IA para acelerar a ciência dos materiais

Francisco Martin-Martinez

Professor sênior, King's College London

martinmartinezlab.com

ESCOLA DE VERÃO VIRTUAL CCAI 2026 · AULA SETORIAL · 13 DE AGOSTO

QUEM SOU EU?

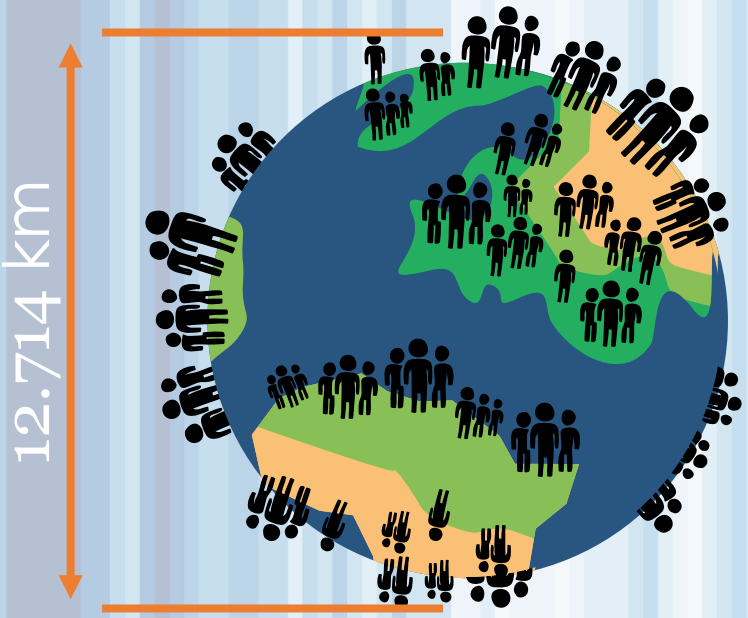
Quem sou eu e por que isso importa para mim



Um cientista que usa a IA como ferramenta ou uma IA que faz ciência de forma autônoma: onde fica o limite?

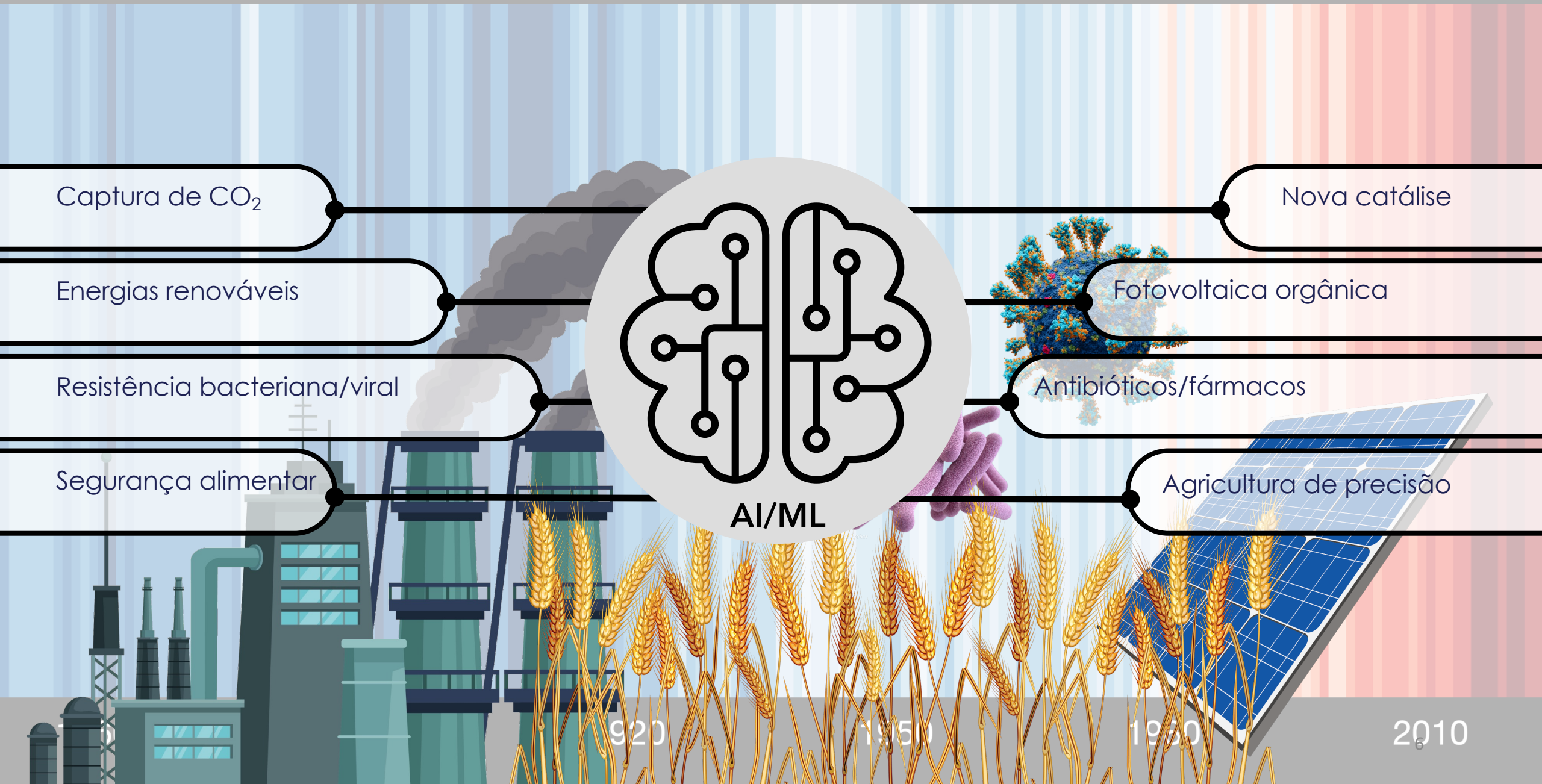
POR QUE ISSO IMPORTA PARA MIM?

Global temperature change (1850-2022)



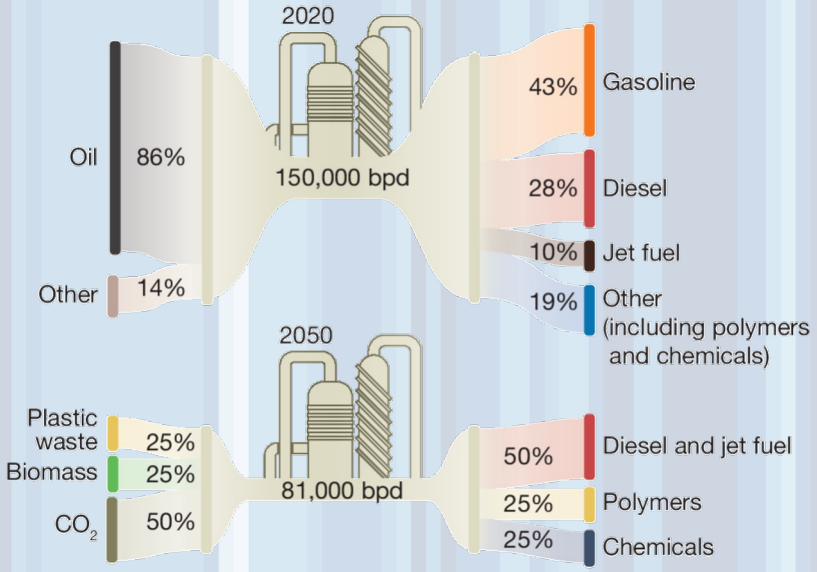
1860 1890 1920 1950 1980 2010

Global temperature change (1850-2022)



Nova mineração de materiais a partir de resíduos

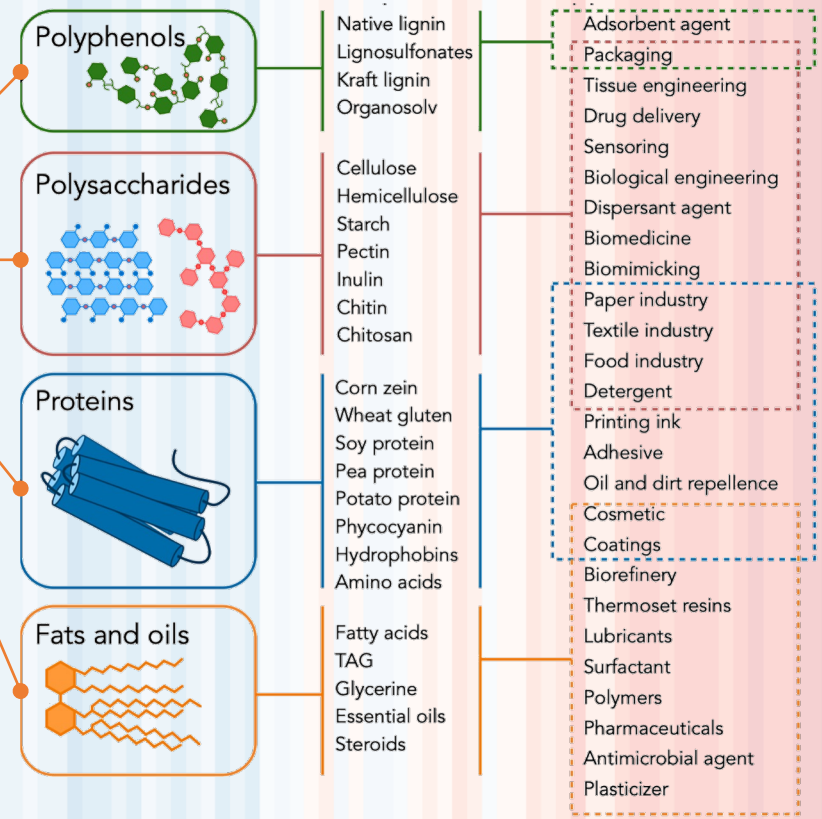
(Bio)refinaria 2050



Nature 629, 295, **2024**

Resíduos de biomassa
140 Gt/ano

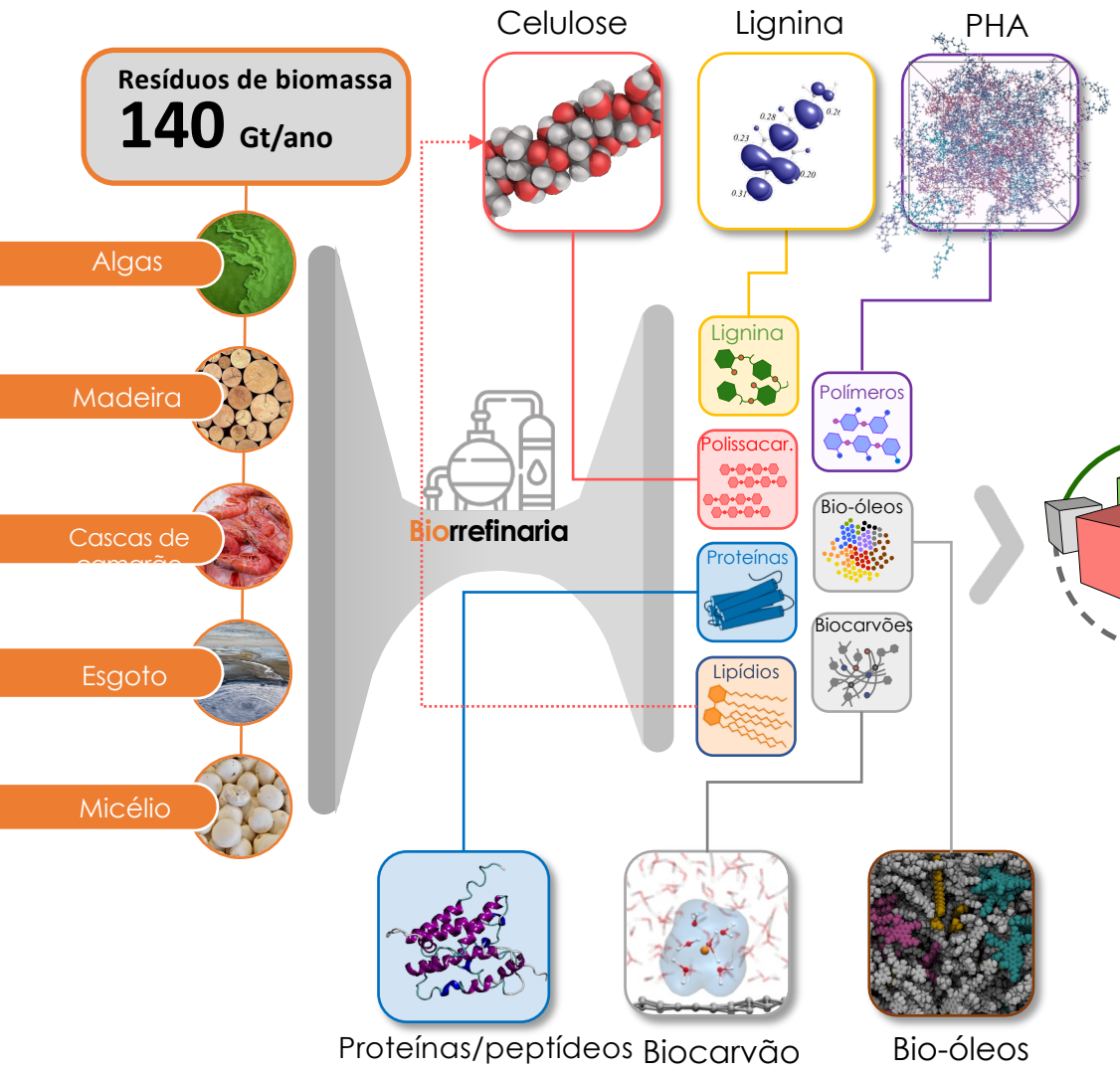
AI/ML

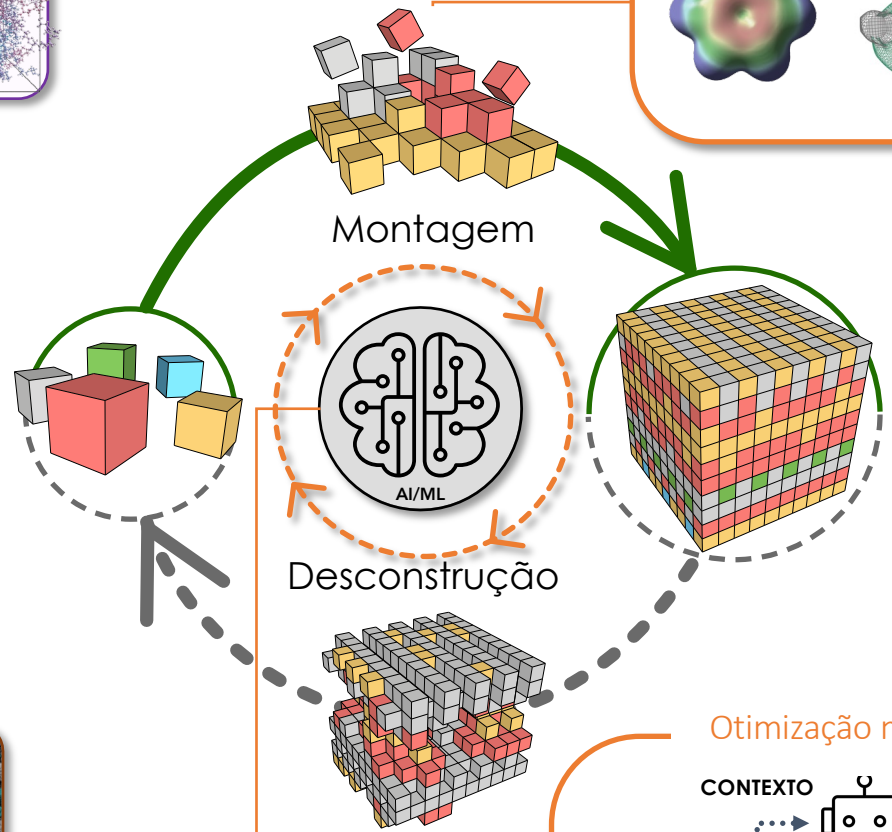


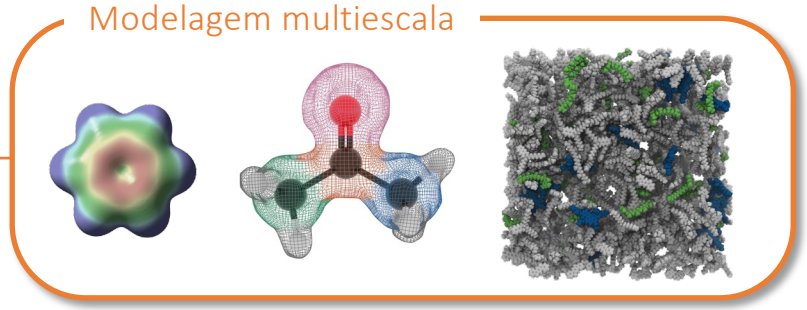
Bachs-Herrera, …, Martin-Martinez **iScience** 26, 106549, **2023**

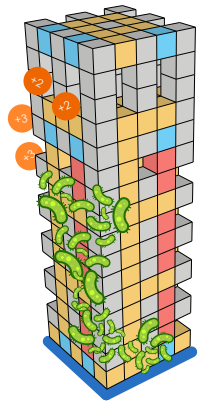
Valorização de resíduos em moléculas e materiais de base biológica

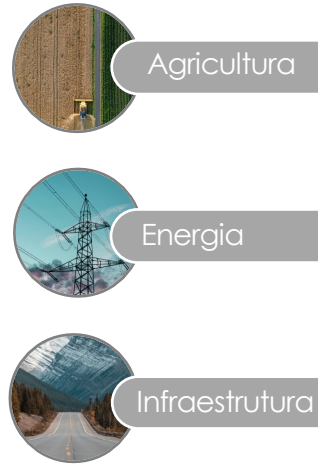
Materiais circulares biobasead ➡ Rumo a uma biorrefinaria quântica movida a IA

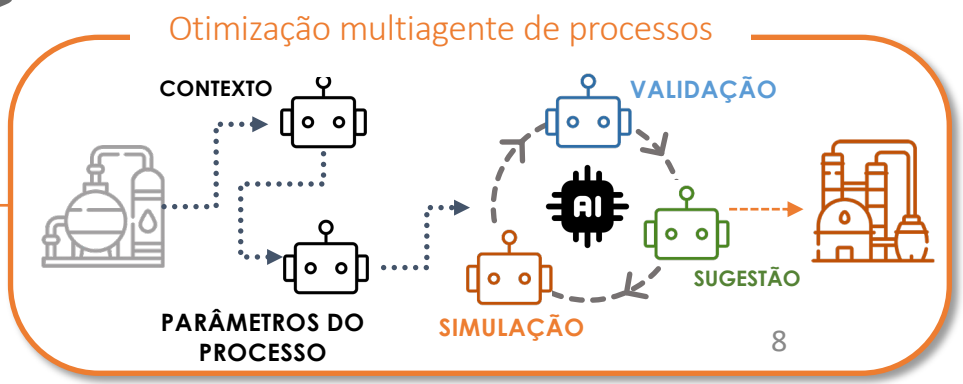












Bachs-Herrera et al., Martin-Martinez **iScience** 26, 106549, **2023**

O QUE VEREMOS HOJE?

O QUE VEREMOS

01 Por que os materiais são um gargalo climático

02 O paradigma tradicional (ainda útil) de modelagem e simulação de materiais

03 Um novo paradigma na descoberta de materiais

04 A caixa de ferramentas da IA

05 Falando de química com os computadores

04 IA responsável

05 IA explicável em química

PARTICIPAÇÃO

Enquete ao vivo – Usem o chat

Qual desafio ligado aos materiais é o mais urgente para o clima?
Não há resposta errada: é para quebrar o gelo

PERGUNTAS?

O GARGALO DOS MATERIAIS

Onde a IA entra na cadeia de materiais

1

Representar

Codificar moléculas/cristais para o AM (impressões digitais, grafos OHE, texto)

2

Prever

Previsão de propriedades, modelos substitutos, **potenciais interatômicos de AM**

3

Gerar

Design inverso/generativo de novos candidatos

4

Sintetizar

Laboratórios autônomos e autoguiados que fecham o ciclo

5

Caracterizar

Análise assistida por IA de espectros, imagens e estruturas

Os materiais, um gargalo climático

Fotovoltaica

Eficiência e estabilidade limitadas pela química dos materiais

Baterias

Densidade de energia, vida útil, dependência de minerais críticos

Combustíveis alternativos

Catalisadores e carreadores para H_2 verde e e-combustíveis

Sorventes de CO_2

Materiais de captura baratos, duráveis e escaláveis

Outros: infraestrutura sustentável, segurança alimentar, descoberta de fármacos, antibióticos…

O problema: o ciclo de descoberta é lento e caro.

Projetar → sintetizar → caracterizar → testar pode levar anos por candidato.

