottimizzazione del controllo delle smart grid.

6 **Accelerare la teoria**

SIMULARE

L'AA rende fattibili simulazioni intrattabili: potenziali e funzionali più economici per scale maggiori.

Adattato da Yao et al., «Machine learning for a sustainable energy future», Nat. Rev. Mater. 2023, Box 1.

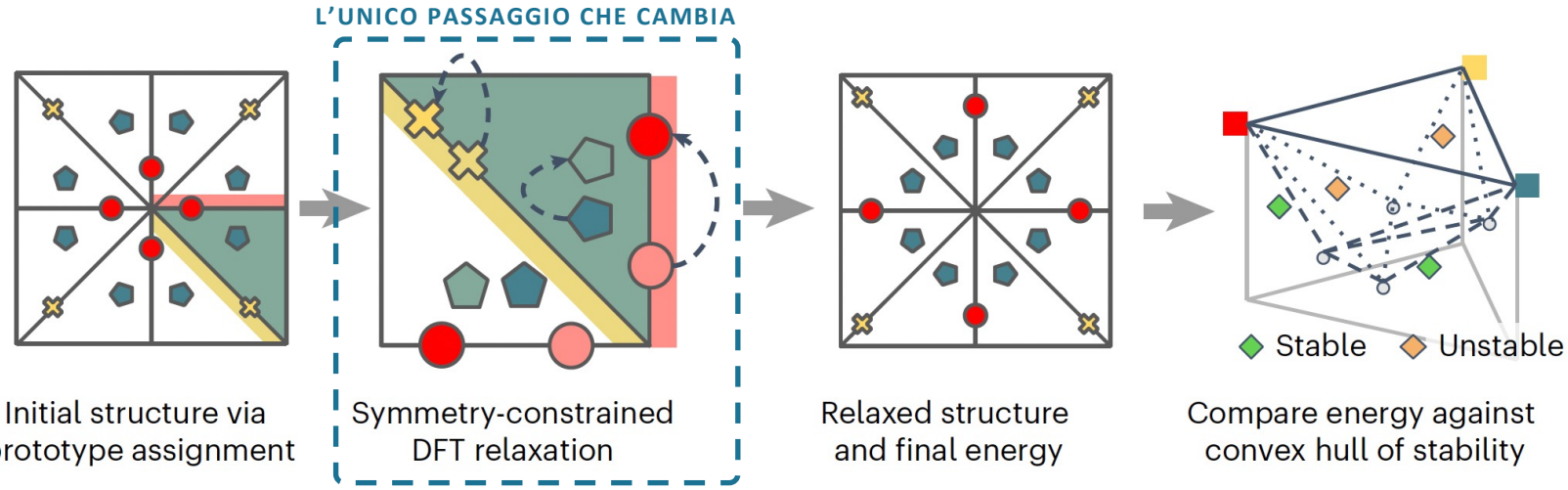
1

Previsione di proprietà

PREVEDERE

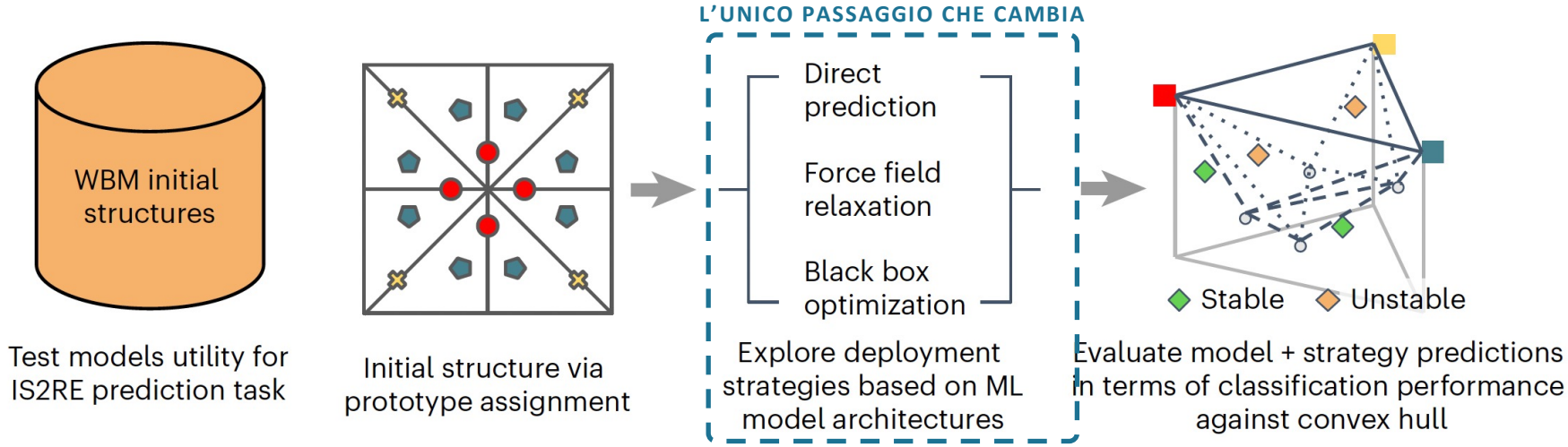
Prevedere proprietà DFT con un surrogato di AA

CONVENZIONALE



MATBENCH DISCOVERY

IS2RE = dalla struttura iniziale all'energia rilassata (Initial Structure to Relaxed Energy)



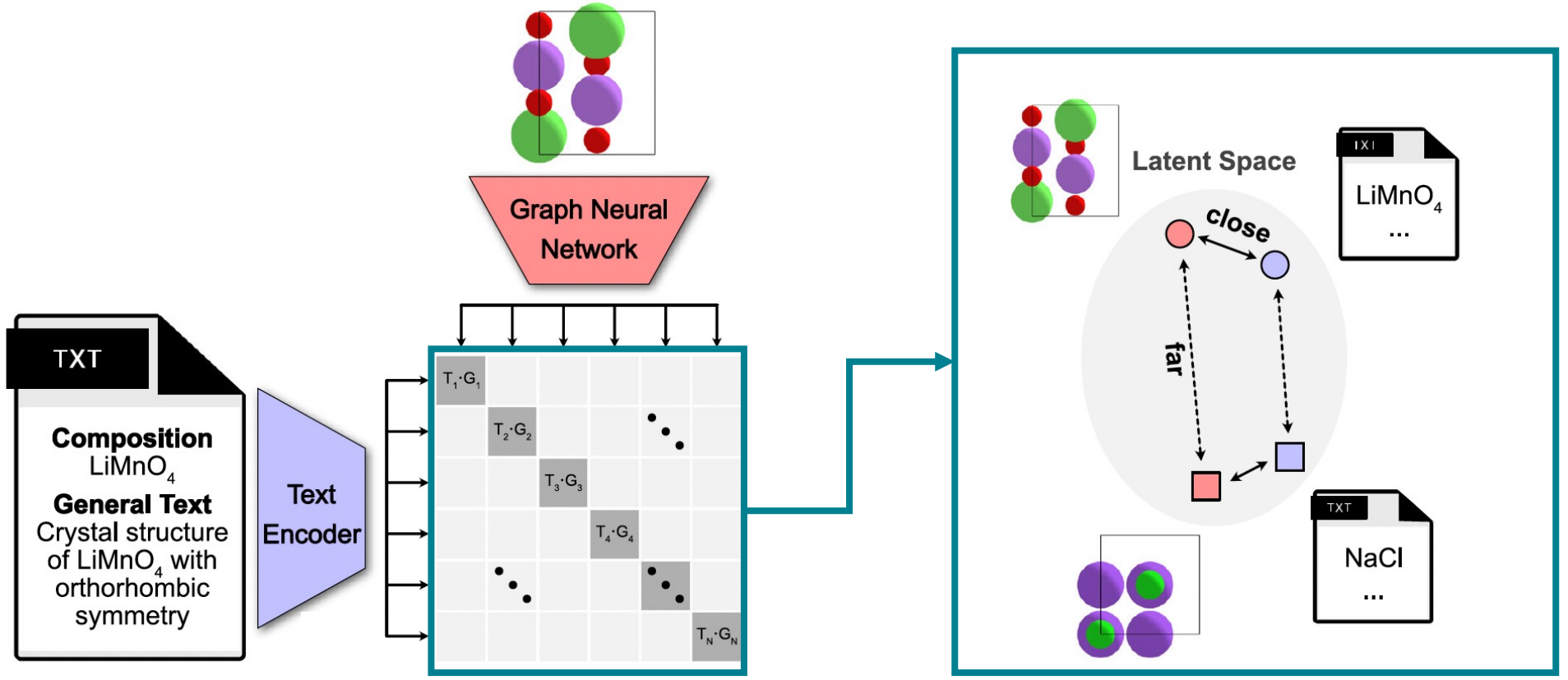
Adattato da Riebesell et al., «A framework to evaluate machine learning crystal stability predictions», Nat. Mach. Intell. 7, 836–847 (2025).

2

Progettazione generativa/inversa

GENERARE

Chemeleon: trasformare un prompt testuale in un cristallo, in tre passaggi



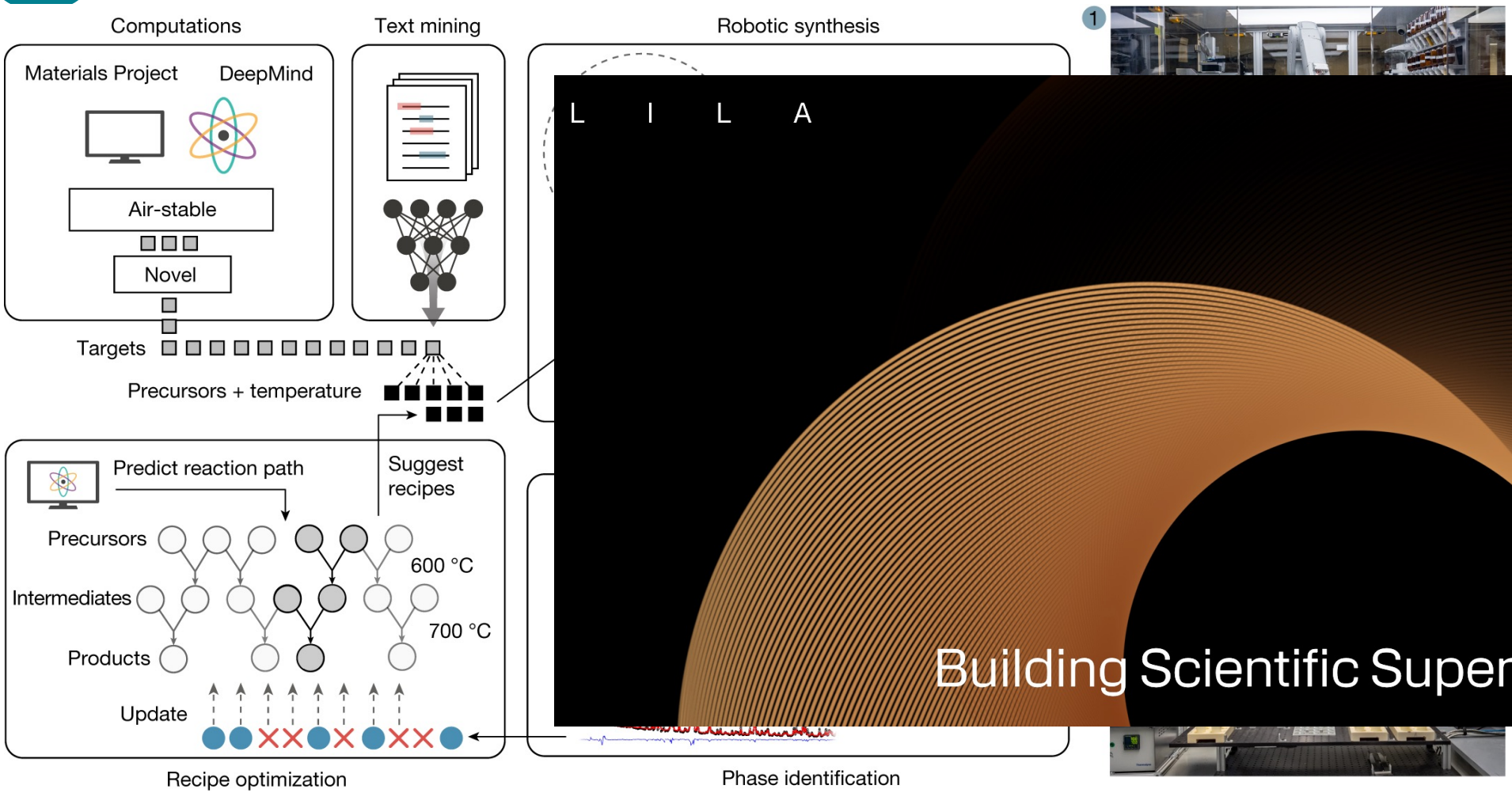
Adattato da Park, Onwuli & Walsh, Nat. Commun. 16 (2025)

