IA para acelerar la ciencia de materiales

Francisco Martin-Martinez

Profesor titular, King's College London

martinmartinezlab.com

ESCUELA DE VERANO VIRTUAL CCAI 2026 · CLASE SECTORIAL · 13 DE AGOSTO

¿QUIÉN SOY?

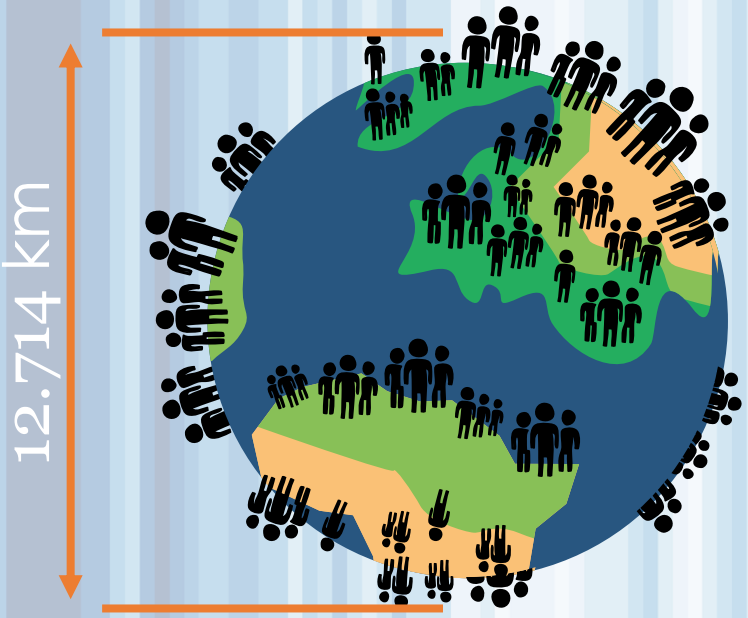
Quién soy y por qué esto me importa



Un científico que usa la IA como herramienta o una IA que hace ciencia de forma autónoma: ¿dónde está el límite?

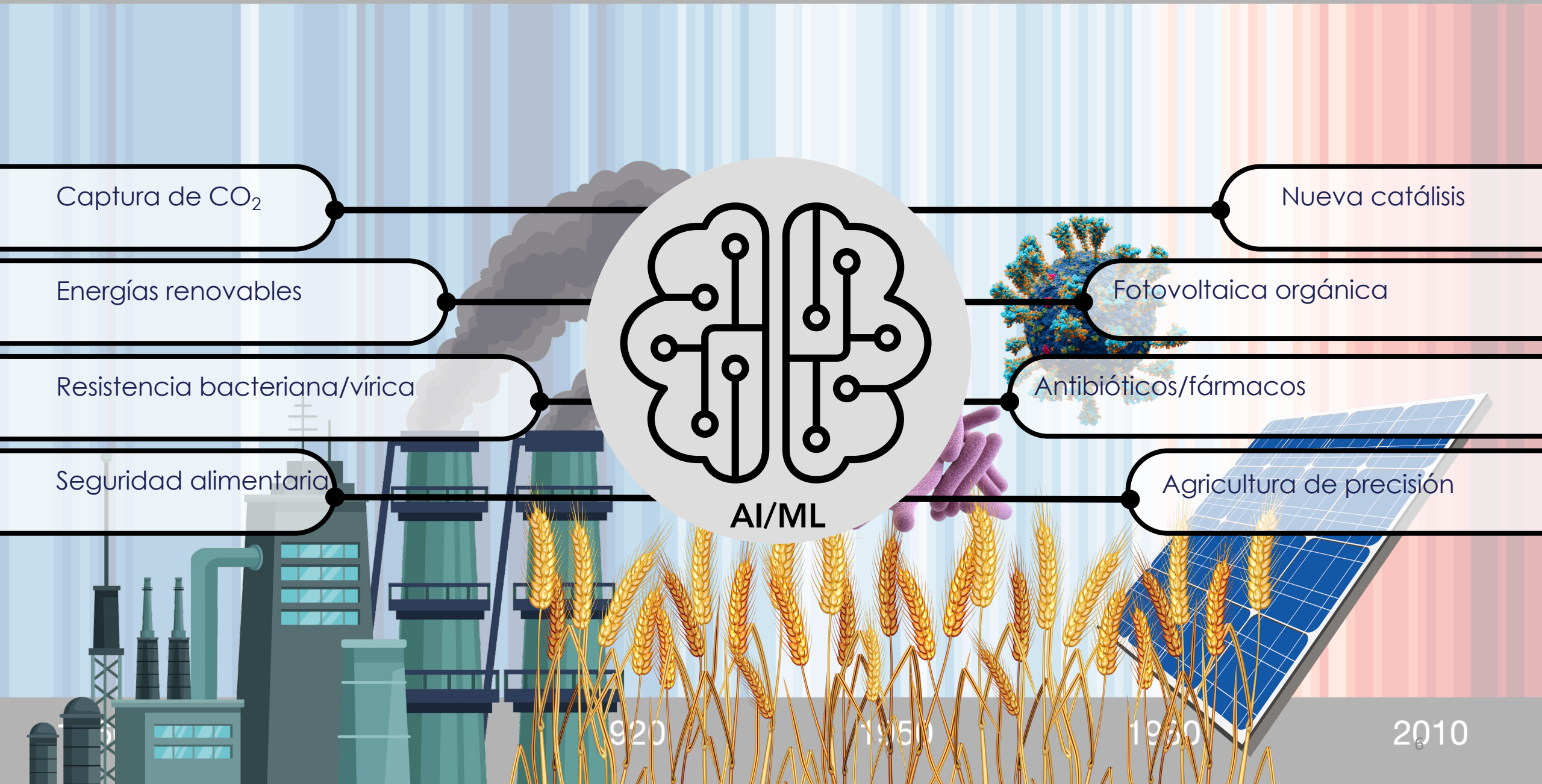
¿POR QUÉ ME IMPORTA?

Global temperature change (1850-2022)



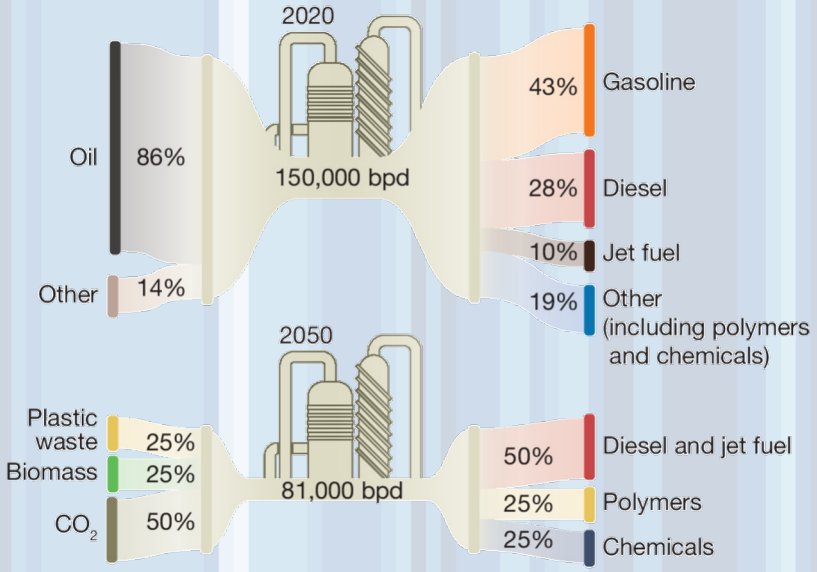
1860 1890 1920 1950 1980 2010

Global temperature change (1850-2022)



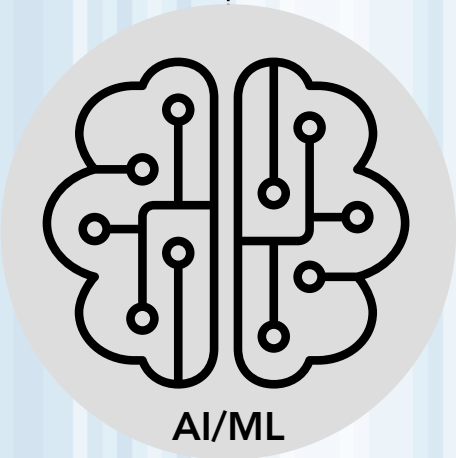
Nueva minería de materiales a partir de residuos

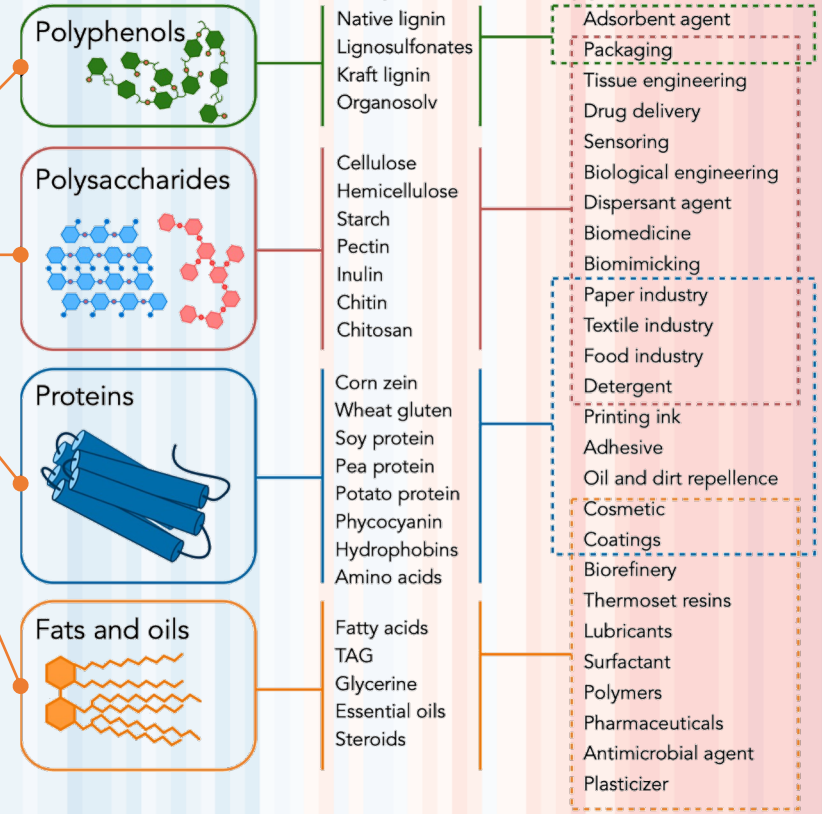
(Bio)rrefinería 2050



Nature 629, 295, **2024**

Residuos de biomasa
140 Gt/año

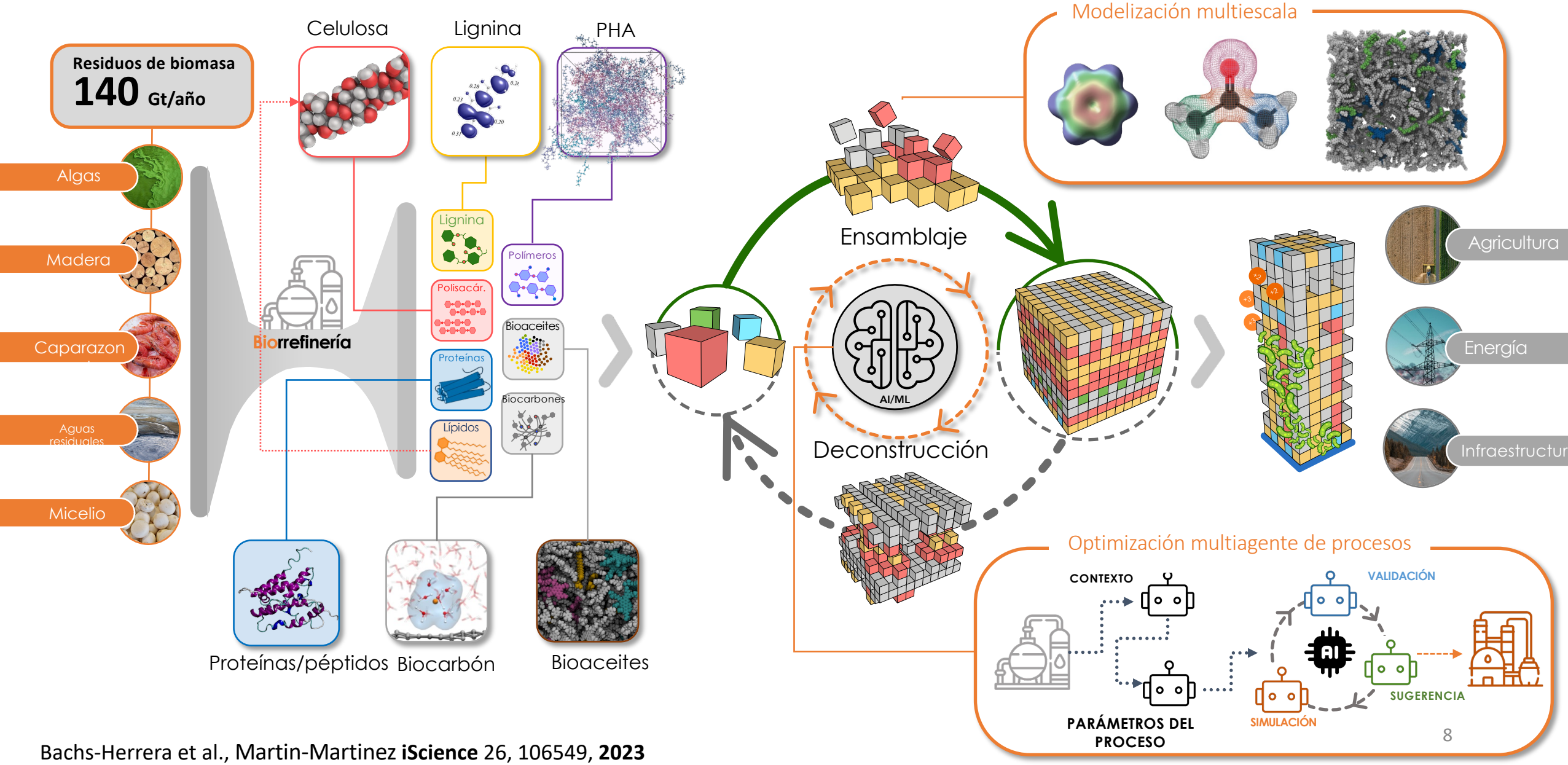




Bachs-Herrera, …, Martin-Martinez **iScience** 26, 106549, **2023**

Valorización de residuos en moléculas y materiales de origen biológico

Materiales circulares biobasados ➡ Hacia una biorrefinería cuántica con IA



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

¿QUÉ VEREMOS HOY?

QUÉ VEREMOS

01 Por qué los materiales son un cuello de botella climático

02 El paradigma tradicional (aún útil) de modelización y simulación de materiales

03 Un nuevo paradigma en el descubrimiento de materiales

04 La caja de herramientas de la IA

05 Hablar de química con los ordenadores

04 IA responsable

05 IA explicable en química

PARTICIPACIÓN

Encuesta en directo – Usen el chat

¿Qué reto relacionado con los materiales les parece más urgente para el clima?
No hay respuestas incorrectas: es para romper el hielo

¿PREGUNTAS?

EL CUELLO DE BOTELLA DE LOS MATERIALES

Dónde interviene la IA en la cadena de materiales

1

Representar

Codificar moléculas/cristales para el AA (huellas, grafos OHE, texto)

2

Predecir

Predicción de propiedades, modelos sustitutos, **potenciales interatómicos de AA**

3

Generar

Diseño inverso/generativo de nuevos candidatos

4

Sintetizar

Laboratorios autónomos y autodirigidos que cierran el ciclo

5

Caracterizar

Análisis asistido por IA de espectros, imágenes y estructuras

Los materiales, un cuello de botella climático

Fotovoltaica

Eficiencia y estabilidad limitadas por la química de los materiales

Baterías

Densidad energética, vida útil, dependencia de minerales críticos

Combustibles alternativos

Catalizadores y portadores para H_2 verde y e-combustibles

Sorbentes de CO_2

Materiales de captura baratos, duraderos y escalables

Otros: infraestructuras sostenibles, seguridad alimentaria, descubrimiento de fármacos, antibióticos…

El problema: el ciclo de descubrimiento es lento y costoso.

Diseñar → sintetizar → caracterizar → probar puede llevar años por candidato.

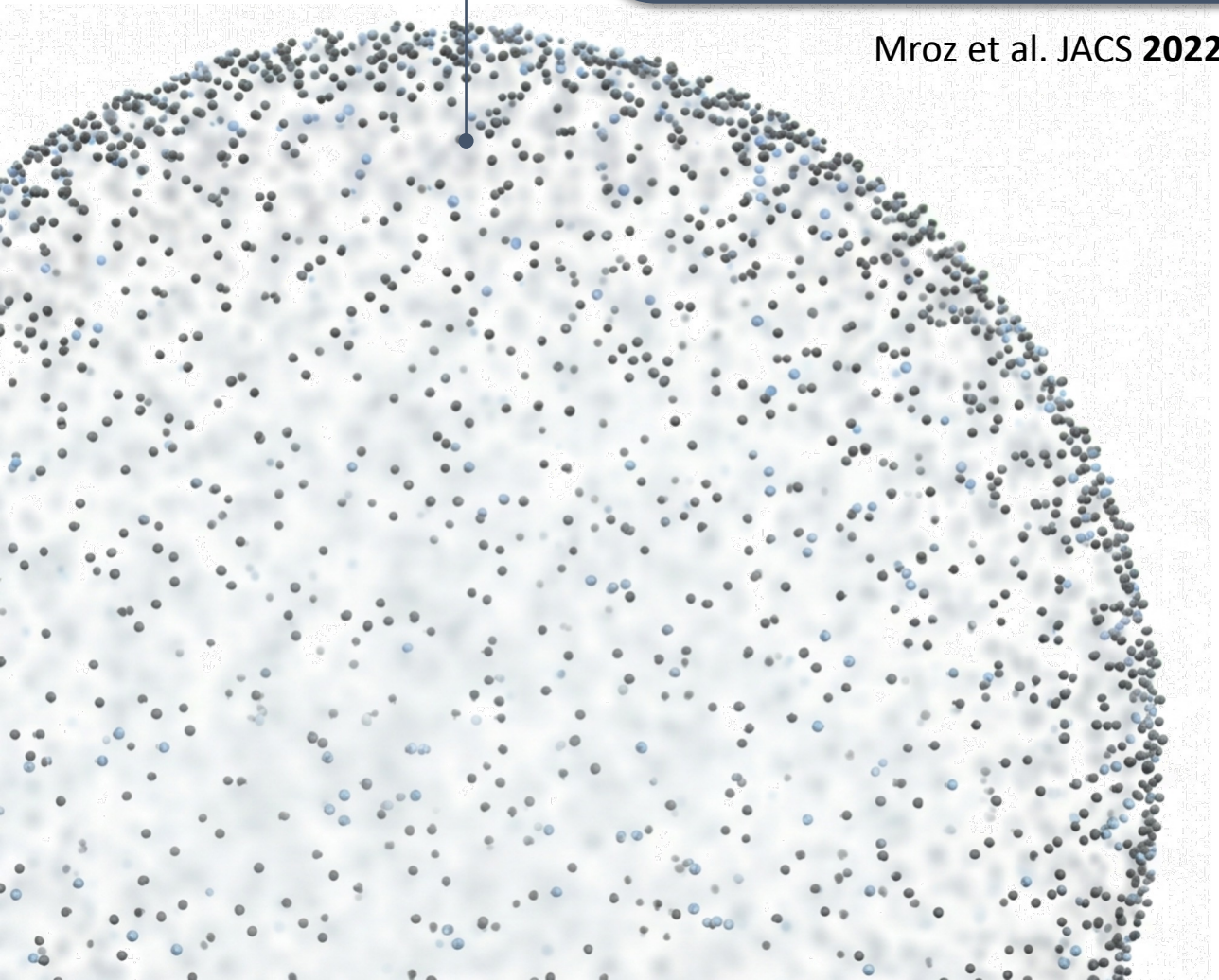
Tanto la mitigación COMO la adaptación dependen de obtener los materiales adecuados, más rápido y a menor coste.