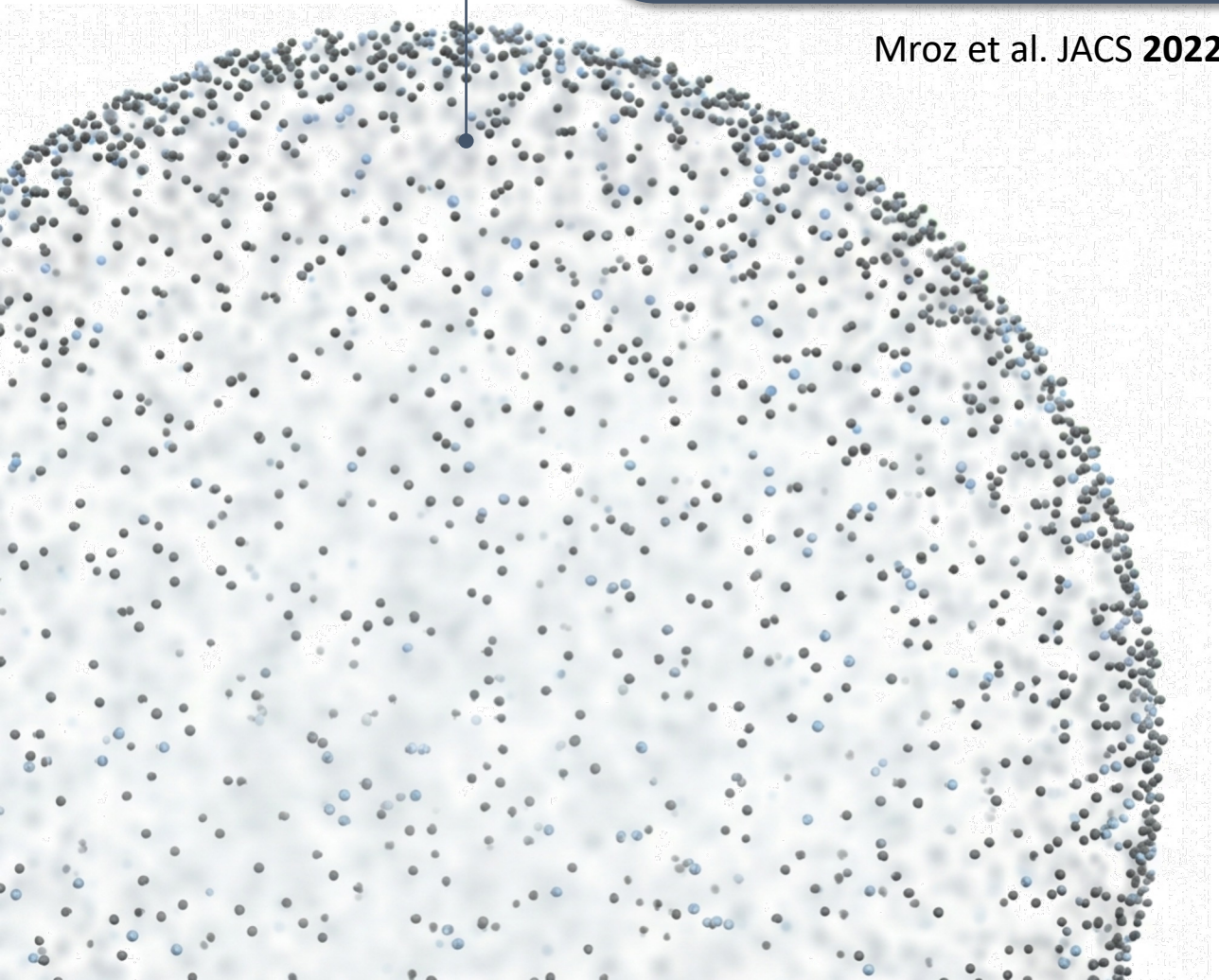
Tanto a mitigação QUANTO a adaptação dependem de obter os materiais certos, mais rápido e mais barato.

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

O gargalo da descoberta

10^{60} Moléculas pequenas (tipo fármaco) para o design de materiais

Mroz et al. JACS **2022**



O espaço químico combinatório é imenso

VS

A validação experimental é lenta e cara

VOCÊ SABIA?

Existem mais moléculas orgânicas pequenas possíveis do que átomos no universo observável.

PARTICIPAÇÃO

Quiz rápido – Verdadeiro/Falso

1. **Existem mais moléculas orgânicas pequenas possíveis do que átomos no universo observável.**

   Verdadeiro | Falso

2. **Um novo material costuma chegar ao mercado em poucas semanas.**

   Verdadeiro | **Falso**

3. **A IA entra na cadeia nas etapas de representar, prever, gerar, sintetizar e caracterizar.**

   Verdadeiro | Falso

Mensagens principais

POR QUE IMPORTA PARA O CLIMA

1. Os materiais são um gargalo climático: tanto a mitigação quanto a adaptação precisam de materiais melhores, mais rápido.

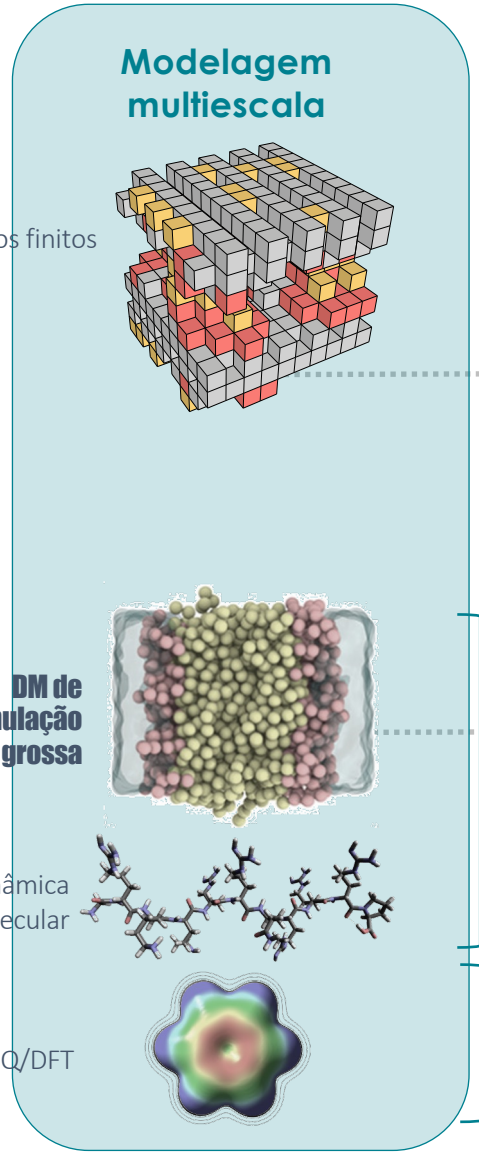
2. O espaço químico não pode ser explorado por força bruta. A IA o transforma em uma busca guiada.

3. As metas climáticas têm prazo fixo. Descobrir mais rápido significa implantar mais rápido.

PERGUNTAS?

MODELAGEM E SIMULAÇÃO DE MATERIAIS

Modelagem multiescala



Simulações de dinâmica molecular

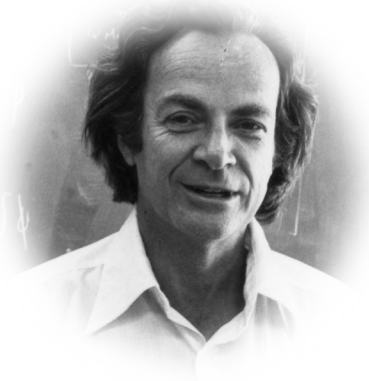


"Todos os modelos estão errados, mas alguns são úteis."

– George E. P. Box



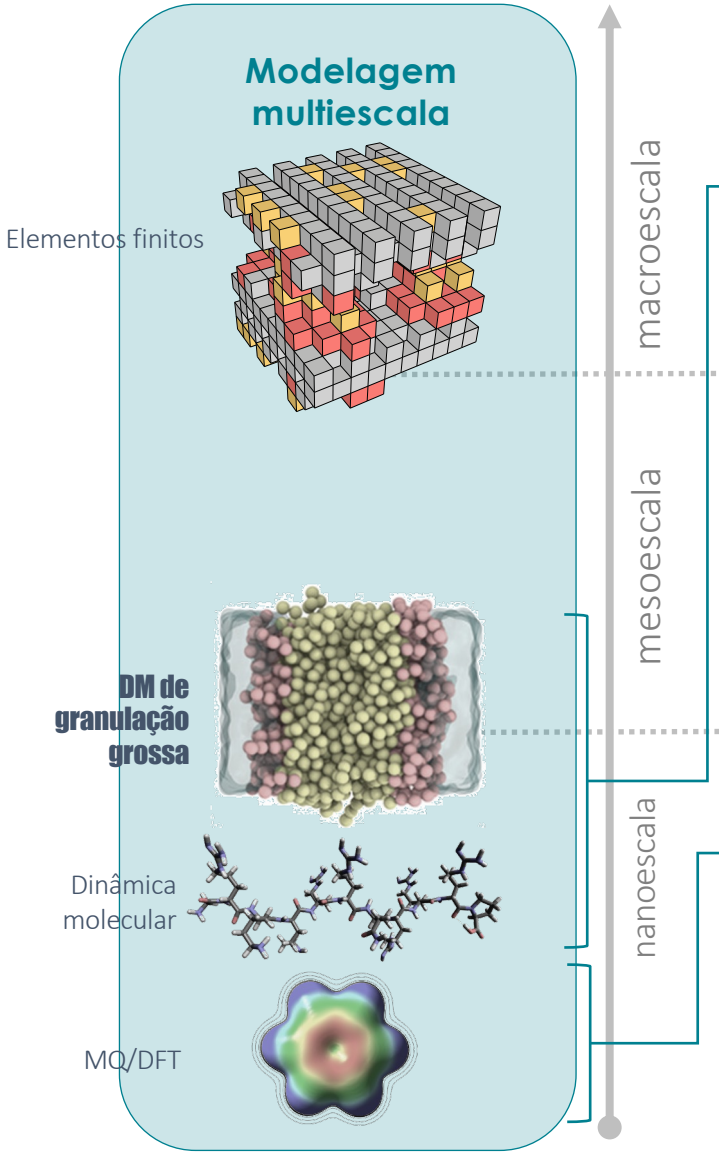
Teoria do funcional da



"Acho que posso dizer com segurança que ninguém entende a mecânica quântica."

– Richard Feynman

Modelagem multiescala



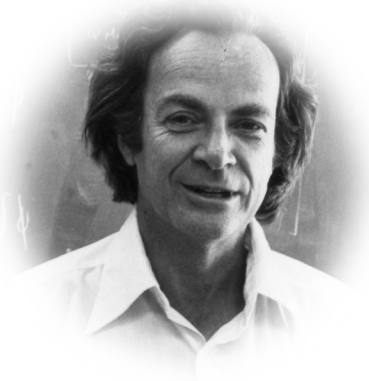
Simulações de dinâmica molecular



"Todos os modelos estão errados, mas alguns são úteis."

– George E. P. Box

Teoria do funcional da densidade (DFT) – Mecânica quântica

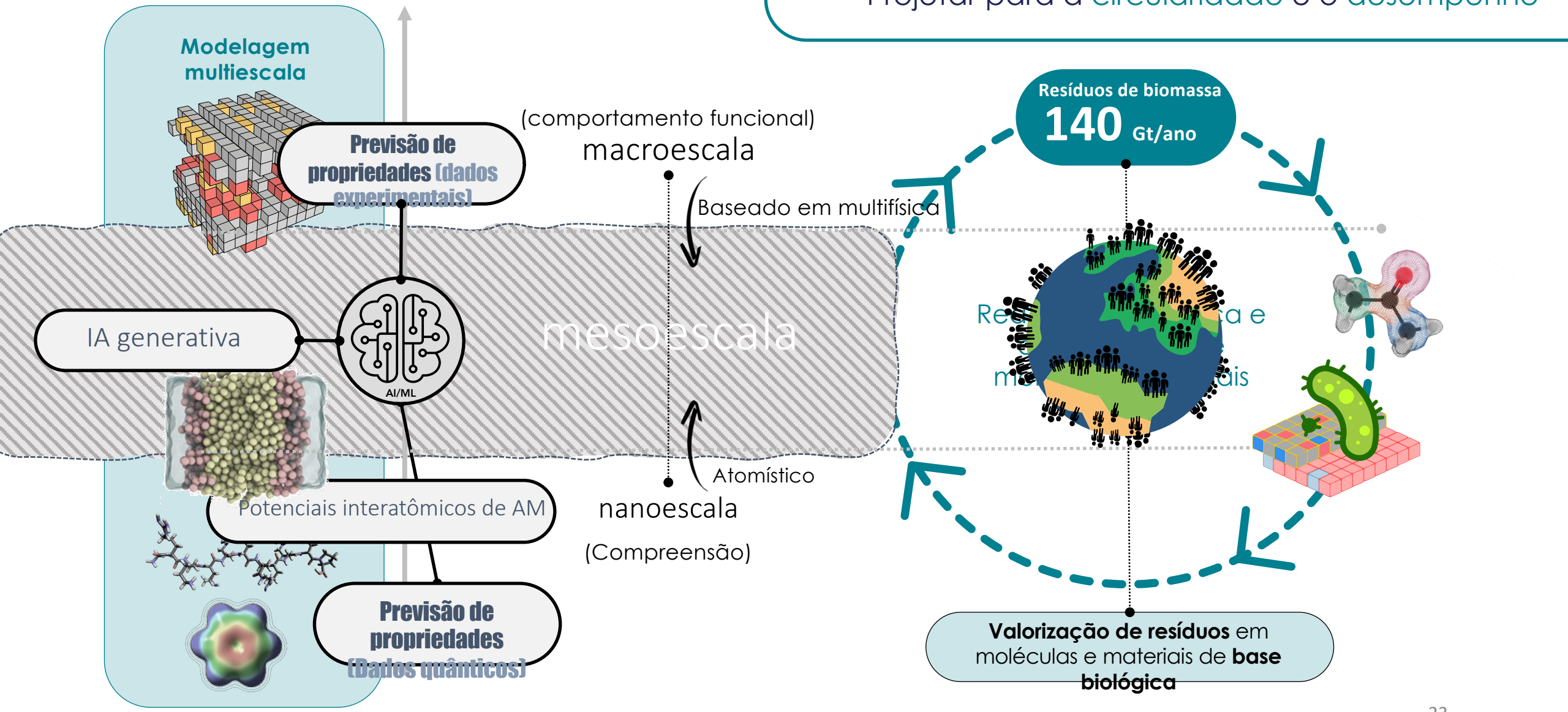


"Acho que posso dizer com segurança que ninguém entende a mecânica quântica."

– Richard Feynman

Modelagem multiescala

Projetar para a circularidade e o desempenho



Software de DFT, DM e AM para química

DFT molecular



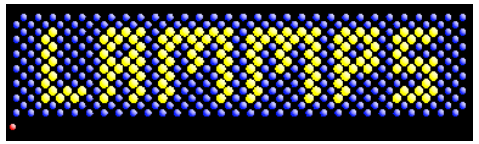
ORCA

DFT de estado sólido



siesta™

DM (todos os átomos e granulação grossa)



GROMACS
FAST. FLEXIBLE. FREE.



NAMD
Scalable Molecular Dynamics



MD ANALYSIS



Integração e quimioinformática

python™

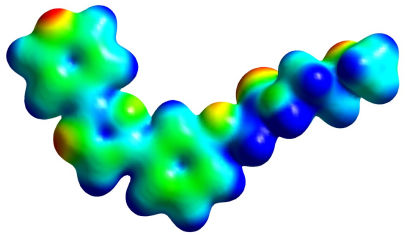
RDKit
Open-Source Cheminformatics and Machine Learning

+ Atomic Simulation Environment (ASE)

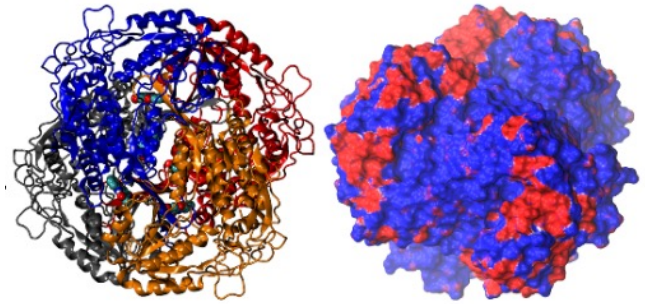
Bancos de dados para a descoberta em química

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

Teoria do funcional da densidade (DFT)



Dinâmica molecular (DM)



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

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) DFT (quântica)

D) Elementos finitos

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) **DFT (quântica)**

D) Elementos finitos

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) DFT (quântica)

D) Elementos finitos

Q2 **Por que usamos a modelagem multiescala?**

A) Uma escala basta

B) Os fenômenos variam conforme a escala

C) Para economizar citações

D) É exigido por lei

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) **DFT (quântica)**

D) Elementos finitos

Q2 **Por que usamos a modelagem multiescala?**

A) Uma escala basta

B) **Os fenômenos variam conforme a escala**

C) Para economizar citações

D) É exigido por lei

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) **DFT (quântica)**

D) Elementos finitos

Q2 **Por que usamos a modelagem multiescala?**

A) Uma escala basta

B) **Os fenômenos variam conforme a escala**

C) Para economizar citações

D) É exigido por lei

Q3 **Em comparação com a DM de todos os átomos, a DM de granulação grossa alcança…**

A) Sistemas menores

B) Sistemas maiores e tempos mais longos

C) Maior precisão quântica

D) Nenhuma diferença

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – múltipla escolha

Q1 **Qual método é o mais preciso, mas também o mais caro?**

A) DM clássica

B) DM de granulação grossa

C) **DFT (quântica)**

D) Elementos finitos

Q2 **Por que usamos a modelagem multiescala?**

A) Uma escala basta

B) **Os fenômenos variam conforme a escala**

C) Para economizar citações

D) É exigido por lei

Q3 **Em comparação com a DM de todos os átomos, a DM de granulação grossa alcança…**

A) Sistemas menores

B) **Sistemas maiores e tempos mais longos**

C) Maior precisão quântica

D) Nenhuma diferença

MODELAGEM E SIMULAÇÃO

Mensagens principais – O lugar mais barato para errar é o computador.

POR QUE IMPORTA PARA O CLIMA

1. Nenhuma escala basta sozinha: a modelagem multiescala é necessária.

2. Precisamos prever a degradação e a reciclabilidade: circularidade, não só desempenho.

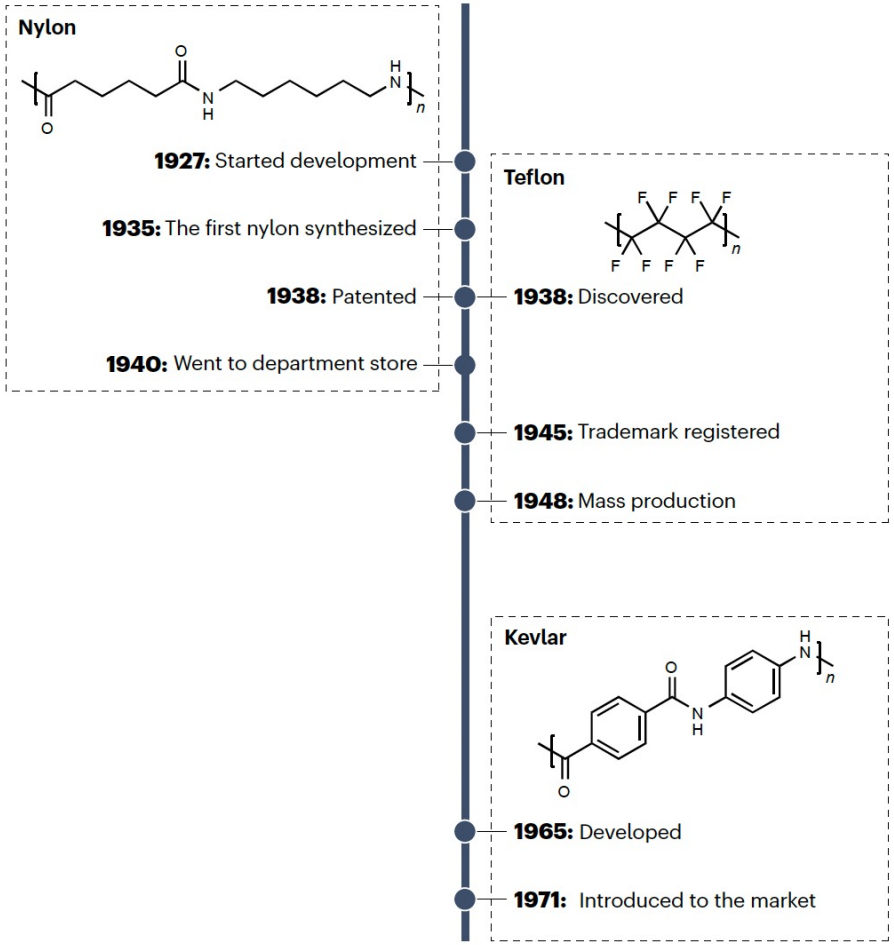
3. Os dados são a base sobre a qual a IA é construída. Bons dados geram bons modelos.