3

Laboratori autonomi

SINTETIZZARE

Un laboratorio autonomo: l'A-Lab



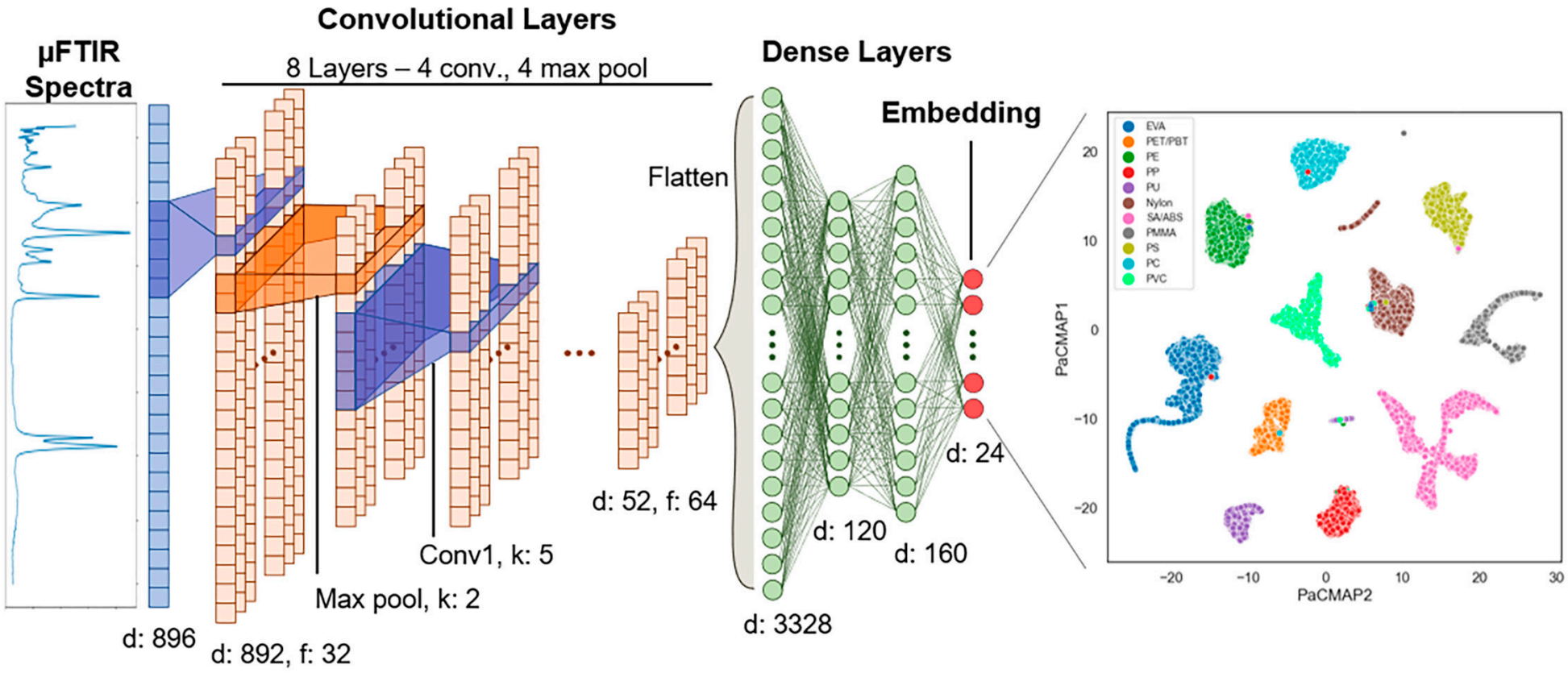
Adattato da Szymanski et al., Nature 624, 86 (2023)

4

Supporto alla caratterizzazione

CARATTERIZZAZIONE

Apprendimento di similarità su immagini spettrali μFTIR di microplastiche



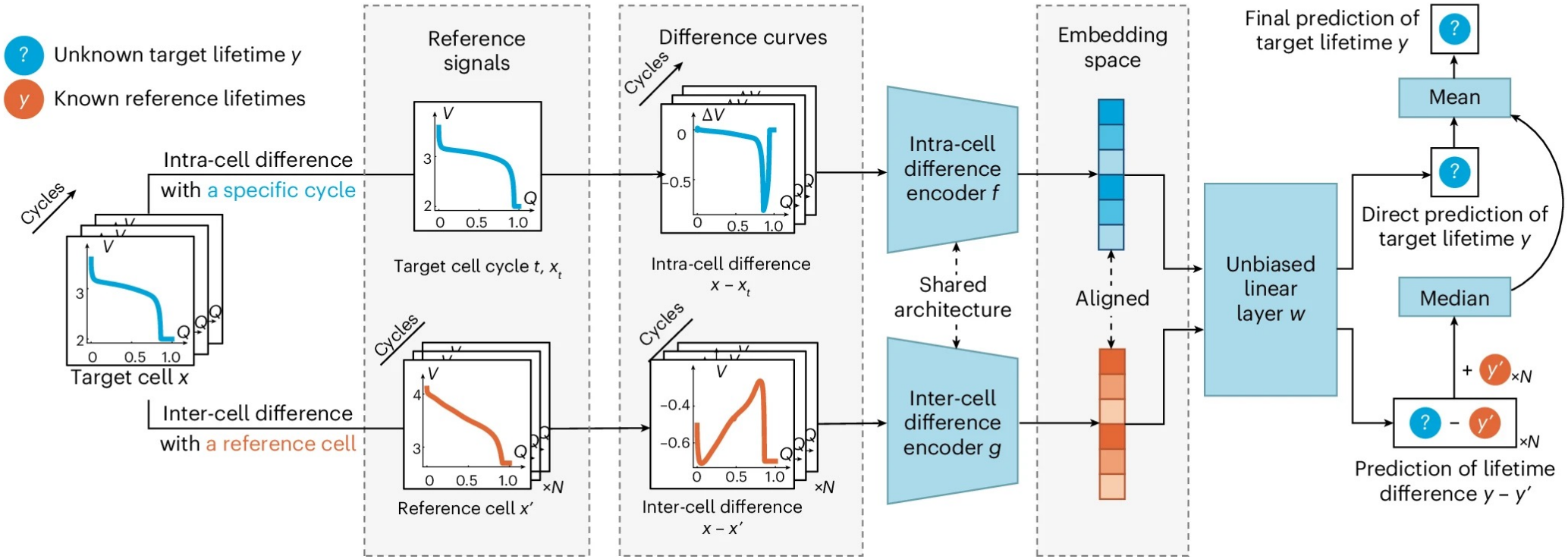
Smolen, Moore, Perez & Wooley, PNAS 122(42), e2509745122 (2025)

5

Gestione dei sistemi

IMPLEMENTARE

Prevedere in anticipo la durata in cicli delle batterie

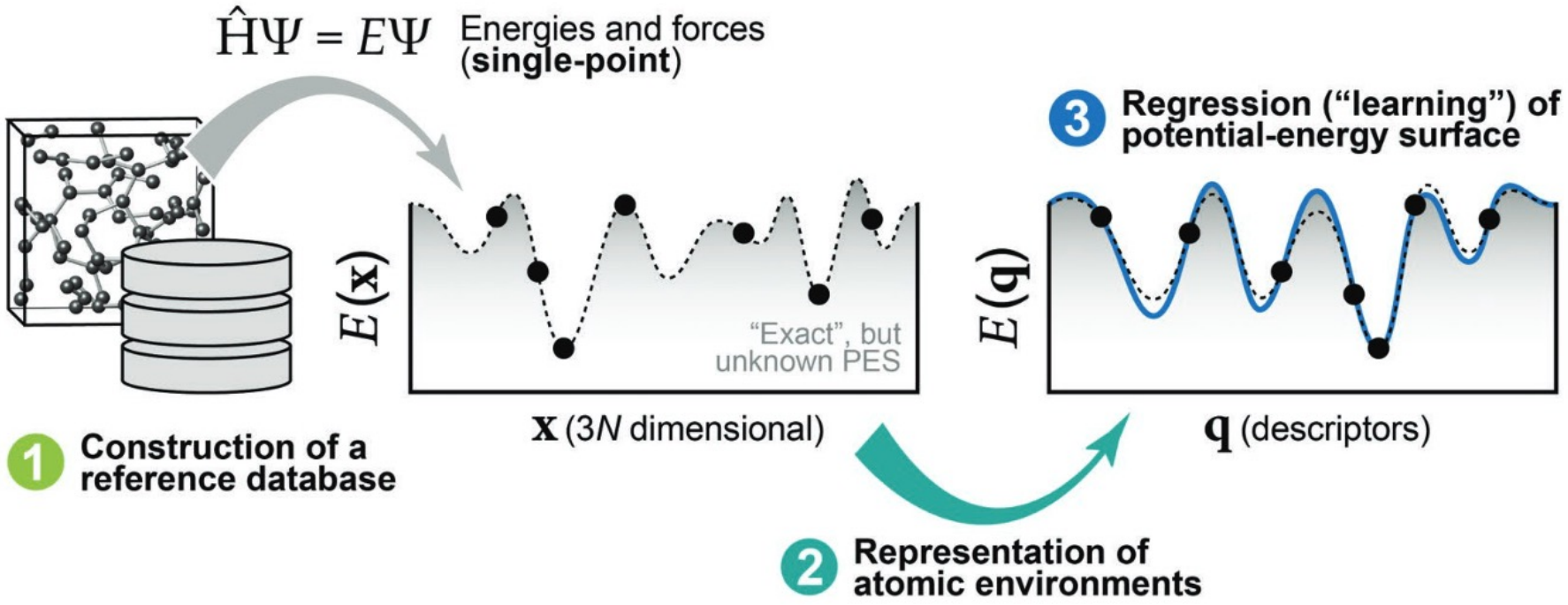


BatLiNet confronta la cella obiettivo con una cella di «riferimento» completamente ciclata, la cui durata è già nota

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

Come funziona un potenziale interatomico di apprendimento automatico

Un'unica ricetta per tutti: ambiente atomico locale → energia, appresa dalla DFT



Deringer, V. L.; Caro, M. A.; Csányi, G. Adv. Mater. 31, 1902765 (2019). 10.1002/adma.201902765

I modelli MLIP in sintesi

Modelli rappresentativi per famiglia: idea chiave, uso tipico e fonte

Modello	Famiglia	Idea chiave	Applicazione tipica	Riferimento
HDNNP	Descrittore + RN	funzioni di simmetria → rete neurale	solidi massivi, acqua	*Behler & Parrinello 2007*
GAP	Descrittore + kernel	SOAP + processo gaussiano	Si, C, amorfi; efficiente nei dati	*Bartók et al. 2010*
DeePMD	Descrittore appreso + RN	potenziale profondo end-to-end	DM su larga scala (acqua, leghe)	*Zhang et al. 2018*
SNAP	Espansione lineare	componenti del bispettro (lineare)	metalli (Ta, W)	*Thompson et al. 2015*
MTP	Espansione lineare	tensori dei momenti	leghe; apprendimento attivo	*Shapeev 2016*
ACE	Espansione per ordine di corpo	espansione in cluster atomici	generale; veloce + accurato	*Drautz 2019*
SchNet	GNN invariante	convoluzioni a filtro continuo	molecole (MD17, QM9)	*Schütt et al. 2018*
NequIP	GNN equivariante	messaggi E(3)-equivarianti	molecole + materiali, efficiente nei dati	*Batzner et al. 2022*
Allegro	GNN equivariante locale	equivariante strettamente locale	DM su scala molto ampia	*Musaelian et al. 2023*
MACE	Equivariante + alto ordine di corpo	messaggi di ordine superiore per strato	DM accurata e veloce	*Batatia et al. 2022*
M3GNet	GNN universale	potenziale a grafo addestrato su MP	rilassamento/screening di cristalli	*Chen & Ong 2022*
CHGNet	Universale (sensibile alla carica)	GNN informata dalla carica	ioni di litio, ossidi, magnetismo	*Deng et al. 2023*
MACE-MP-0	Fondazionale (equivariante)	modello fondazionale per tutta la tavola periodica	DM pronta all'uso per tutte le chimiche	*Batatia et al. 2024*

SOAP = sovrapposizione regolare delle posizioni atomiche (smooth overlap of atomic positions) · E(3) = rotazioni, traslazioni e riflessioni in 3D · MP = Materials Project · GNN = rete neurale a grafo.