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

El cuello de botella del descubrimiento

10^{60} Moléculas pequeñas (tipo fármaco) para el diseño de materiales

Mroz et al. JACS **2022**



El espacio químico combinatorio es inmenso

VS

La validación experimental es lenta y costosa

¿SABÍAS QUE…?

Hay más moléculas orgánicas pequeñas posibles que átomos en el universo observable.

PARTICIPACIÓN

Test rápido – Verdadero/Falso

1. **Hay más moléculas orgánicas pequeñas posibles que átomos en el universo observable.**

   Verdadero | Falso

2. **Un material nuevo suele llegar al mercado en unas semanas.**

   Verdadero | **Falso**

3. **La IA interviene en la cadena en las etapas de representar, predecir, generar, sintetizar y caracterizar.**

   Verdadero | Falso

Mensajes clave

POR QUÉ IMPORTA PARA EL CLIMA

1. Los materiales son un cuello de botella climático: tanto la mitigación como la adaptación necesitan mejores materiales, más rápido.

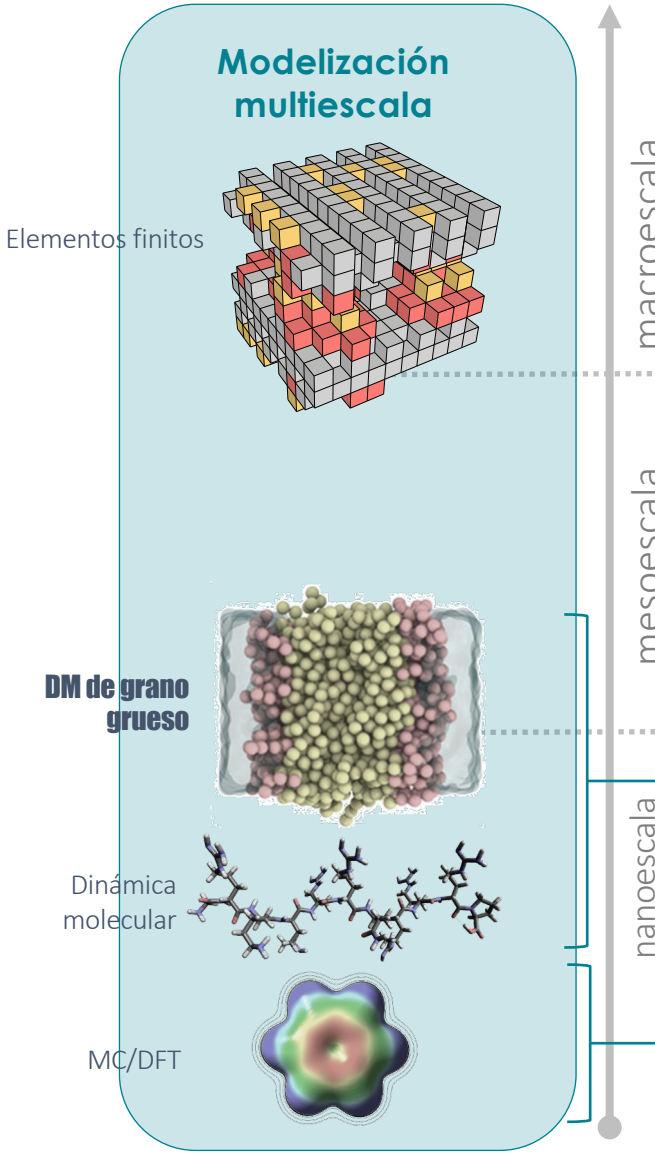
2. El espacio químico no puede explorarse por fuerza bruta. La IA lo convierte en una búsqueda guiada.

3. Los objetivos climáticos tienen un plazo fijo. Descubrir más rápido es desplegar más rápido.

¿PREGUNTAS?

MODELIZACIÓN Y SIMULACIÓN DE MATERIALES

Modelización multiescala



Simulaciones de dinámica molecular

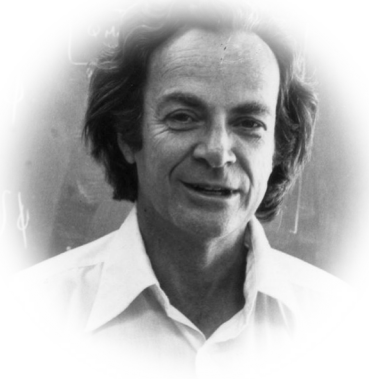


«Todos los modelos son erróneos, pero algunos son útiles».

– George E. P. Box



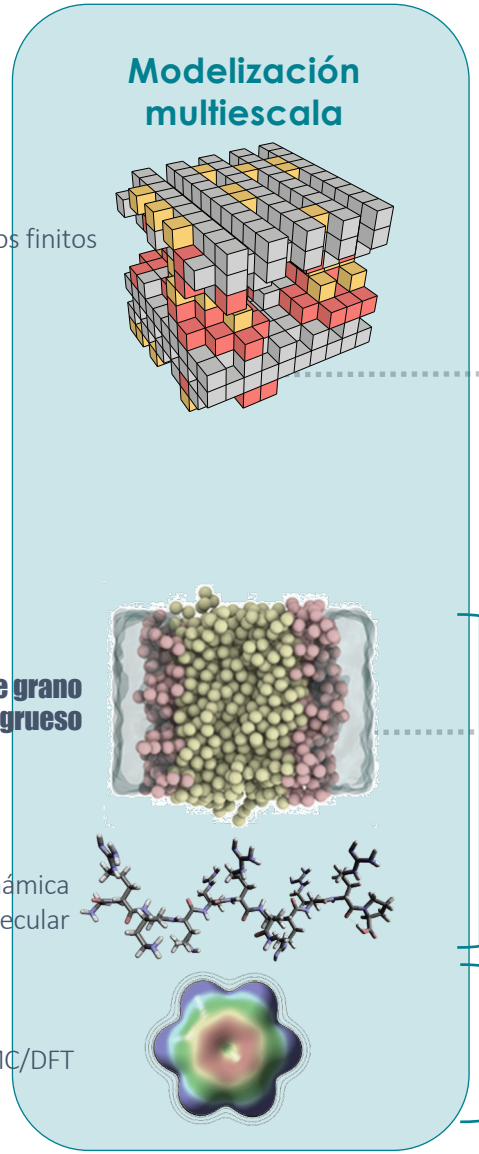
Teoría del funcional de



«Creo que puedo decir con seguridad que nadie entiende la mecánica cuántica».

– Richard Feynman

Modelización multiescala



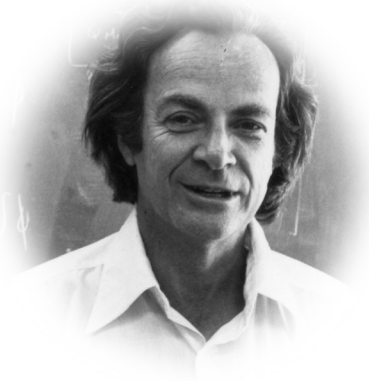
Simulaciones de dinámica molecular



«Todos los modelos son erróneos, pero algunos son útiles».

– George E. P. Box

Teoría del funcional de la densidad (DFT) – Mecánica cuántica

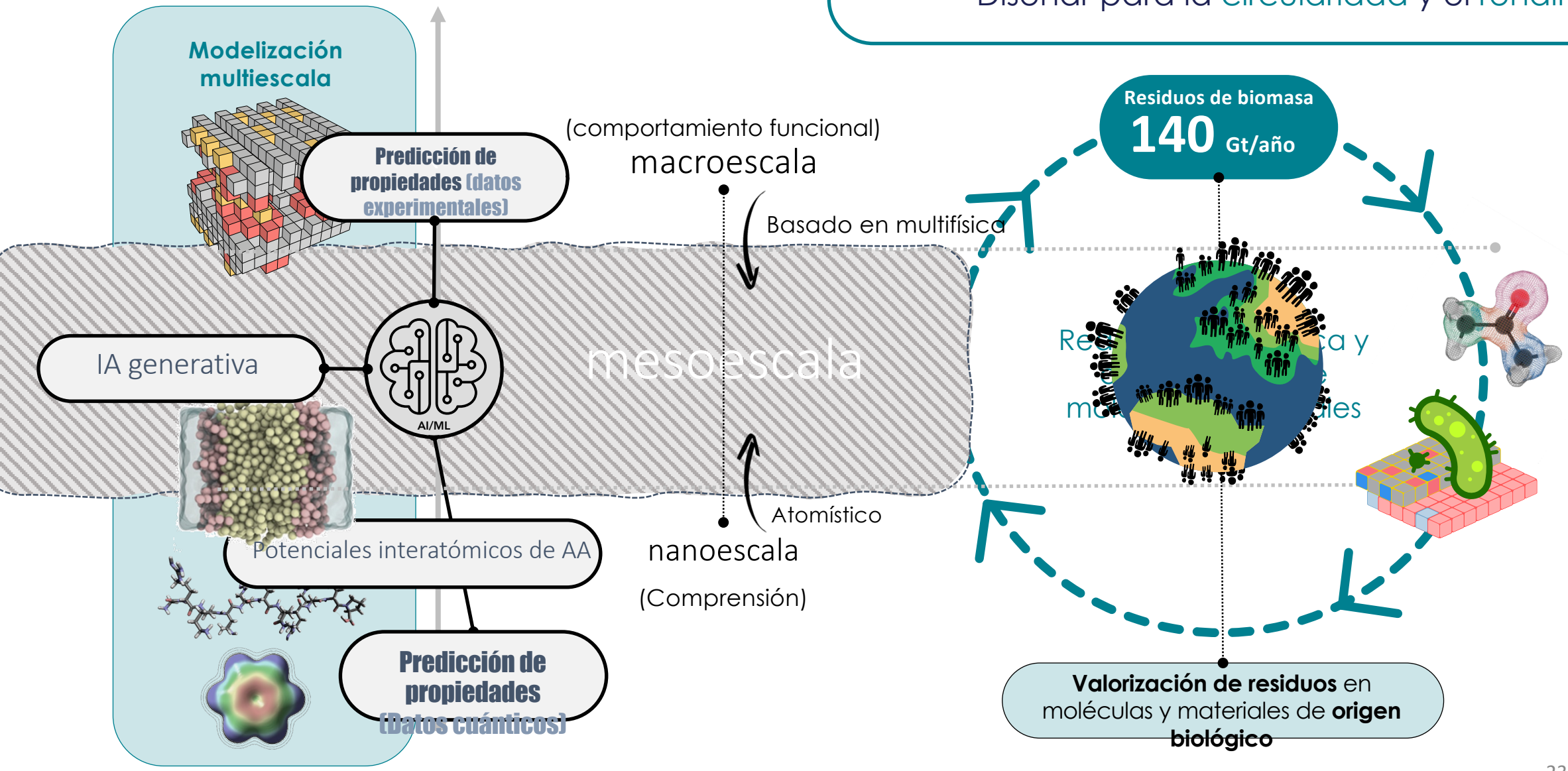


*«Creo que puedo decir con seguridad que **nadie entiende la mecánica cuántica**».*

– Richard Feynman

Modelización multiescala

Diseñar para la circularidad y el rendimiento



Software de DFT, DM y AA para química

DFT molecular



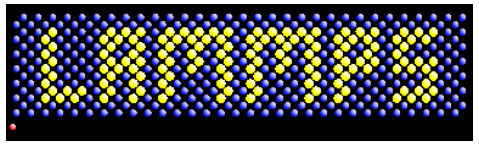
ORCA

DFT de estado sólido



siesta TM

DM (todos los átomos y grano grueso)



GROMACS
FAST. FLEXIBLE. FREE.



NAMD
Scalable Molecular Dynamics



MD ANALYSIS



Integración y quimioinformática

python™



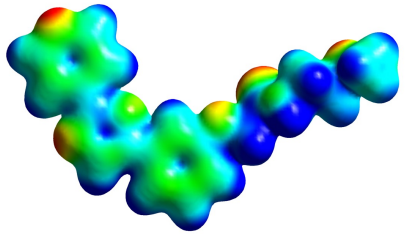
Open-Source Cheminformatics and Machine Learning

+ Atomic Simulation Environment (ASE)

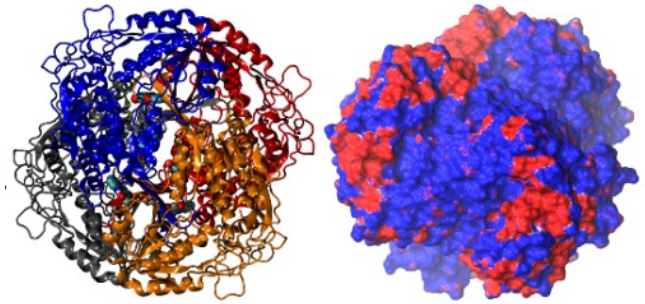
Bases de datos para facilitar el descubrimiento en química

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

Teoría del funcional de la densidad (DFT)



Dinámica molecular (DM)



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

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) DFT (cuántica)

D) Elementos finitos

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) **DFT (cuántica)**

D) Elementos finitos

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) **DFT (cuántica)**

D) Elementos finitos

Q2 **¿Por qué usamos la modelización multiescala?**

A) Basta con una escala

B) Los fenómenos varían según la escala

C) Para ahorrar citas

D) Lo exige la ley

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) **DFT (cuántica)**

D) Elementos finitos

Q2 **¿Por qué usamos la modelización multiescala?**

A) Basta con una escala

B) **Los fenómenos varían según la escala**

C) Para ahorrar citas

D) Lo exige la ley

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) DFT (cuántica)

D) Elementos finitos

Q2 **¿Por qué usamos la modelización multiescala?**

A) Basta con una escala

B) Los fenómenos varían según la escala

C) Para ahorrar citas

D) Lo exige la ley

Q3 **En comparación con la DM de todos los átomos, la DM de grano grueso alcanza…**

A) Sistemas más pequeños

B) Sistemas más grandes y tiempos más largos

C) Mayor precisión cuántica

D) Ninguna diferencia

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – opción múltiple

Q1 **¿Qué método es el más preciso, pero también el más costoso?**

A) DM clásica

B) DM de grano grueso

C) **DFT (cuántica)**

D) Elementos finitos

Q2 **¿Por qué usamos la modelización multiescala?**

A) Basta con una escala

B) **Los fenómenos varían según la escala**

C) Para ahorrar citas

D) Lo exige la ley

Q3 **En comparación con la DM de todos los átomos, la DM de grano grueso alcanza…**

A) Sistemas más pequeños

B) **Sistemas más grandes y tiempos más largos**

C) Mayor precisión cuántica

D) Ninguna diferencia

MODELIZACIÓN Y SIMULACIÓN

Mensajes clave – El lugar más barato para fallar es un ordenador.

POR QUÉ IMPORTA PARA EL CLIMA

1. Ninguna escala basta por sí sola: hace falta la modelización multiescala.

2. Debemos predecir la degradación y la reciclabilidad: circularidad, no solo rendimiento.

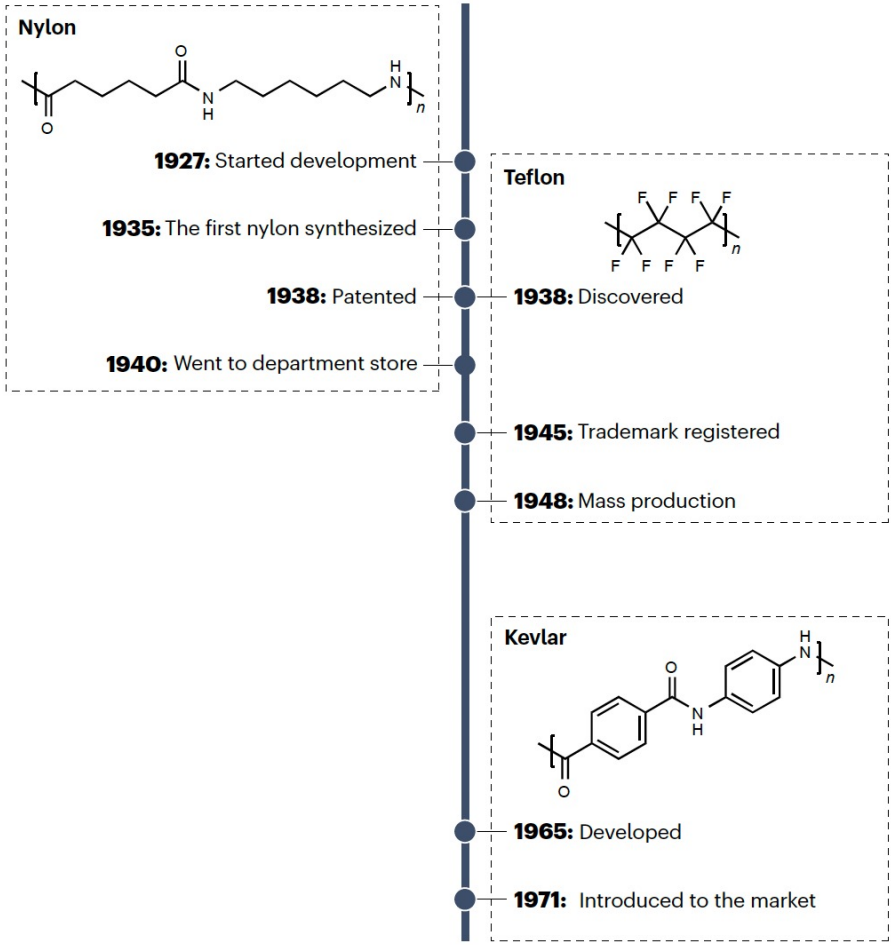
3. Los datos son la base sobre la que se construye la IA. Buenos datos hacen buenos modelos.

¿PREGUNTAS?

UN NUEVO PARADIGMA EN EL DESCUBRIMIENTO DE MOLÉCULAS Y MATERIALES

Materiales lentos, mundo rápido

Innovaciones en polímeros durante los dos últimos siglos.



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

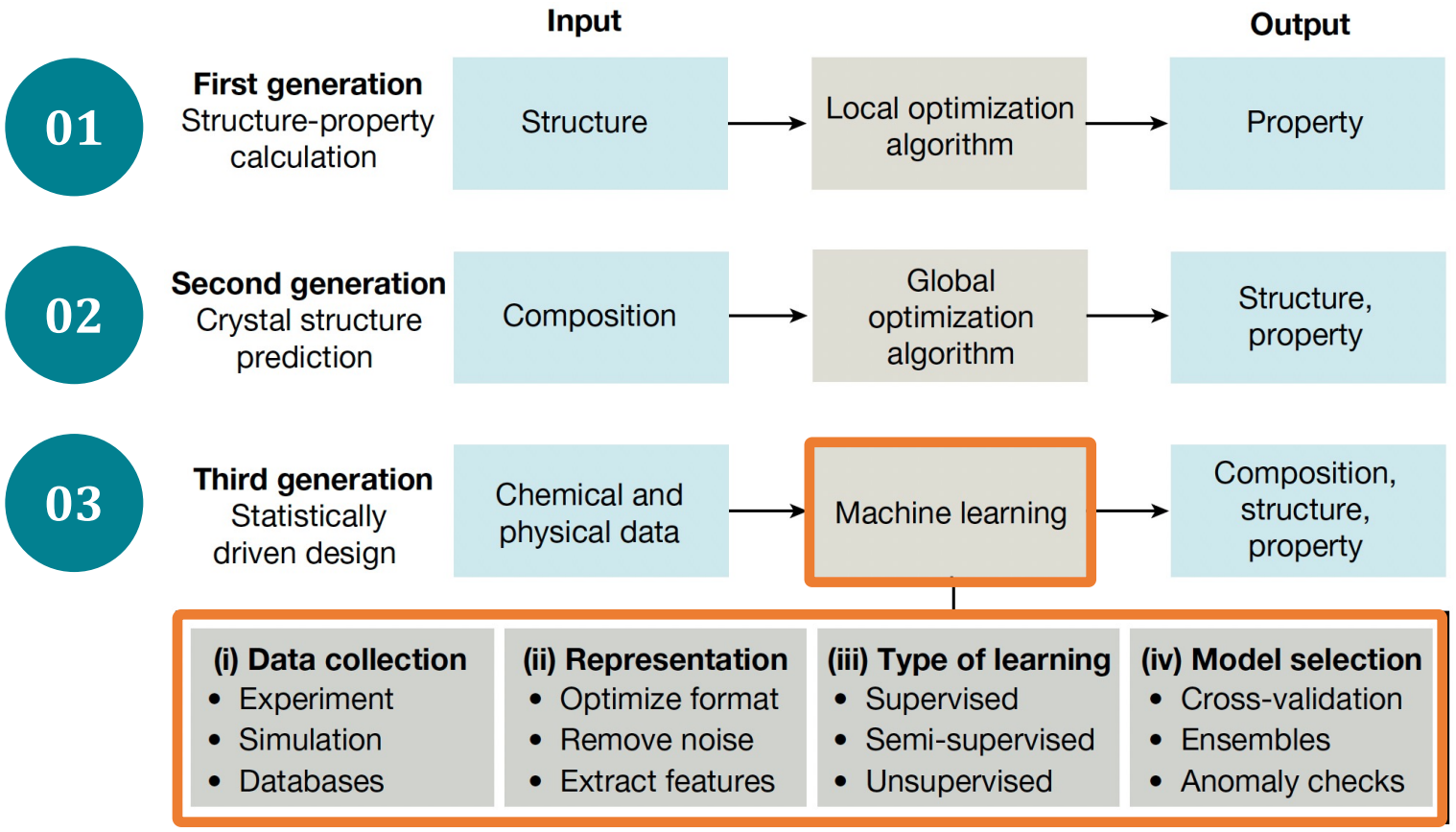
1977: Heeger, MacDiarmid and Shirakawa develop polyacetylene, the first conductive polymer

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).