PERGUNTAS?

UM NOVO PARADIGMA NA DESCOBERTA DE MOLÉCULAS E MATERIAIS

Materiais lentos, mundo rápido

Inovações em polímeros nos últimos dois séculos.



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

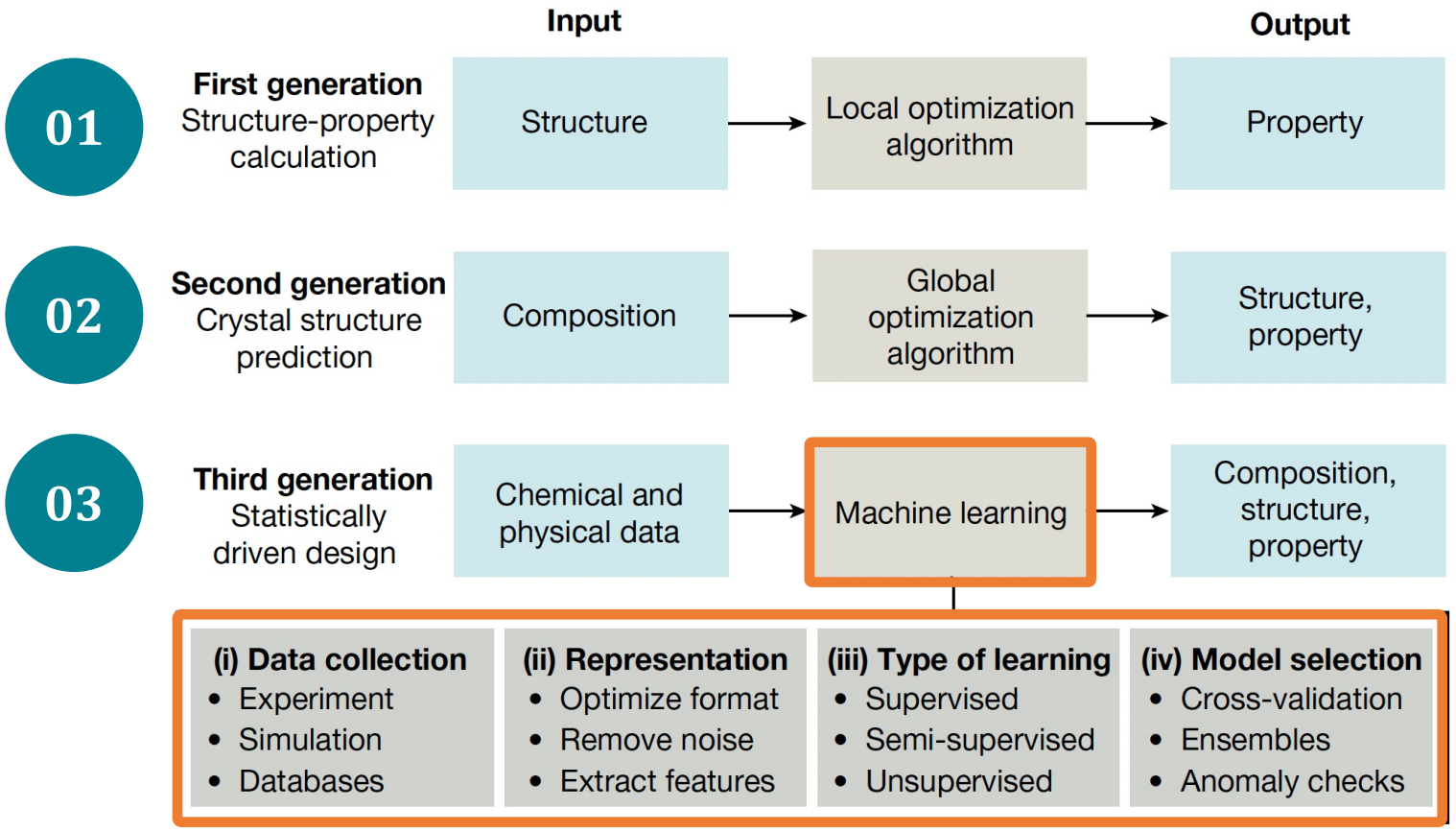
1977: Heeger, MacDiarmid and Shirakawa develop polyacetylene, the first conductive polymer

2011: Launch of Materials Genome Initiative

2020: 100-year anniversary of modern polymer science

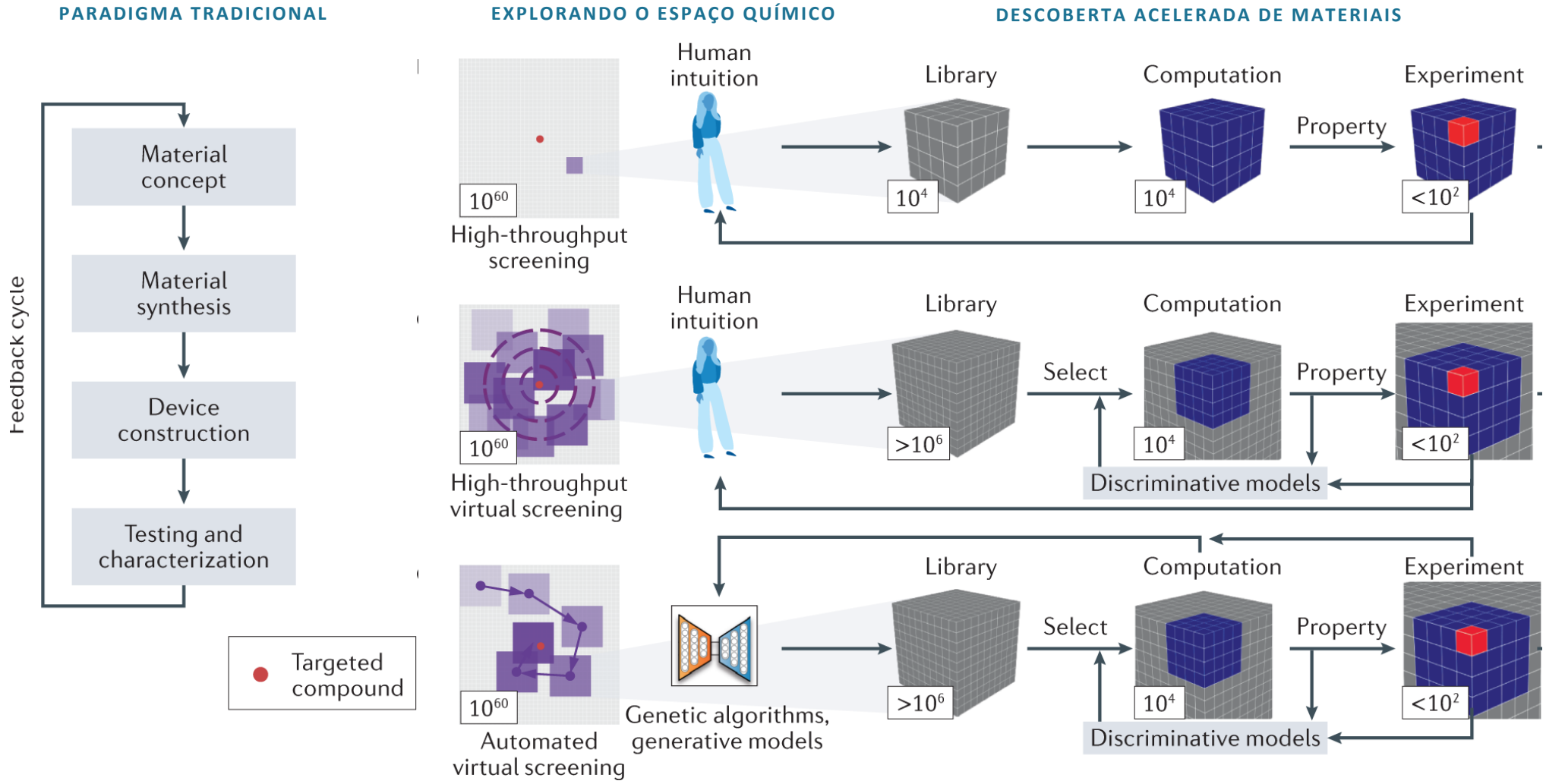
Tran, Gurnani, Kim, Pilania, Kwon, Lively, Ramprasad, Nat. Rev. Mater. 9, 866–886 (2024).

Da tentativa e erro ao design orientado por dados:
evolução do fluxo de trabalho de pesquisa em química computacional



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

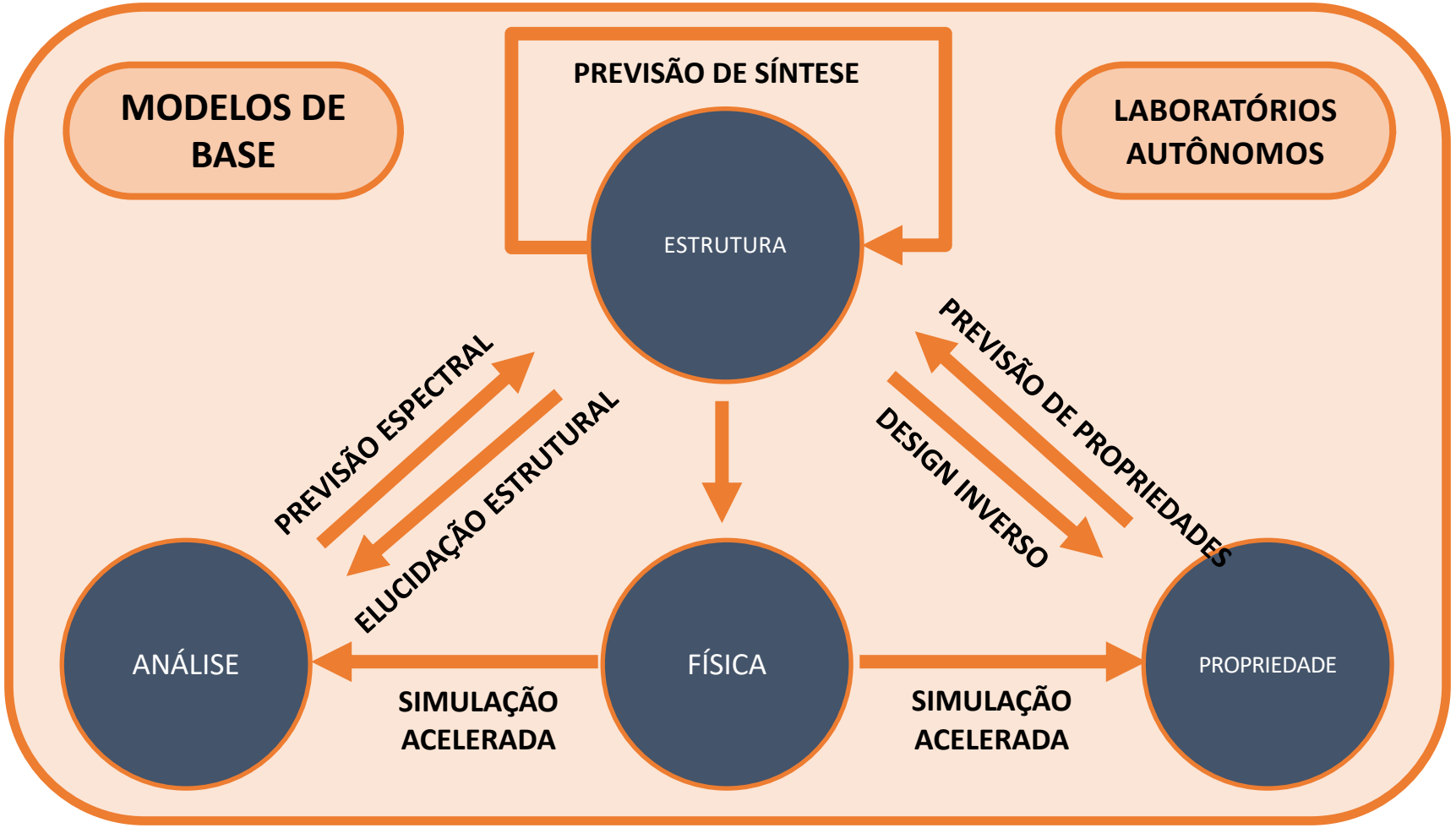
Novo paradigma na descoberta de materiais



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

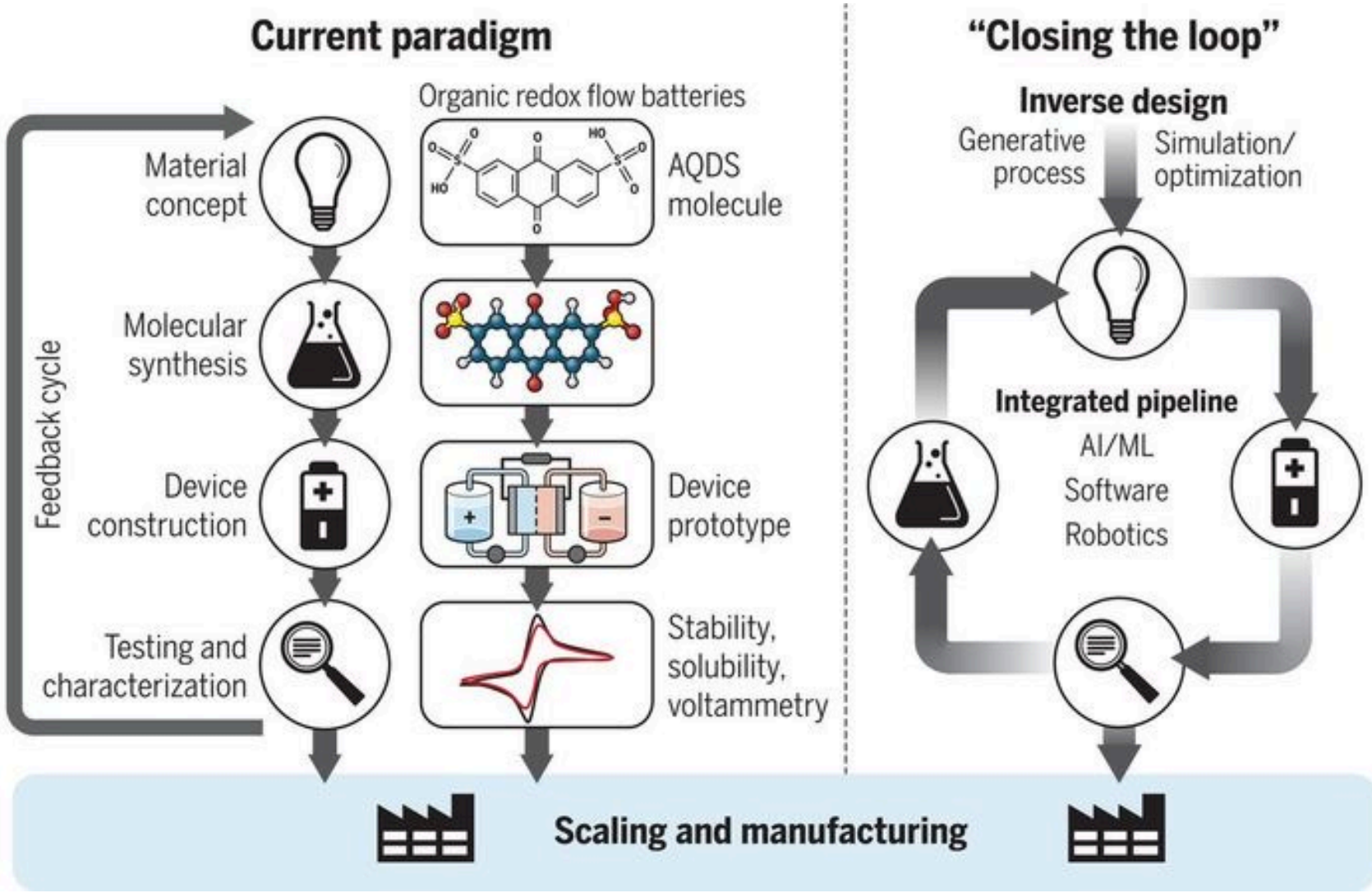
Aprendizado de máquina para química e materiais

NOVO PARADIGMA



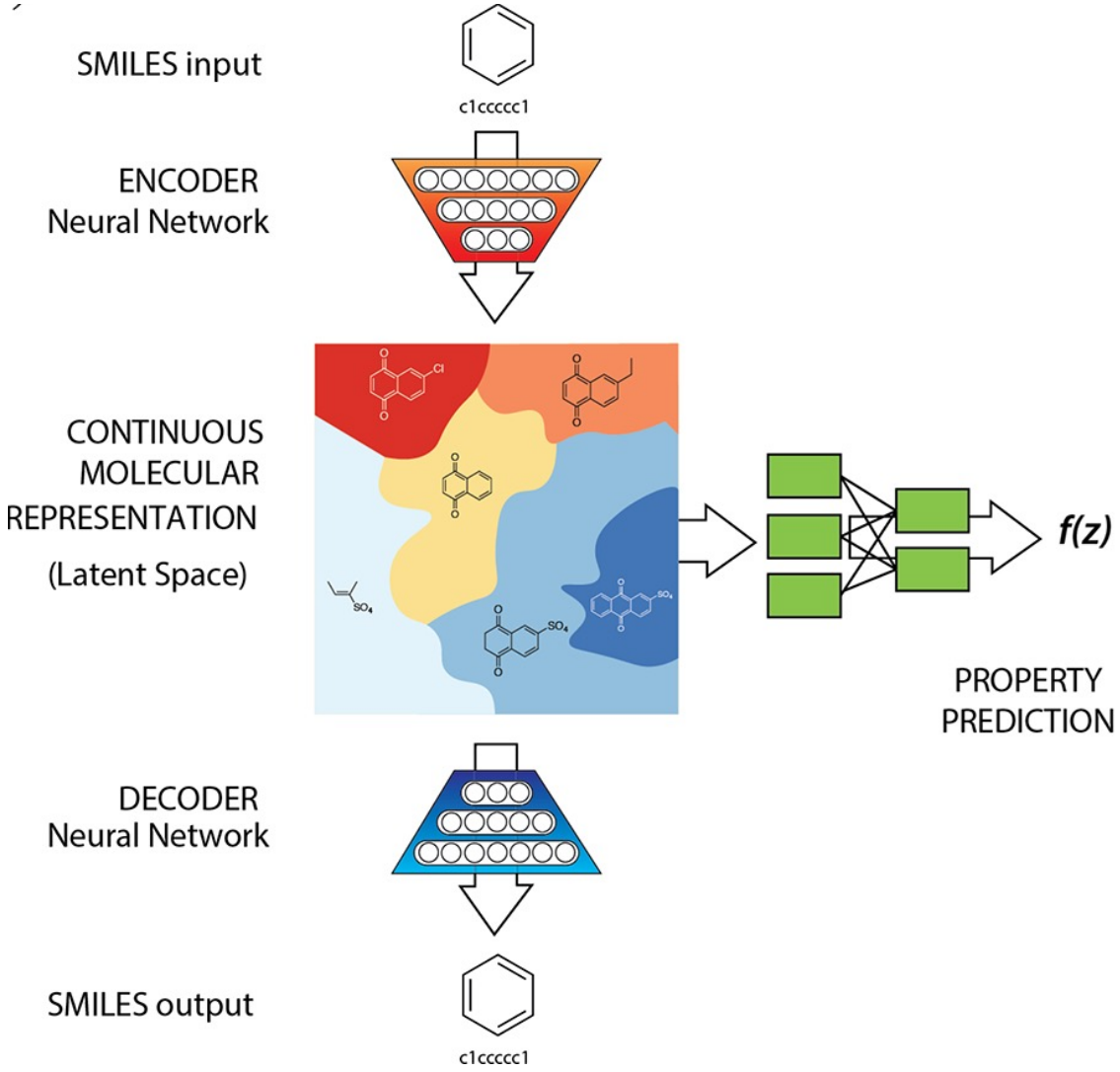
Cheng, Ser, Skreta, et al., Aspuru-Guzik, Faraday Discuss. 256, 10–60 (2025).