Potenziali interatomici di apprendimento automatico, in ordine all'incirca cronologico

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

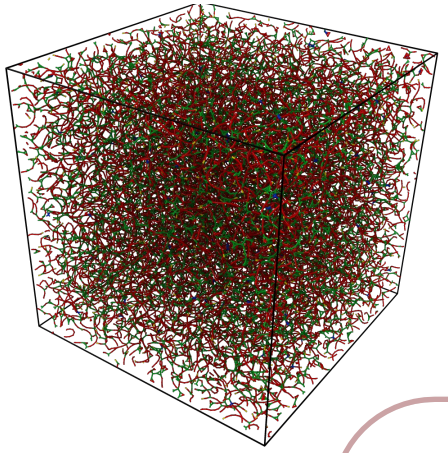
12. Deng, B. et al. CHGNet: charge-informed universal neural network potential. Nat. Mach. Intell. 5, 1031–1041 (2023). 10.1038/s42256-023-00716-3

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

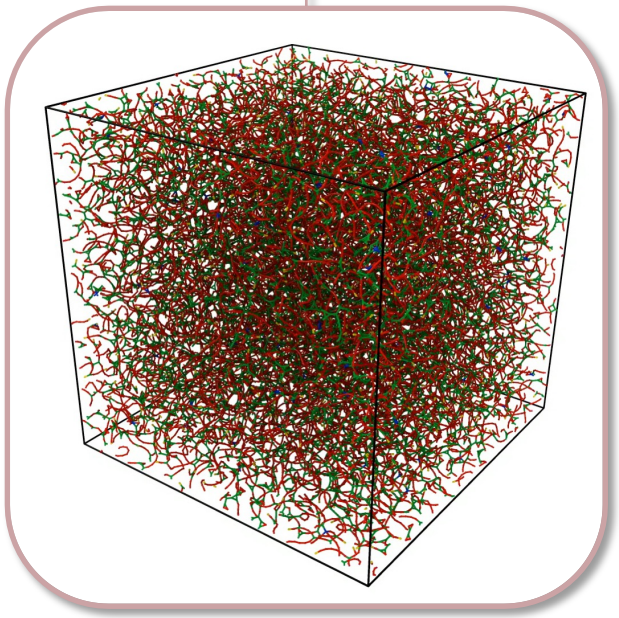
14. Batatia, I. et al. A foundation model for atomistic materials chemistry (MACE-MP-0). arXiv:2401.00096 (2024).

Modellare il biochar con i MLIP (GAP20)

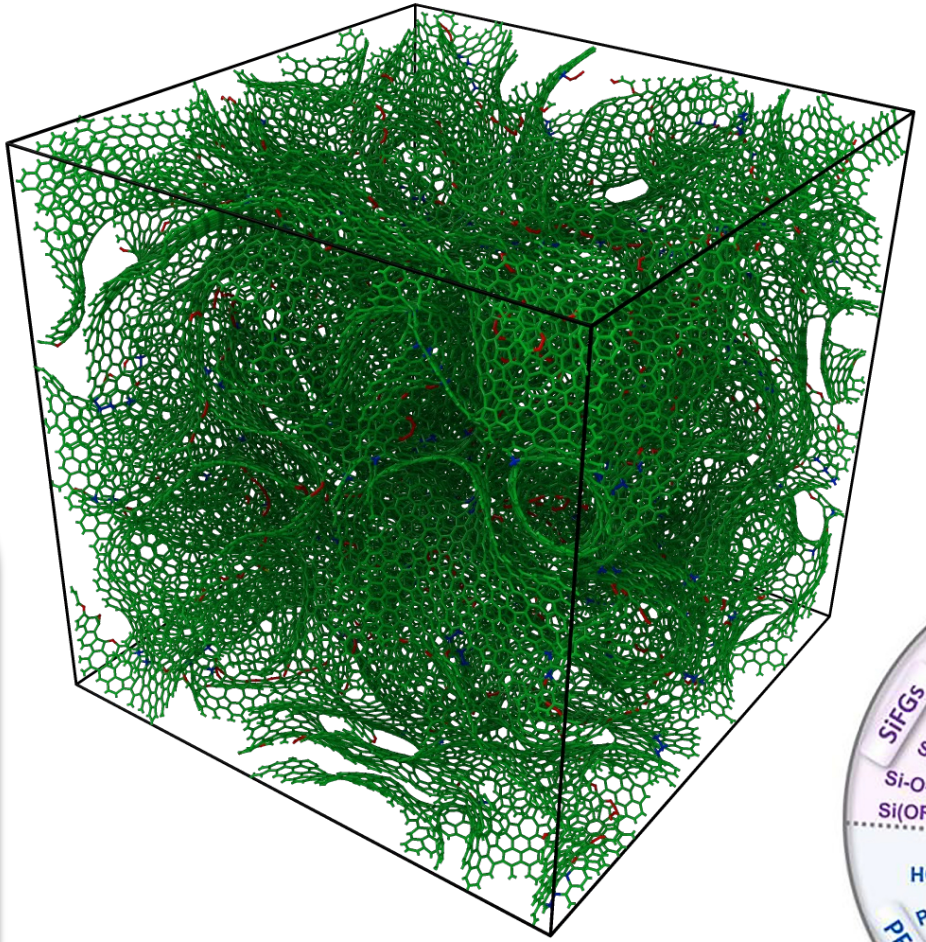
Scala atomica

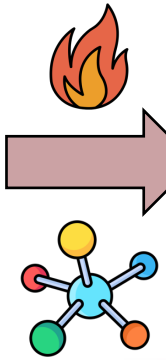


Simulazione di DM

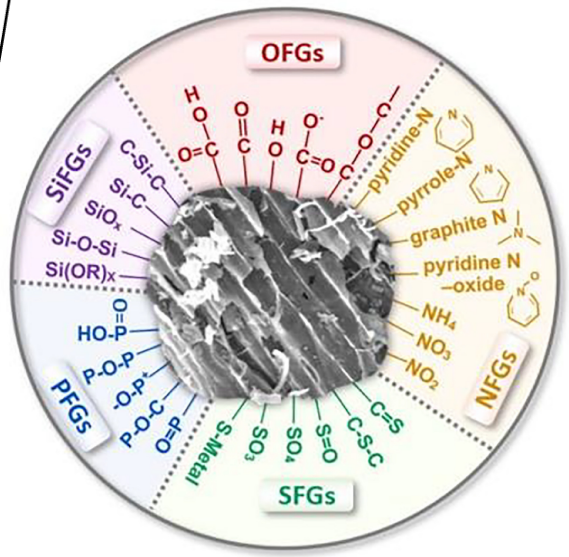


Caratterizzazione strutturale:
Porosità, struttura atomica, ibridazione del carbonio, ordine strutturale





Biochar Modello



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

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ

Quiz lampo – Vero/Falso

1 **Un potenziale interatomico di AA punta all'accuratezza DFT a un costo molto inferiore.**

Vero | Falso

2 **Un modello con il 100% di accuratezza in addestramento è sempre il più affidabile su dati nuovi.**

Vero | **Falso**

3 **I candidati generati dall'IA richiedono comunque una validazione DFT/sperimentale.**

Vero | Falso

PERCHÉ È IMPORTANTE PER IL CLIMA

Un’unica cassetta degli attrezzi per tutti i materiali dell’energia pulita.

- I potenziali di AA offrono un’accuratezza vicina alla DFT per una frazione dell’energia e del tempo: si possono vagliare milioni di candidati.
- I laboratori autonomi trasformano rapidamente le previsioni migliori in campioni validati.
- La validazione evita di sprecare il prezioso tempo di laboratorio sui falsi positivi.

Messaggi chiave

PERCHÉ È IMPORTANTE PER IL CLIMA

1. I laboratori autonomi trasformano rapidamente le previsioni migliori in campioni validati.

2. I MLIP offrono l'accuratezza della DFT alla velocità della DM: la casella «accelerare la simulazione» diventata realtà.

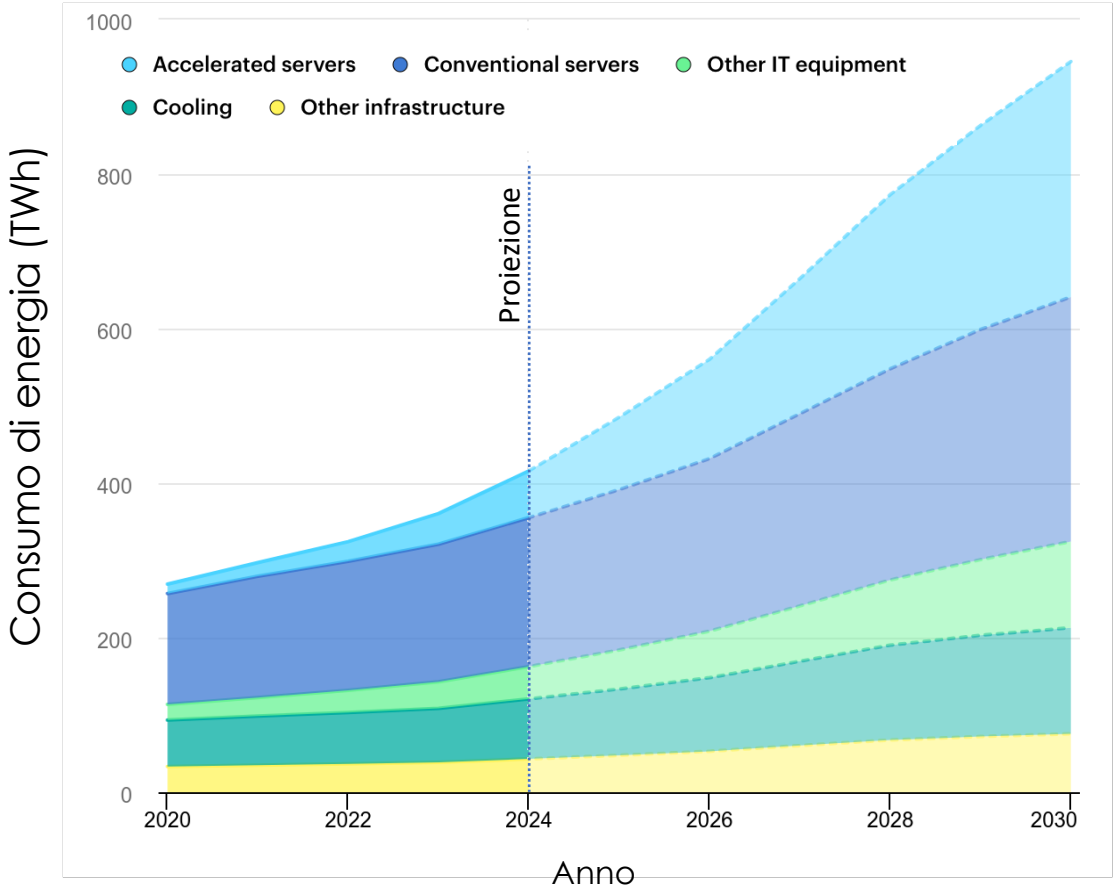
3. Generare è facile; validare è difficile

DOMANDE?

PARLARE DI CHIMICA AI COMPUTER

Costo energetico crescente dell'IA e dei data center

Serve un'informatica più verde



Fonte: Agenzia internazionale dell'energia (AIE)

Dove si trova il supercalcolo?