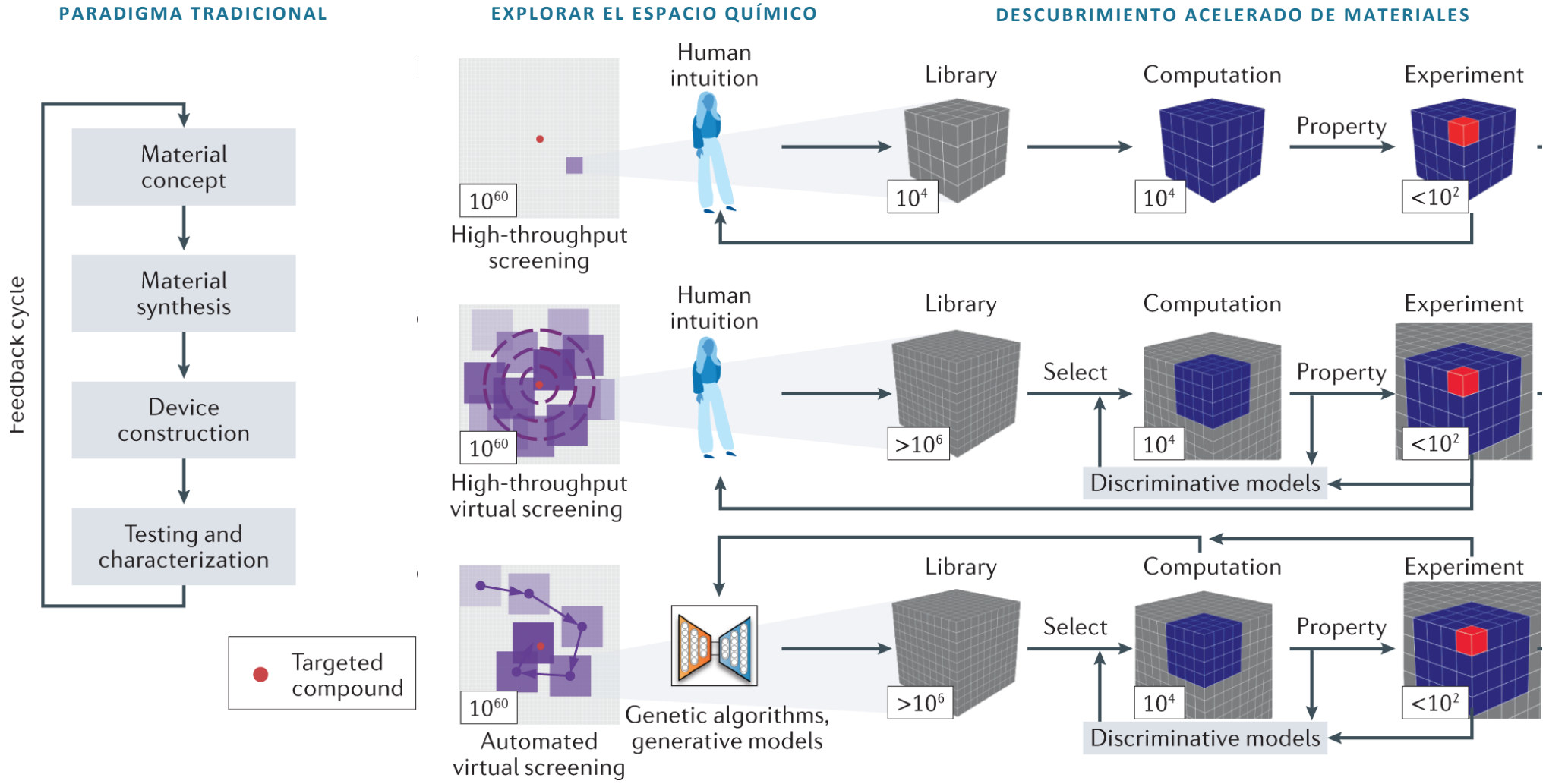
Del ensayo y error al diseño basado en datos:
evolución del flujo de trabajo de investigación en química computacional



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

Nuevo paradigma en el descubrimiento de materiales

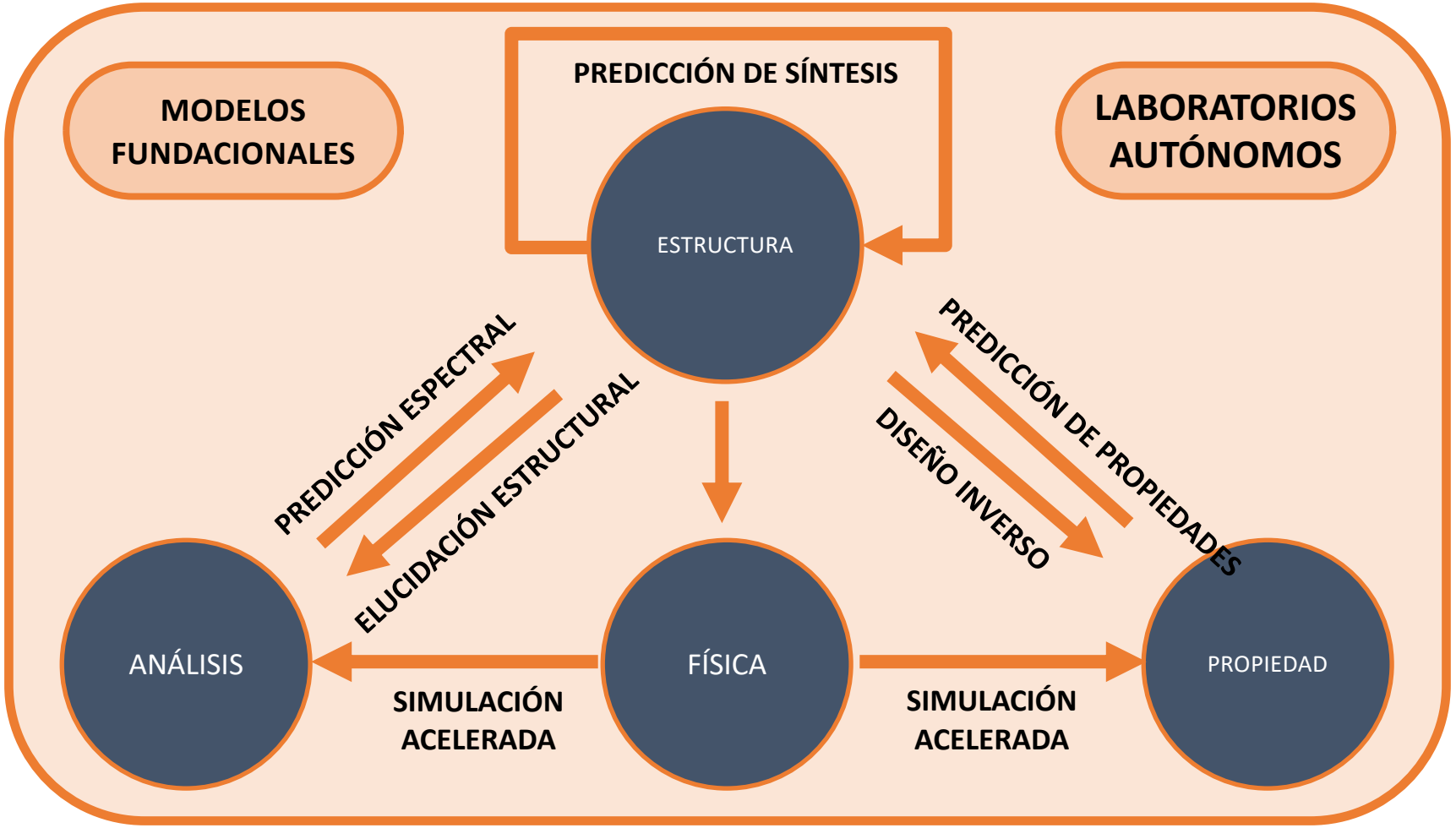
NUEVO PARADIGMA



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

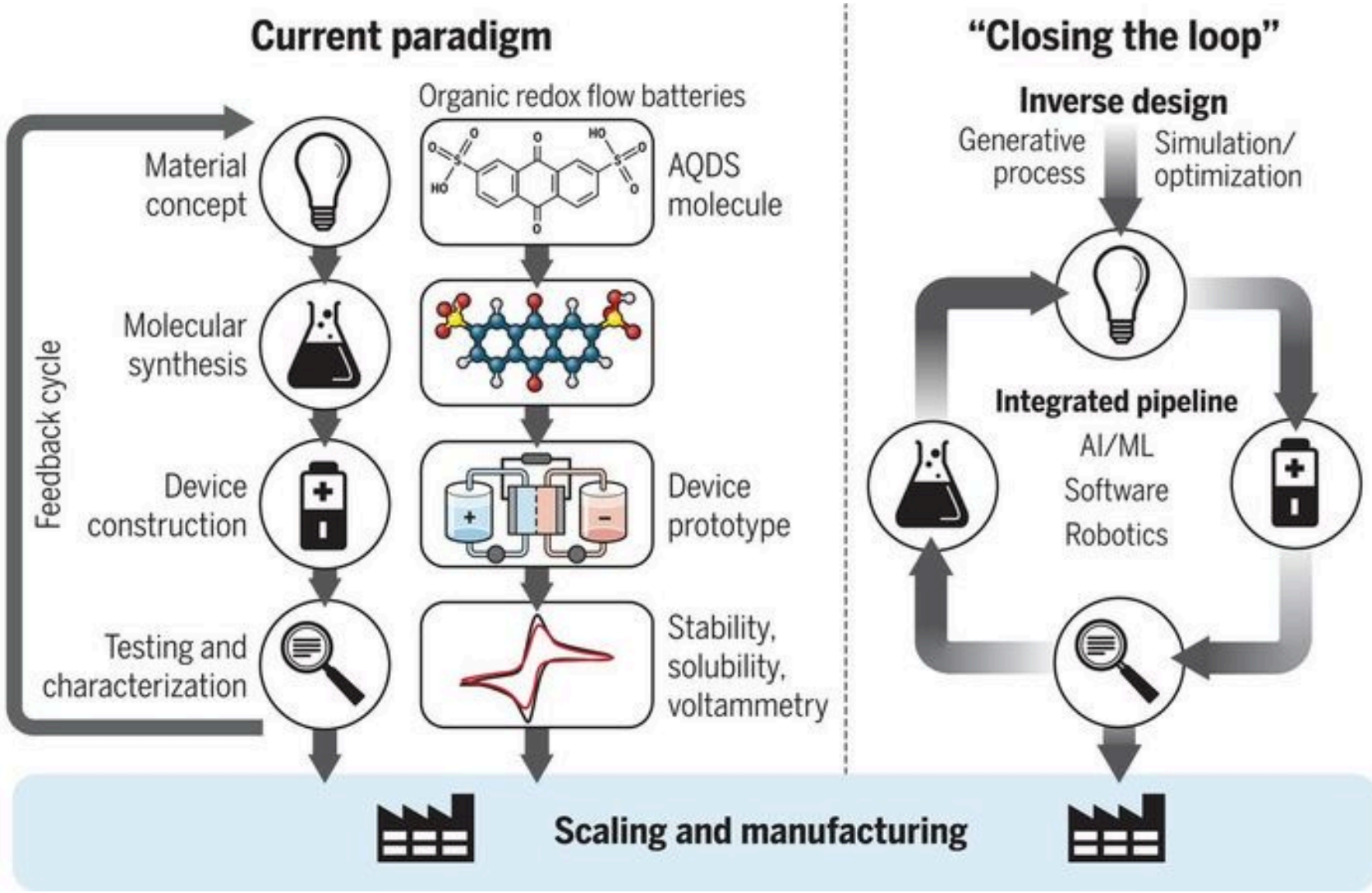
Aprendizaje automático para química y materiales

NUEVO PARADIGMA



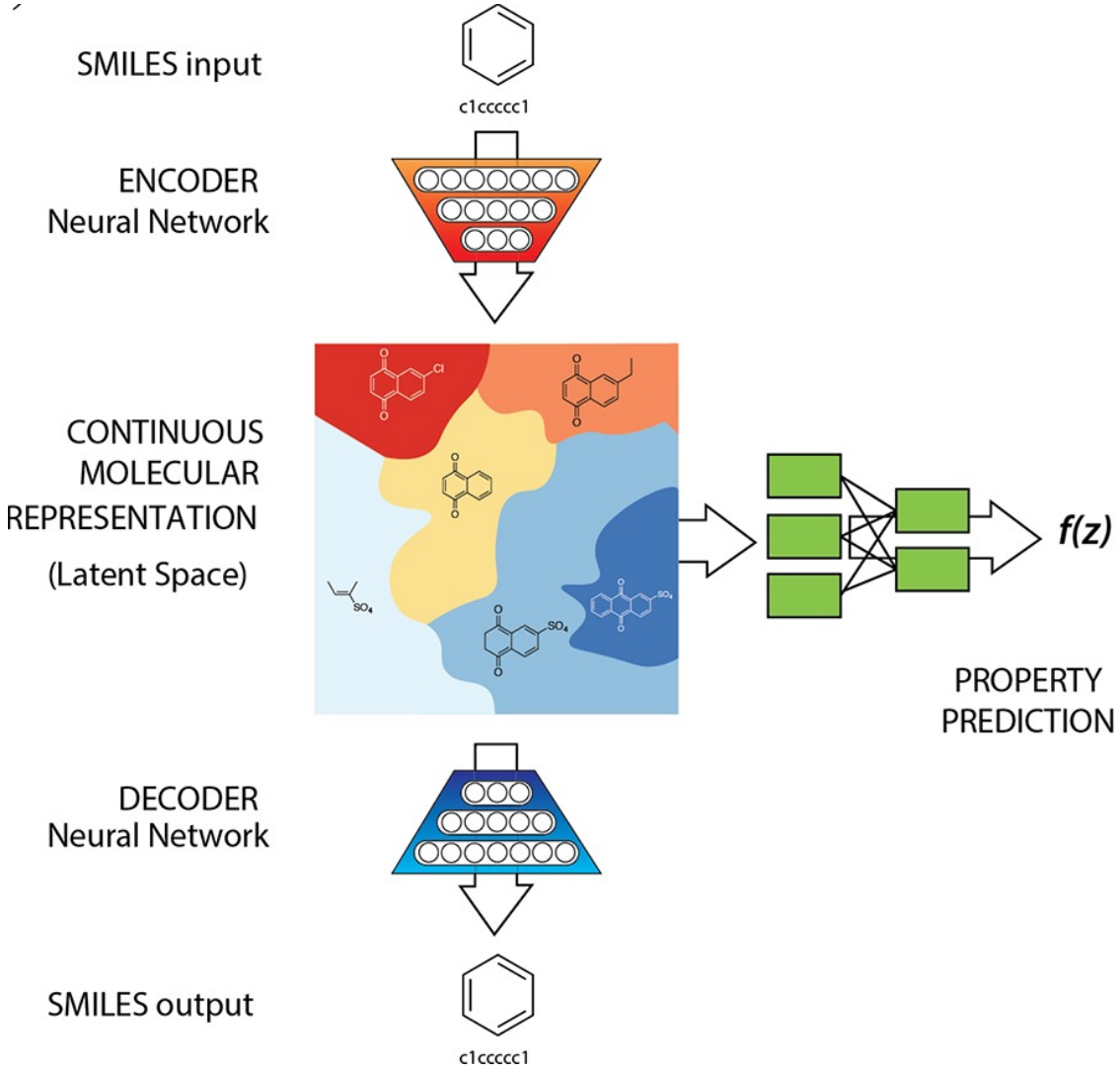
Cheng, Ser, Skreta, et al., Aspuru-Guzik, Faraday Discuss. 256, 10–60 (2025).

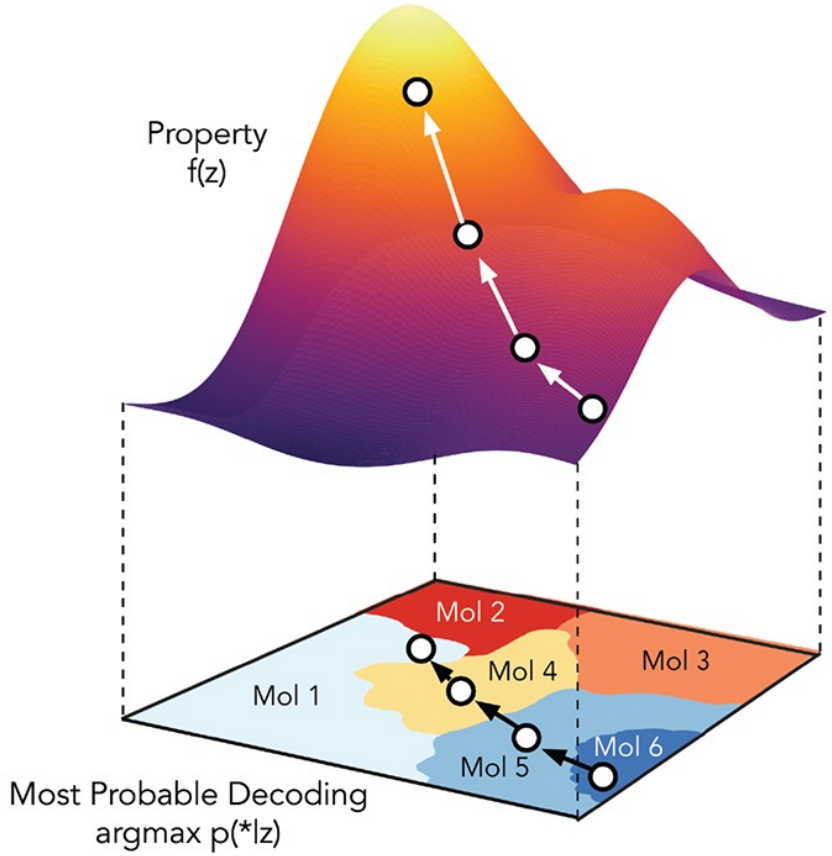
Diseño directo frente a inverso: el ejemplo de las baterías de flujo redox



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

VAE químico: convertir las moléculas en un mapa que optimizar





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

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – Verdadero/Falso

1. **En el diseño inverso se especifica la propiedad deseada y el modelo propone una estructura.**

   Verdadero | Falso

2. **Un VAE proporciona un espacio latente continuo en el que se puede optimizar.**

   Verdadero | Falso

3. **La «tercera generación» del diseño de materiales es el ensayo y error manual.**

   Verdadero | **Falso**

NUEVO PARADIGMA

Mensajes clave – Diseñar según una especificación en lugar de dar con ella por casualidad.

POR QUÉ IMPORTA PARA EL CLIMA

1. Estamos en una generación de diseño de materiales guiada por el AA.

2. Diseño inverso = optimizar en un espacio latente aprendido (espacio de diseño continuo).

3. Hay dos facilitadores transversales: los modelos fundacionales y los laboratorios autónomos.

¿PREGUNTAS?