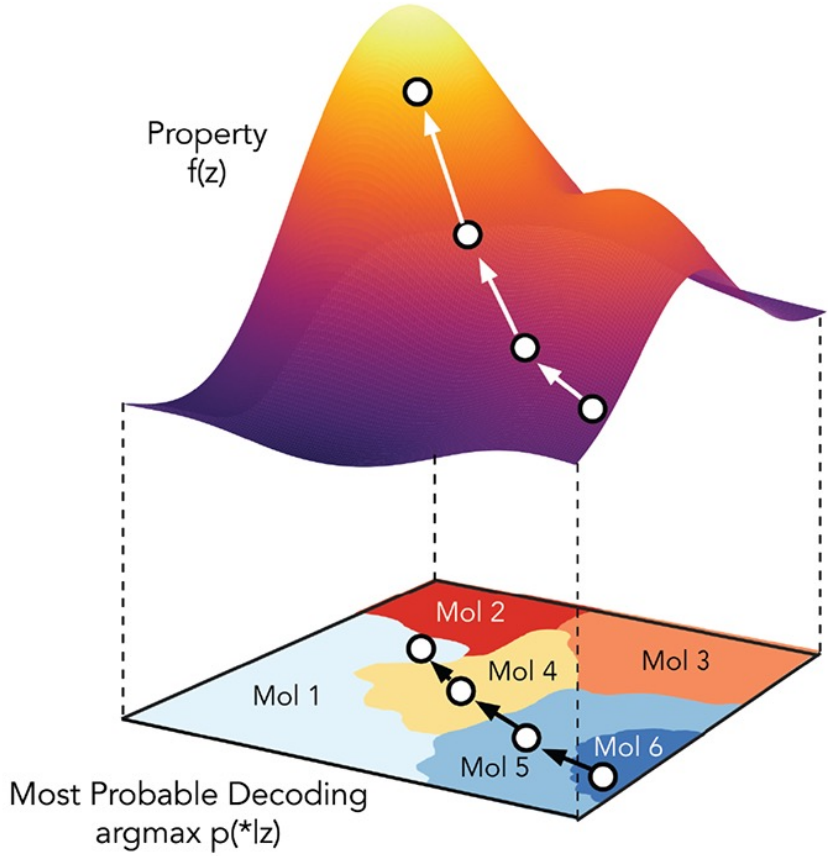
Design direto versus inverso: o exemplo das baterias de fluxo redox



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

VAE químico: transformar moléculas em um mapa a ser otimizado





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

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – Verdadeiro/Falso

1 **No design inverso, você especifica a propriedade desejada e o modelo propõe uma estrutura.**

Verdadeiro | Falso

2 **Um VAE fornece um espaço latente contínuo no qual é possível otimizar.**

Verdadeiro | Falso

3 **A “terceira geração” do design de materiais é a tentativa e erro manual.**

Verdadeiro | **Falso**

NOVO PARADIGMA

Mensagens principais – Projetar segundo uma especificação em vez de topar com ela por acaso.

POR QUE IMPORTA PARA O CLIMA

1. Estamos em uma geração de design de materiais guiada pelo AM.

2. Design inverso = otimizar em um espaço latente aprendido (espaço de design contínuo).

3. Há dois facilitadores transversais: os modelos de base e os laboratórios autônomos.

PERGUNTAS?

A CAIXA DE FERRAMENTAS DA IA

Uma caixa de ferramentas da IA

Seis coisas que o AM realmente faz ao longo da cadeia de materiais

1 **Previsão de propriedades**

PREVER

Modelos supervisionados preveem uma propriedade y (p. ex., o band gap) a partir de uma estrutura x, auxiliando ou substituindo simulações caras.

2 **Design generativo**

GERAR

Modelos não supervisionados/generativos produzem novos candidatos x', condicionados a propriedades-alvo: a rota do design inverso.

3 **Laboratórios autônomos**

SINTETIZAR

Laboratórios autônomos planejam e executam experimentos: retrossíntese, previsão de reações, otimização por aprendizado ativo.

4 **Apoio à caracterização**

CARACTERIZAR

O AM interpreta medições: estrutura cristalina por MET, coordenação por XAS, simetria por difração de elétrons.

5 **Gestão de sistemas**

IMPLANTAR

O AM gerencia dispositivos e redes: previsão da vida útil de baterias, previsão de carga, otimização do controle de redes inteligentes.

6 **Acelerar a teoria**

SIMULAR

O AM torna viáveis simulações intratáveis: potenciais e funcionais mais baratos para escalas maiores.

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

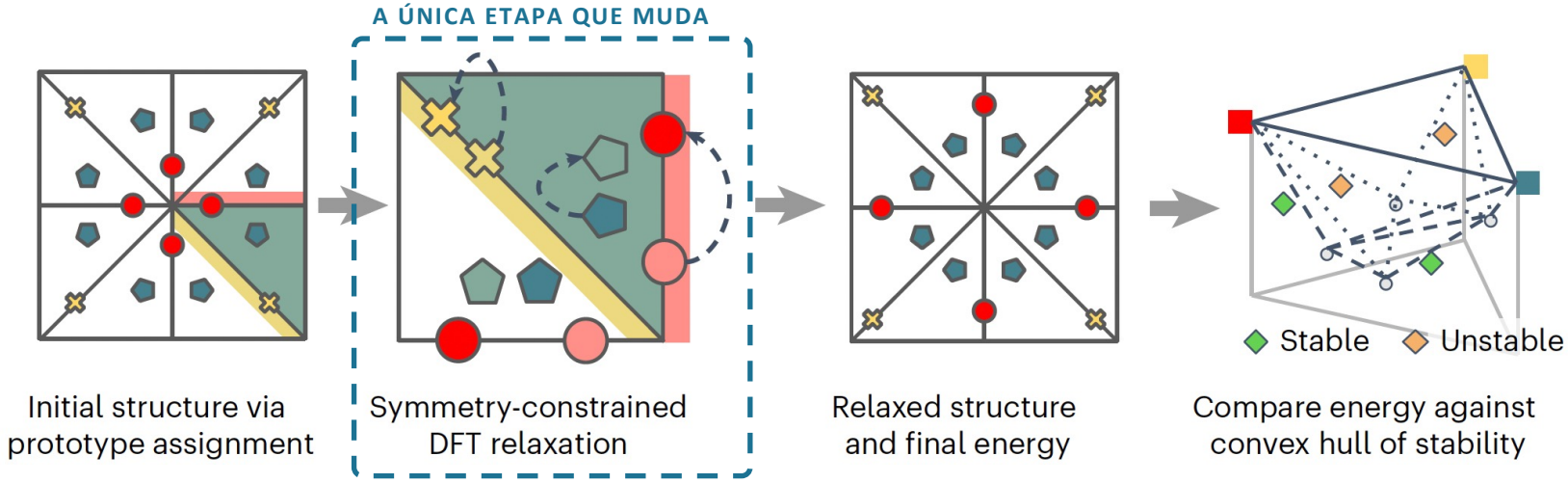
1

Previsão de propriedades

PREVER

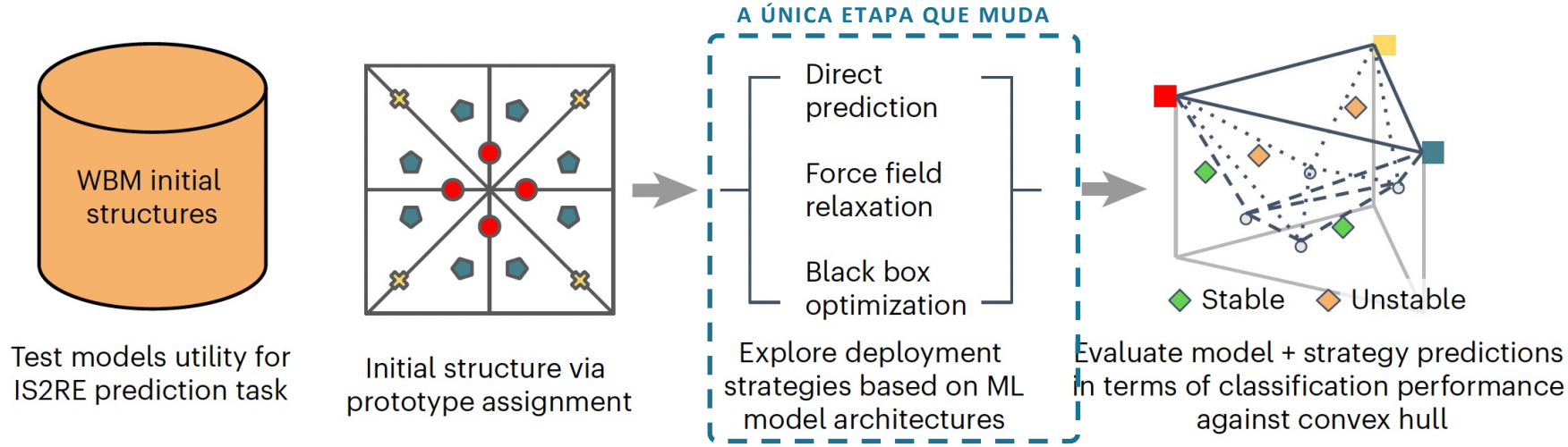
Prever propriedades DFT com um substituto de AM

CONVENCIONAL



MATBENCH DISCOVERY

IS2RE = da estrutura inicial à energia relaxada (Initial Structure to Relaxed Energy)



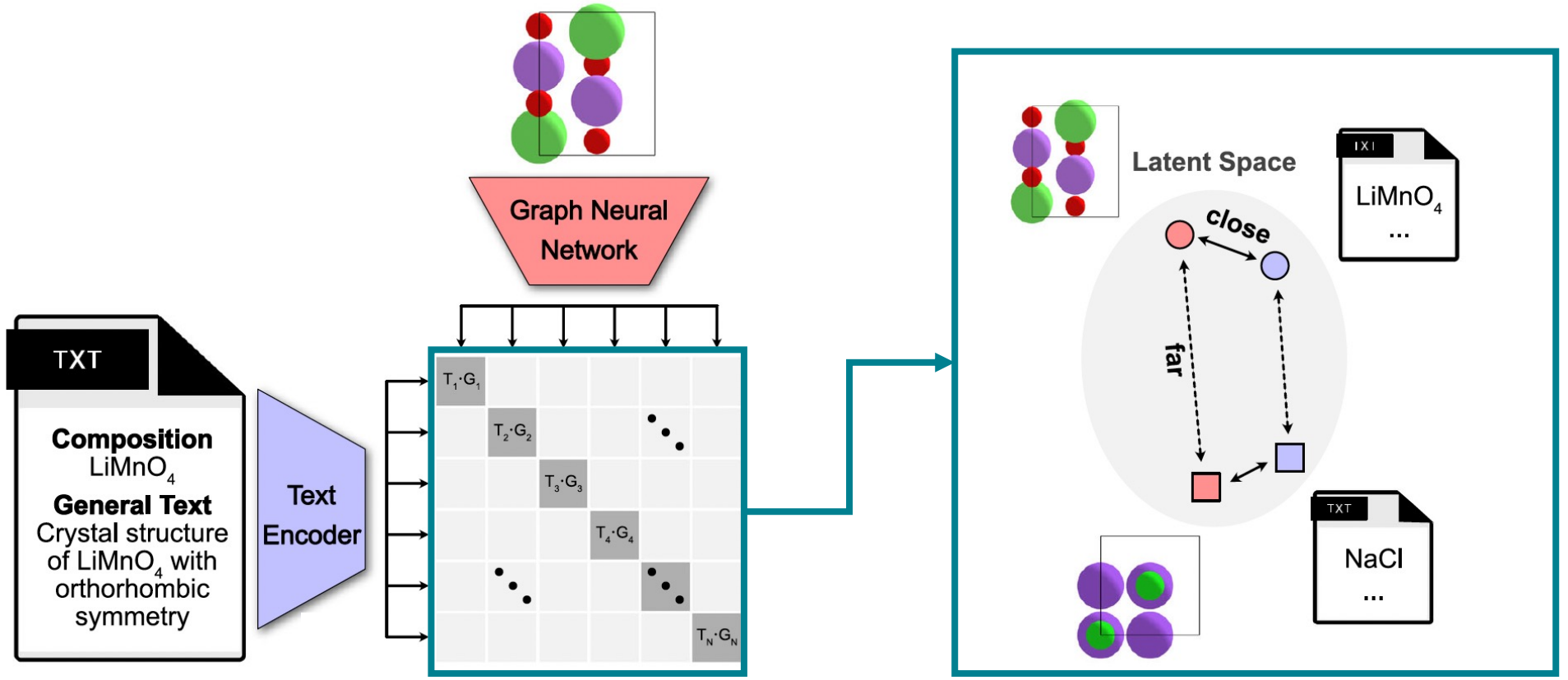
Adaptado de Riebesell et al., “A framework to evaluate machine learning crystal stability predictions”, Nat. Mach. Intell. 7, 836–847 (2025).

2

Design generativo/inverso

GERAR

Chemeleon: transformando um prompt de texto em um cristal, em três etapas



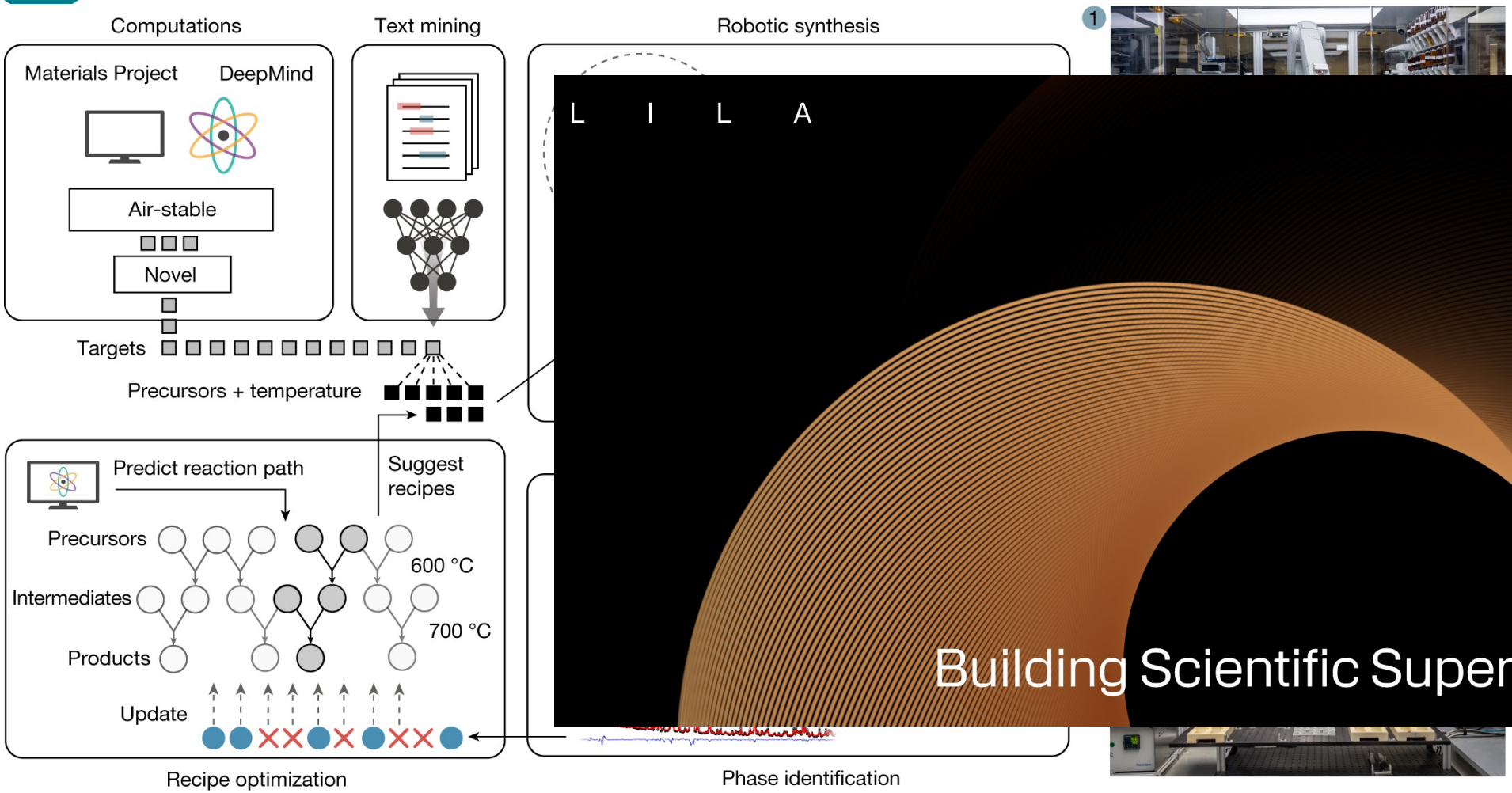
Adaptado de Park, Onwuli & Walsh, Nat. Commun. 16 (2025)

3

Laboratórios autônomos

SINTETIZAR

Um laboratório autônomo: o A-Lab



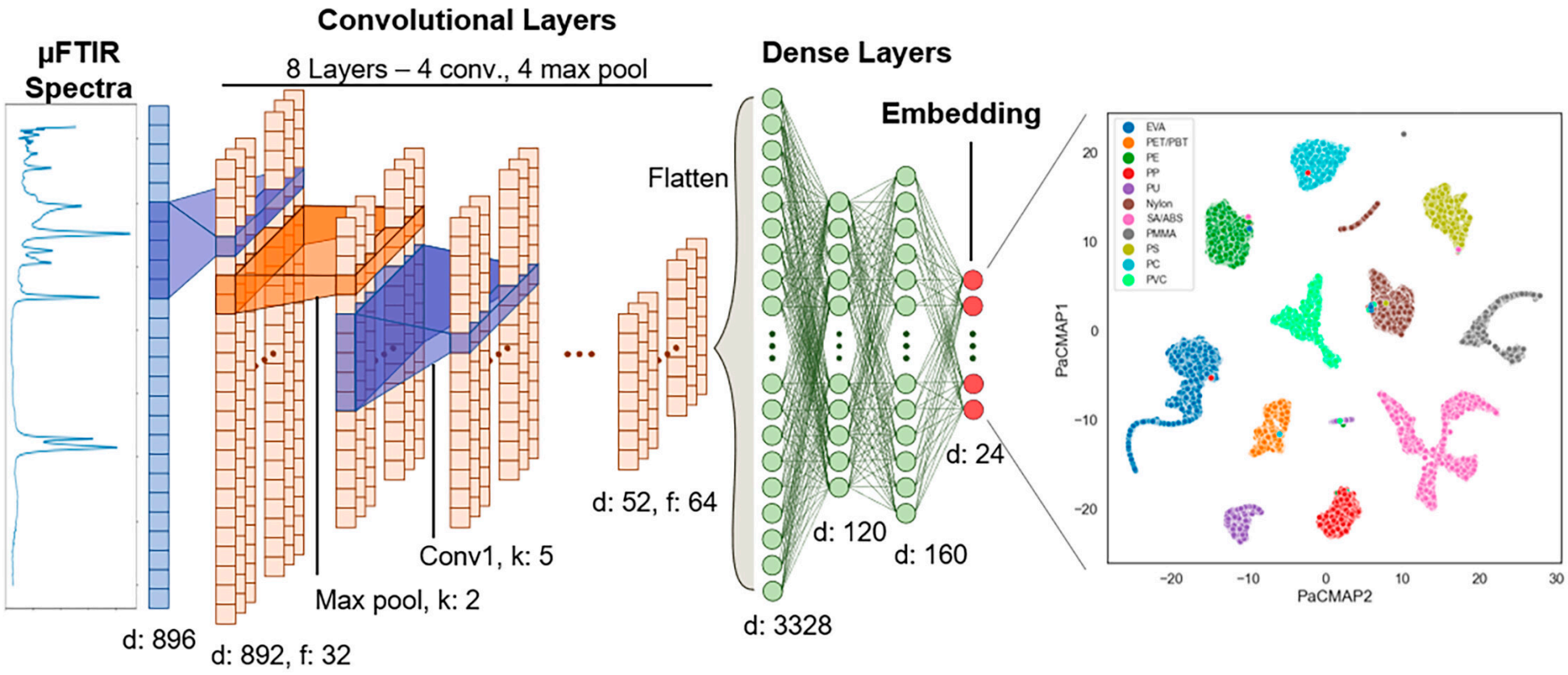
Adaptado de Szymanski et al., Nature 624, 86 (2023)