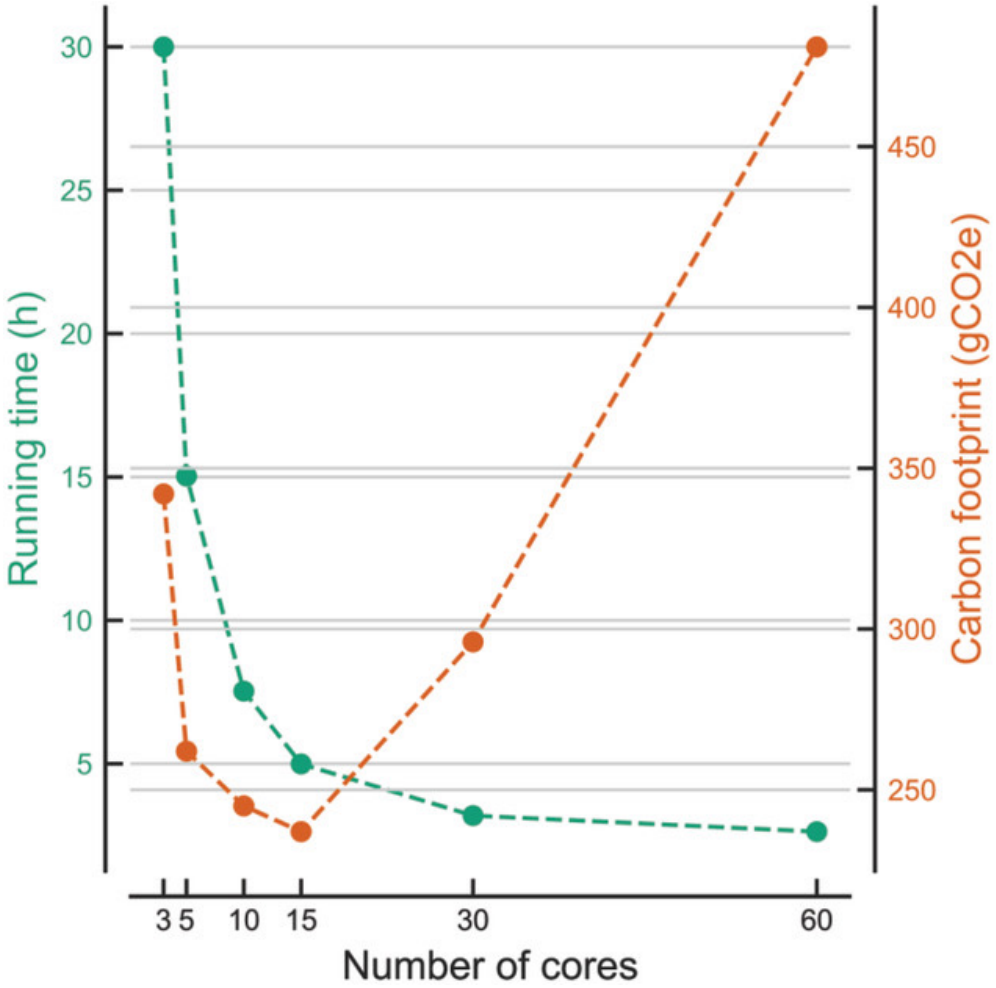


TOP 500
The List.

Migliorare l'accesso al supercalcolo su dispositivi o in contesti con risorse limitate: **democratizzazione**

Fonte: top500.org

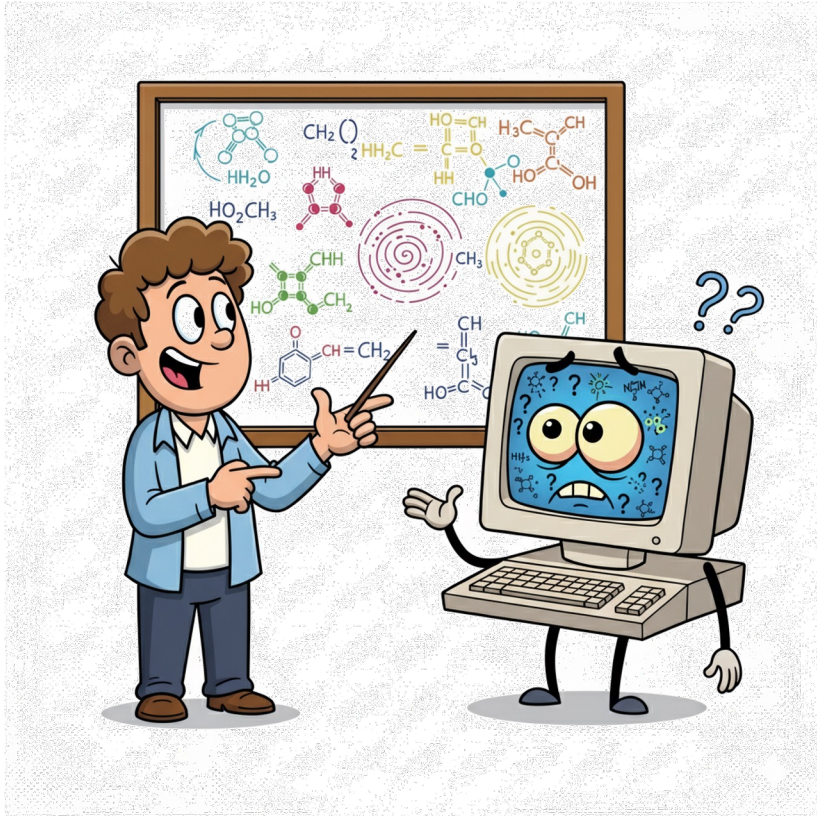
Algoritmi verdi





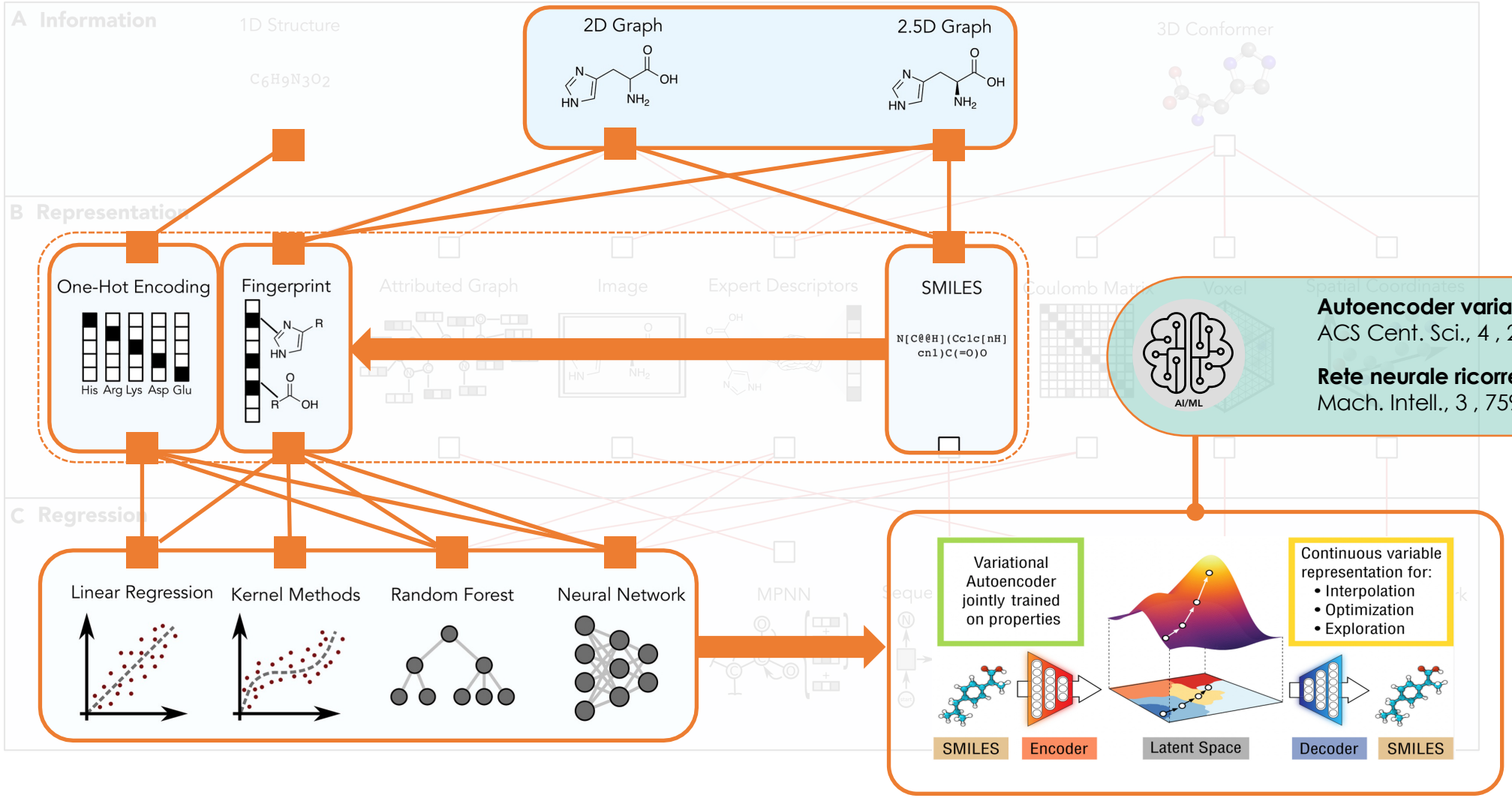
Lannelongue et al., Advanced Science, 8, 2021

Un passo indietro: come parliamo di chimica ai computer?



L'apprendimento automatico per la chimica richiede di **vettorizzare le molecole**

L'AA per la chimica richiede di vettorizzare le molecole



Autoencoder variazionali (VAE)
ACS Cent. Sci., 4 , 268, 2018`

Rete neurale ricorrente (RNN) Nat. Mach. Intell., 3 , 759, 2021

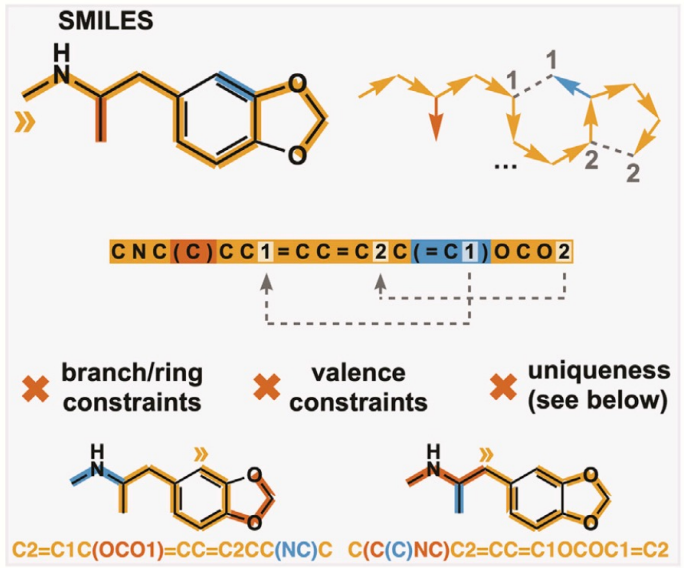
Pattanaik and Coley, **Chem 6, 1204-1207, 2020**

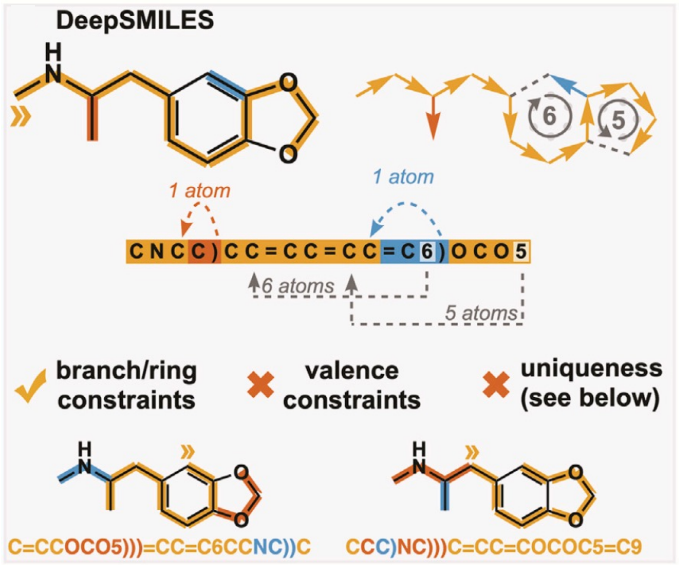
Skinnider et al. **Nat. Mach. Intell.**, 3 , 759, **2021**