LA CAJA DE HERRAMIENTAS DE LA IA

Una caja de herramientas de la IA

Seis cosas que el AA hace realmente a lo largo de la cadena de materiales

1 **Predicción de propiedades**

PREDECIR

Los modelos supervisados predicen una propiedad y (p. ej., la banda prohibida) a partir de una estructura x, complementando o sustituyendo simulaciones costosas.

2 **Diseño generativo**

GENERAR

Los modelos no supervisados/generativos producen nuevos candidatos x', condicionados a propiedades objetivo: la vía del diseño inverso.

3 **Laboratorios autónomos**

SINTETIZAR

Laboratorios autónomos planifican y ejecutan experimentos: retrosíntesis, predicción de reacciones, optimización por aprendizaje activo.

4 **Ayuda a la caracterización**

CARACTERIZAR

El AA interpreta mediciones: estructura cristalina por TEM, coordinación por XAS, simetría por difracción de electrones.

5 **Gestión de sistemas**

DESPLEGAR

El AA gestiona dispositivos y redes: predicción de la vida útil de baterías, previsión de la demanda, optimización del control de redes inteligentes.

6 **Acelerar la teoría**

SIMULAR

El AA hace viables simulaciones intratables: potenciales y funcionales más baratos para escalas mayores.

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

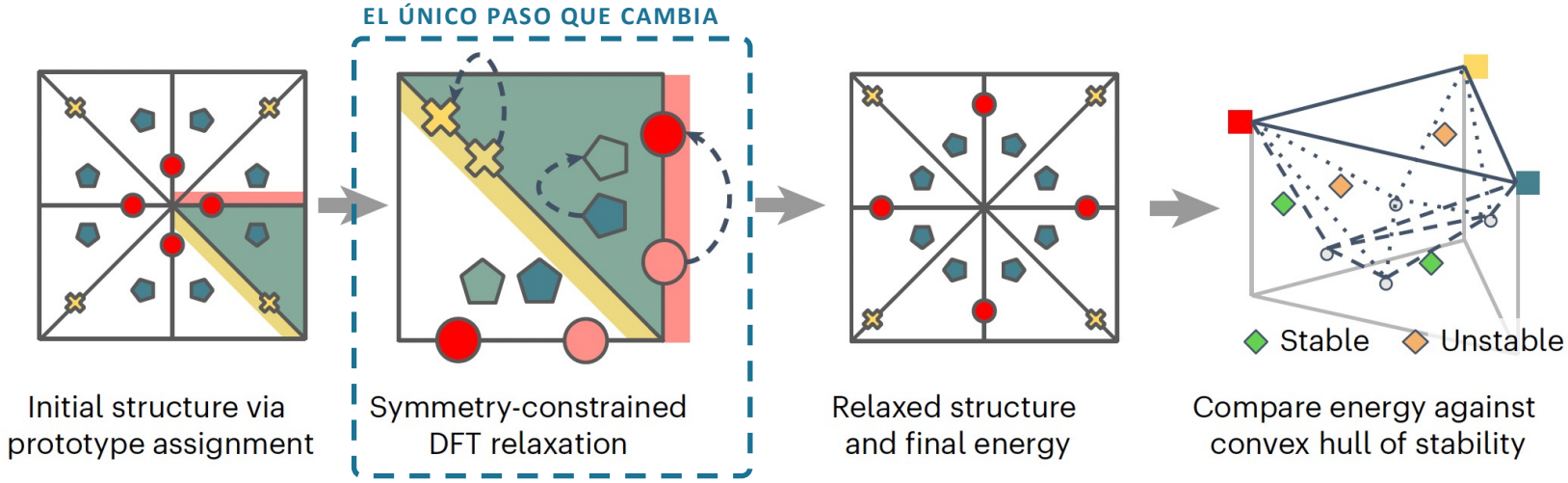
1

Predicción de propiedades

PREDECIR

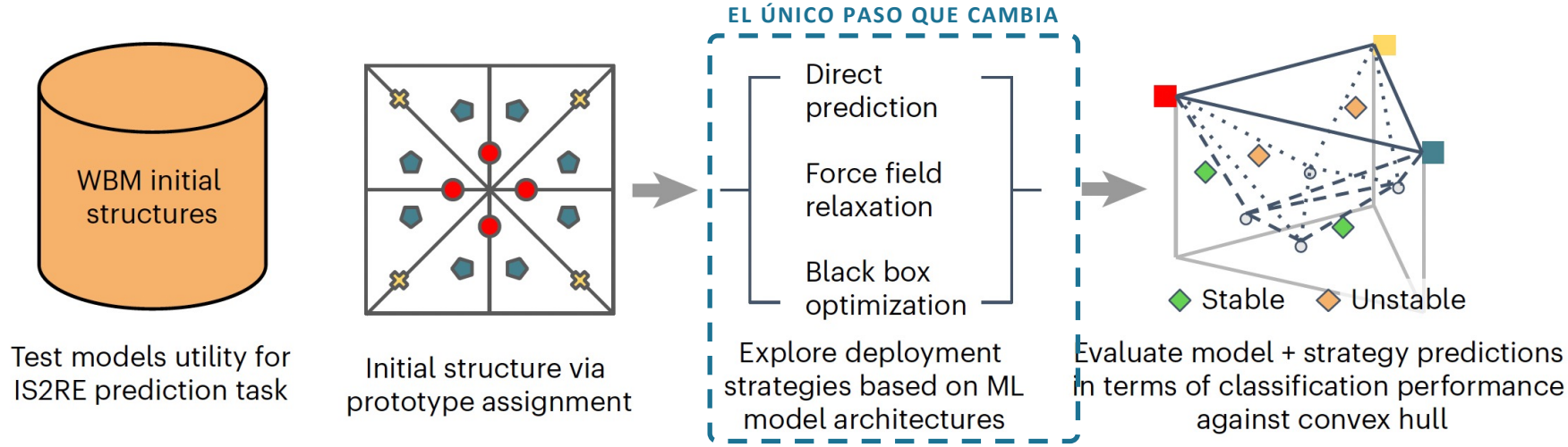
Predecir propiedades DFT con un sustituto de AA

CONVENCIONAL



MATBENCH DISCOVERY

IS2RE = de la estructura inicial a la energía relajada (Initial Structure to Relaxed Energy)



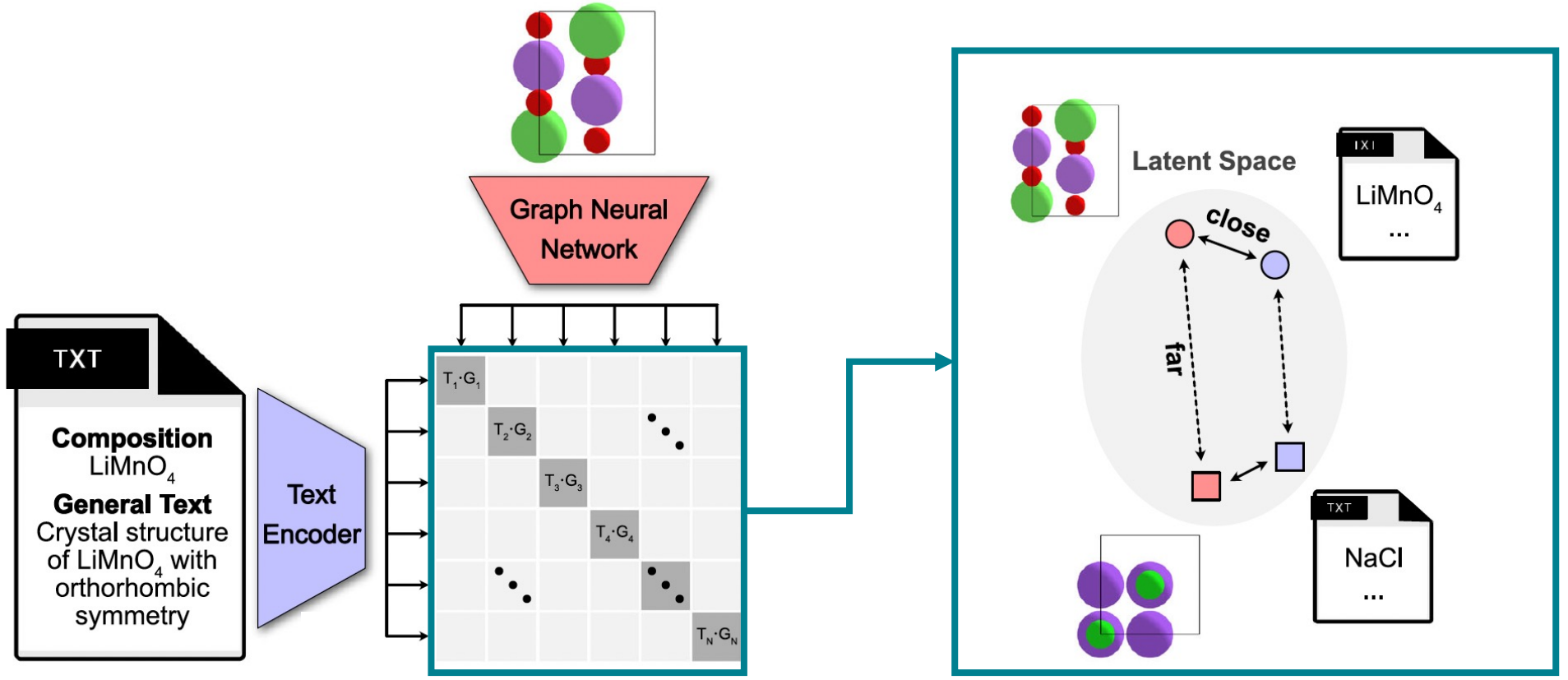
Adaptado de Riebesell et al., «A framework to evaluate machine learning crystal stability predictions», Nat. Mach. Intell. 7, 836–847 (2025).

2

Diseño generativo/inverso

GENERAR

Chemeleon: convertir una instrucción de texto en un cristal, en tres pasos



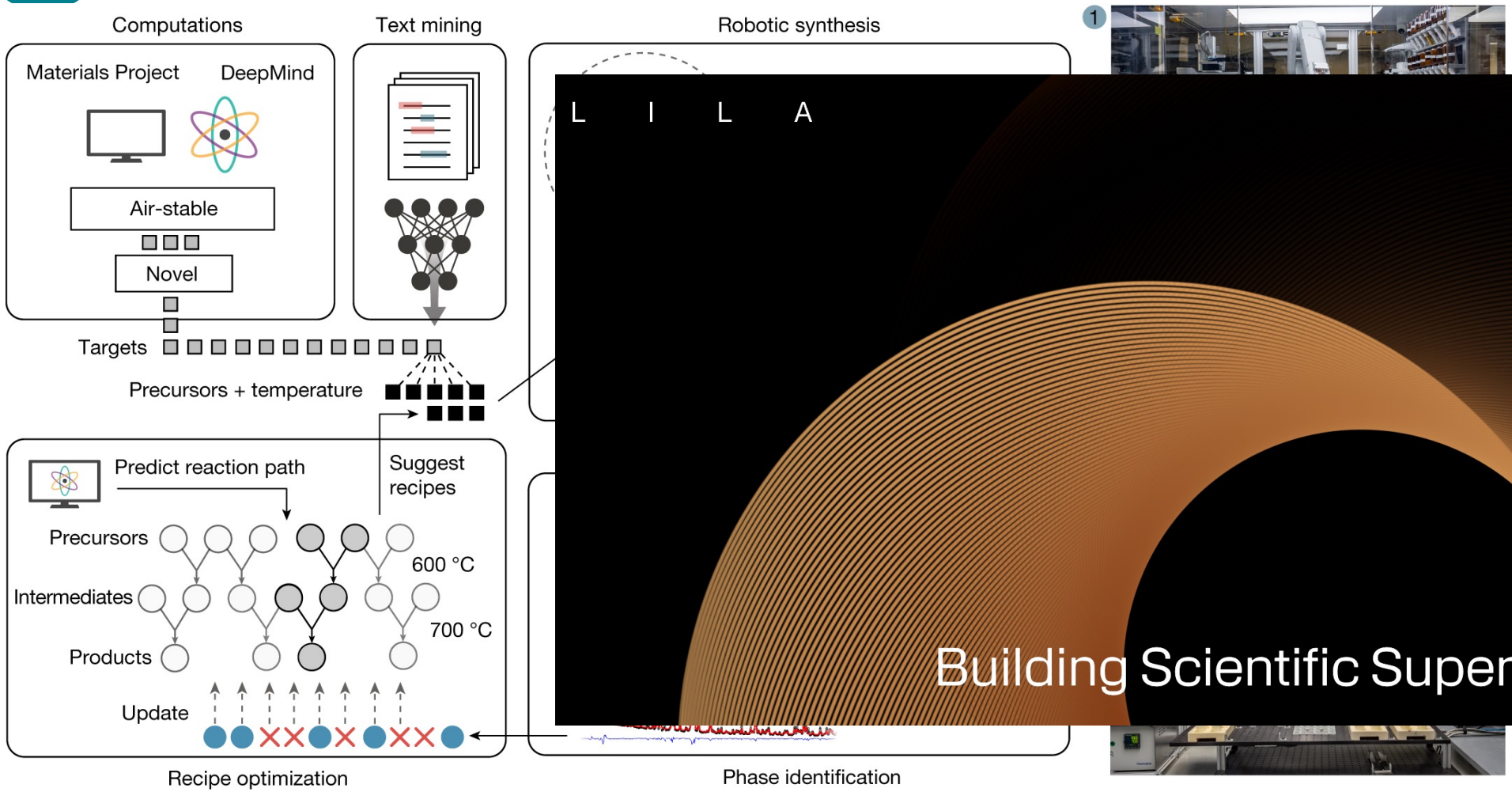
Adaptado de Park, Onwuli & Walsh, Nat. Commun. 16 (2025)

3

Laboratorios autónomos

SINTETIZAR

Un laboratorio autónomo: el A-Lab



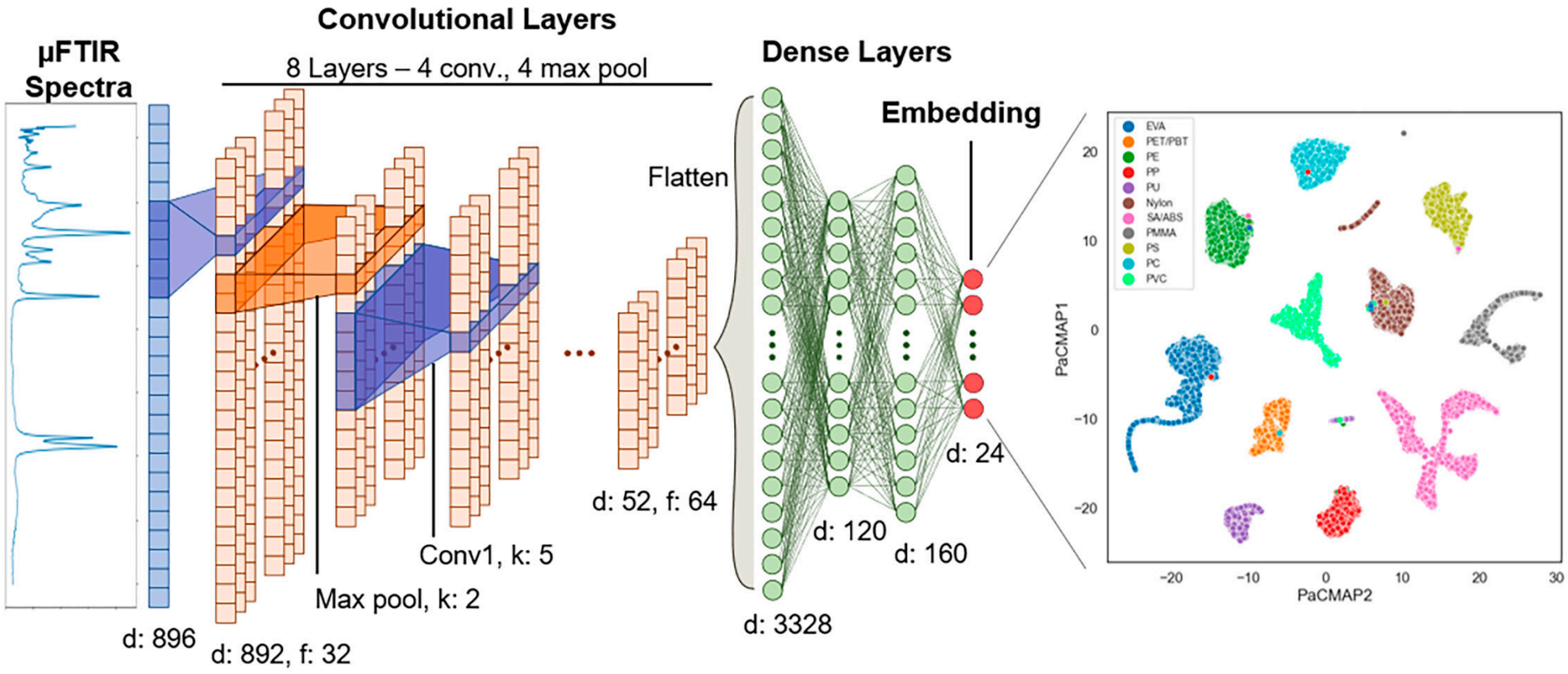
Adaptado de Szymanski et al., Nature 624, 86 (2023)

4

Ayuda a la caracterización

CARACTERIZACIÓN

Aprendizaje de similitud en imágenes espectrales μFTIR de microplásticos



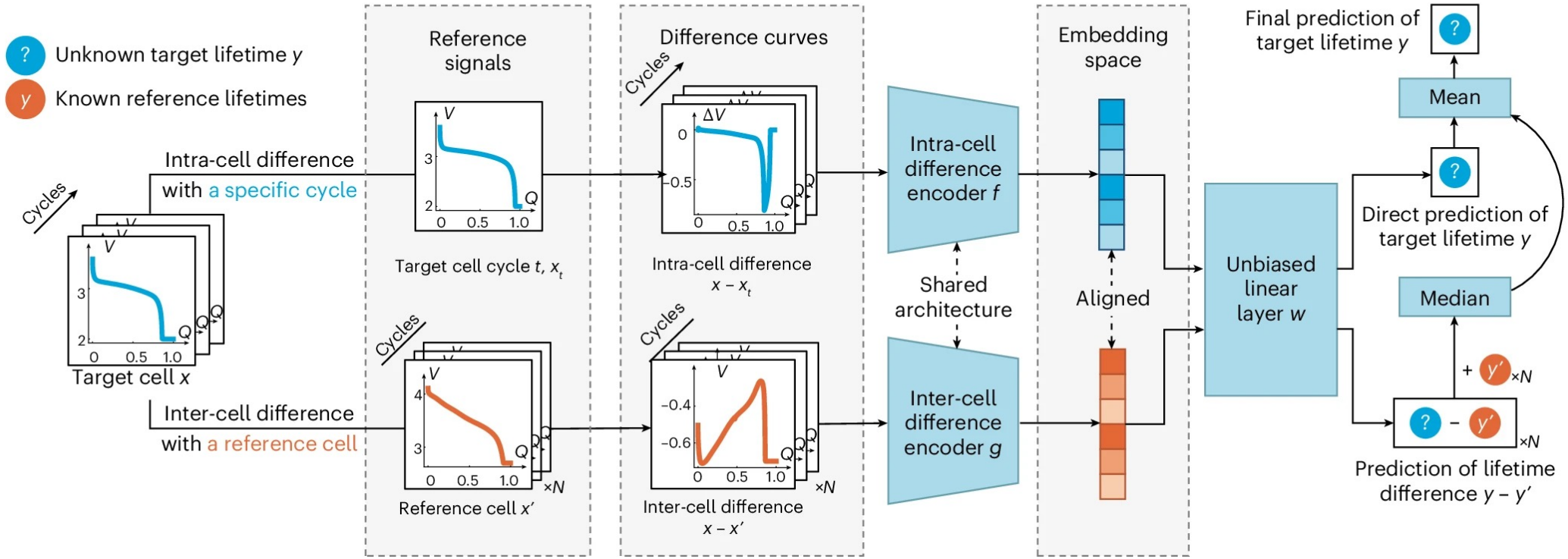
Smolen, Moore, Perez & Wooley, PNAS 122(42), e2509745122 (2025)