4

Apoio à caracterização

CARACTERIZAÇÃO

Aprendizado de similaridade em imagens espectrais μFTIR de microplásticos



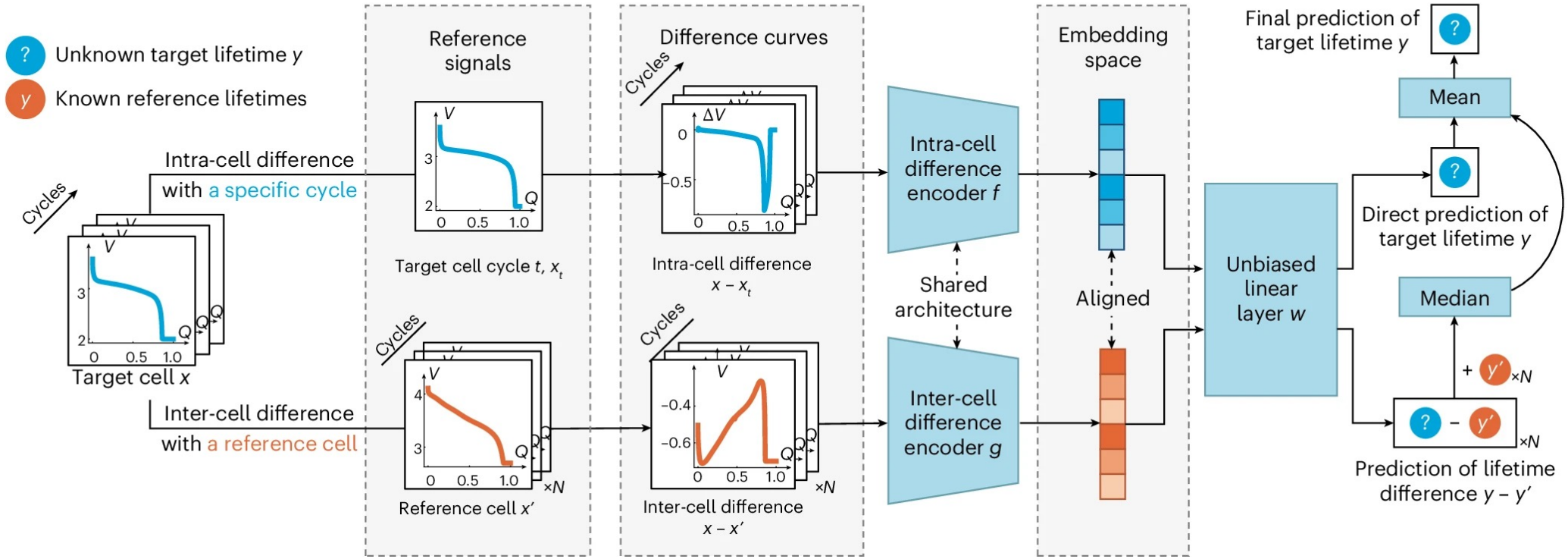
Smolen, Moore, Perez & Wooley, PNAS 122(42), e2509745122 (2025)

5

Gestão de sistemas

IMPLANTAR

Prever cedo a vida útil em ciclos das baterias

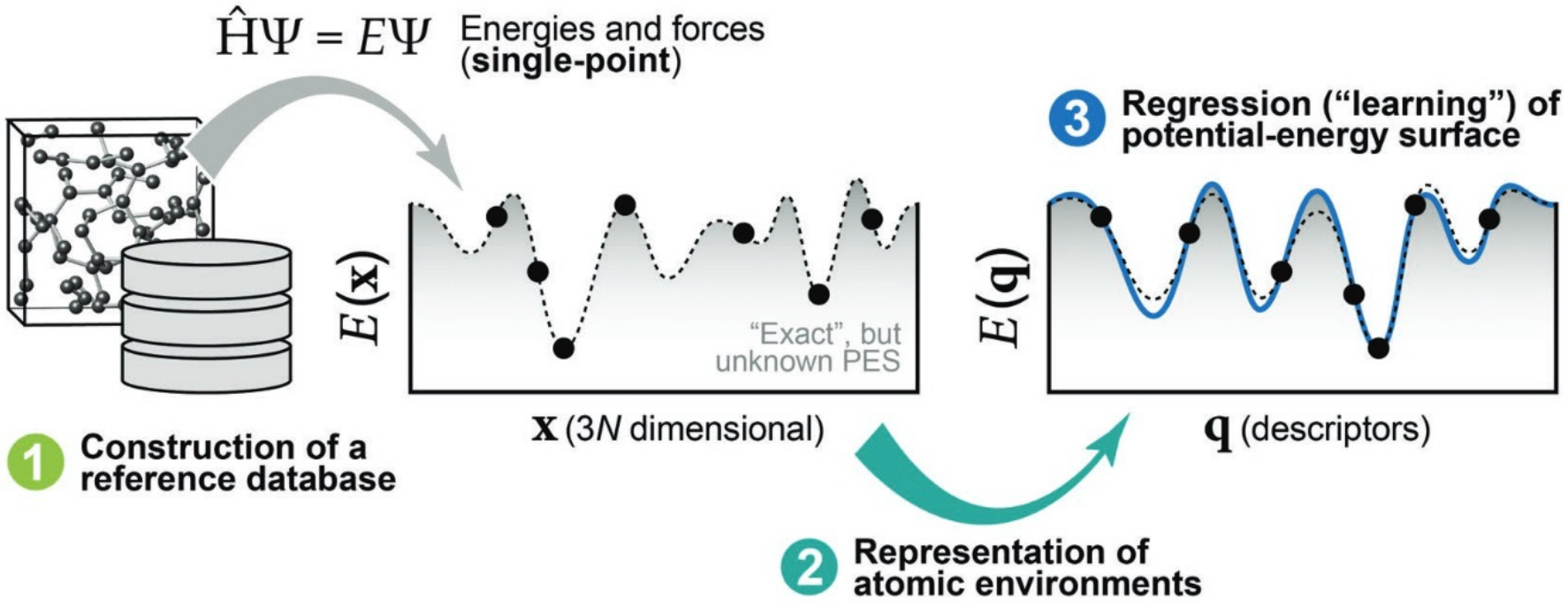


O BatLiNet compara a célula-alvo com uma célula de “referência” totalmente ciclada cuja vida útil já é conhecida

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

Como funciona um potencial interatômico de aprendizado de máquina

Uma receita comum a todos: ambiente atômico local → energia, aprendida a partir da DFT



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

Modelos MLIP em resumo

Modelos representativos por família: ideia-chave, uso típico e fonte

Modelo	Família	Ideia-chave	Aplicação típica	Referência
HDNNP	Descritor + RN	funções de simetria → rede neural	sólidos maciços, água	*Behler & Parrinello 2007*
GAP	Descritor + kernel	SOAP + processo gaussiano	Si, C, amorfos; eficiente em dados	*Bartók et al. 2010*
DeePMD	Descritor aprendido + RN	potencial profundo de ponta a ponta	DM em larga escala (água, ligas)	*Zhang et al. 2018*
SNAP	Expansão linear	componentes do biespectro (linear)	metais (Ta, W)	*Thompson et al. 2015*
MTP	Expansão linear	tensores de momentos	ligas; aprendizado ativo	*Shapeev 2016*
ACE	Expansão por ordem de corpo	expansão em clusters atômicos	geral; rápido + preciso	*Drautz 2019*
SchNet	GNN invariante	convoluções de filtro contínuo	moléculas (MD17, QM9)	*Schütt et al. 2018*
NequIP	GNN equivariante	mensagens E(3)-equivariantes	moléculas + materiais, eficiente em dados	*Batzner et al. 2022*
Allegro	GNN equivariante local	equivariante estritamente local	DM em escala muito grande	*Musaelian et al. 2023*
MACE	Equivariante + alta ordem de corpo	mensagens de ordem superior por camada	DM precisa e rápida	*Batatia et al. 2022*
M3GNet	GNN universal	potencial de grafos treinado no MP	relaxação/triagem de cristais	*Chen & Ong 2022*
CHGNet	Universal (sensível à carga)	GNN informada pela carga	íons de lítio, óxidos, magnetismo	*Deng et al. 2023*
MACE-MP-0	De base (equivariante)	modelo de base para toda a tabela periódica	DM pronta para uso em todas as químicas	*Batatia et al. 2024*

SOAP = sobreposição suave de posições atômicas (smooth overlap of atomic positions) · E(3) = rotações, translações e reflexões em 3D · MP = Materials Project · GNN = rede neural de grafos.

Potenciais interatômicos de aprendizado de máquina, em ordem aproximadamente cronológica

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

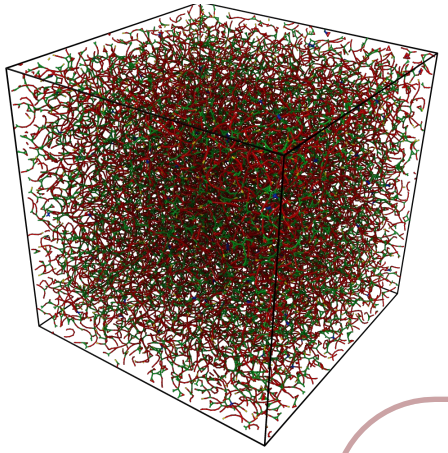
13. Merchant, A. et al. Scaling deep learning for materials discovery (GNoME). Nature 624, 80–85 (2023). 10.1038/s41586-023-06735-9

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

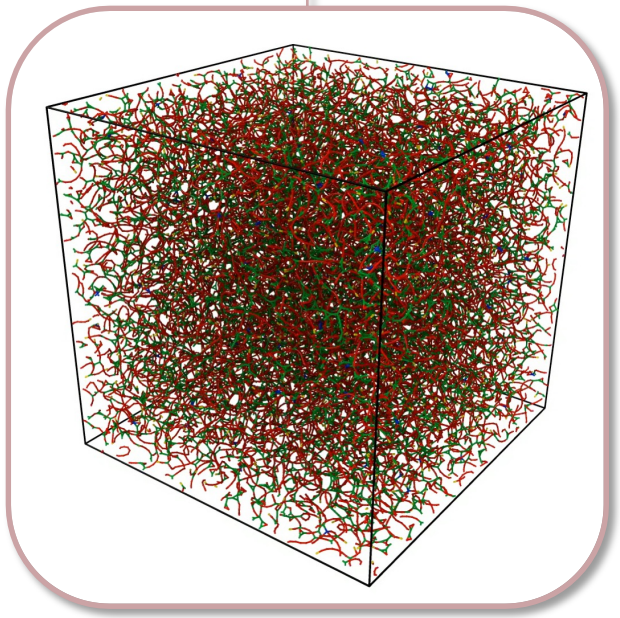
Modelagem do biocarvão com MLIPs (GAP20)

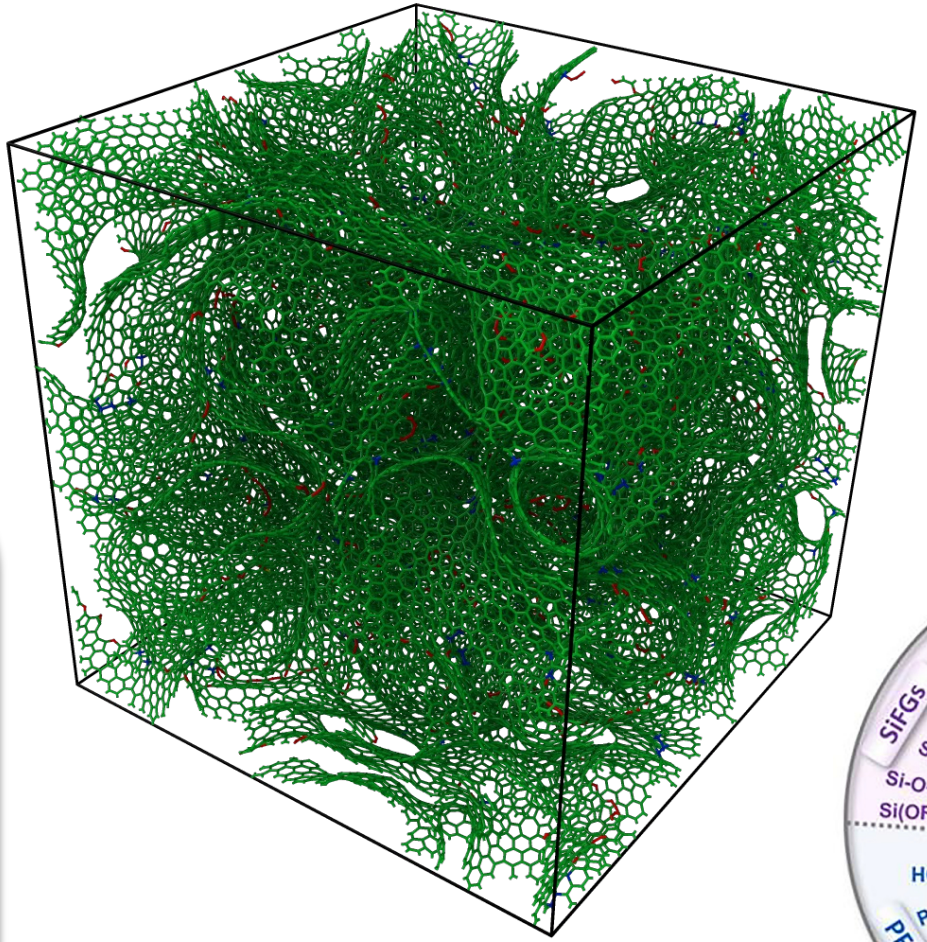
Escala atômica

Caracterização estrutural:
Porosidade, estrutura atômica, hibridização do carbono, ordem estrutural

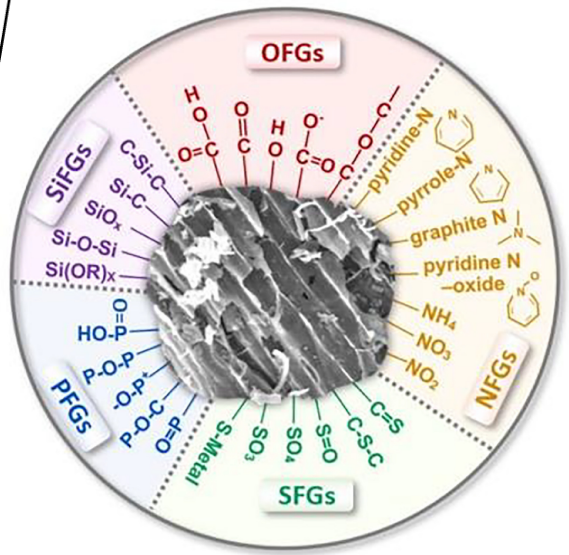


Simulação de DM





Biocarvão Modelo



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

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ

Quiz rápido – Verdadeiro/Falso

1. **Um potencial interatômico de AM busca a precisão da DFT a um custo muito menor.**

   Verdadeiro | Falso

2. **Um modelo com 100% de acurácia no treinamento é sempre o mais confiável com dados novos.**

   Verdadeiro | **Falso**

3. **Candidatos gerados por IA ainda precisam de validação por DFT/experimental.**

   Verdadeiro | Falso

POR QUE IMPORTA PARA O CLIMA

Uma só caixa de ferramentas para todos os materiais de energia limpa.

- Potenciais de AM oferecem precisão próxima à da DFT por uma fração da energia e do tempo: é possível triar milhões de candidatos.
- Laboratórios autônomos transformam rapidamente as melhores previsões em amostras validadas.
- A validação evita desperdiçar o escasso tempo de laboratório com falsos positivos.

Mensagens principais

POR QUE IMPORTA PARA O CLIMA

1. Laboratórios autônomos transformam rapidamente as melhores previsões em amostras validadas.

2. Os MLIPs oferecem a precisão da DFT na velocidade da DM: a caixa “acelerar a simulação” virou realidade.

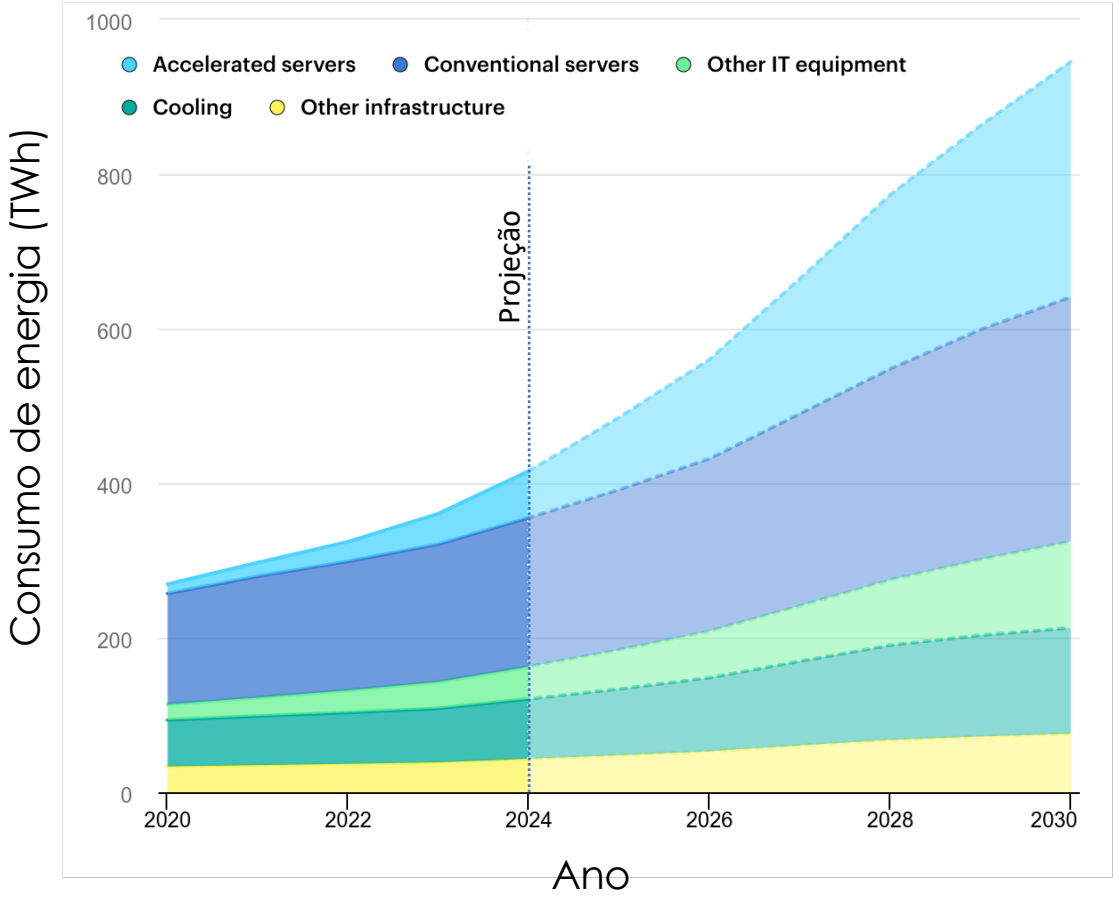
3. Gerar é fácil; validar é difícil

PERGUNTAS?

FALANDO DE QUÍMICA COM OS COMPUTADORES

Custo energético crescente da IA e dos data centers

Por uma computação mais verde



Fonte: Agência Internacional de Energia (AIE)

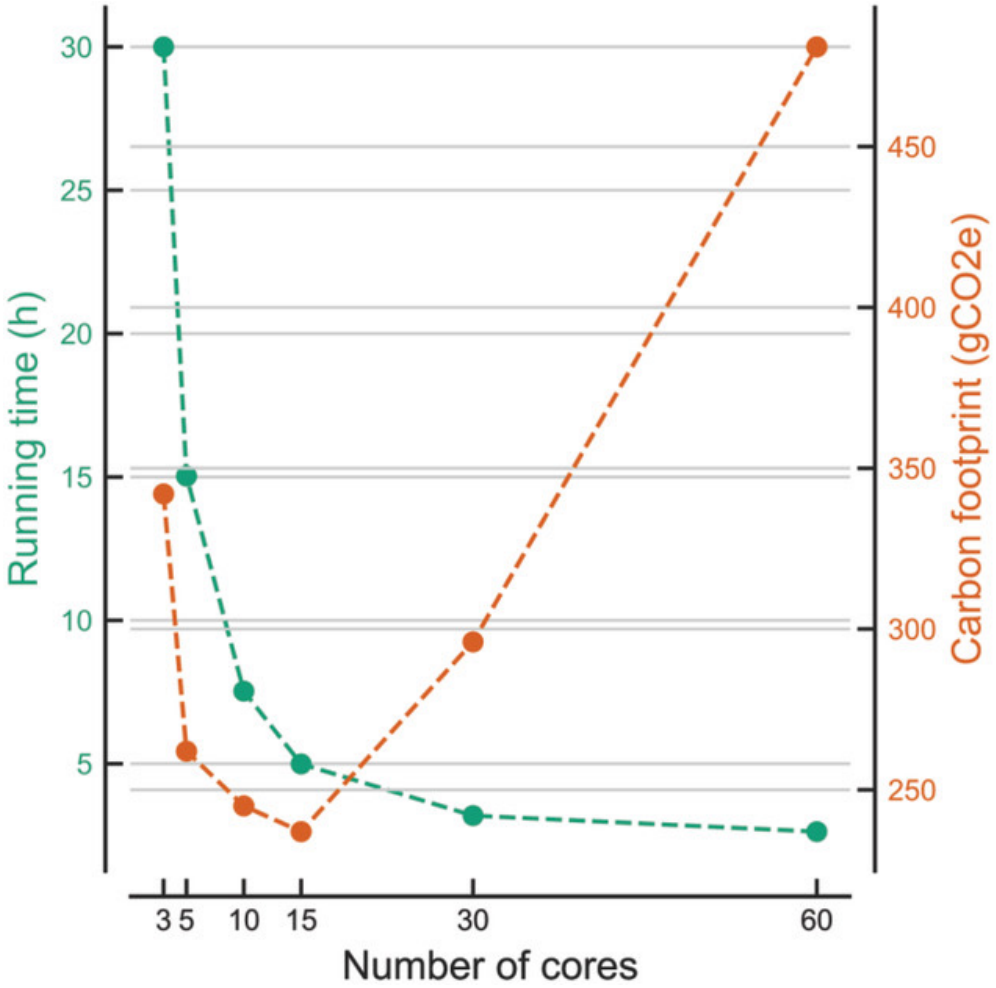
Onde está a supercomputação?