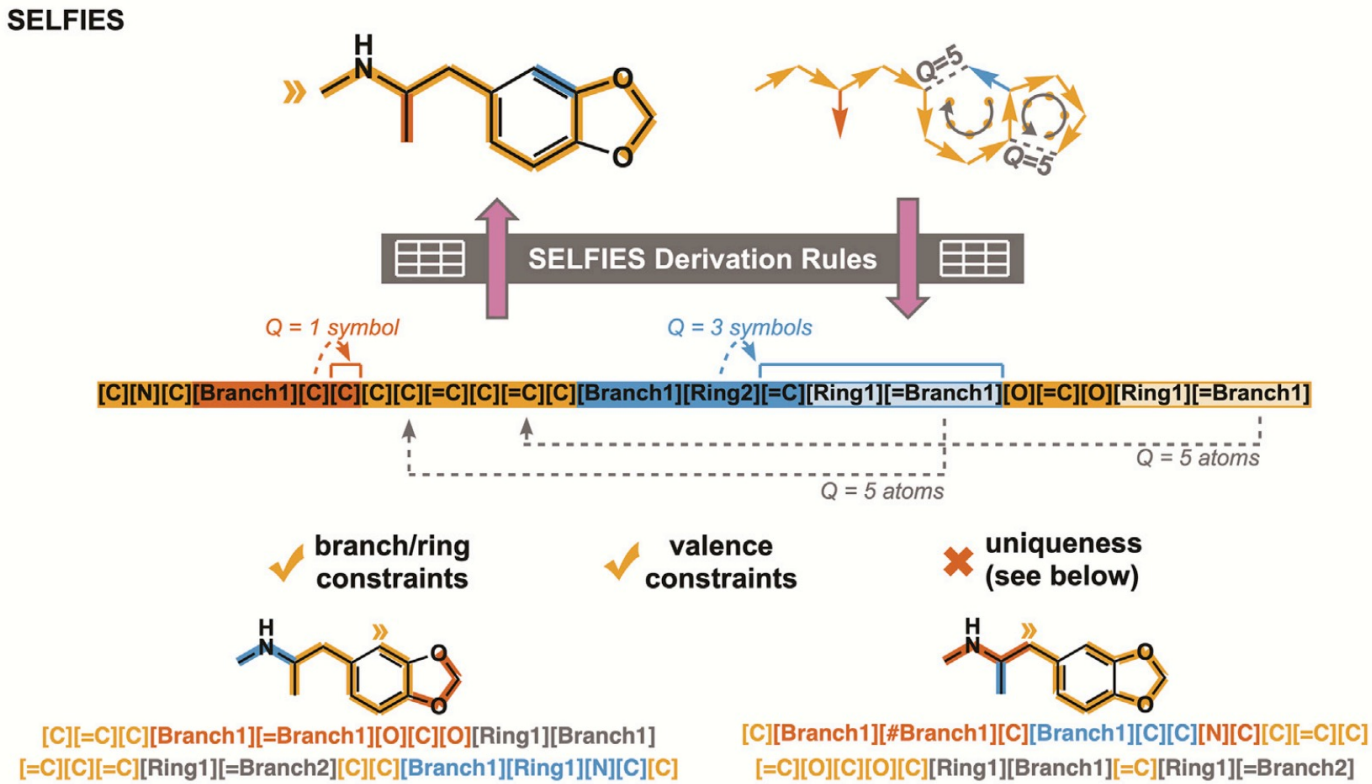
Gomez-Bombarelli et al. **ACS Cent Sci.** 4, 268–76, **2018**

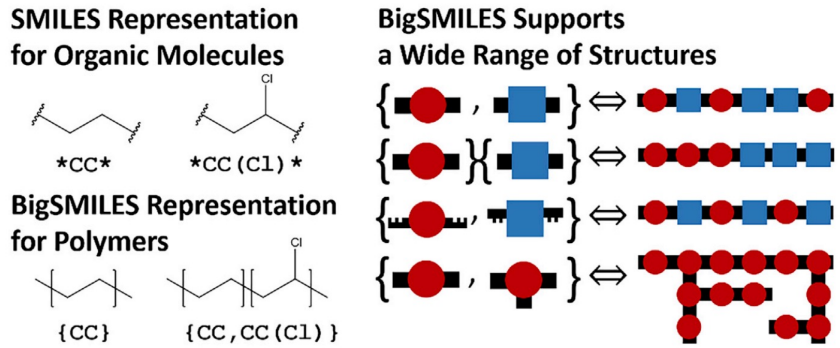
Codifica dei dati leggibile dalla macchina

PARLARE DI CHIMICA



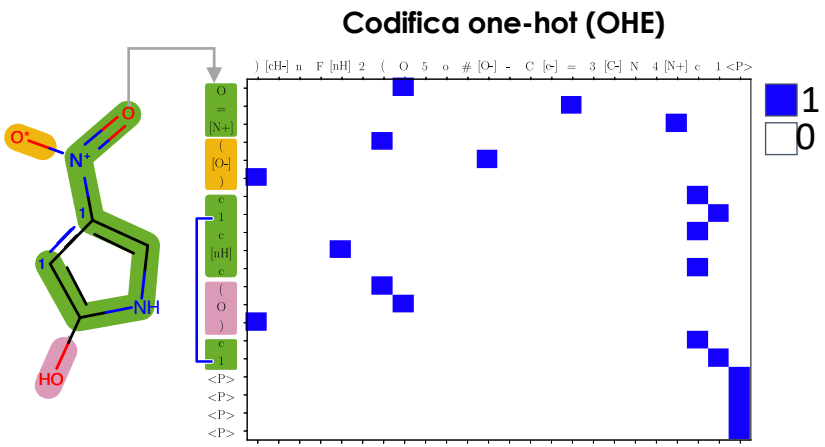


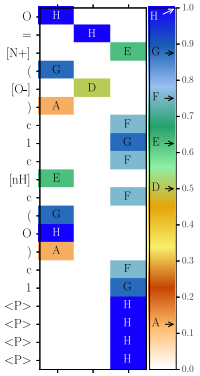




Krenn et al. Patterns 3, 10, 100588, 2022

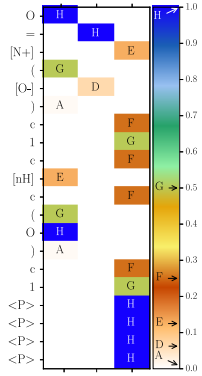
Codifiche one-hot incorporate (eOHE)





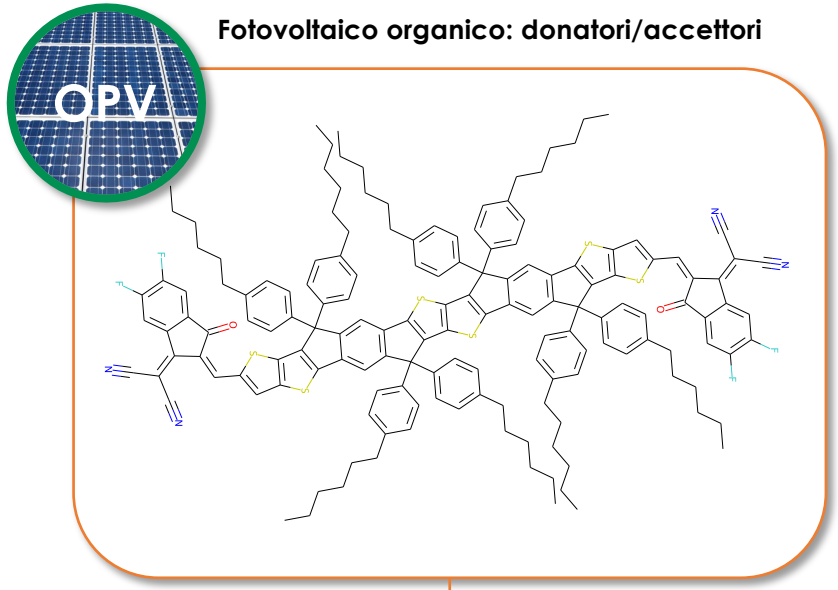
Riduzione lineare
$q = 8$

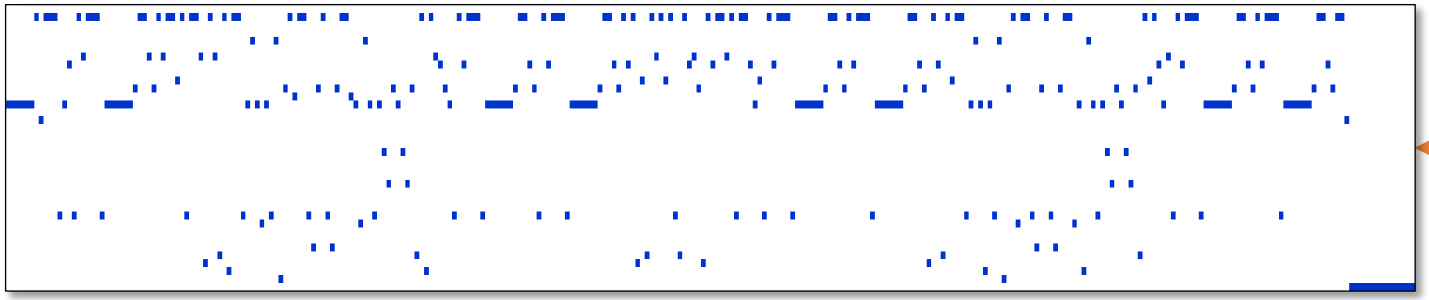
$$v_1(r,q) = \frac{r+1}{q}$$



Riduzione per potenze di 2
$q = 8$

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



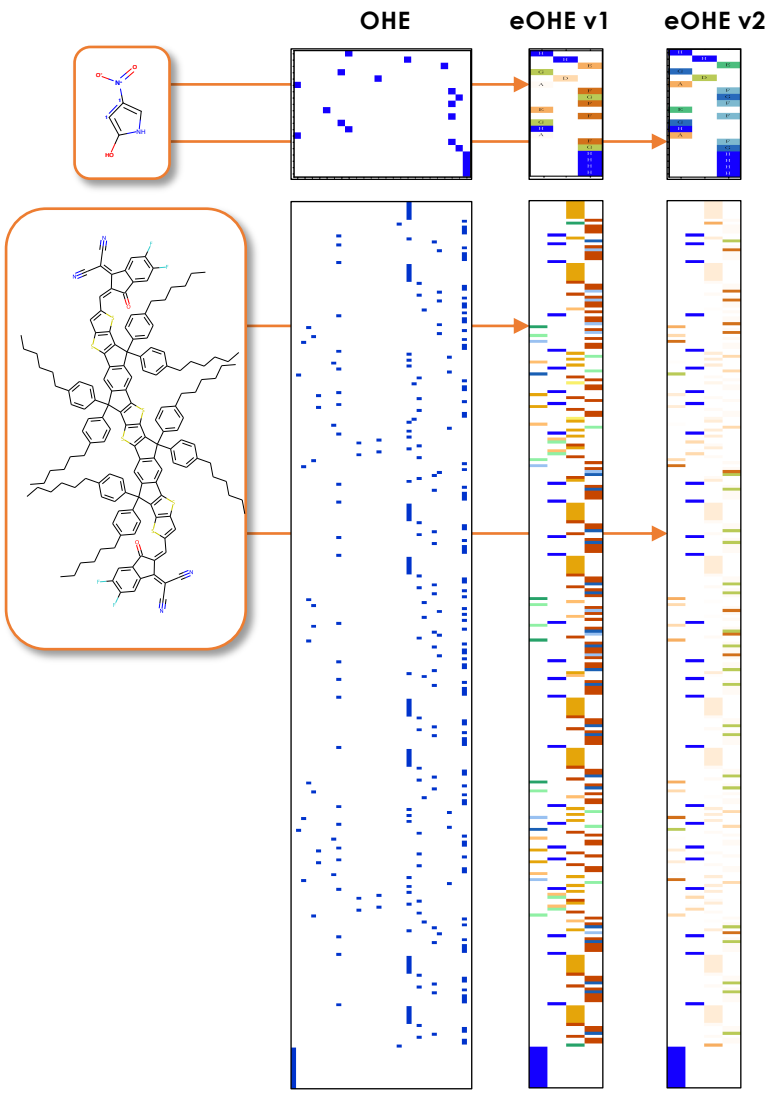


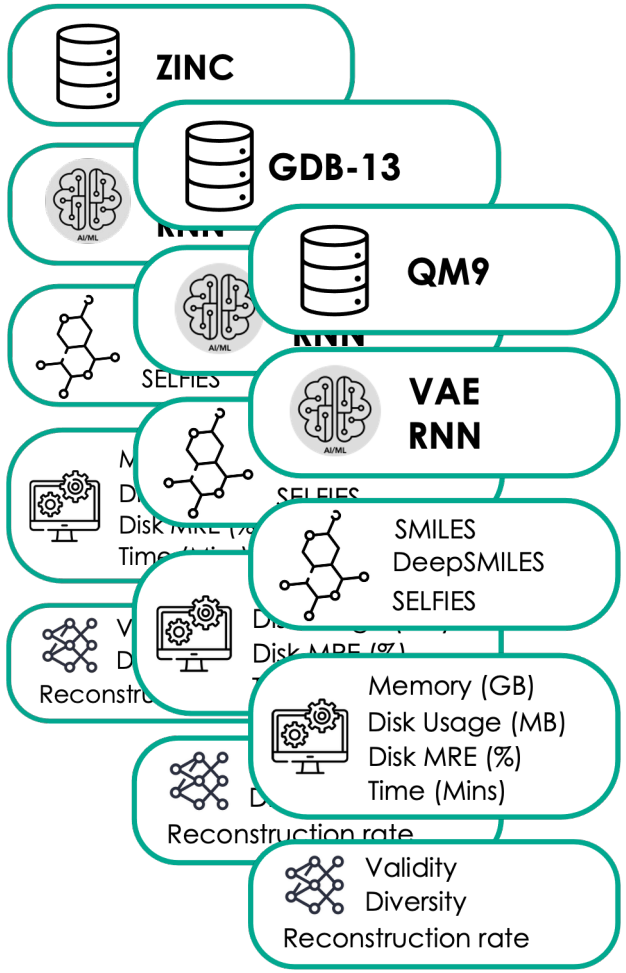
CCCCCCc1ccc(C2(c3ccc(CCCCCC)cc3)c3cc4c(cc3-c3sc5cc(C=C6C(=O)c7cc(F)c(F)cc7C6=C(C#N)C#N)sc5c32)C(c2ccc(CCCCCC)cc2)(c2ccc(CCCCCC)cc2)c2c-4sc3c4c(sc23)-c2cc3c(cc2C4(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)-c2sc4cc(C=C5C(=O)c6cc(F)c(F)cc6C5=C(C#N)C#N)sc4c2C3(c2ccc(CCCCCC) cc2)c2ccc(CCCCCC)cc2)cc1