5

Gestión de sistemas

DESPLEGAR

Predecir de forma temprana la vida útil en ciclos de las baterías

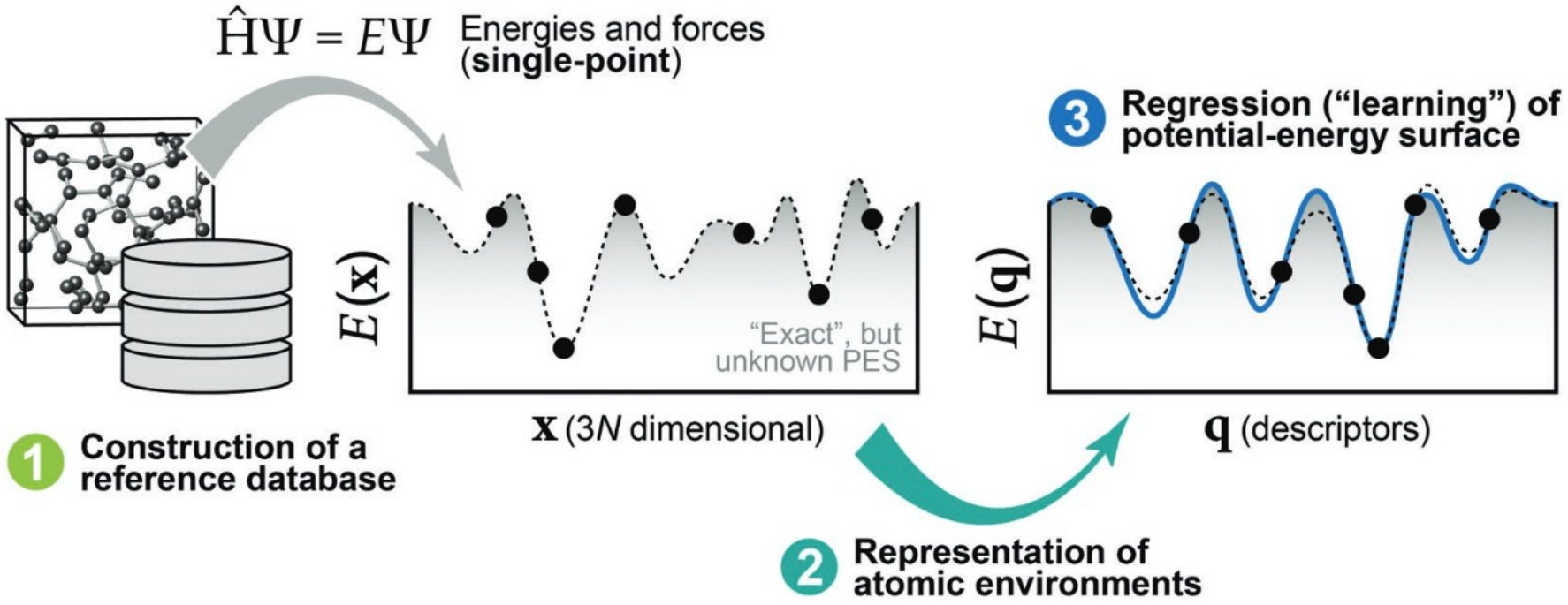


BatLiNet compara la celda objetivo con una celda de «referencia» completamente ciclada cuya vida útil ya se conoce

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

Cómo funciona un potencial interatómico de aprendizaje automático

Una misma receta para todos: entorno atómico local → energía, aprendida a partir de la DFT



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

Los modelos MLIP de un vistazo

Modelos representativos por familia: idea clave, uso típico y fuente

Modelo	Familia	Idea clave	Aplicación típica	Referencia
HDNNP	Descriptor + RN	funciones de simetría → red neuronal	sólidos masivos, agua	*Behler & Parrinello 2007*
GAP	Descriptor + núcleo	SOAP + proceso gaussiano	Si, C, amorfos; eficiente en datos	*Bartók et al. 2010*
DeePMD	Descriptor aprendido + RN	potencial profundo de extremo a extremo	DM a gran escala (agua, aleaciones)	*Zhang et al. 2018*
SNAP	Expansión lineal	componentes del biespectro (lineal)	metales (Ta, W)	*Thompson et al. 2015*
MTP	Expansión lineal	tensores de momentos	aleaciones; aprendizaje activo	*Shapeev 2016*
ACE	Expansión por orden de cuerpo	expansión en clústeres atómicos	general; rápido + preciso	*Drautz 2019*
SchNet	GNN invariante	convoluciones de filtro continuo	moléculas (MD17, QM9)	*Schütt et al. 2018*
NequIP	GNN equivariante	mensajes E(3)-equivariantes	moléculas + materiales, eficiente en datos	*Batzner et al. 2022*
Allegro	GNN equivariante local	equivariante estrictamente local	DM a muy gran escala	*Musaelian et al. 2023*
MACE	Equivariante + orden de cuerpo alto	mensajes de orden superior por capa	DM precisa y rápida	*Batatia et al. 2022*
M3GNet	GNN universal	potencial de grafos entrenado con MP	relajación/cribado de cristales	*Chen & Ong 2022*
CHGNet	Universal (sensible a la carga)	GNN informada por la carga	iones de litio, óxidos, magnetismo	*Deng et al. 2023*
MACE-MP-0	Fundacional (equivariante)	modelo fundacional para toda la tabla periódica	DM lista para usar en todas las químicas	*Batatia et al. 2024*

SOAP = solapamiento suave de posiciones atómicas (smooth overlap of atomic positions) · E(3) = rotaciones, traslaciones y reflexiones en 3D · MP = Materials Project · GNN = red neuronal de grafos.

Potenciales interatómicos de aprendizaje automático, en orden aproximadamente cronológico

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

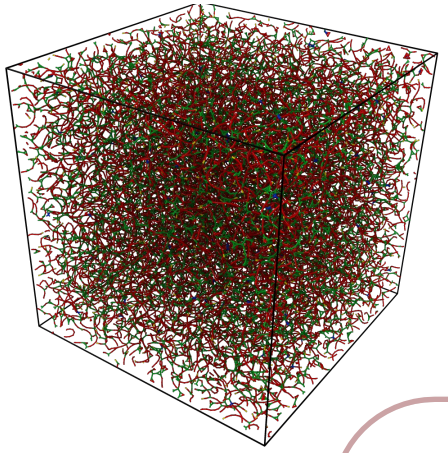
13. Merchant, A. et al. Scaling deep learning for materials discovery (GNoME). Nature 624, 80–85 (2023). 10.1038/s41586-023-06735-9

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

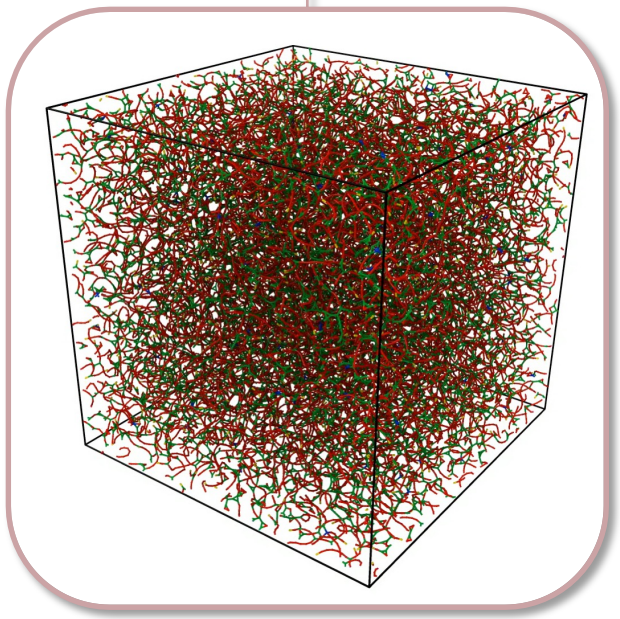
Modelización del biocarbón con MLIP (GAP20)

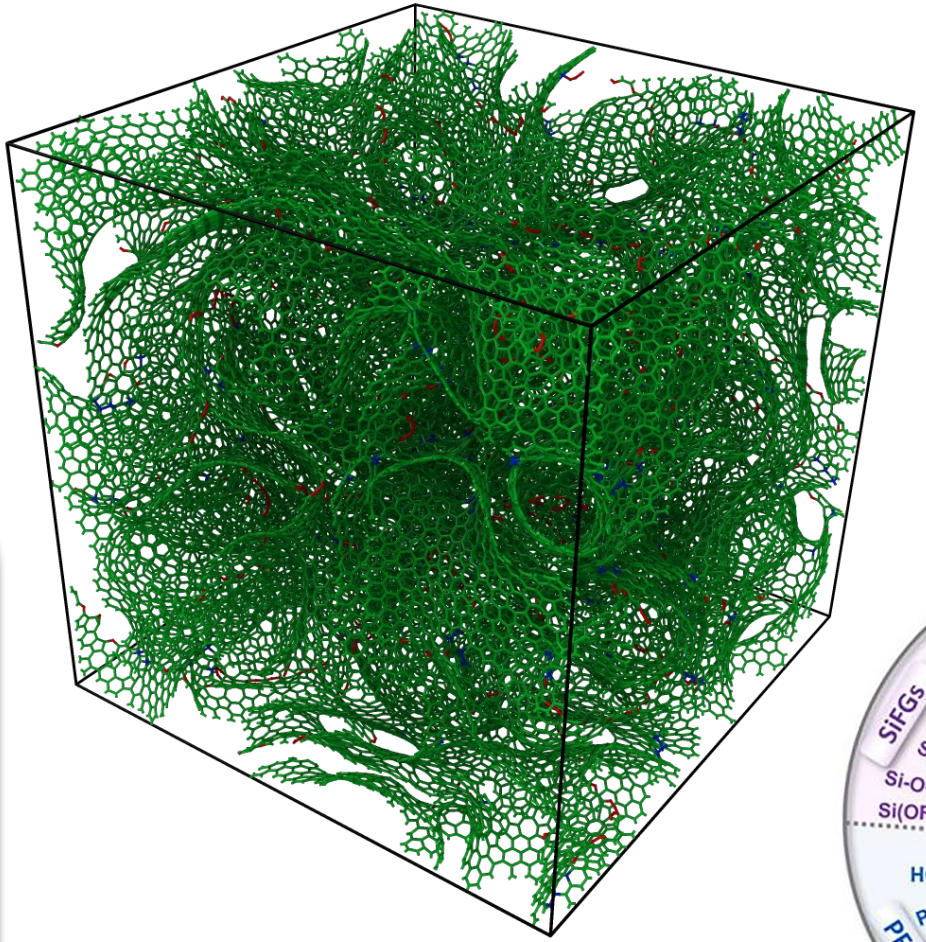
Escala atómica

Caracterización estructural:
Porosidad, estructura atómica, hibridación del carbono, orden estructural

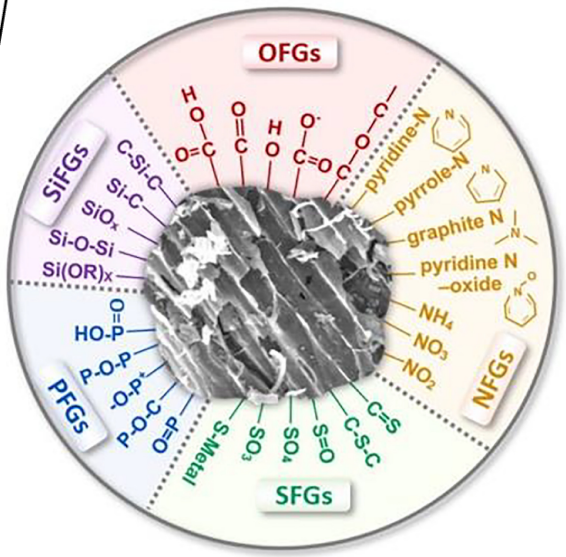


Simulación de DM





Biocarbón Modelo



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

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST

Test rápido – Verdadero/Falso

1. **Un potencial interatómico de AA busca la precisión de la DFT a un coste mucho menor.**

   Verdadero | Falso

2. **Un modelo con un 100 % de precisión en el entrenamiento es siempre el más fiable con datos nuevos.**

   Verdadero | **Falso**

3. **Los candidatos generados por IA siguen necesitando validación por DFT/experimental.**

   Verdadero | Falso

POR QUÉ IMPORTA PARA EL CLIMA

Una sola caja de herramientas para todos los materiales de energía limpia.

- Los potenciales de AA ofrecen una precisión cercana a la DFT por una fracción de la energía y del tiempo: se pueden cribar millones de candidatos.
- Los laboratorios autónomos convierten rápidamente las mejores predicciones en muestras validadas.
- La validación evita malgastar el escaso tiempo de laboratorio en falsos positivos.

Mensajes clave

POR QUÉ IMPORTA PARA EL CLIMA

1. Los laboratorios autónomos convierten rápidamente las mejores predicciones en muestras validadas.

2. Los MLIP ofrecen la precisión de la DFT a la velocidad de la DM: la casilla «acelerar la simulación» hecha realidad.

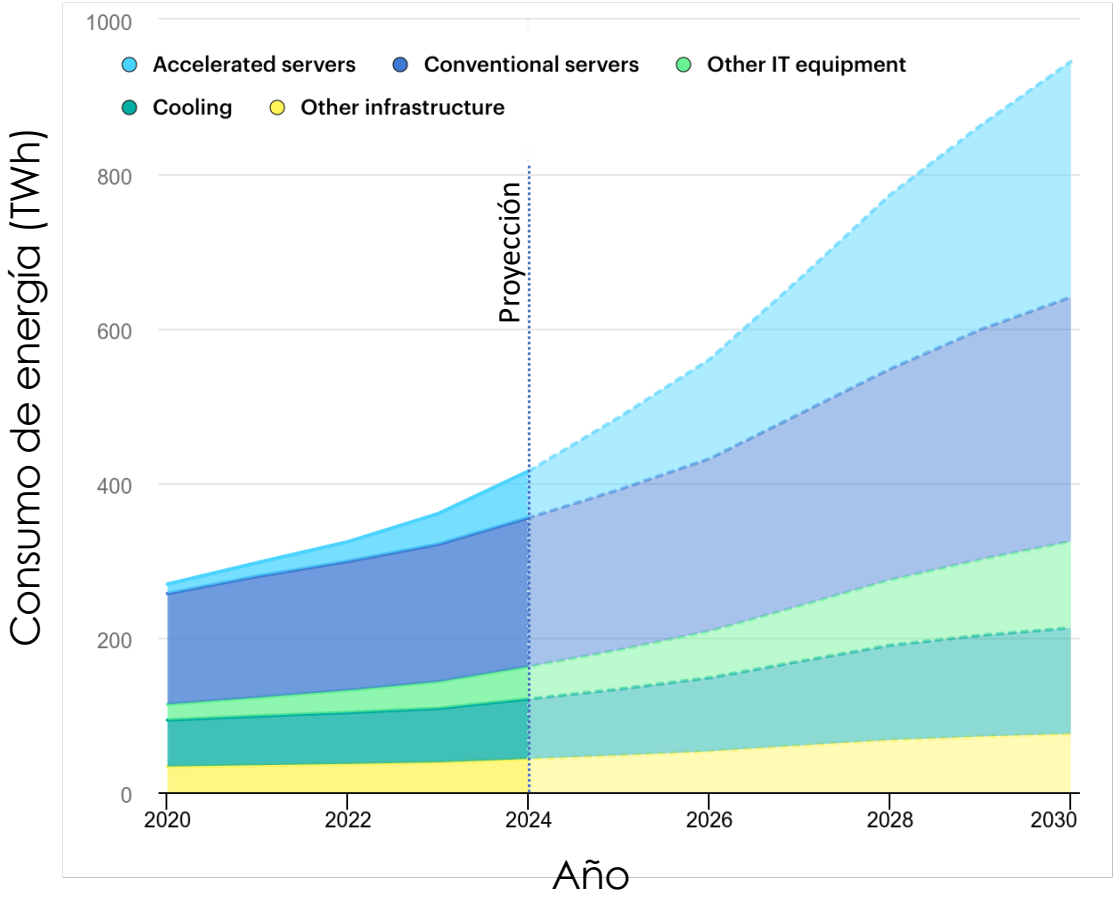
3. Generar es fácil; validar es difícil

¿PREGUNTAS?

HABLAR DE QUÍMICA CON LOS ORDENADORES

Coste energético creciente de la IA y los centros de datos

Necesidad de una informática más verde



Fuente: Agencia Internacional de la Energía (AIE)

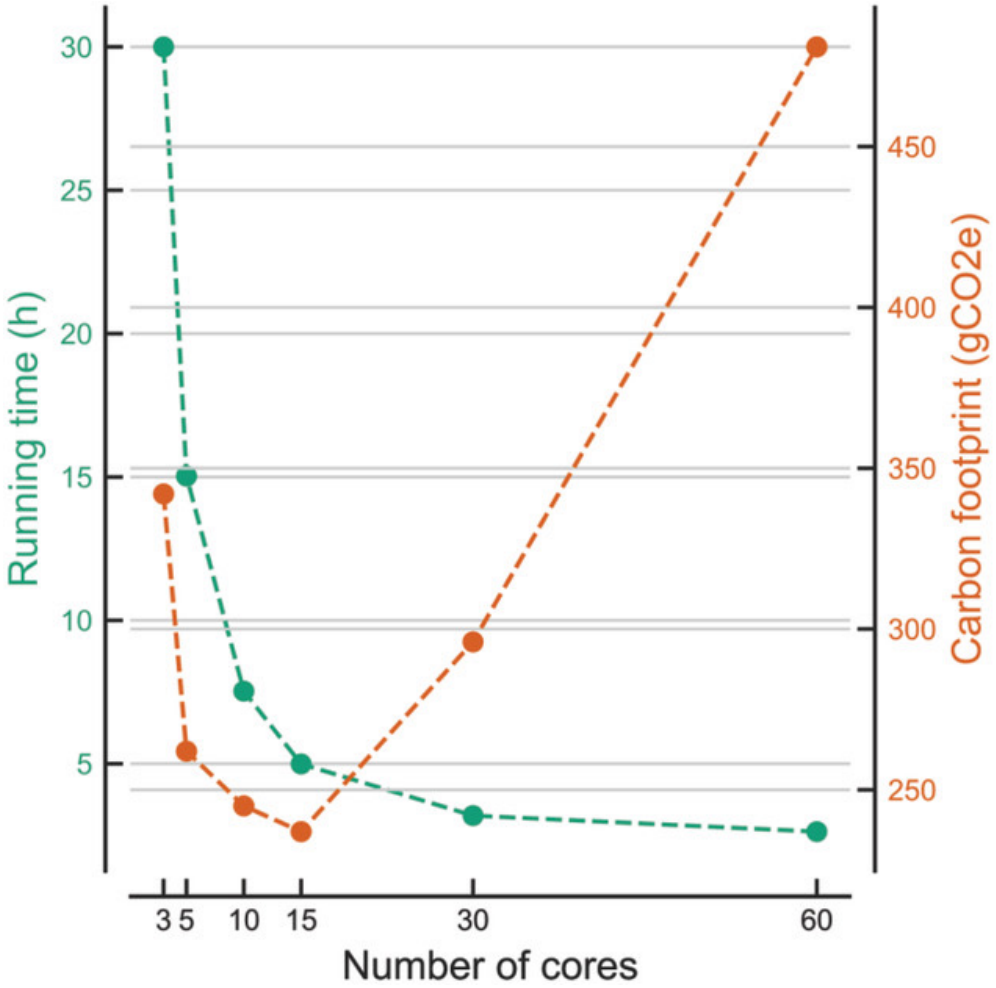
¿Dónde está la supercomputación?