Melhorar o acesso à supercomputação em dispositivos ou contextos com recursos limitados: **democratização**

Fonte: top500.org

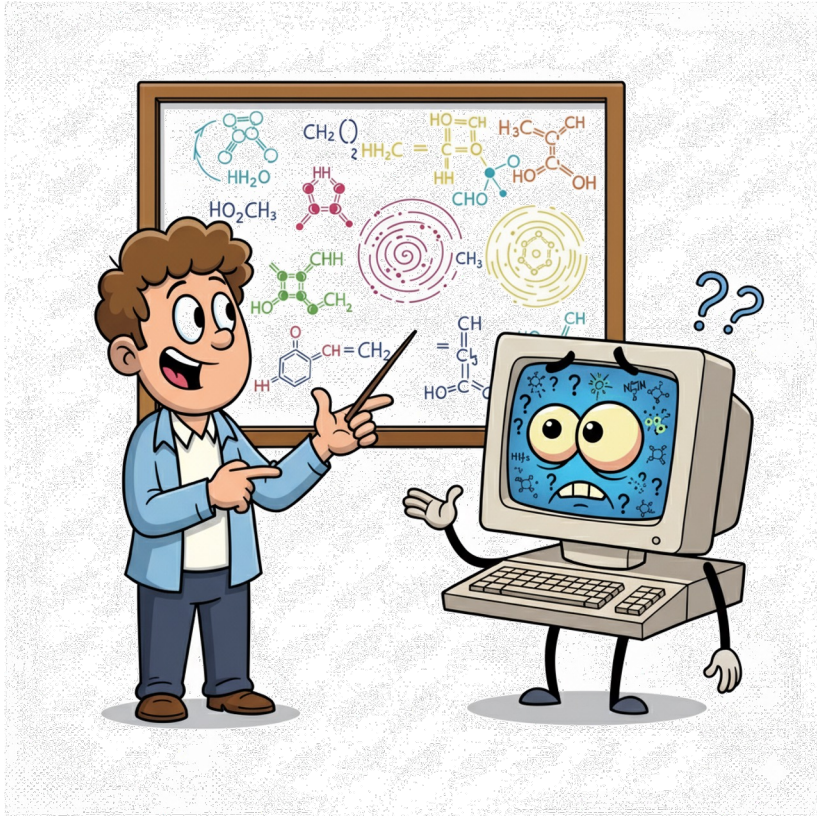
Algoritmos verdes





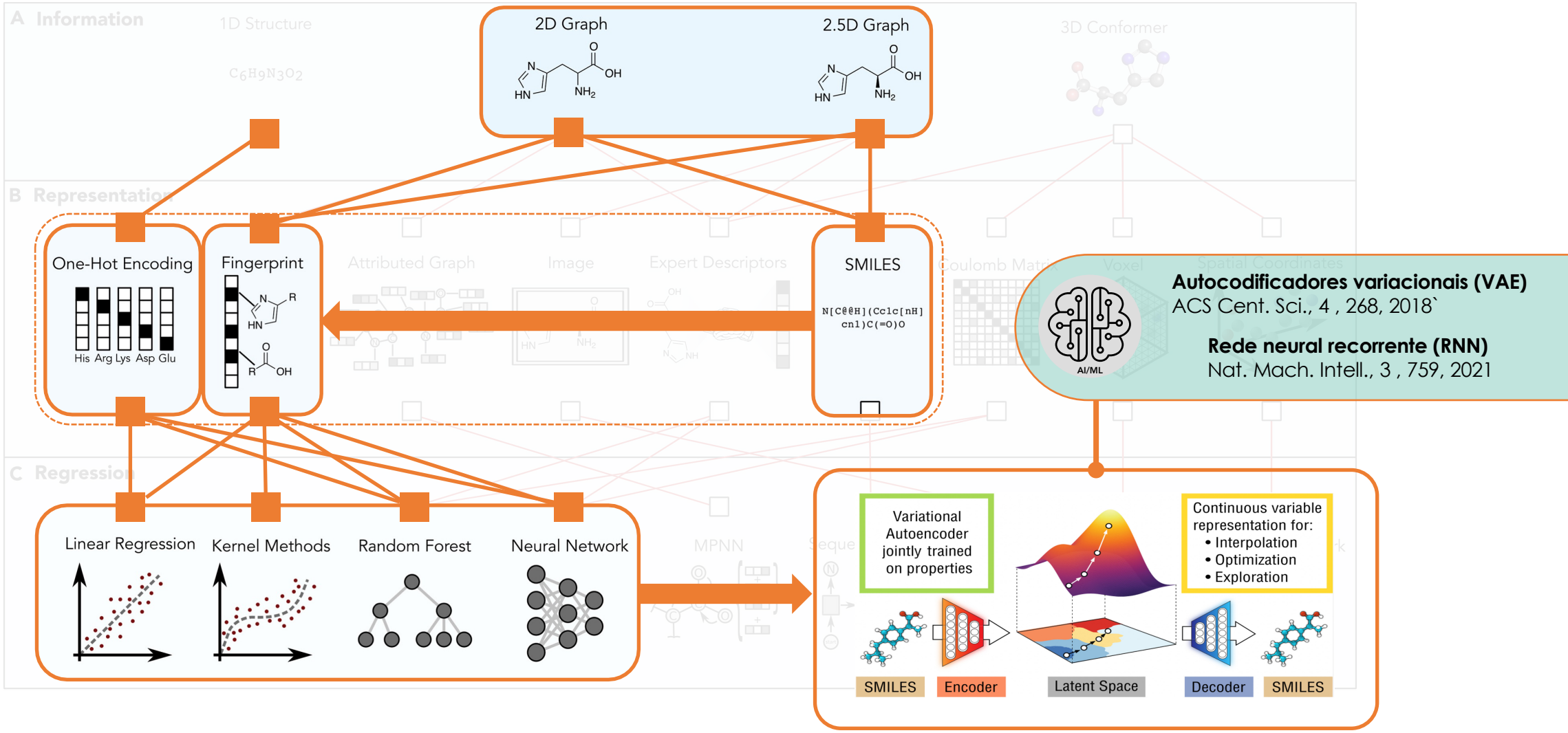
Lannelongue et al., Advanced Science, 8, 2021

Um passo atrás: como falamos de química com os computadores?



O aprendizado de máquina para química exige **vetorizar as moléculas**

O AM para química exige vetorizar as moléculas



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

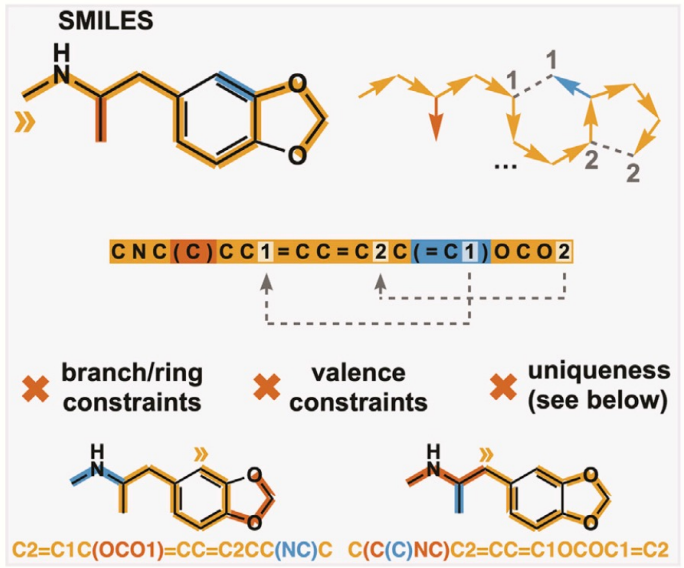
Rede neural recorrente (RNN)
Nat. Mach. Intell., 3 , 759, 2021

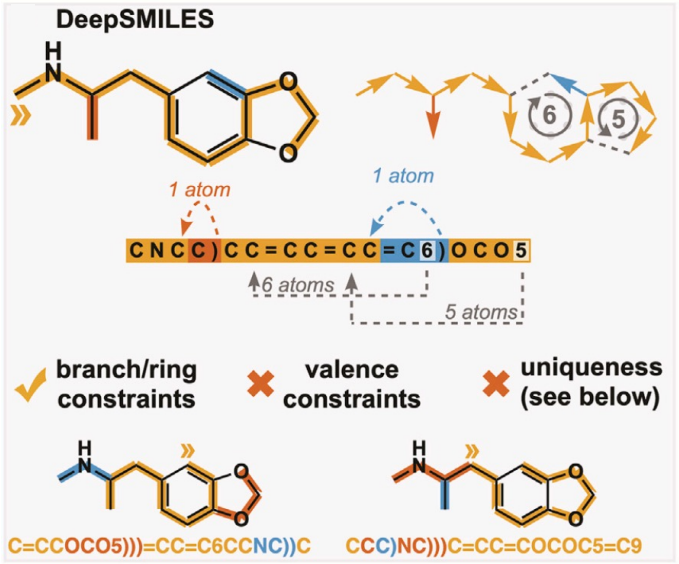
Pattanaik and Coley, **Chem 6, 1204-1207, 2020**

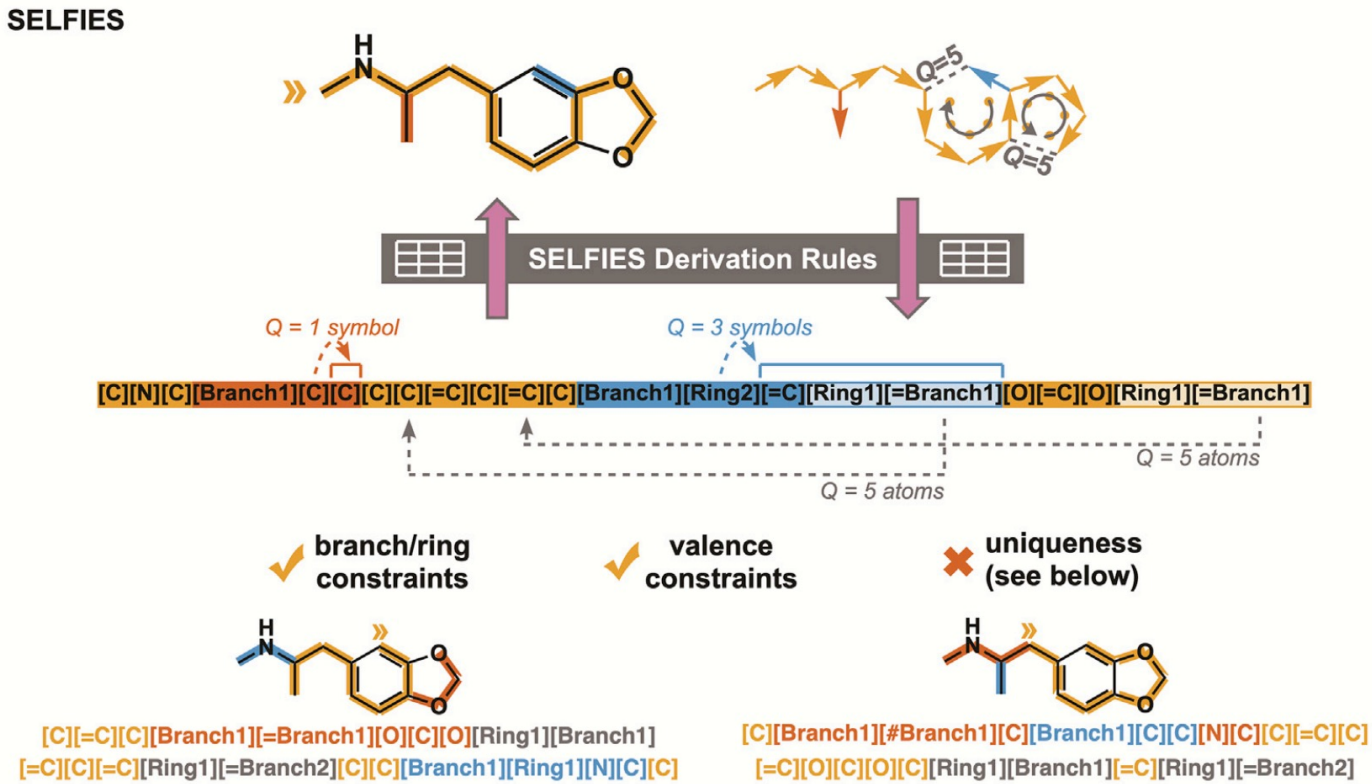
Skinnider et al. **Nat. Mach. Intell.**, 3 , 759, **2021**

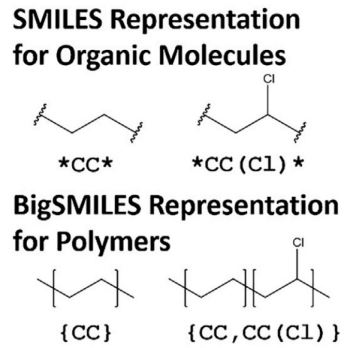
Gomez-Bombarelli et al. **ACS Cent Sci.** 4, 268–76, **2018**

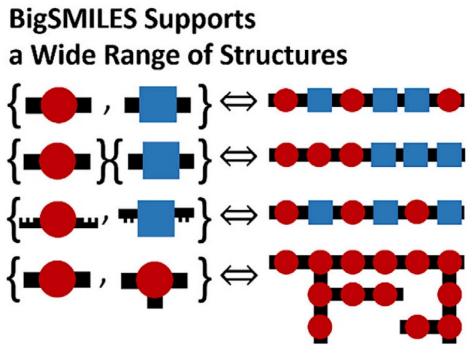
Codificação de dados legível por máquina







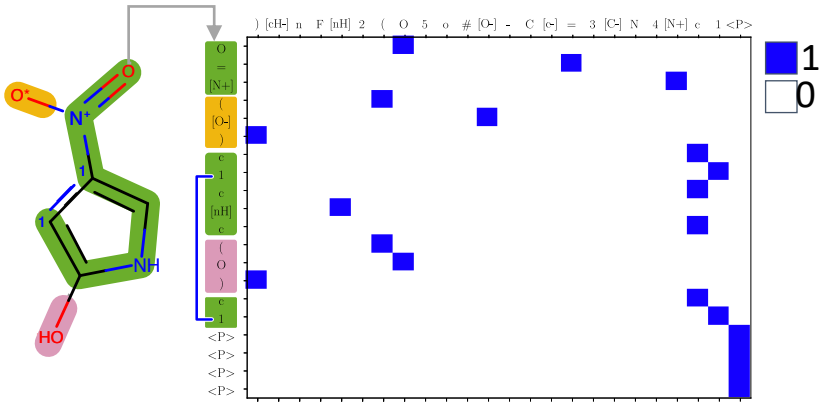




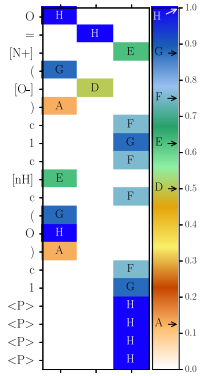
Krenn et al. Patterns 3, 10, 100588, 2022

Codificações one-hot incorporadas (eOHE)

Codificação one-hot (OHE)