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

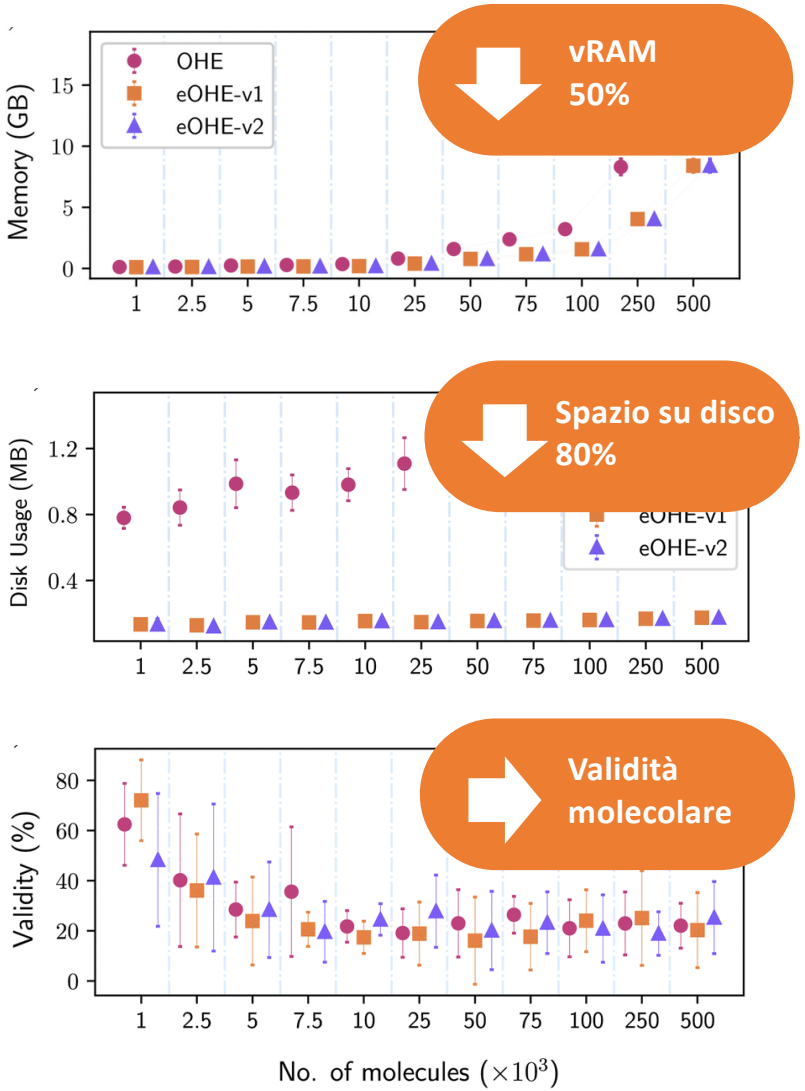
Codifica leggibile dalla macchina più efficiente

PARLARE DI CHIMICA





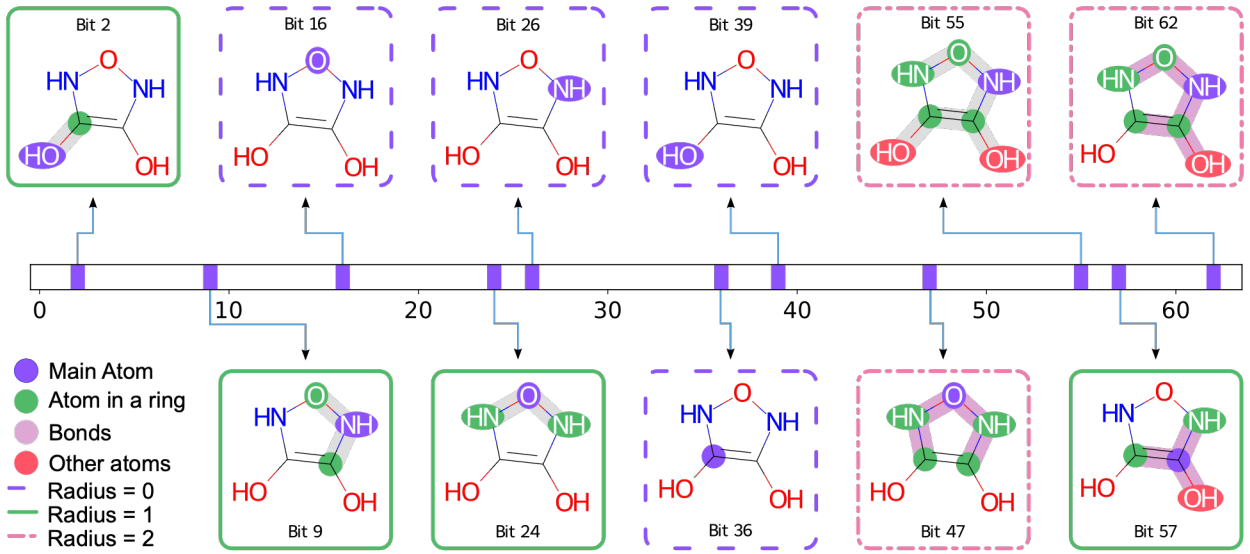
Prestazioni negli autoencoder variazionali



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

Fingerprint di Morgan incorporati (eMP)

Un fingerprint di Morgan è un vettore binario di dimensione fissa che codifica la presenza (1) o l'assenza (0) di sottostrutture in una molecola.



QM9: oltre 134.000 molecole
NFA: 50.000 molecole
RedDB: 15.000 molecole

Random Forest (RF), Percettrone multistrato (MLP), Regressore KNN (KNR), regressore gradient boosting (GBR), rete neurale profonda (DNN)

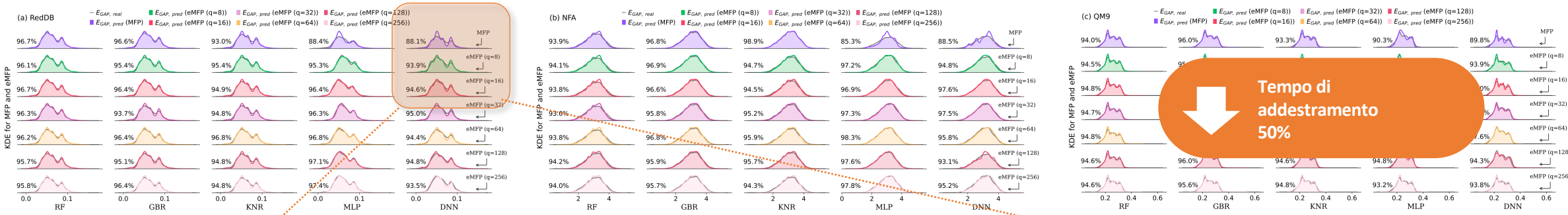
MFP
eMFP (q=8, 16, 32, 64, 128, 256)

150 prove per ottimizzare 15.750 iperparametri

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

Gap energetico con fingerprint di Morgan incorporati (eMP)

PARLARE DI CHIMICA



Tempo di addestramento 50%

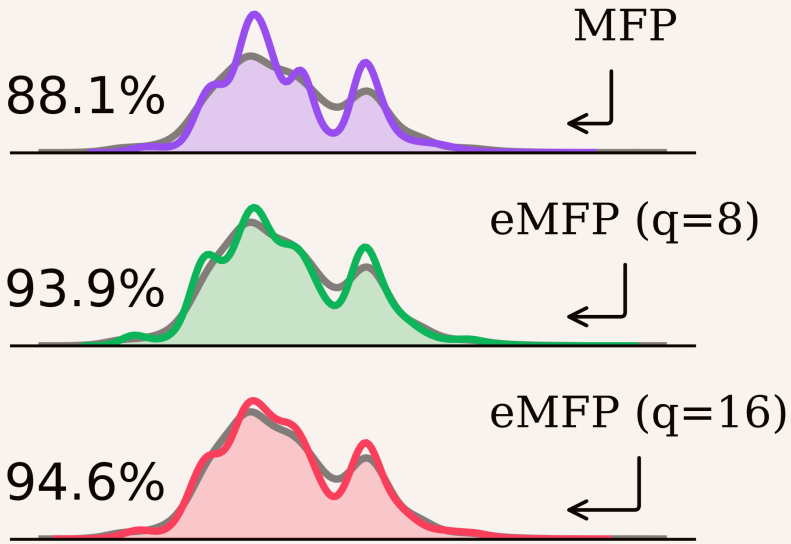
Base RedDB

$-\ E_{GAP,\ real}$

$E_{GAP,\ pred}$ (MFP)

$E_{GAP,\ pred}$ (eMFP (q=8))

$E_{GAP,\ pred}$ (eMFP (q=16))



Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, *in revisione,* **2026**

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

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES

B) SELFIES

C) InChI

D) Nessuna di queste

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES

B) SELFIES

C) InChI

D) Nessuna di queste

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES

B) SELFIES

C) InChI

D) Nessuna di queste

Q2 **eOHE/eMFP migliorano soprattutto…**

A) Solo l’accuratezza

B) Memoria/disco/tempo a pari accuratezza

C) Il numero di pubblicazioni

D) Niente

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES
B) SELFIES
C) InChI
D) Nessuna di queste

Q2 **eOHE/eMFP migliorano soprattutto…**

A) Solo l'accuratezza
B) Memoria/disco/tempo a pari accuratezza
C) Il numero di pubblicazioni
D) Niente

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES
B) SELFIES
C) InChI
D) Nessuna di queste

Q2 **eOHE/eMFP migliorano soprattutto…**

A) Solo l'accuratezza
B) Memoria/disco/tempo a pari accuratezza
C) Il numero di pubblicazioni
D) Niente

Q3 **Perché il costo energetico dell'IA è importante?**

A) Non lo è
B) Sostenibilità e democratizzazione
C) Il calcolo è gratuito
D) Solo per i videogiochi

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Quale rappresentazione garantisce una molecola valida per ogni stringa?**

A) SMILES
B) **SELFIES**
C) InChI
D) Nessuna di queste

Q2 **eOHE/eMFP migliorano soprattutto…**

A) Solo l'accuratezza
B) **Memoria/disco/tempo a pari accuratezza**
C) Il numero di pubblicazioni
D) Niente

Q3 **Perché il costo energetico dell'IA è importante?**

A) Non lo è
B) **Sostenibilità e democratizzazione**
C) Il calcolo è gratuito
D) Solo per i videogiochi

PARLARE DI CHIMICA

Messaggi chiave – La rappresentazione è una scelta di sostenibilità.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. La vettorizzazione è una scelta di progettazione di primo piano, non un passaggio secondario di pre-elaborazione.

2. eOHE/eMFP riducono memoria, disco e tempo a pari accuratezza
efficienza = calcolo più verde e più democratico.

3. Un'informatica verde e accessibile fa parte di un'accelerazione responsabile.

DOMANDE?

SVILUPPO DI MATERIALI GUIDATO DAGLI LLM

Sviluppo di materiali guidato dagli LLM: un nuovo livello nello stack

Oltre la previsione di numeri: modelli che leggono, ragionano, pianificano e usano strumenti

La svolta: l'AA specializzato associa una struttura a una proprietà. Un modello linguistico di grandi dimensioni aggiunge un'interfaccia in linguaggio naturale, ampie conoscenze pregresse, una pianificazione passo dopo passo e la capacità di richiamare strumenti esterni, **trasformando un predittore in un collaboratore di ricerca**.

1

Interfaccia in linguaggio naturale

Si interroga, si adatta con pochi esempi; forte quando i dati sono scarsi, dove l'AA classico fatica.

2

Conoscenza del mondo

Pre-addestrato sulla letteratura scientifica: porta conoscenze chimiche pregresse e contesto, non solo adattamento di curve.

3

Pianificazione e ragionamento

Scomporre un obiettivo («trovare un catodo agli ioni di sodio stabile») in passaggi ordinati e verificabili.

4

Uso di strumenti

Richiamare simulatori, banche dati e robot, basando le risposte su calcoli reali invece di tirare a indovinare.