Mejorar el acceso a la supercomputación en dispositivos o contextos con recursos limitados: **democratización**

Fuente: top500.org

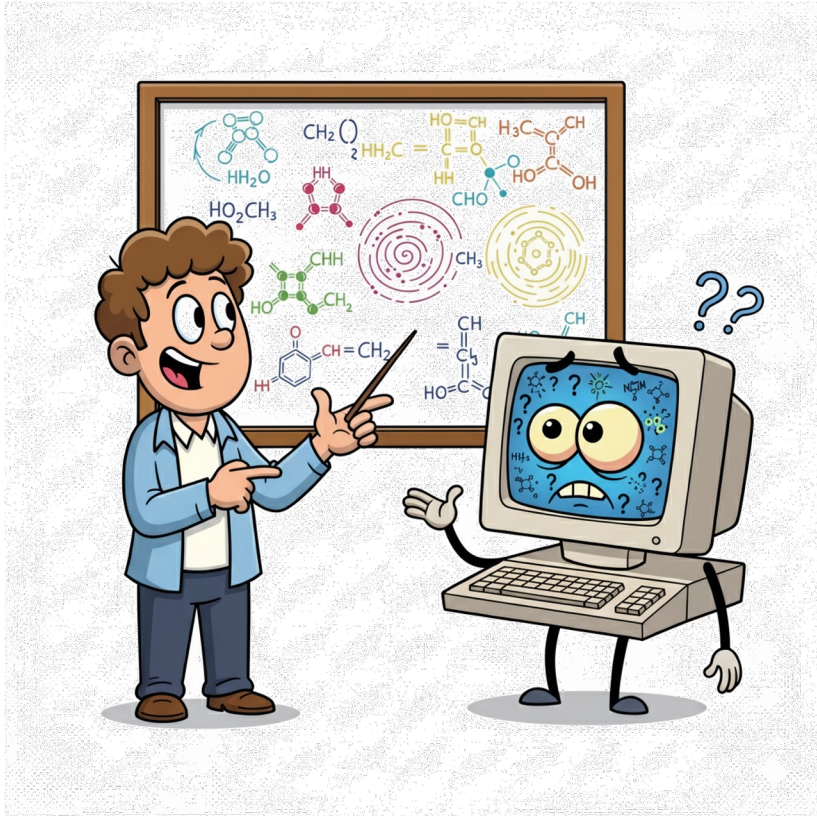
Algoritmos verdes





Lannelongue et al., Advanced Science, 8, 2021

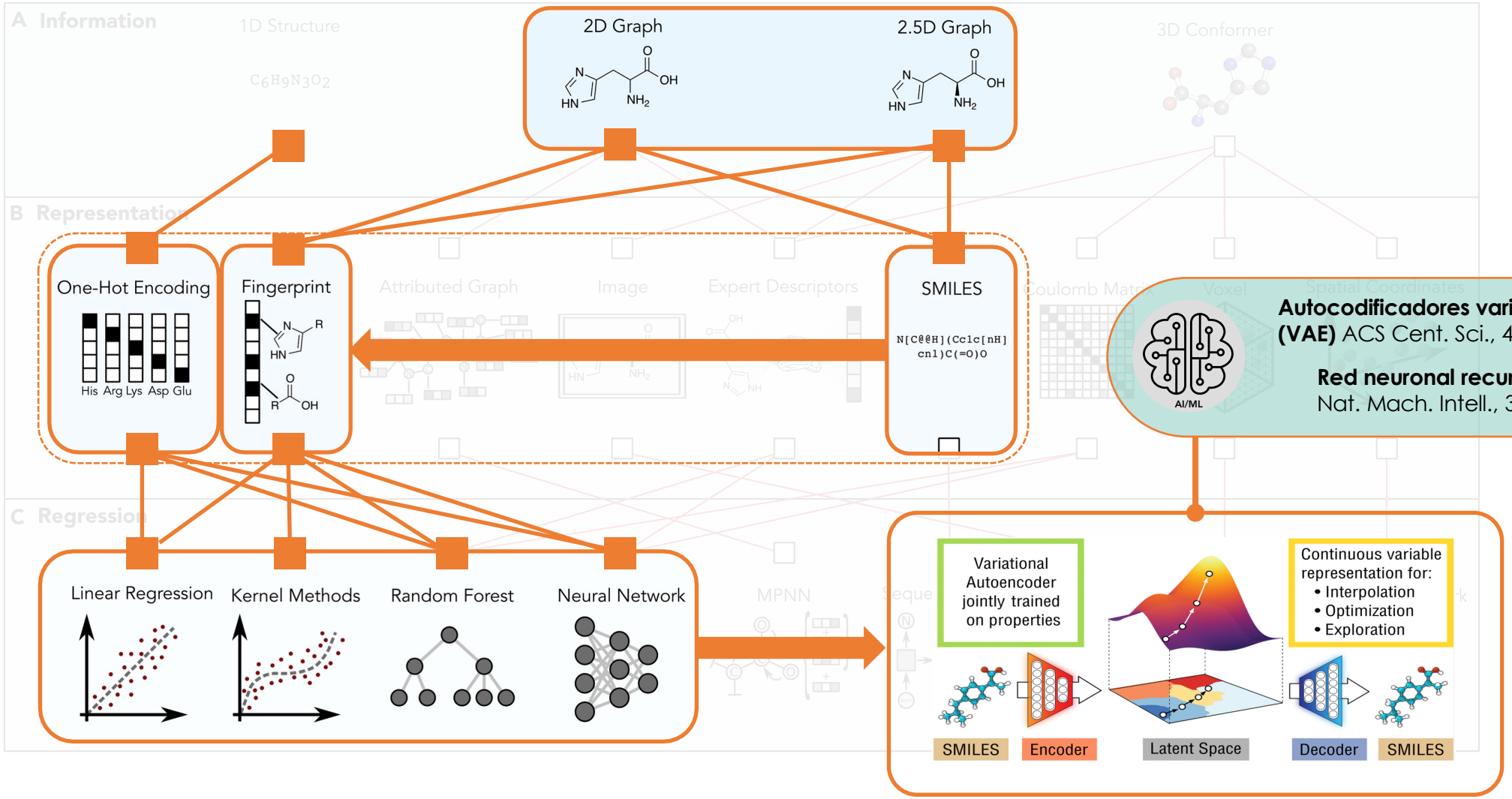
Un paso atrás: ¿cómo hablamos de química con los ordenadores?



El aprendizaje automático para química exige **vectorizar las moléculas**

El AA para química exige vectorizar las moléculas

HABLAR DE QUÍMICA



Autocodificadores variacionales (VAE) ACS Cent. Sci., 4 , 268, 2018`

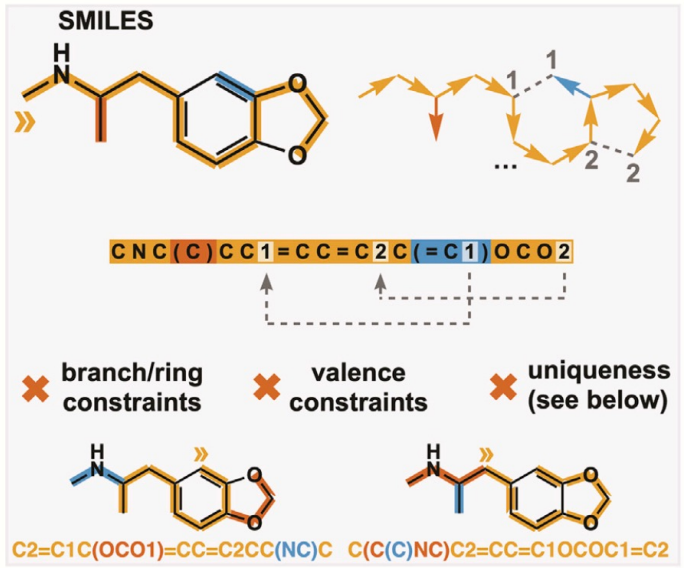
Red neuronal recurrente (RNN) Nat. Mach. Intell., 3 , 759, 2021

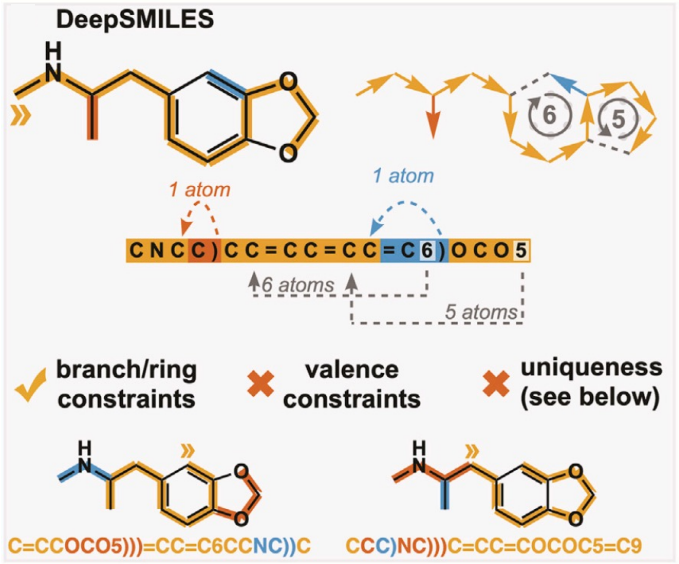
Pattanaik and Coley, **Chem 6, 1204-1207, 2020**

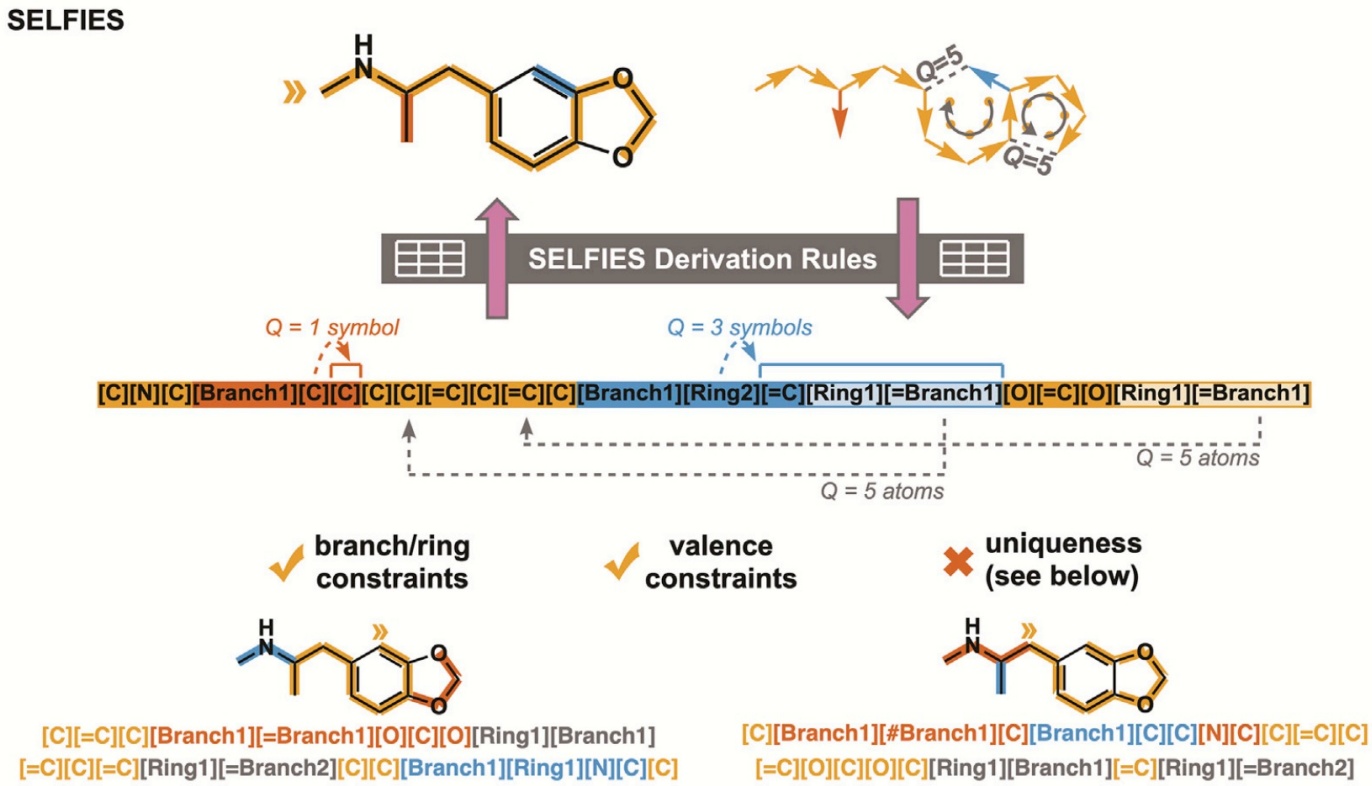
Skinnider et al. **Nat. Mach. Intell.**, 3 , 759, **2021**

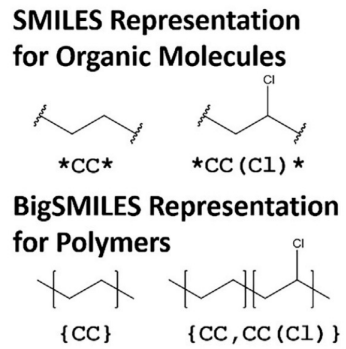
Gomez-Bombarelli et al. **ACS Cent Sci.** 4, 268–76, **2018**

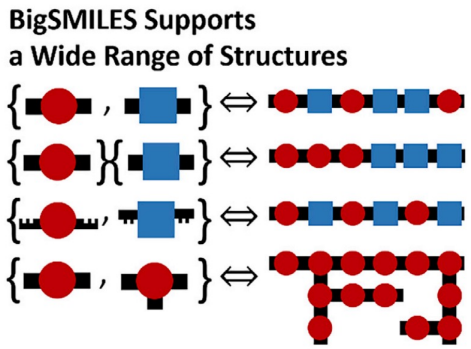
Codificación de datos legible por máquina







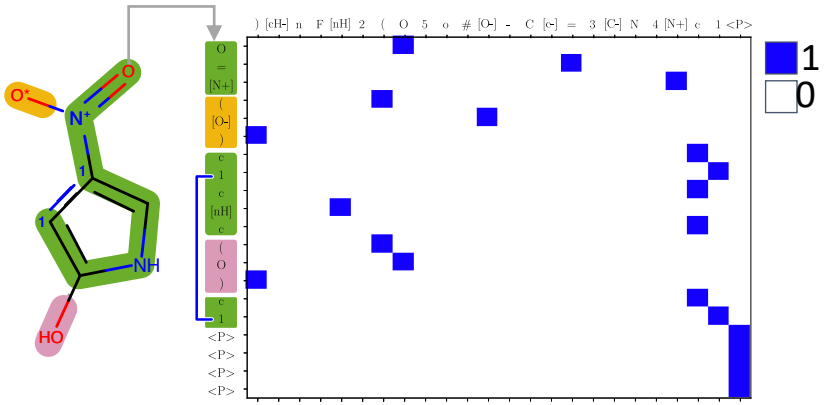




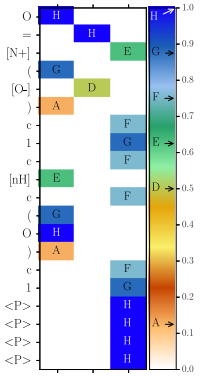
Krenn et al. Patterns 3, 10, 100588, 2022

Codificaciones one-hot incrustadas (eOHE)

Codificación one-hot (OHE)



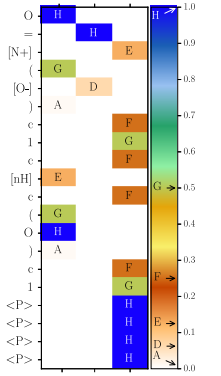
eOHE v1



Reducción lineal
$q = 8$

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

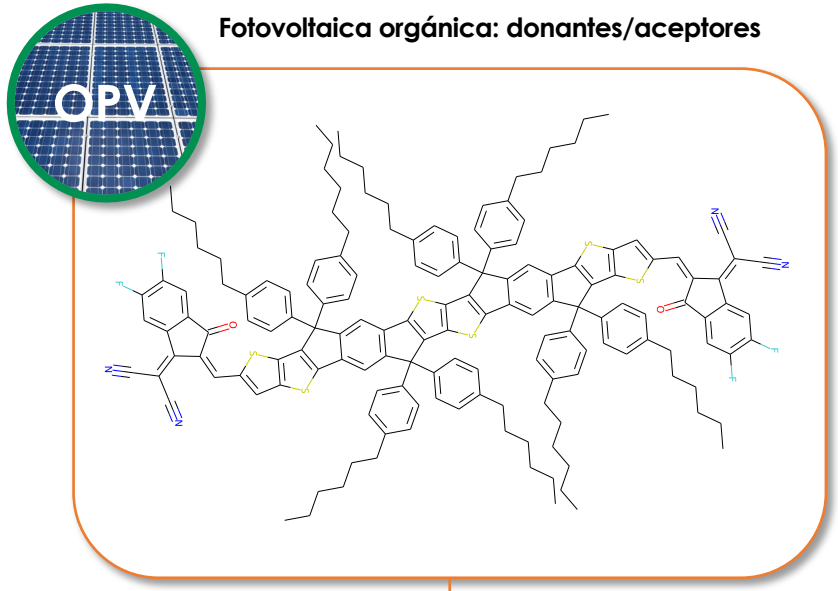
eOHE v2

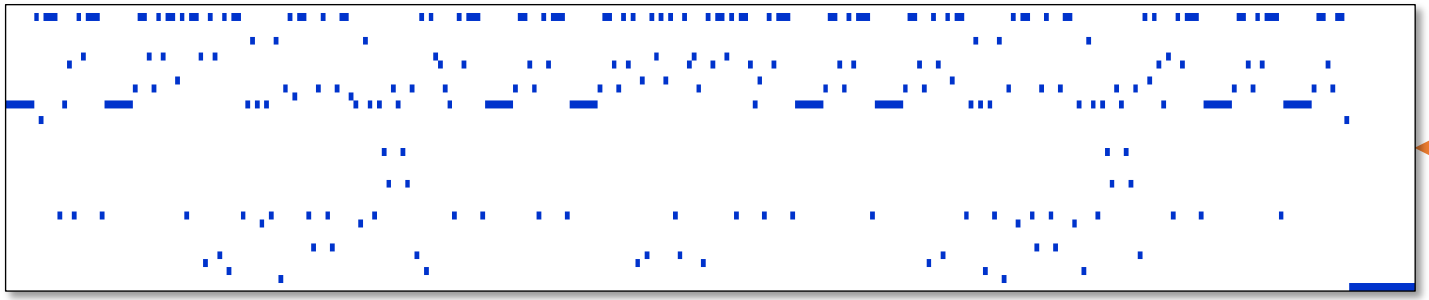


Reducción por potencias de 2
$q = 8$

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

Fotovoltaica orgánica: donantes/aceptores

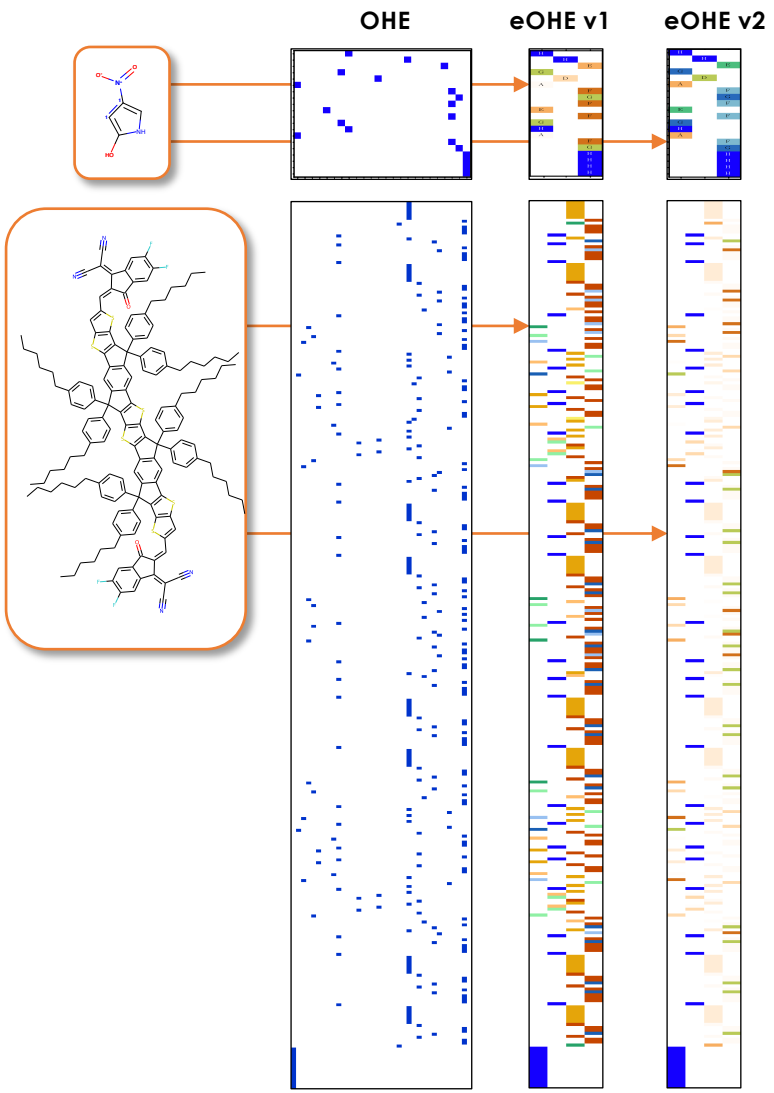


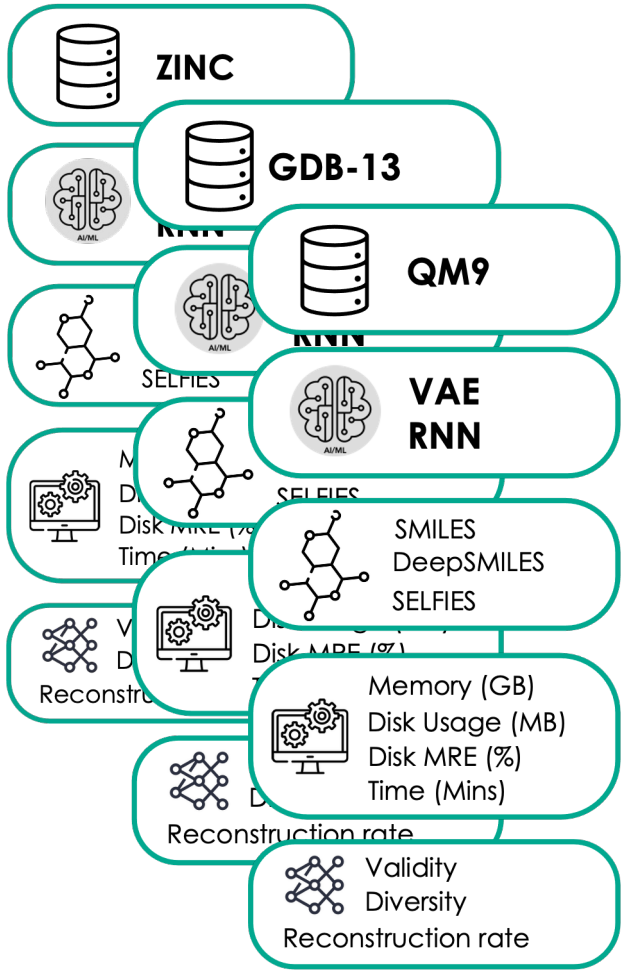


CCCCCCc1ccc(C2(c3ccc(CCCCCC)cc3)c3cc4c(cc3-c3sc5cc(C=C6C(=O)c7cc(F)c(F)cc7C6=C(C#N)C#N)sc5c32)C(c2ccc(CCCCCC)cc2)(c2ccc(CCCCCC)cc2)c2c-4sc3c4c(sc23)-c2cc3c(cc2C4(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)-c2sc4cc(C=C5C(=O)c6cc(F)c(F)cc6C5=C(C#N)C#N)sc4c2C3(c2ccc(CCCCCC) cc2)c2ccc(CCCCCC)cc2)cc1