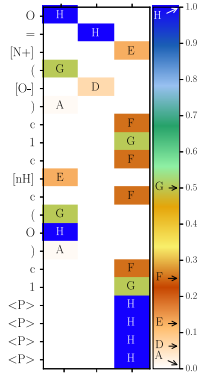
eOHE v1



Redução linear
$q = 8$

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

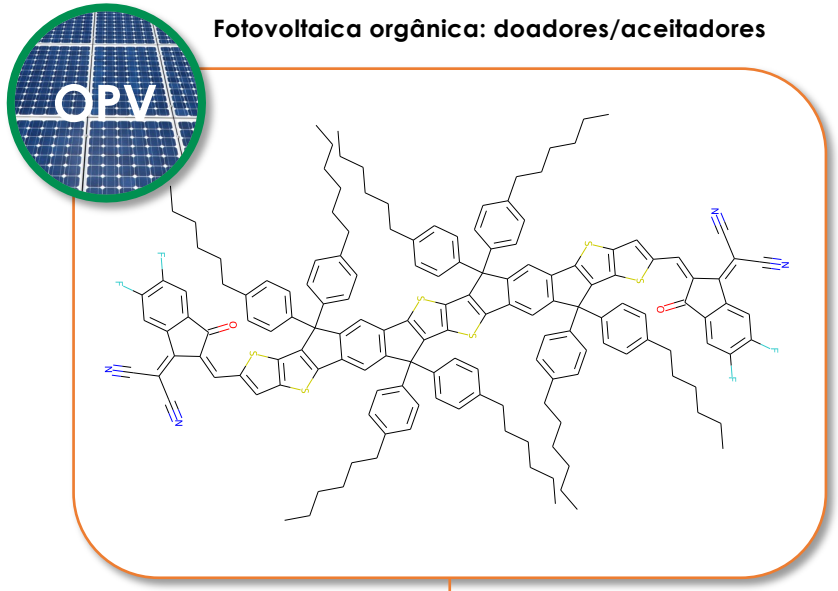
eOHE v2

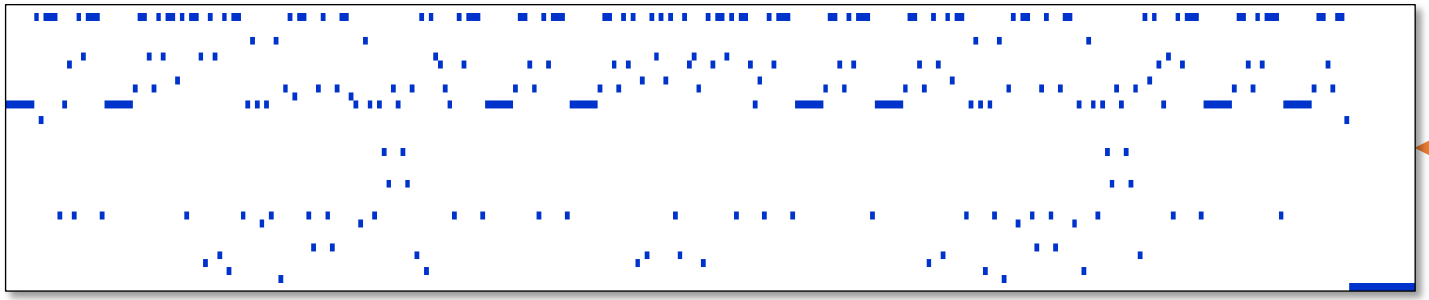


Redução por potências de 2
$q = 8$

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

Fotovoltaica orgânica: doadores/aceitadores

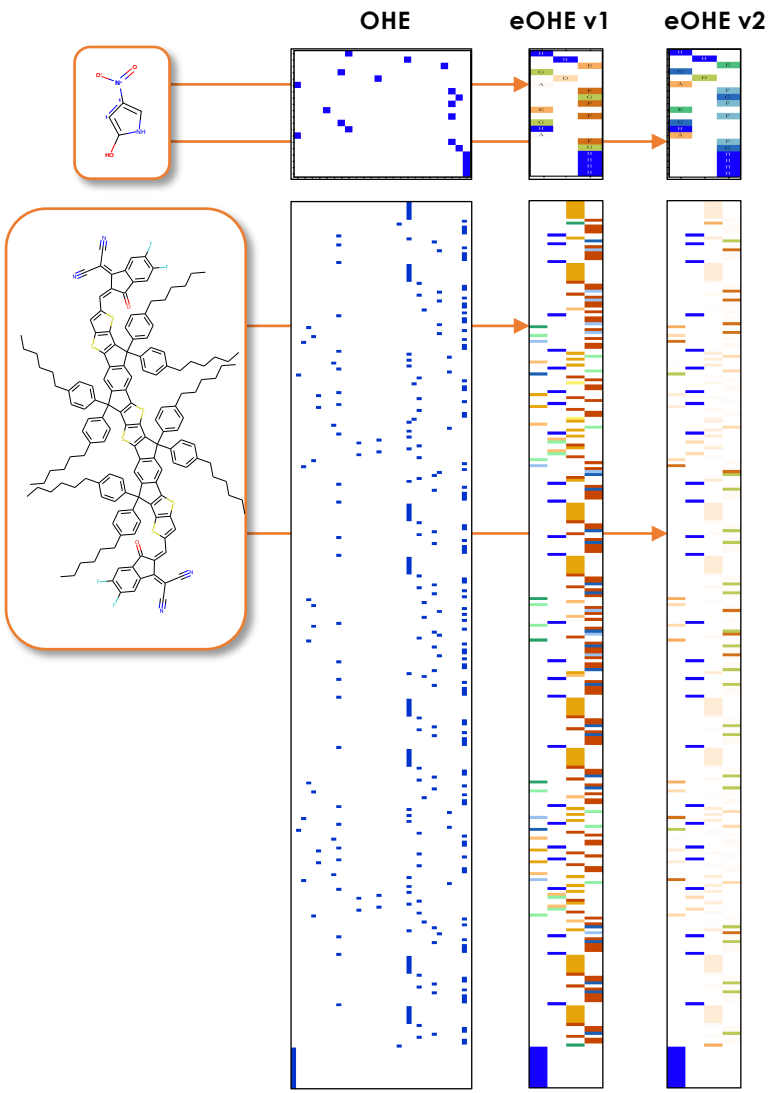


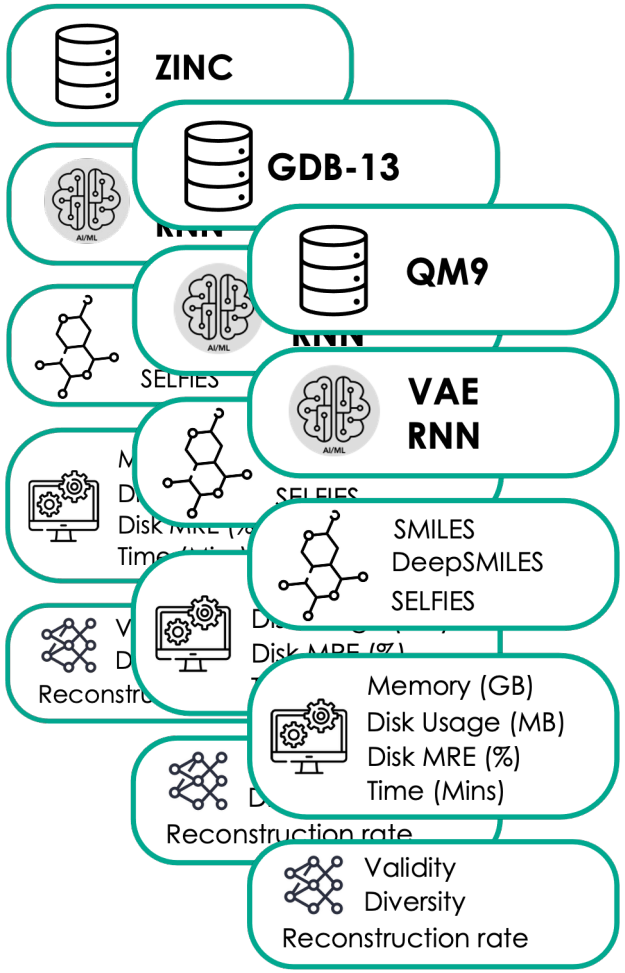


CCCCCCc1ccc(C2(c3ccc(CCCCCC)cc3)c3cc4c(cc3-c3sc5cc(C=C6C(=O)c7cc(F)c(F)cc7C6=C(C#N)C#N)sc5c32)C(c2ccc(CCCCCC)cc2)(c2ccc(CCCCCC)cc2)c2c-4sc3c4c(sc23)-c2cc3c(cc2C4(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)-c2sc4cc(C=C5C(=O)c6cc(F)c(F)cc6C5=C(C#N)C#N)sc4c2C3(c2ccc(CCCCCC) cc2)c2ccc(CCCCCC)cc2)cc1

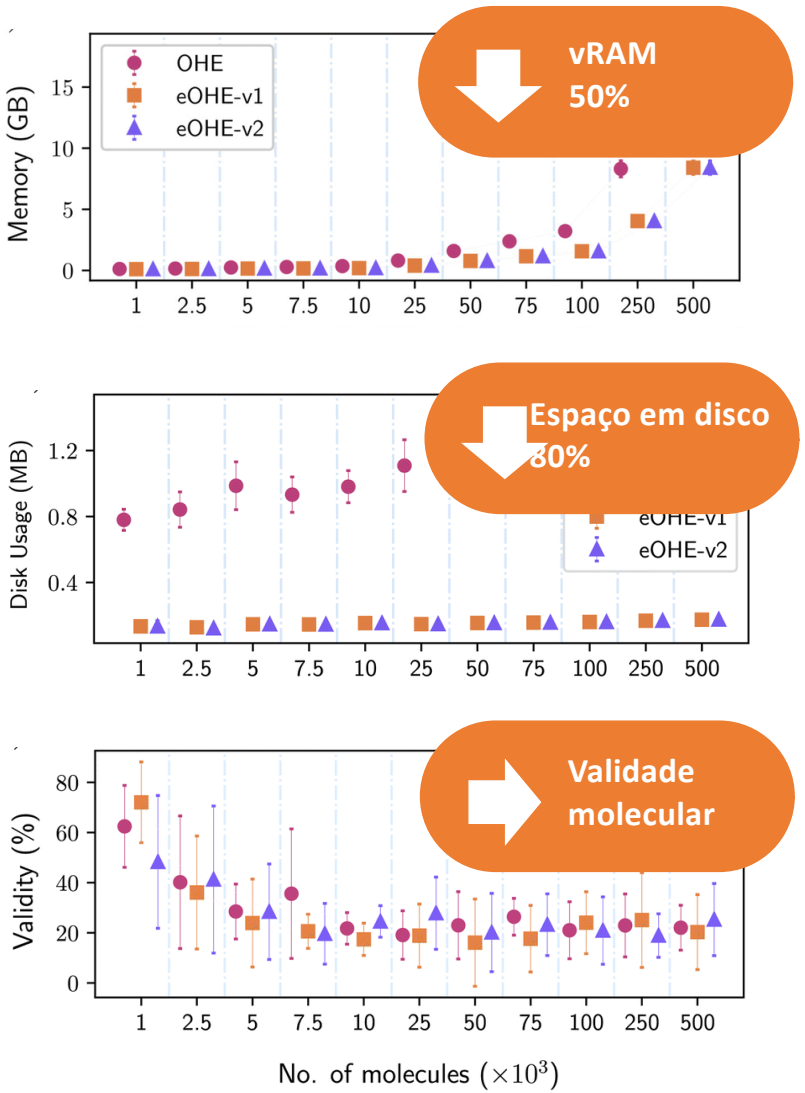
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez,. **Digital Discovery**, 4, 776-789, **2025**

Codificação legível por máquina mais eficiente





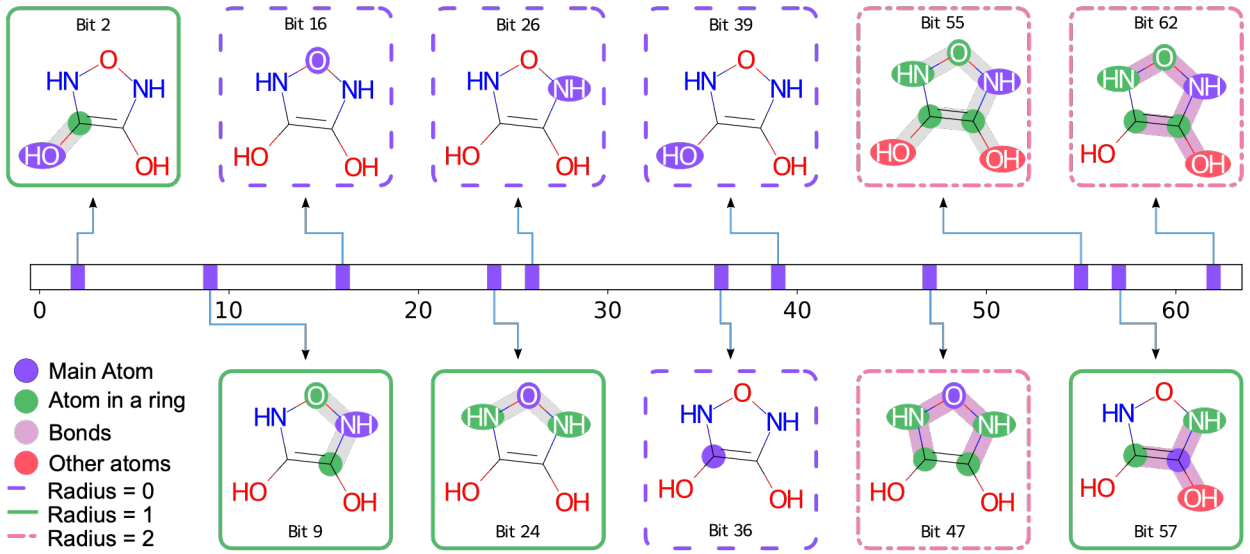
Desempenho em autocodificadores variacionais



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

Impressões digitais de Morgan incorporadas (eMP)

Uma impressão digital de Morgan é um vetor binário de tamanho fixo que codifica a presença (1) ou ausência (0) de subestruturas em uma molécula.



QM9: mais de 134.000 moléculas
NFA: 50.000 moléculas
RedDB: 15.000 moléculas

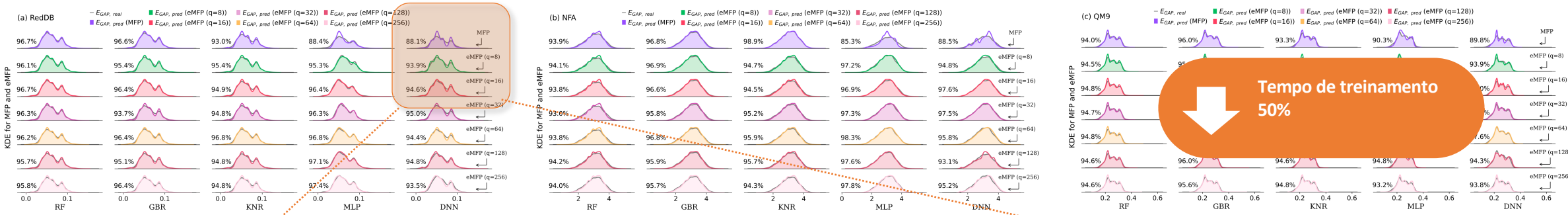
Floresta aleatória (RF), Perceptron multicamadas (MLP), Regressor KNN (KNR), regressor de gradient boosting (GBR), rede neural profunda (DNN)

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

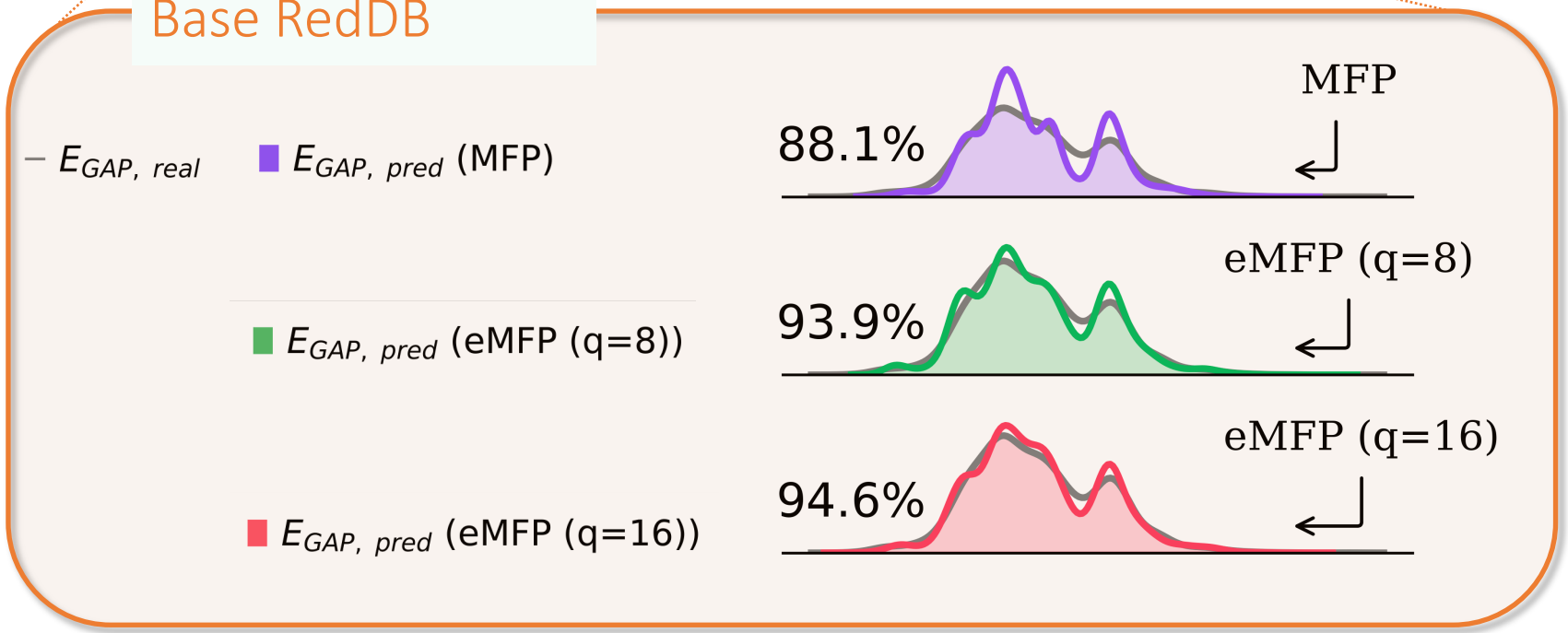
150 ensaios para otimizar 15.750 hiperparâmetros

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

Gap de energia com impressões digitais de Morgan incorporadas (eMP)



Base RedDB



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

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES

B) SELFIES

C) InChI

D) Nenhuma delas

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES

B) SELFIES

C) InChI

D) Nenhuma delas

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES

B) SELFIES

C) InChI

D) Nenhuma delas

Q2 **eOHE/eMFP melhoram principalmente…**

A) Apenas a acurácia

B) Memória/disco/tempo com a mesma acurácia

C) O número de publicações

D) Nada

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES

B) SELFIES

C) InChI

D) Nenhuma delas

Q2 **eOHE/eMFP melhoram principalmente…**

A) Apenas a acurácia

B) Memória/disco/tempo com a mesma acurácia

C) O número de publicações

D) Nada

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES
B) SELFIES
C) InChI
D) Nenhuma delas

Q2 **eOHE/eMFP melhoram principalmente…**

A) Apenas a acurácia
B) Memória/disco/tempo com a mesma acurácia
C) O número de publicações
D) Nada

Q3 **Por que o custo energético da IA importa?**

A) Não importa
B) Sustentabilidade e democratização
C) A computação é gratuita
D) Só para jogos

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Qual representação garante uma molécula válida para cada string?**

A) SMILES
B) SELFIES
C) InChI
D) Nenhuma delas

Q2 **eOHE/eMFP melhoram principalmente…**

A) Apenas a acurácia
B) Memória/disco/tempo com a mesma acurácia
C) O número de publicações
D) Nada

Q3 **Por que o custo energético da IA importa?**

A) Não importa
B) Sustentabilidade e democratização
C) A computação é gratuita
D) Só para jogos

FALANDO DE QUÍMICA

Mensagens principais – A representação é uma decisão de sustentabilidade.

POR QUE IMPORTA PARA O CLIMA

1. A vetorização é uma decisão de design de primeira linha, não uma etapa secundária de pré-processamento.

2. eOHE/eMFP reduzem memória, disco e tempo com a mesma acurácia eficiência = computação mais verde e mais democrática.

3. Uma computação verde e acessível faz parte de uma aceleração responsável.

PERGUNTAS?

DESENVOLVIMENTO DE MATERIAIS GUIADO POR LLMs

Desenvolvimento de materiais guiado por LLMs: uma nova camada na pilha

Além de prever números: modelos que leem, raciocinam, planejam e usam ferramentas

A mudança: o AM especializado mapeia uma estrutura para uma propriedade. Um grande modelo de linguagem acrescenta uma interface em linguagem natural, amplo conhecimento prévio, planejamento passo a passo e a capacidade de chamar ferramentas externas, **transformando um preditor em um colaborador de pesquisa**.

1

Interface em linguagem natural

Pergunte, adapte com alguns exemplos; forte quando há poucos dados, onde o AM clássico tem dificuldade.

2

Conhecimento de mundo

Pré-treinado na literatura científica: traz conhecimento químico prévio e contexto, não apenas ajuste de curvas.

3

Planejamento e raciocínio

Decompor um objetivo ("encontrar um catodo de íons de sódio estável") em etapas ordenadas e verificáveis.

4

Uso de ferramentas

Chamar simuladores, bancos de dados e robôs, fundamentando as respostas em cálculos reais em vez de adivinhar.

…mas cuidado. LLMs são fluentes, mas pouco confiáveis. Inventam fatos, citam errado e podem errar com total confiança. Tudo depende de **ancorá-los em ferramentas e validá-los**

O que os LLMs realmente fazem no desenvolvimento de materiais

1 Ler a literatura

Transformar milhares de artigos em bancos de dados estruturados e consultáveis (composições, MOFs, dopantes).

p. ex. **Dagdelen et al. 2024**

2 Prever propriedades

Responder a perguntas de química em linguagem natural; competitivo com o AM, forte quando há poucos dados.

p. ex. **Jablonka et al. 2024**

3 Gerar e fazer design inverso

Escrever moléculas e cristais como texto; geração orientada por propriedades.

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

4 Planejar a síntese

Propor rotas sintéticas e receitas; retrossíntese e design de procedimentos.

p. ex. **ChemCrow 2024**

5 Raciocinar com ferramentas

Dar ao modelo ferramentas especializadas (RDKit, calculadoras, busca) para que ele pare de adivinhar.

p. ex. **ChemCrow – 18 ferramentas**

6 Realizar experimentos de forma autônoma

Ciclos de planejar-agir-observar que controlam instrumentos e robôs para testar hipóteses.

p. ex. **Coscientist 2023**

Sistemas LLM de referência para química e materiais

Sistemas representativos: o que fazem, a abordagem e a fonte

Sistema	O que faz	LLM/abordagem	Referência
Coscientist	projeta, planeja e executa experimentos (p. ex., acoplamentos cruzados com Pd)	GPT-4 + web/documentos/código + robótica	*Boiko et al., Nature 2023*
ChemCrow	síntese orgânica, design de fármacos e materiais; realizou sínteses reais	GPT-4 + 18 ferramentas especializadas	*Bran et al., Nat. Mach. Intell. 2024*
LLM de química preditiva	previsão e classificação de propriedades a partir de texto; poucos dados	GPT-3 ajustado; em contexto	*Jablonka et al., Nat. Mach. Intell. 2024*
Extração estruturada	lê artigos → registros estruturados (dopantes, MOFs, composições)	GPT-3/Llama-2 ajustados	*Dagdelen et al., Nat. Commun. 2024*
LLM cristal-texto	gera cristais inorgânicos estáveis escritos como texto	LLaMA-2 70B ajustado	*Gruver et al., ICLR 2024*
Chemeleon	geração de estruturas cristalinas guiada por texto	contrastivo texto–cristal + difusão	*Park et al., Nat. Commun. 2025*

Referências

LLMs e agentes autônomos para química e materiais

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

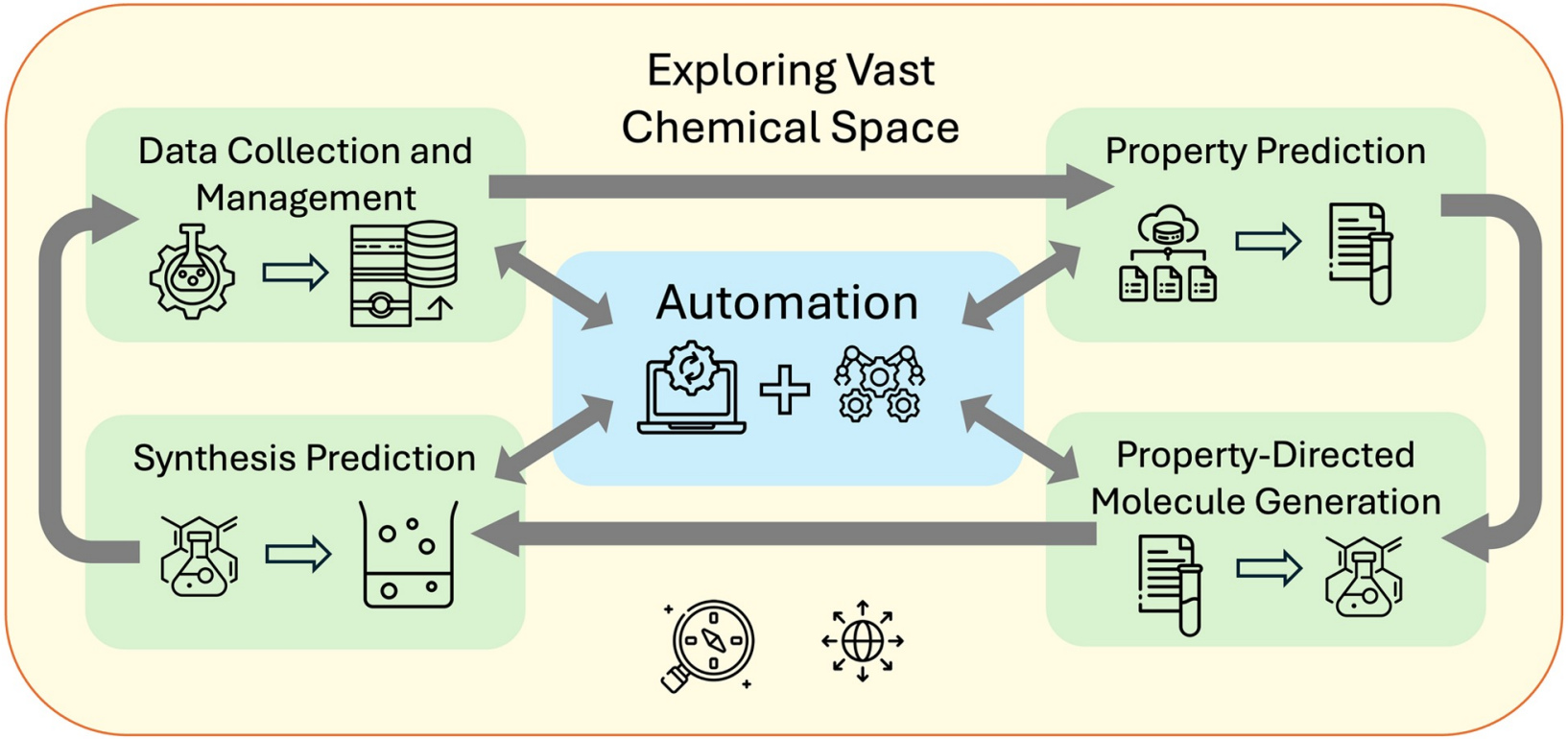
5. Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-Tuned Language Models Generate Stable Inorganic Materials as Text. ICLR 2024. arXiv:2402.04379

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

O que LLMs e agentes realmente podem fazer em química?