…ma attenzione. Gli LLM sono fluenti ma inaffidabili. Inventano fatti, citano male e possono sbagliare con sicurezza. Tutto dipende da **ancorarli a strumenti e validarli**

Che cosa fanno davvero gli LLM nello sviluppo di materiali

1 Leggere la letteratura

Trasformare migliaia di articoli in banche dati strutturate e interrogabili (composizioni, MOF, droganti).

es. **Dagdelen et al. 2024**

2 Prevedere proprietà

Rispondere a domande di chimica in linguaggio naturale; competitivo con l'AA, forte quando i dati sono scarsi.

es. **Jablonka et al. 2024**

3 Generare e progettare in modo inverso

Scrivere molecole e cristalli come testo; generazione orientata alle proprietà.

es. **Gruver et al. 2024 · Chemeleon**

4 Pianificare la sintesi

Proporre vie sintetiche e ricette; retrosintesi e progettazione di procedure.

es. **ChemCrow 2024**

5 Ragionare con strumenti

Dotare il modello di strumenti specialistici (RDKit, calcolatori, ricerca) perché smetta di tirare a indovinare.

es. **ChemCrow – 18 strumenti**

6 Condurre esperimenti in autonomia

Cicli pianificare-agire-osservare che guidano strumenti e robot per verificare ipotesi.

es. **Coscientist 2023**

Sistemi LLM di riferimento per chimica e materiali

Sistemi rappresentativi: che cosa fanno, l'approccio e la fonte

Sistema	Che cosa fa	LLM/approccio	Riferimento
Coscientist	progetta, pianifica ed esegue esperimenti (ad es. accoppiamenti incrociati al Pd)	GPT-4 + web/documenti/codice + robotica	*Boiko et al., Nature 2023*
ChemCrow	sintesi organica, progettazione di farmaci e materiali; ha eseguito sintesi reali	GPT-4 + 18 strumenti specialistici	*Bran et al., Nat. Mach. Intell. 2024*
LLM di chimica predittiva	previsione e classificazione di proprietà dal testo; pochi dati	GPT-3 con fine-tuning; in contesto	*Jablonka et al., Nat. Mach. Intell. 2024*
Estrazione strutturata	legge articoli → record strutturati (droganti, MOF, composizioni)	GPT-3/Llama-2 con fine-tuning	*Dagdelen et al., Nat. Commun. 2024*
LLM cristallo-testo	genera cristalli inorganici stabili scritti come testo	LLaMA-2 70B con fine-tuning	*Gruver et al., ICLR 2024*
Chemeleon	generazione di strutture cristalline guidata dal testo	contrastivo testo–cristallo + diffusione	*Park et al., Nat. Commun. 2025*

Riferimenti

LLM e agenti autonomi per chimica e materiali

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

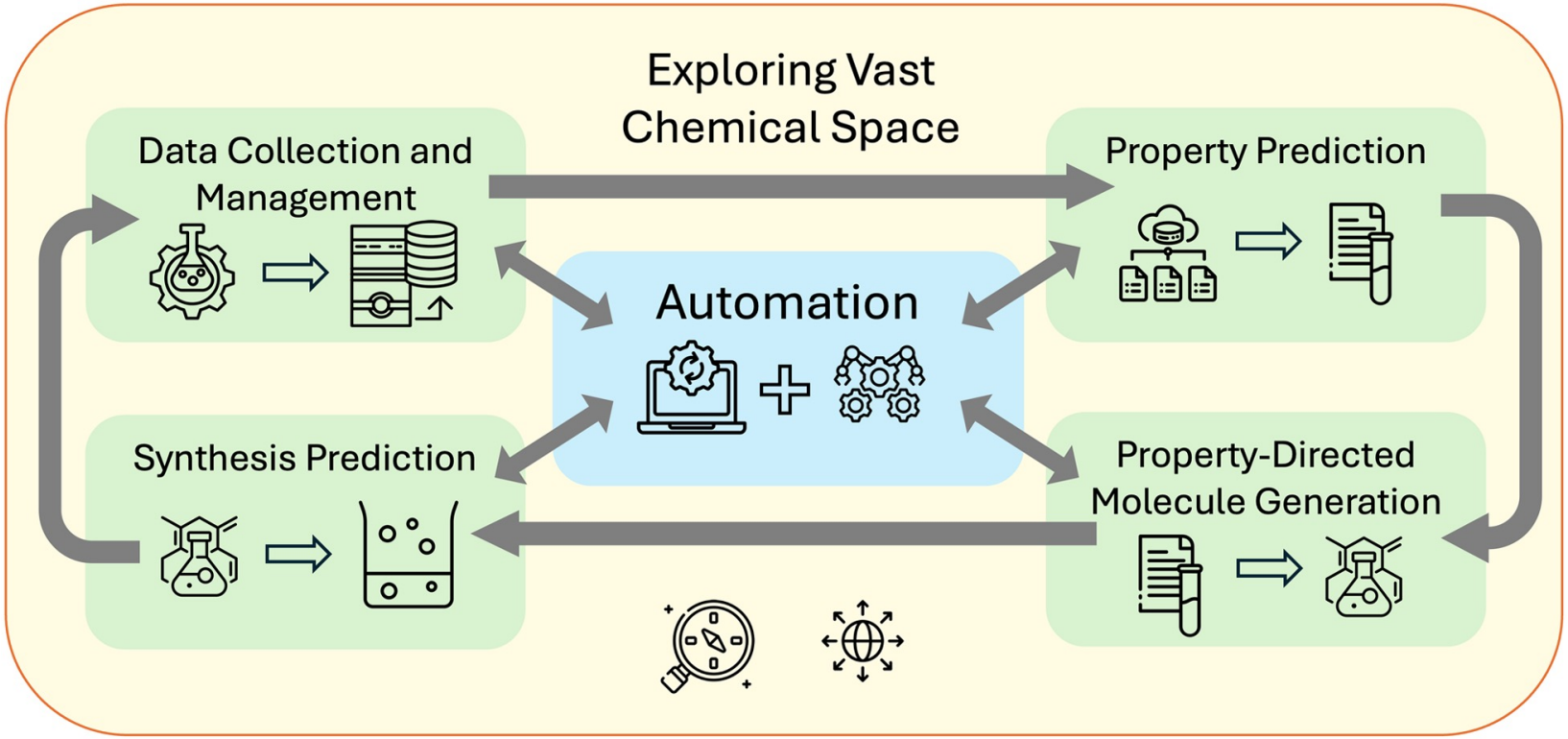
5. Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-Tuned Language Models Generate Stable Inorganic Materials as Text. ICLR 2024. arXiv:2402.04379

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

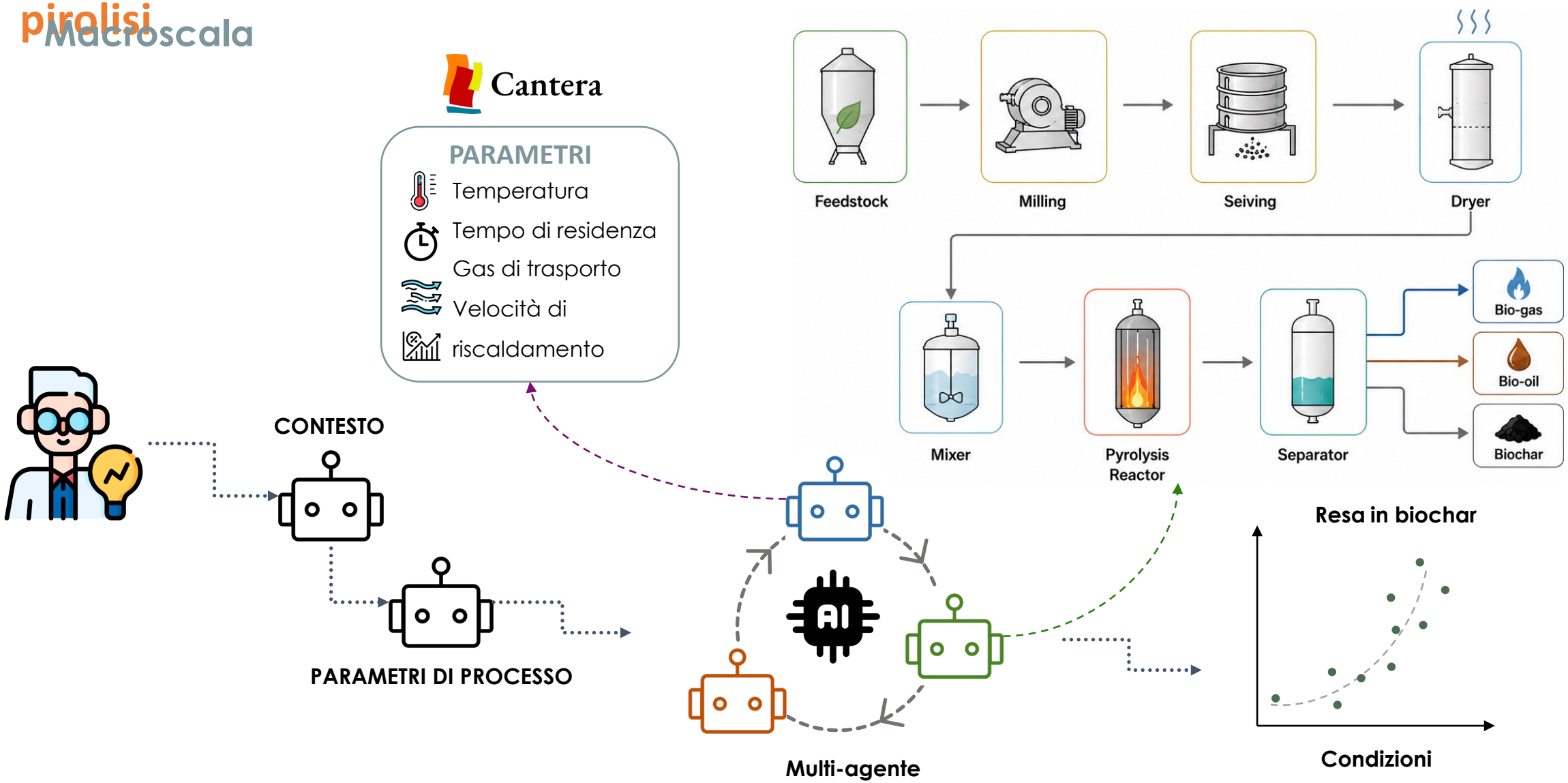
Che cosa possono fare davvero gli LLM e gli agenti in chimica?



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

Modellare il biochar – ingegneria del processo di pirolisi

Macroscala



PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande

B) Pianifica e agisce tramite strumenti, in ciclo

C) Non usa mai strumenti

D) Funziona solo offline

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande

B) Pianifica e agisce tramite strumenti, in ciclo

C) Non usa mai strumenti

D) Funziona solo offline

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande

B) Pianifica e agisce tramite strumenti, in ciclo

C) Non usa mai strumenti

D) Funziona solo offline

Q2 **Dotare un LLM di strumenti (RDKit, ricerca, simulatori)…**

A) Lo peggiora

B) Lo ancora e ne migliora l'affidabilità

C) Elimina tutti gli errori

D) È impossibile

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande

B) Pianifica e agisce tramite strumenti, in ciclo

C) Non usa mai strumenti

D) Funziona solo offline

Q2 **Dotare un LLM di strumenti (RDKit, ricerca, simulatori)…**

A) Lo peggiora

B) Lo ancora e ne migliora l'affidabilità

C) Elimina tutti gli errori

D) È impossibile

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande
B) Pianifica e agisce tramite strumenti, in ciclo
C) Non usa mai strumenti
D) Funziona solo offline

Q2 **Dotare un LLM di strumenti (RDKit, ricerca, simulatori)…**

A) Lo peggiora
B) Lo ancora e ne migliora l'affidabilità
C) Elimina tutti gli errori
D) È impossibile

Q3 **Gli LLM sono particolarmente utili quando…**

A) I dati sono enormi
B) I dati sono scarsi
C) Mai
D) Solo per le immagini

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (scelta multipla)

Scelta multipla

Q1 **Un «agente» LLM (rispetto a un assistente) è uno che…**

A) Risponde solo alle domande
B) Pianifica e agisce tramite strumenti, in ciclo
C) Non usa mai strumenti
D) Funziona solo offline

Q2 **Dotare un LLM di strumenti (RDKit, ricerca, simulatori)…**

A) Lo peggiora
B) Lo ancora e ne migliora l'affidabilità
C) Elimina tutti gli errori
D) È impossibile

Q3 **Gli LLM sono particolarmente utili quando…**

A) I dati sono enormi
B) I dati sono scarsi
C) Mai
D) Solo per le immagini

Messaggi chiave – Liberare la conoscenza intrappolata nei testi.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. Gli LLM aggiungono un livello di linguaggio + ragionamento + uso di strumenti all'AA specializzato.