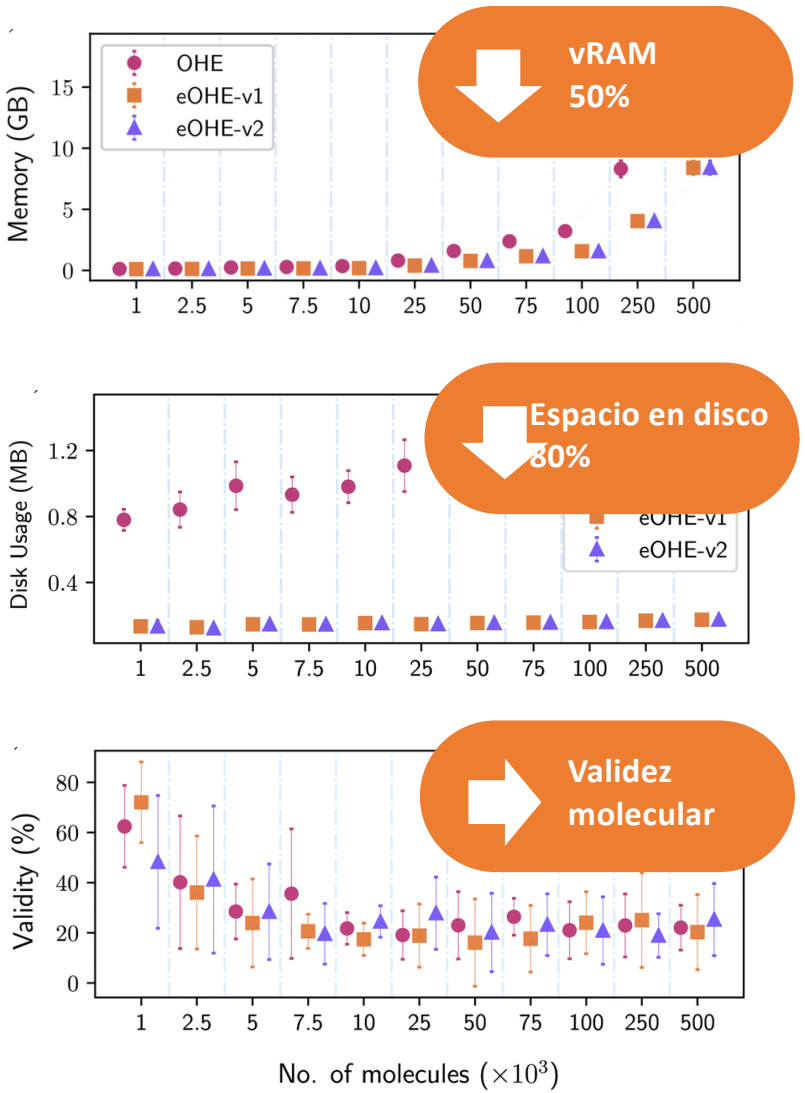
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez,. **Digital Discovery**, 4, 776-789, **2025**

Codificación legible por máquina más eficiente





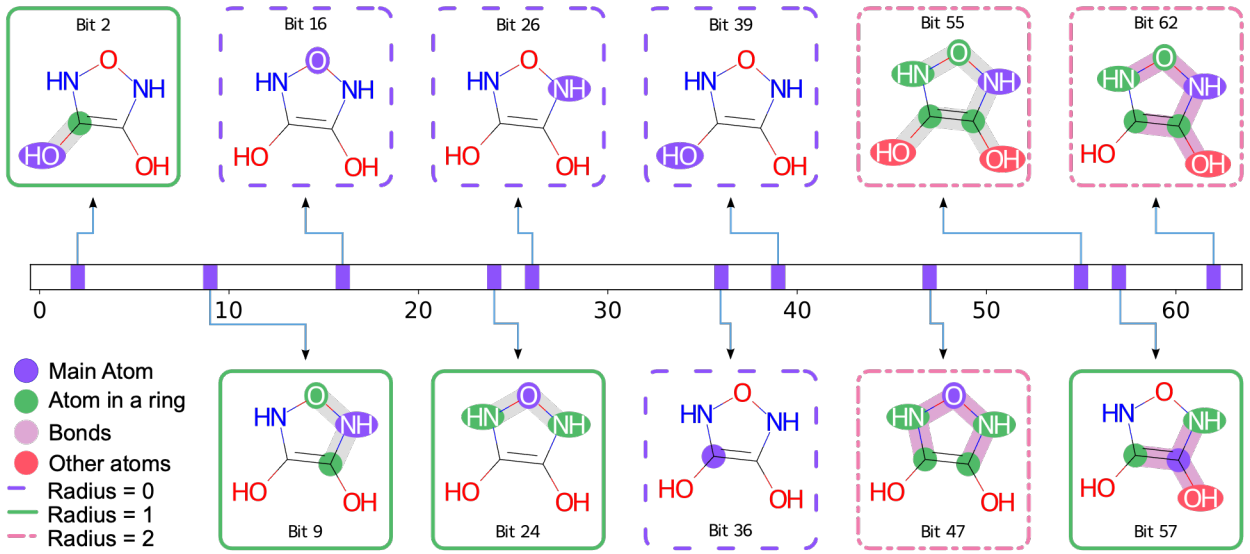
Rendimiento en autocodificadores variacionales



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

Huellas de Morgan incrustadas (eMP)

Una huella de Morgan es un vector binario de tamaño fijo que codifica la presencia (1) o ausencia (0) de subestructuras en una molécula.

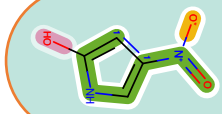




QM9: más de 134 000 moléculas
NFA: 50 000 moléculas
RedDB: 15 000 moléculas



Bosque aleatorio (RF), Perceptrón multicapa (MLP), Regresor KNN (KNR), regresor de gradient boosting (GBR), red neuronal profunda (DNN)



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

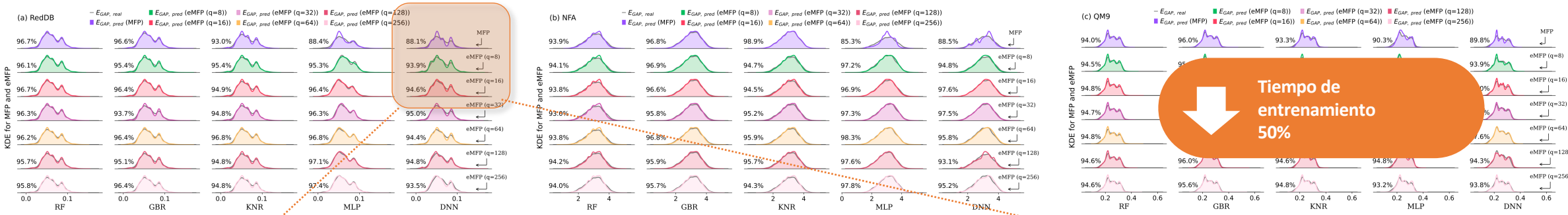


150 ensayos para optimizar 15 750 hiperparámetros

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

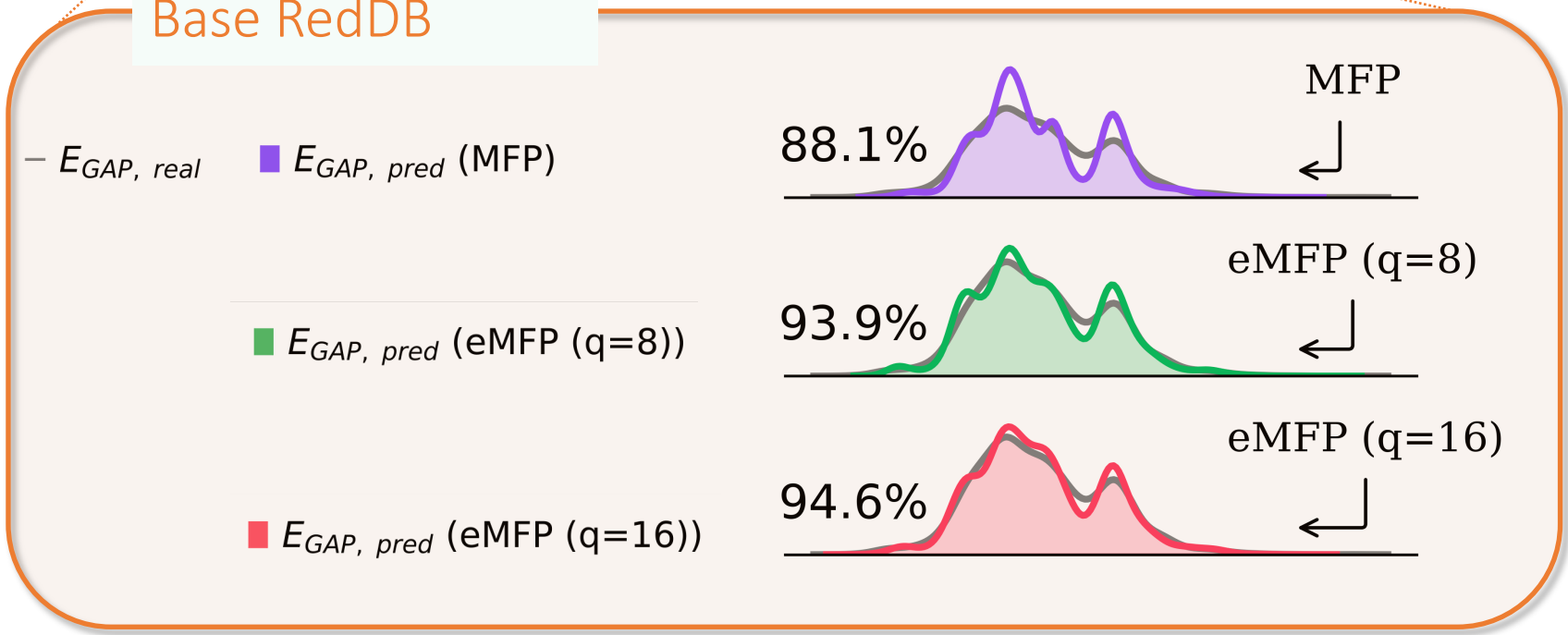
Brecha energética con huellas de Morgan incrustadas (eMP)

HABLAR DE QUÍMICA



Tiempo de entrenamiento 50%

Base RedDB



Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, *en revisión,* **2026**

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

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES

B) SELFIES

C) InChI

D) Ninguna de ellas

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES

B) SELFIES

C) InChI

D) Ninguna de ellas

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES

B) SELFIES

C) InChI

D) Ninguna de ellas

Q2 **eOHE/eMFP mejoran principalmente…**

A) Solo la precisión

B) Memoria/disco/tiempo con igual precisión

C) El número de publicaciones

D) Nada

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES

B) SELFIES

C) InChI

D) Ninguna de ellas

Q2 **eOHE/eMFP mejoran principalmente…**

A) Solo la precisión

B) Memoria/disco/tiempo con igual precisión

C) El número de publicaciones

D) Nada

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES
B) SELFIES
C) InChI
D) Ninguna de ellas

Q2 **eOHE/eMFP mejoran principalmente…**

A) Solo la precisión
B) Memoria/disco/tiempo con igual precisión
C) El número de publicaciones
D) Nada

Q3 **¿Por qué importa el coste energético de la IA?**

A) No importa
B) Sostenibilidad y democratización
C) La computación es gratuita
D) Solo para videojuegos

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **¿Qué representación garantiza una molécula válida para cada cadena?**

A) SMILES
B) SELFIES
C) InChI
D) Ninguna de ellas

Q2 **eOHE/eMFP mejoran principalmente…**

A) Solo la precisión
B) Memoria/disco/tiempo con igual precisión
C) El número de publicaciones
D) Nada

Q3 **¿Por qué importa el coste energético de la IA?**

A) No importa
B) Sostenibilidad y democratización
C) La computación es gratuita
D) Solo para videojuegos

HABLAR DE QUÍMICA

Mensajes clave – La representación es una decisión de sostenibilidad.

POR QUÉ IMPORTA PARA EL CLIMA

1. La vectorización es una decisión de diseño de primer orden, no un paso de preprocesamiento secundario.

2. eOHE/eMFP reducen memoria, disco y tiempo con igual precisión eficiencia = computación más verde y más democrática.

3. Una informática verde y accesible forma parte de una aceleración responsable.

¿PREGUNTAS?

DESARROLLO DE MATERIALES GUIADO POR LLM

Desarrollo de materiales guiado por LLM: una nueva capa en la pila

Más allá de predecir cifras: modelos que leen, razonan, planifican y usan herramientas

El cambio: el AA especializado asigna una estructura a una propiedad. Un gran modelo de lenguaje añade una interfaz en lenguaje natural, amplios conocimientos previos, planificación paso a paso y la capacidad de llamar a herramientas externas, **lo que convierte un predictor en un colaborador de investigación**.

1

Interfaz en lenguaje natural

Se le pregunta, se adapta con unos pocos ejemplos; destaca cuando hay pocos datos, donde el AA clásico flaquea.

2

Conocimiento del mundo

Preentrenado con la bibliografía científica: aporta conocimientos químicos previos y contexto, no solo ajuste de curvas.

3

Planificación y razonamiento

Descomponer un objetivo («encontrar un cátodo de iones de sodio estable») en pasos ordenados y verificables.

4

Uso de herramientas

Llamar a simuladores, bases de datos y robots, basando las respuestas en cálculos reales en lugar de adivinar.

…pero cuidado. Los LLM son fluidos pero poco fiables. Se inventan hechos, citan mal y pueden equivocarse con total seguridad. Todo depende de **anclarlos en herramientas y validarlos**

Lo que hacen realmente los LLM en el desarrollo de materiales

1 Leer la bibliografía

Convertir miles de artículos en bases de datos estructuradas y consultables (composiciones, MOF, dopantes).

p. ej. **Dagdelen et al. 2024**

2 Predecir propiedades

Responder preguntas de química en lenguaje natural; compite con el AA y destaca cuando hay pocos datos.

p. ej. **Jablonka et al. 2024**

3 Generar y diseñar a la inversa

Escribir moléculas y cristales como texto; generación orientada por propiedades.

p. ej. **Gruver et al. 2024** · **Chemeleon**

4 Planificar la síntesis

Proponer rutas sintéticas y recetas; retrosíntesis y diseño de procedimientos.

p. ej. **ChemCrow 2024**

5 Razonar con herramientas

Dotar al modelo de herramientas expertas (RDKit, calculadoras, búsqueda) para que deje de adivinar.

p. ej. **ChemCrow – 18 herramientas**

6 Realizar experimentos de forma autónoma

Ciclos de planificar-actuar-observar que controlan instrumentos y robots para poner a prueba hipótesis.

p. ej. **Coscientist 2023**

Sistemas LLM de referencia para química y materiales

Sistemas representativos: qué hacen, el enfoque y la fuente

Sistema	Qué hace	LLM/enfoque	Referencia
Coscientist	diseña, planifica y ejecuta experimentos (p. ej., acoplamientos cruzados con Pd)	GPT-4 + web/documentos/código + robótica	*Boiko et al., Nature 2023*
ChemCrow	síntesis orgánica, diseño de fármacos y materiales; realizó síntesis reales	GPT-4 + 18 herramientas expertas	*Bran et al., Nat. Mach. Intell. 2024*
LLM de química predictiva	predicción y clasificación de propiedades a partir de texto; pocos datos	GPT-3 ajustado; en contexto	*Jablonka et al., Nat. Mach. Intell. 2024*
Extracción estructurada	lee artículos → registros estructurados (dopantes, MOF, composiciones)	GPT-3/Llama-2 ajustados	*Dagdelen et al., Nat. Commun. 2024*
LLM cristal-texto	genera cristales inorgánicos estables escritos como texto	LLaMA-2 70B ajustado	*Gruver et al., ICLR 2024*
Chemeleon	generación de estructuras cristalinas guiada por texto	contrastivo texto–cristal + difusión	*Park et al., Nat. Commun. 2025*

Referencias

LLM y agentes autónomos para química y materiales

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

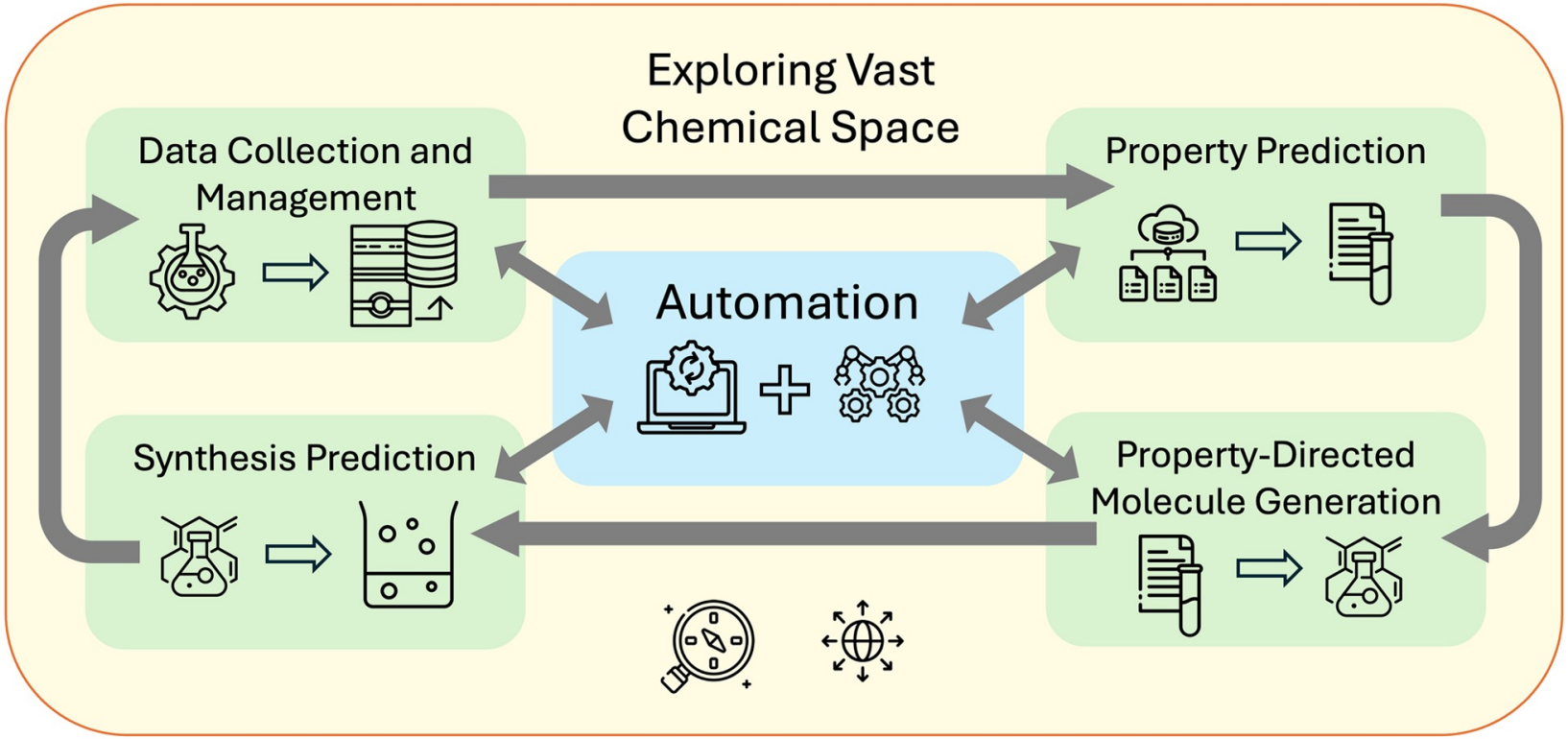
5. Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-Tuned Language Models Generate Stable Inorganic Materials as Text. ICLR 2024. arXiv:2402.04379

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

¿Qué pueden hacer realmente los LLM y los agentes en química?