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

Modelagem do biocarvão – engenharia do processo de pirólise

Macroescala

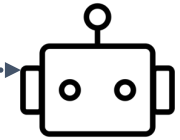
Cantera

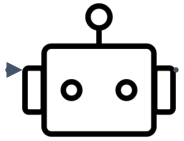
PARÂMETROS

- Temperatura
- Tempo de residência
- Gás de arraste
- Taxa de aquecimento

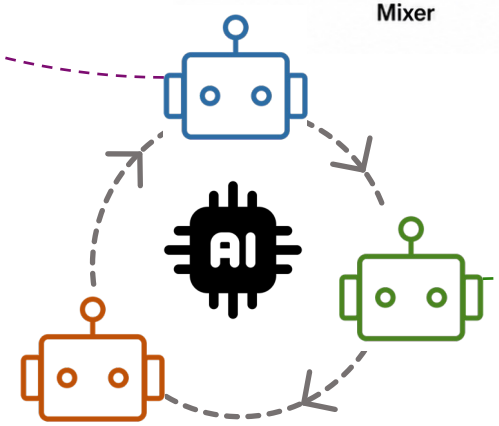


CONTEXTO

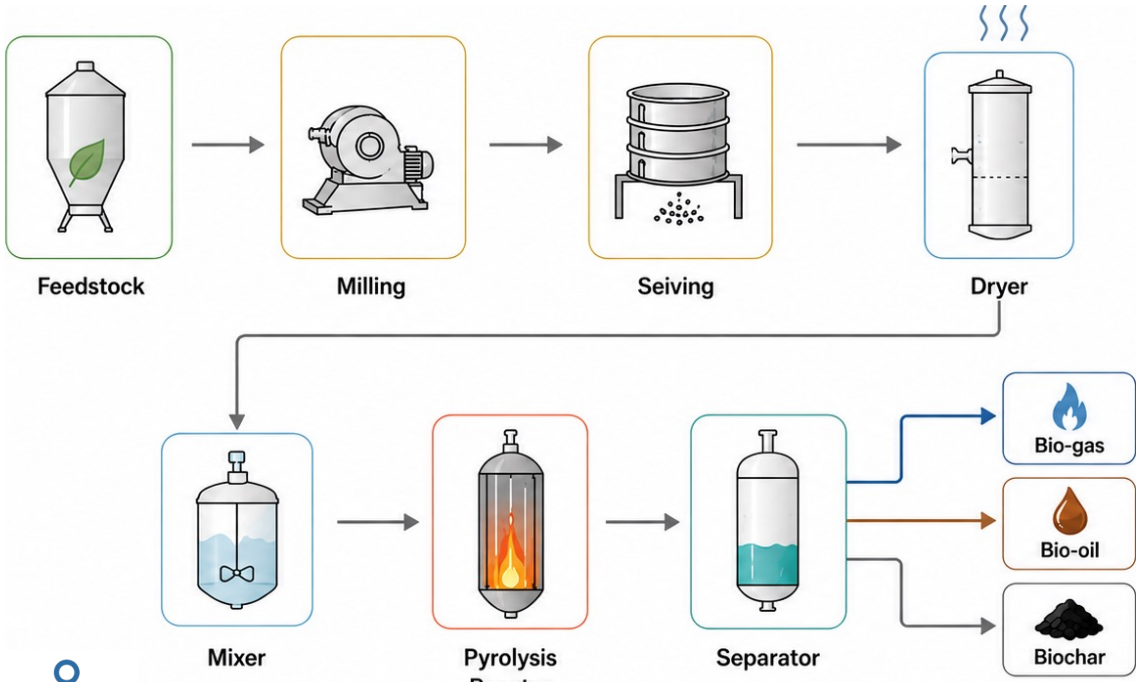




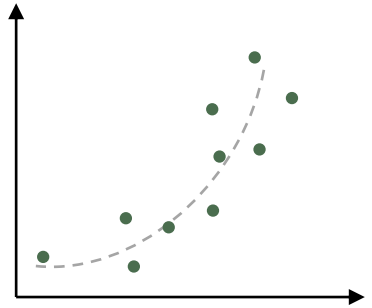
PARÂMETROS DO PROCESSO



Multiagente



Rendimento de biocarvão



Condições

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas

B) Planeja e age com ferramentas, em ciclo

C) Nunca usa ferramentas

D) Só funciona off-line

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas

B) Planeja e age com ferramentas, em ciclo

C) Nunca usa ferramentas

D) Só funciona off-line

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas

B) Planeja e age com ferramentas, em ciclo

C) Nunca usa ferramentas

D) Só funciona off-line

Q2 **Dar ferramentas (RDKit, busca, simuladores) a um LLM…**

A) Piora o modelo

B) Ancora o modelo e melhora a confiabilidade

C) Elimina todos os erros

D) É impossível

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas

B) Planeja e age com ferramentas, em ciclo

C) Nunca usa ferramentas

D) Só funciona off-line

Q2 **Dar ferramentas (RDKit, busca, simuladores) a um LLM…**

A) Piora o modelo

B) Ancora o modelo e melhora a confiabilidade

C) Elimina todos os erros

D) É impossível

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas
B) Planeja e age com ferramentas, em ciclo
C) Nunca usa ferramentas
D) Só funciona off-line

Q2 **Dar ferramentas (RDKit, busca, simuladores) a um LLM…**

A) Piora o modelo
B) Ancora o modelo e melhora a confiabilidade
C) Elimina todos os erros
D) É impossível

Q3 **LLMs são especialmente úteis quando…**

A) Há dados em enorme quantidade
B) Há poucos dados
C) Nunca
D) Só para imagens

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (múltipla escolha)

Múltipla escolha

Q1 **Um “agente” LLM (versus assistente) é aquele que…**

A) Só responde a perguntas
B) Planeja e age com ferramentas, em ciclo
C) Nunca usa ferramentas
D) Só funciona off-line

Q2 **Dar ferramentas (RDKit, busca, simuladores) a um LLM…**

A) Piora o modelo
B) Ancora o modelo e melhora a confiabilidade
C) Elimina todos os erros
D) É impossível

Q3 **LLMs são especialmente úteis quando…**

A) Há dados em enorme quantidade
B) Há poucos dados
C) Nunca
D) Só para imagens

Mensagens principais – Liberar o conhecimento preso nos textos.

POR QUE IMPORTA PARA O CLIMA

1. LLMs acrescentam uma camada de linguagem + raciocínio + uso de ferramentas ao AM especializado.

2. Transformar décadas de literatura em dados utilizáveis; planejar sínteses; orquestrar ferramentas.

3. Manter um humano responsável e registrar cada ação. Rastreabilidade e supervisão tornam a autonomia confiável (e reprodutível).

PERGUNTAS?

IA RESPONSÁVEL



IA responsável para materiais e química

Ferramentas poderosas trazem responsabilidades. A mesma aceleração corta dos dois lados

1 Integridade e reprodutibilidade

Vazamento de dados e a corrida por rankings inflam os resultados.
Documente o suficiente para que outros possam reproduzi-los.

2 Previsões confiáveis

Informe a incerteza, respeite o domínio de aplicabilidade e valide depois (DFT → experimento).

3 Qualidade dos dados e vieses

Os conjuntos de dados pendem para sistemas bem estudados e bem-sucedidos.
Lixo entra, lixo sai

4 Abertura e FAIR

Compartilhe dados, código e pesos dos modelos
Localizáveis, acessíveis, interoperáveis, reutilizáveis (FAIR).

5 Segurança e duplo uso

Ferramentas de design poderosas podem ser invertidas para causar danos.
Controle de acesso e supervisão humana.

6 Sustentabilidade e computação

Treinamento e inferência consomem energia.
Dimensione bem os modelos e reutilize modelos pré-treinados.

Vieses dos dados

Desafios

- Crenças implícitas e vieses de contextos ocidentais podem ser normalizados na IA, tornando-a menos responsável em escala global.
- É preciso adotar uma perspectiva global nos pressupostos centrais da IA: dados de treinamento, desempenho dos modelos e **explicabilidade** dos sistemas
- Fazer cedo as perguntas difíceis sobre impactos positivos e negativos, em vez de recorrer a reparos e correções a posteriori.

Ações

- Ser robusto ao multilinguismo e à diversidade de letramentos
- Personalizar experiências com o modelo de identidade compartilhada
- **A explicabilidade** e a transparência devem fazer sentido para novos usuários de tecnologia
- Tratar de forma proativa a equidade diante dos desequilíbrios dos dados

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

Referências

IA responsável e confiável para a ciência

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).

PARTICIPAÇÃO

FECHAMENTO DA SEÇÃO · QUIZ (verdadeiro/falso)

Verdadeiro ou falso?

1. **Os conjuntos de dados de química publicados são enviesados, em parte porque experimentos fracassados raramente são relatados.**

   Verdadeiro | Falso

2. **Um modelo de design de fármacos pode ser invertido para projetar toxinas.**

   Verdadeiro | Falso

3. **FAIR significa Findable, Accessible, Interoperable, Reusable (localizáveis, acessíveis, interoperáveis, reutilizáveis).**

   Verdadeiro | Falso

Mensagens principais – Acelerar só ajuda se for confiável e compartilhado.

POR QUE IMPORTA PARA O CLIMA

1. Os dados são enviesados: viés de publicação, viés humano, pressupostos do contexto ocidental.

2. Ferramentas generativas/agênticas poderosas são intrinsecamente de duplo uso.

3. Reprodutibilidade, abertura (FAIR) e custo computacional fazem parte de fazer isso bem.

PERGUNTAS?

IA EXPLICÁVEL

IA explicável: da previsão à compreensão

Na ciência, geralmente queremos o "porquê", não só o número

A lacuna: um modelo pode prever uma propriedade com precisão e ainda assim não nos ensinar nada.
A explicabilidade pergunta o que o modelo aprendeu, se aprendeu pelos motivos certos e que nova ciência ele sugere.

1 Compreensão

Transformar um modelo ajustado em regras de design e mecanismos. Recuperar a física/química por trás dos dados.

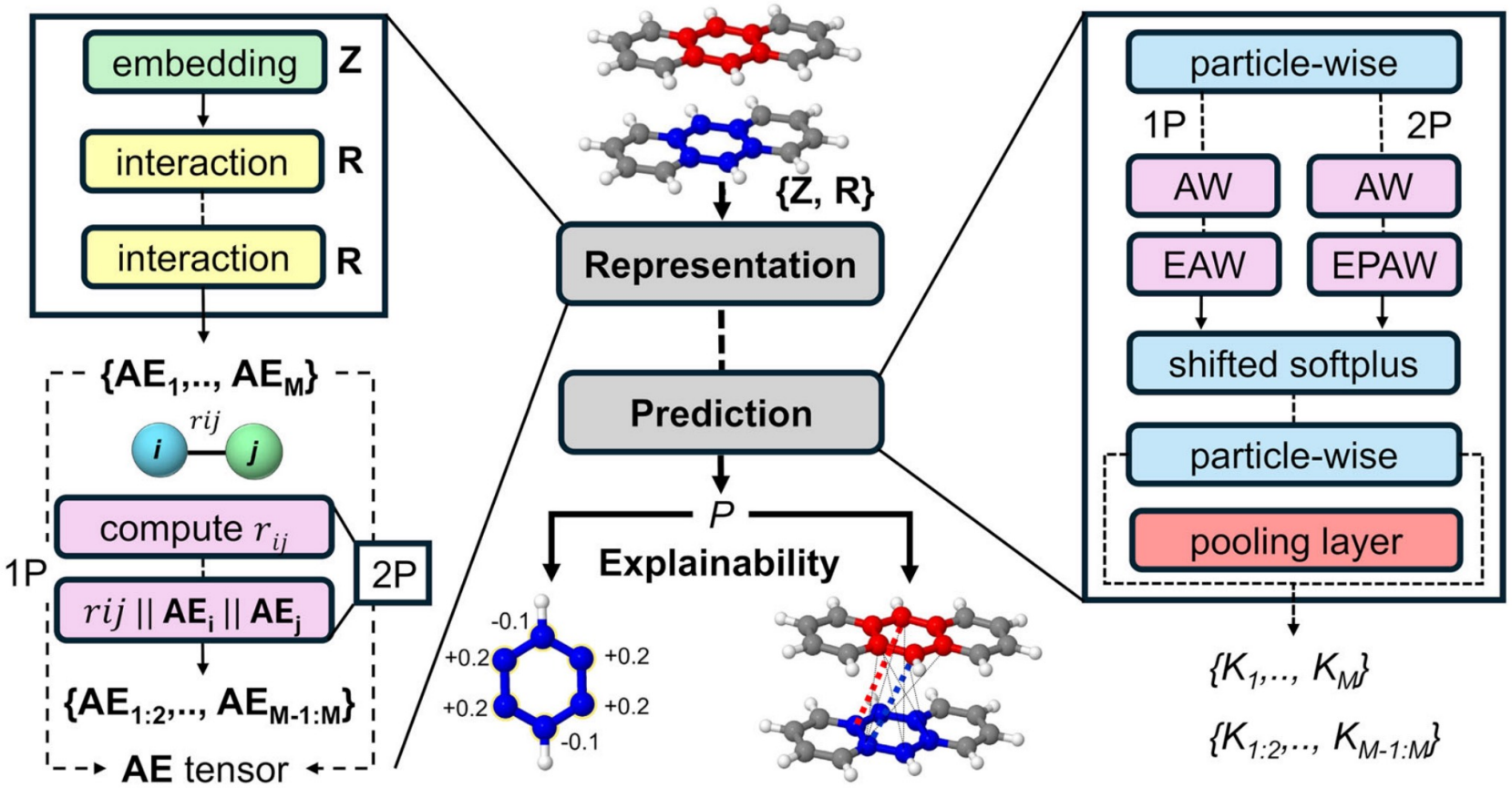
2 Confiança e validação

Verificar se o modelo prevê pelos motivos certos, e não por correlações falsas ou artefatos do conjunto de dados.

3 Descoberta

A interpretação revela novas relações estrutura–propriedade e hipóteses testáveis.

IA química explicável: 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

Referências

IA explicável e interpretável para a ciência e os materiais

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 EXPLICÁVEL

Mensagens principais – Transformar previsões rápidas em regras de design confiáveis.

POR QUE IMPORTA PARA O CLIMA

1. Na ciência, queremos o porquê, não só o número.

2. Métodos interpretáveis desmascaram modelos que acertam pelos motivos errados.

3. Trate cada explicação como uma hipótese a verificar contra a física e o experimento.

PERGUNTAS?

REFLEXÕES FINAIS

EDITORIAL | 19 May 2026

nature

Why AI cannot do good science without humans

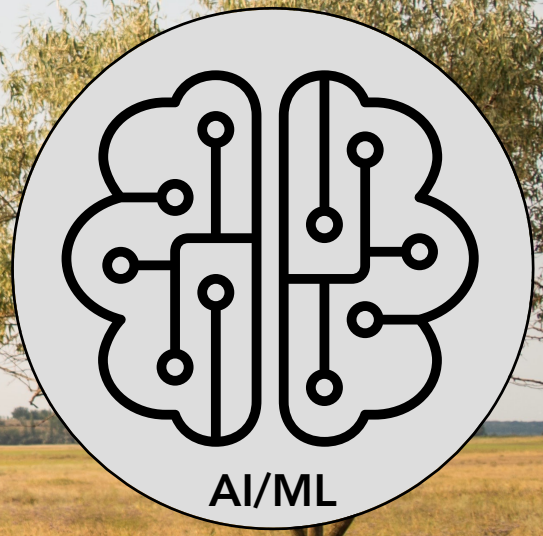
With the arrival of 'AI scientists', it's as well to remember that human wisdom, empathy and sheer messiness are as much part of progress as are process and efficiency.

Projetar para a circularidade

Projetar para o desempenho

e o desempenho





Repositórios on-line: ferramentas de código aberto (materiais e IA)

Um kit inicial para compartilhar. As ferramentas do nosso grupo estão destacadas; verifique se os links estão atualizados antes de distribuir.

Área	Ferramenta	Para que serve	Repositório
Dados e ferramentas	**pymatgen**	Análise de materiais, E/S e API do Materials Project	*github.com/materialsproject/pymatgen*
Dados e ferramentas	**matminer**	Descritores/vetorização prontos para AM	*github.com/hackingmaterials/matminer*
Dados e ferramentas	**RDKit**	Quimioinformática: moléculas, impressões digitais, descritores	*github.com/rdkit/rdkit*
Dados e ferramentas	**DeepChem**	Kit de AM para química: conjuntos de dados, modelos, benchmarks	*github.com/deepchem/deepchem*
Dados e ferramentas	**ASE**	Atomic Simulation Environment: construir/executar/analisar	*gitlab.com/ase/ase*
Representações	**SELFIES**	Representação molecular em string sempre válida	*github.com/aspuru-guzik-group/selfies*
Representações	**eOHE**	Codificações one-hot incorporadas (entradas eficientes)	*github.com/MMLabCodes/eOHE*
Representações	**eMFP**	Impressões digitais de Morgan incorporadas (compactas, mais rápidas)	*github.com/MMLabCodes/eMFP*
Generativo	**chemical_vae**	Design molecular em espaço latente contínuo	*github.com/aspuru-guzik-group/chemical_vae*
Generativo	**MatterGen**	Modelo generativo de difusão para cristais inorgânicos	*github.com/microsoft/mattergen*
Potenciais de AM	**MACE**	Potenciais interatômicos de AM (+ modelo de base MACE-MP-0)	*github.com/ACEsuit/mace*
Potenciais de AM	**CHGNet**	Potencial universal de rede neural informado pela carga	*github.com/CederGroupHub/chgnet*
Potenciais de AM	**MatGL (M3GNet)**	Potenciais universais de grafos + modelos de propriedades	*github.com/materialsvirtuallab/matgl*
Autonomia	**Ax / BoTorch**	Otimização bayesiana para laboratórios autônomos	*github.com/facebook/Ax*
Agentes LLM	**ChemCrow**	Agente LLM de química conectado a ferramentas especializadas	*github.com/ur-whitelab/chemcrow-public*
Explicabilidade	**SHAP**	Explicações por atribuição de características para qualquer modelo	*github.com/shap/shap*
Explicabilidade	**SISSO**	Descoberta de descritores/fórmulas interpretáveis	*github.com/rouyang2017/SISSO*
Computação verde	**CodeCarbon**	Acompanhe o custo de carbono da sua computação	*github.com/mlco2/codecarbon*

Linhas destacadas = repositórios do nosso grupo (github.com/MMLabCodes). Lista representativa, não exaustiva.

Leituras complementares

GARGALO DOS MATERIAIS

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

MODELAGEM E SIMULAÇÃO

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

NOVO 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

CAIXA DE FERRAMENTAS 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

FALANDO 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 RESPONSÁVEL

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 EXPLICÁVEL

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

Obrigado. Alguma pergunta final?