2. Trasformare decenni di letteratura in dati utilizzabili; pianificare sintesi; orchestrare strumenti.

3. Mantenere un essere umano responsabile e registrare ogni azione. Tracciabilità e supervisione rendono l'autonomia affidabile (e riproducibile).

DOMANDE?

IA RESPONSABILE



IA responsabile per materiali e chimica

Strumenti potenti comportano responsabilità. La stessa accelerazione è un'arma a doppio taglio

1 Integrità e riproducibilità

Fughe di dati e rincorsa alle classifiche gonfiano i risultati.
Documentare abbastanza perché altri possano riprodurli.

2 Previsioni affidabili

Indicare l'incertezza, rispettare il dominio di applicabilità e validare a valle (DFT → esperimento).

3 Qualità dei dati e bias

I set di dati privilegiano sistemi ben studiati e di successo.
Dati spazzatura, risultati spazzatura

4 Apertura e FAIR

Condividere dati, codice e pesi dei modelli
Reperibili, accessibili, interoperabili, riutilizzabili (FAIR).

5 Sicurezza e duplice uso

Potenti strumenti di progettazione possono essere ribaltati per nuocere.
Controllo degli accessi e supervisione umana.

6 Sostenibilità e calcolo

Addestramento e inferenza consumano energia.
Dimensionare bene i modelli e riutilizzare quelli pre-addestrati.

Bias nei dati

Sfide

- Convinzioni implicite e bias dei contesti occidentali possono essere normalizzati nell'IA, rendendola meno responsabile su scala globale.
- Serve una prospettiva globale nei presupposti fondamentali dell'IA: dati di addestramento, prestazioni dei modelli e **spiegabilità** dei sistemi
- Porre subito le domande difficili sugli impatti positivi e negativi, invece di introdurre riparazioni e correzioni a posteriori.

Azioni

- Essere robusti rispetto al multilinguismo e ai diversi livelli di alfabetizzazione
- Personalizzare le esperienze con il modello di identità condivisa
- **Spiegabilità** e trasparenza devono avere senso per i nuovi utenti della tecnologia
- Affrontare in modo proattivo l'equità di fronte agli squilibri dei dati

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

Riferimenti

IA responsabile e affidabile per la scienza

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).

PARTECIPAZIONE

CHIUSURA DELLA SEZIONE · QUIZ (vero/falso)

Vero o falso?

1. **I set di dati di chimica pubblicati sono distorti, in parte perché gli esperimenti falliti vengono riportati raramente.**

   Vero | Falso

2. **Un modello di progettazione di farmaci può essere ribaltato per progettare tossine.**

   Vero | Falso

3. **FAIR significa Findable, Accessible, Interoperable, Reusable (reperibili, accessibili, interoperabili, riutilizzabili).**

   Vero | Falso

Messaggi chiave – Accelerare serve solo se è affidabile e condiviso.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. I dati sono distorti: bias di pubblicazione, bias umani, presupposti del contesto occidentale.

2. Gli strumenti generativi/agentici potenti sono intrinsecamente a duplice uso.

3. Riproducibilità, apertura (FAIR) e costo computazionale fanno parte del lavorare bene.

DOMANDE?

IA SPIEGABILE

IA spiegabile: dalla previsione alla comprensione

Nella scienza di solito vogliamo il «perché», non solo il numero

Il divario: un modello può prevedere una proprietà con precisione e non insegnarci comunque nulla.
La spiegabilità si chiede che cosa ha imparato il modello, se l'ha imparato per le ragioni giuste e quale nuova scienza suggerisce.

1 Comprensione

Trasformare un modello adattato in regole di progettazione e meccanismi. Recuperare la fisica/chimica dietro i dati.

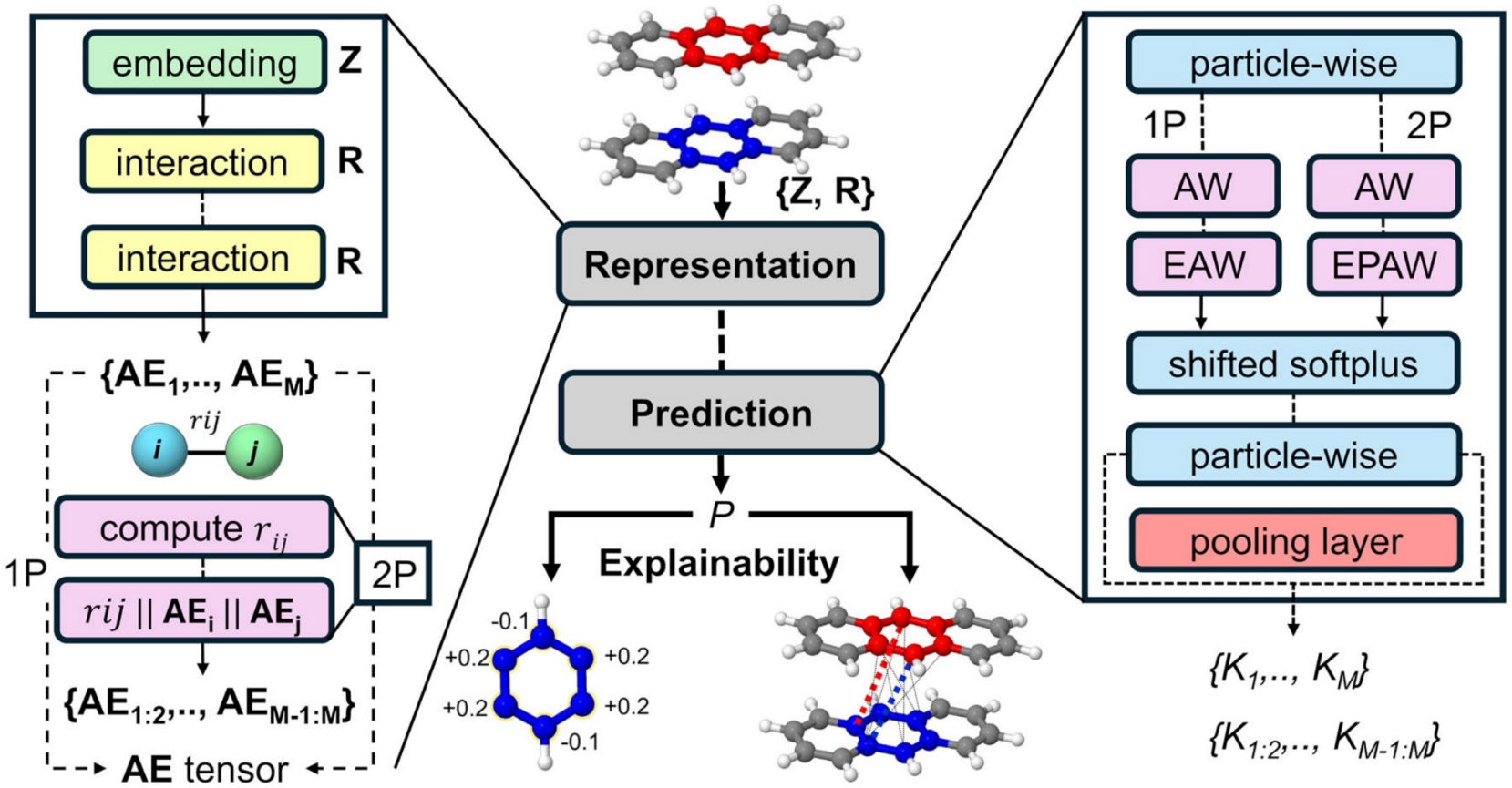
2 Fiducia e validazione

Verificare che il modello preveda per le ragioni giuste e non per false correlazioni o artefatti del set di dati.

3 Scoperta

L'interpretazione fa emergere nuove relazioni struttura–proprietà e ipotesi verificabili.

IA chimica spiegabile: 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

Riferimenti

IA spiegabile e interpretabile per la scienza e i materiali

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

IA SPIEGABILE

Messaggi chiave – Trasformare previsioni rapide in regole di progettazione affidabili.

PERCHÉ È IMPORTANTE PER IL CLIMA

1. Nella scienza vogliamo il perché, non solo il numero.

2. I metodi interpretabili smascherano i modelli che hanno ragione per i motivi sbagliati.

3. Trattare ogni spiegazione come un'ipotesi da verificare rispetto alla fisica e all'esperimento.

DOMANDE?

RIFLESSIONI FINALI

EDITORIAL | 19 May 2026

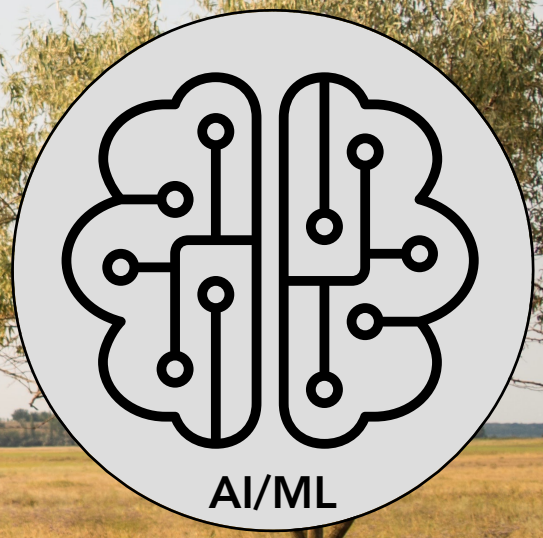
nature

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.

Progettare per la circolarità
Progettare per le prestazioni
e le prestazioni





Repository online: strumenti open source (materiali e IA)

Un kit di partenza da condividere. Gli strumenti del nostro gruppo sono evidenziati; verificare che i link siano aggiornati prima della distribuzione.

Ambito	Strumento	A che cosa serve	Repository
Dati e strumenti	**pymatgen**	Analisi dei materiali, I/O e API di Materials Project	*github.com/materialsproject/pymatgen*
Dati e strumenti	**matminer**	Descrittori/vettorizzazione pronti per l'AA	*github.com/hackingmaterials/matminer*
Dati e strumenti	**RDKit**	Chemioinformatica: molecole, fingerprint, descrittori	*github.com/rdkit/rdkit*
Dati e strumenti	**DeepChem**	Toolkit di AA per la chimica: set di dati, modelli, benchmark	*github.com/deepchem/deepchem*
Dati e strumenti	**ASE**	Atomic Simulation Environment: costruire/eseguire/analizzare	*gitlab.com/ase/ase*
Rappresentazioni	**SELFIES**	Rappresentazione molecolare a stringa sempre valida	*github.com/aspuru-guzik-group/selfies*
Rappresentazioni	**eOHE**	Codifiche one-hot incorporate (input efficienti)	*github.com/MMLabCodes/eOHE*
Rappresentazioni	**eMFP**	Fingerprint di Morgan incorporati (compatti, più veloci)	*github.com/MMLabCodes/eMFP*
Generativi	**chemical_vae**	Progettazione molecolare in uno spazio latente continuo	*github.com/aspuru-guzik-group/chemical_vae*
Generativi	**MatterGen**	Modello generativo a diffusione per cristalli inorganici	*github.com/microsoft/mattergen*
Potenziali di AA	**MACE**	Potenziali interatomici di AA (+ modello fondazionale MACE-MP-0)	*github.com/ACEsuit/mace*
Potenziali di AA	**CHGNet**	Potenziale universale a rete neurale informato dalla carica	*github.com/CederGroupHub/chgnet*
Potenziali di AA	**MatGL (M3GNet)**	Potenziali universali a grafo + modelli di proprietà	*github.com/materialsvirtuallab/matgl*
Autonomia	**Ax / BoTorch**	Ottimizzazione bayesiana per laboratori autonomi	*github.com/facebook/Ax*
Agenti LLM	**ChemCrow**	Agente LLM per la chimica collegato a strumenti specialistici	*github.com/ur-whitelab/chemcrow-public*
Spiegabilità	**SHAP**	Spiegazioni per attribuzione delle caratteristiche per qualsiasi modello	*github.com/shap/shap*
Spiegabilità	**SISSO**	Scoperta di descrittori/formule interpretabili	*github.com/rouyang2017/SISSO*
Informatica verde	**CodeCarbon**	Monitorare il costo in carbonio del proprio calcolo	*github.com/mlco2/codecarbon*

Righe evidenziate = repository del nostro gruppo (github.com/MMLabCodes). Elenco rappresentativo, non esaustivo.

Letture di approfondimento

COLLO DI BOTTIGLIA

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

MODELLAZIONE E SIMULAZIONE

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

NUOVO PARADIGMA

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

CASSETTA ATTREZZI IA

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

PARLARE DI CHIMICA

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

IA RESPONSABILE

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

IA SPIEGABILE

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

Grazie. Ci sono domande finali?

[MM] LAB