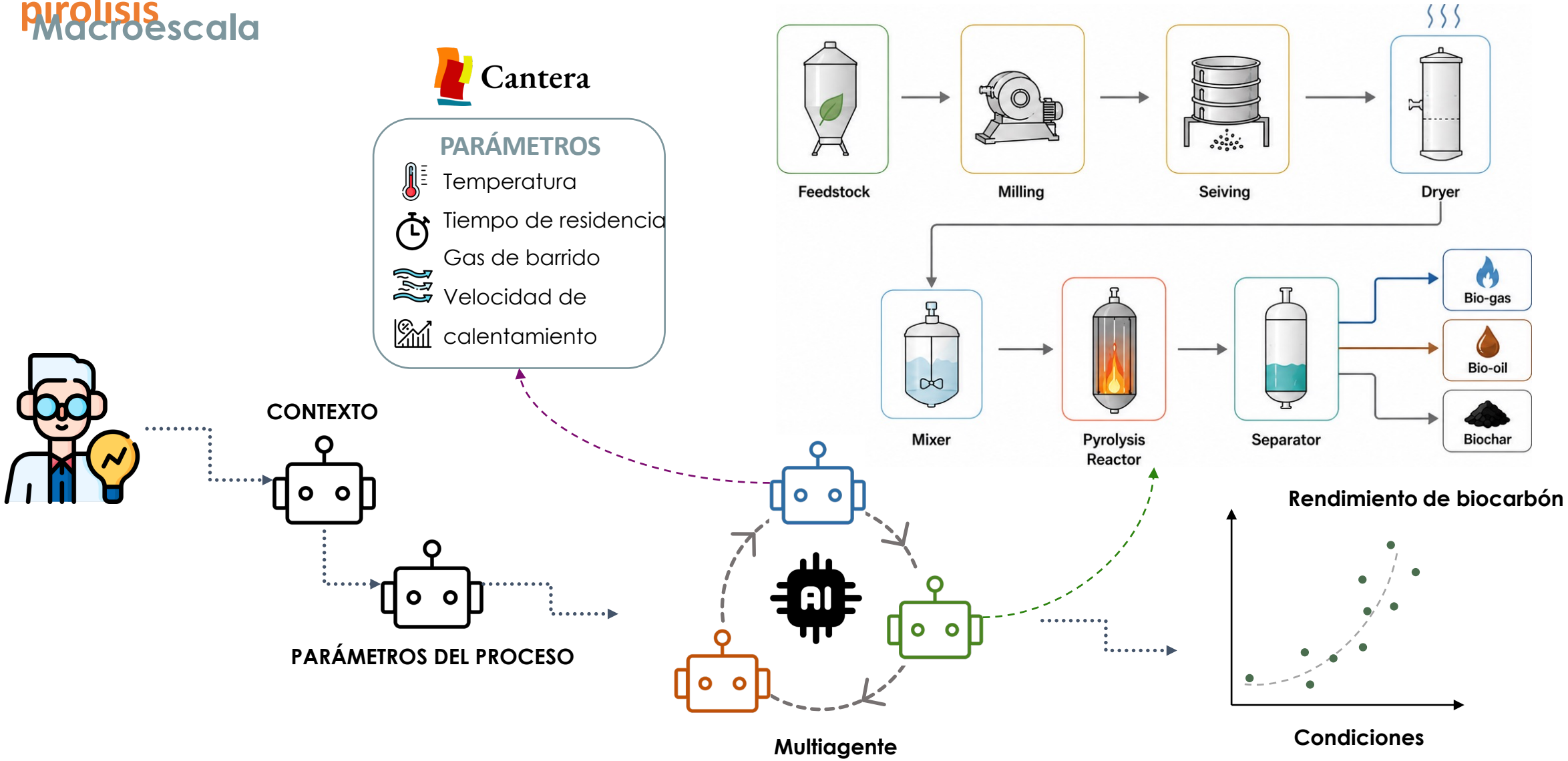


Butler, Davies, Cartwright, Isayev, Walsh, Nature 559, 547–555 (2018).

Modelización del biocarbón – ingeniería del proceso de pirólisis

DESARROLLO CON LLM

Macroescala



PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas

B) Planifica y actúa con herramientas, en bucle

C) Nunca usa herramientas

D) Solo funciona sin conexión

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas

B) Planifica y actúa con herramientas, en bucle

C) Nunca usa herramientas

D) Solo funciona sin conexión

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas

B) Planifica y actúa con herramientas, en bucle

C) Nunca usa herramientas

D) Solo funciona sin conexión

Q2 **Dotar a un LLM de herramientas (RDKit, búsqueda, simuladores)…**

A) Lo empeora

B) Lo ancla y mejora su fiabilidad

C) Elimina todos los errores

D) Es imposible

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas

B) Planifica y actúa con herramientas, en bucle

C) Nunca usa herramientas

D) Solo funciona sin conexión

Q2 **Dotar a un LLM de herramientas (RDKit, búsqueda, simuladores)…**

A) Lo empeora

B) Lo ancla y mejora su fiabilidad

C) Elimina todos los errores

D) Es imposible

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas
B) Planifica y actúa con herramientas, en bucle
C) Nunca usa herramientas
D) Solo funciona sin conexión

Q2 **Dotar a un LLM de herramientas (RDKit, búsqueda, simuladores)…**

A) Lo empeora
B) Lo ancla y mejora su fiabilidad
C) Elimina todos los errores
D) Es imposible

Q3 **Los LLM son especialmente útiles cuando…**

A) Hay muchísimos datos
B) Hay pocos datos
C) Nunca
D) Solo con imágenes

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (opción múltiple)

Opción múltiple

Q1 **Un «agente» LLM (frente a un asistente) es aquel que…**

A) Solo responde preguntas
B) Planifica y actúa con herramientas, en bucle
C) Nunca usa herramientas
D) Solo funciona sin conexión

Q2 **Dotar a un LLM de herramientas (RDKit, búsqueda, simuladores)…**

A) Lo empeora
B) Lo ancla y mejora su fiabilidad
C) Elimina todos los errores
D) Es imposible

Q3 **Los LLM son especialmente útiles cuando…**

A) Hay muchísimos datos
B) Hay pocos datos
C) Nunca
D) Solo con imágenes

Mensajes clave – Liberar el conocimiento atrapado en los textos.

POR QUÉ IMPORTA PARA EL CLIMA

1. Los LLM añaden una capa de lenguaje + razonamiento + uso de herramientas al AA especializado.

2. Convertir décadas de bibliografía en datos utilizables; planificar síntesis; orquestar herramientas.

3. Mantener a una persona responsable y registrar cada acción. La trazabilidad y la supervisión hacen que la autonomía sea fiable (y reproducible).

¿PREGUNTAS?

IA RESPONSABLE



IA responsable para materiales y química

Las herramientas potentes conllevan responsabilidades. La misma aceleración es un arma de doble filo

1 Integridad y reproducibilidad

La fuga de datos y la obsesión por las clasificaciones inflan los resultados.
Documentar lo suficiente para que otros puedan reproducirlos.

2 Predicciones fiables

Indicar la incertidumbre, respetar el dominio de aplicabilidad y validar después (DFT → experimento).

3 Calidad de los datos y sesgos

Los conjuntos de datos se inclinan hacia sistemas bien estudiados y exitosos.
Si entra basura, sale basura

4 Apertura y FAIR

Compartir datos, código y pesos de los modelos
Localizables, accesibles, interoperables, reutilizables (FAIR).

5 Seguridad y doble uso

Las herramientas de diseño potentes pueden invertirse para hacer daño.
Control de acceso y supervisión humana.

6 Sostenibilidad y computación

El entrenamiento y la inferencia consumen energía.
Dimensionar bien los modelos y reutilizar modelos preentrenados.

Sesgos de los datos

Retos

- Las creencias implícitas y los sesgos de contextos occidentales pueden normalizarse en la IA y hacerla menos responsable a escala mundial.
- Es necesario aplicar una perspectiva global a los supuestos básicos de la IA: datos de entrenamiento, rendimiento de los modelos y **explicabilidad** de los sistemas
- Plantear desde el principio las preguntas difíciles sobre los impactos positivos y negativos, en lugar de recurrir a reparaciones y parches a posteriori.

Acciones

- Ser robustos frente al multilingüismo y la diversidad de alfabetizaciones
- Personalizar las experiencias con el modelo de identidad compartida
- **La explicabilidad** y la transparencia deben tener sentido para los nuevos usuarios de la tecnología
- Abordar de forma proactiva la equidad ante los desequilibrios de los datos

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

Referencias

IA responsable y fiable para la ciencia

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).

PARTICIPACIÓN

CIERRE DE SECCIÓN · TEST (verdadero/falso)

¿Verdadero o falso?

1. **Los conjuntos de datos de química publicados están sesgados, en parte porque rara vez se publican los experimentos fallidos.**

   Verdadero | Falso

2. **Un modelo de diseño de fármacos puede invertirse para diseñar toxinas.**

   Verdadero | Falso

3. **FAIR significa Findable, Accessible, Interoperable, Reusable (localizables, accesibles, interoperables, reutilizables).**

   Verdadero | Falso

Mensajes clave – Acelerar solo sirve si es fiable y compartido.

POR QUÉ IMPORTA PARA EL CLIMA

1. Los datos están sesgados: sesgo de publicación, sesgo humano, supuestos propios del contexto occidental.

2. Las herramientas generativas/agénticas potentes son intrínsecamente de doble uso.

3. La reproducibilidad, la apertura (FAIR) y el coste computacional forman parte de hacerlo bien.

¿PREGUNTAS?

IA EXPLICABLE

IA explicable: de la predicción a la comprensión

En ciencia solemos querer el «porqué», no solo la cifra

La brecha: un modelo puede predecir una propiedad con precisión y aun así no enseñarnos nada.
La explicabilidad se pregunta qué ha aprendido el modelo, si lo ha aprendido por las razones correctas y qué ciencia nueva sugiere.

1 Comprensión

Convertir un modelo ajustado en reglas de diseño y mecanismos. Recuperar la física/química que hay detrás de los datos.

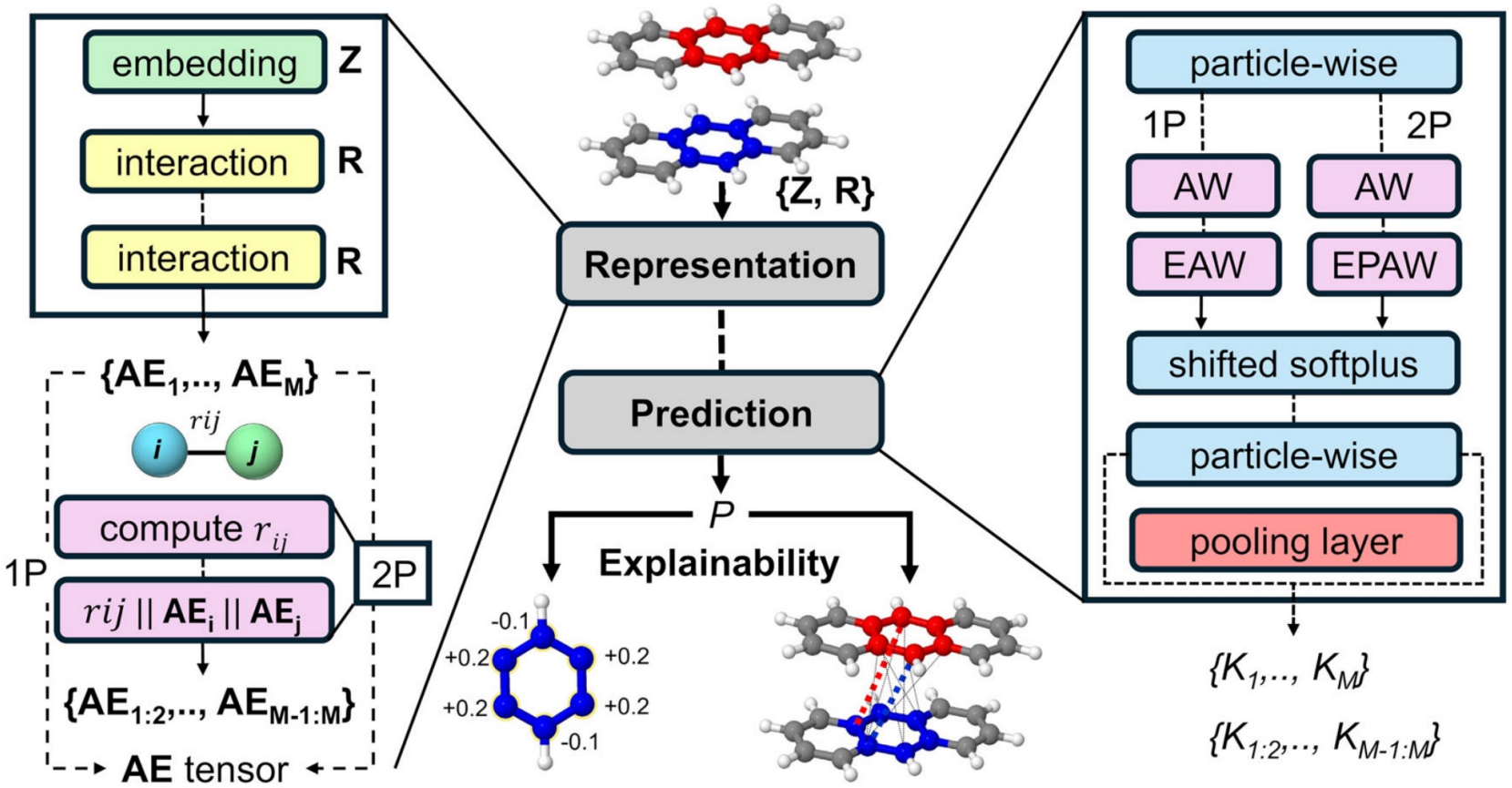
2 Confianza y validación

Comprobar que el modelo predice por las razones correctas, no por correlaciones falsas o artefactos del conjunto de datos.

3 Descubrimiento

La interpretación saca a la luz nuevas relaciones estructura–propiedad e hipótesis comprobables.

IA química explicable: 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

Referencias

IA explicable e interpretable para la ciencia y los materiales

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 EXPLICABLE

Mensajes clave – Convertir predicciones rápidas en reglas de diseño fiables.

POR QUÉ IMPORTA PARA EL CLIMA

1. En ciencia queremos el porqué, no solo la cifra.

2. Los métodos interpretables desenmascaran los modelos que aciertan por las razones equivocadas.

3. Tratar cada explicación como una hipótesis que verificar frente a la física y el experimento.

¿PREGUNTAS?

REFLEXIONES FINALES

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.

Diseñar para la circularidad
Diseñar para el rendimiento
y el rendimiento



AI/ML

Repositorios en línea: herramientas de código abierto (materiales e IA)

Un kit de inicio para compartir. Las herramientas de nuestro grupo aparecen resaltadas; comprobar que los enlaces estén actualizados antes de distribuirlo.

Área	Herramienta	Para qué sirve	Repositorio
Datos y herramientas	**pymatgen**	Análisis de materiales, E/S y API de Materials Project	*github.com/materialsproject/pymatgen*
Datos y herramientas	**matminer**	Descriptores/vectorización listos para usar en AA	*github.com/hackingmaterials/matminer*
Datos y herramientas	**RDKit**	Quimioinformática: moléculas, huellas, descriptores	*github.com/rdkit/rdkit*
Datos y herramientas	**DeepChem**	Herramientas de AA para química: conjuntos de datos, modelos, bancos de pruebas	*github.com/deepchem/deepchem*
Datos y herramientas	**ASE**	Atomic Simulation Environment: construir/ejecutar/analizar	*gitlab.com/ase/ase*
Representaciones	**SELFIES**	Representación molecular en cadena siempre válida	*github.com/aspuru-guzik-group/selfies*
Representaciones	**eOHE**	Codificaciones one-hot incrustadas (entradas eficientes)	*github.com/MMLabCodes/eOHE*
Representaciones	**eMFP**	Huellas de Morgan incrustadas (compactas, más rápidas)	*github.com/MMLabCodes/eMFP*
Generativo	**chemical_vae**	Diseño molecular en un espacio latente continuo	*github.com/aspuru-guzik-group/chemical_vae*
Generativo	**MatterGen**	Modelo generativo de difusión para cristales inorgánicos	*github.com/microsoft/mattergen*
Potenciales de AA	**MACE**	Potenciales interatómicos de AA (+ modelo fundacional MACE-MP-0)	*github.com/ACEsuit/mace*
Potenciales de AA	**CHGNet**	Potencial universal de red neuronal informado por la carga	*github.com/CederGroupHub/chgnet*
Potenciales de AA	**MatGL (M3GNet)**	Potenciales universales de grafos + modelos de propiedades	*github.com/materialsvirtuallab/matgl*
Autonomía	**Ax / BoTorch**	Optimización bayesiana para laboratorios autónomos	*github.com/facebook/Ax*
Agentes LLM	**ChemCrow**	Agente LLM de química conectado a herramientas expertas	*github.com/ur-whitelab/chemcrow-public*
Explicabilidad	**SHAP**	Explicaciones por atribución de características para cualquier modelo	*github.com/shap/shap*
Explicabilidad	**SISSO**	Descubrimiento de descriptores/fórmulas interpretables	*github.com/rouyang2017/SISSO*
Informática verde	**CodeCarbon**	Seguir el coste de carbono de tu computación	*github.com/mlco2/codecarbon*

Filas resaltadas = repositorios de nuestro grupo (github.com/MMLabCodes). Lista representativa, no exhaustiva.

Lecturas complementarias

CUELLO DE BOTELLA

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

MODELIZACIÓN Y SIMULACIÓN

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

NUEVO 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

CAJA DE HERRAMIENTAS 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

HABLAR DE QUÍMICA

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 RESPONSABLE

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 EXPLICABLE

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

Gracias. ¿Alguna pregunta final?

