L'IA au service de l'accélération de la science des matériaux

Francisco Martin-Martinez

Maître de conférences, King's College London

martinmartinezlab.com

ÉCOLE D'ÉTÉ VIRTUELLE CCAI 2026 · COURS SECTORIEL · 13 AOÛT 2026

QUI SUIS-JE ?

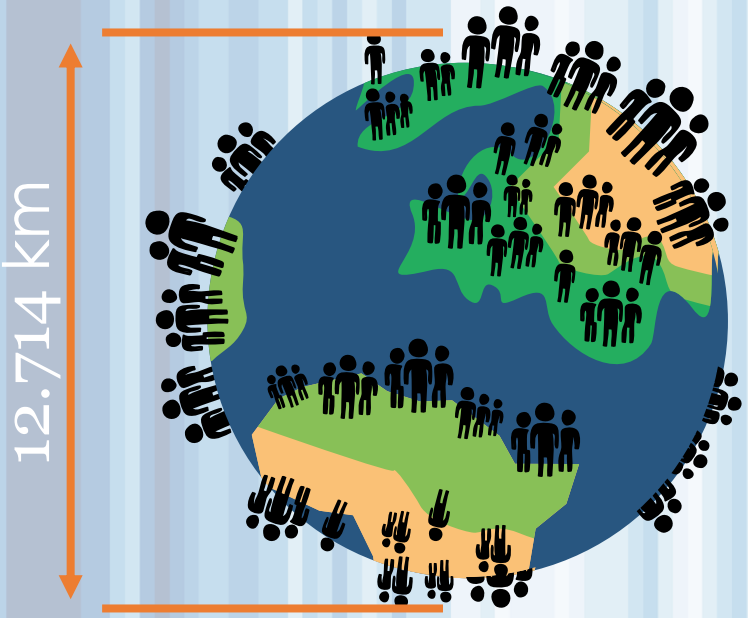
Qui suis-je, et pourquoi ce sujet me tient à cœur



Un scientifique qui utilise l'IA comme outil ou une IA qui fait de la science de manière autonome : où se situe la limite ?

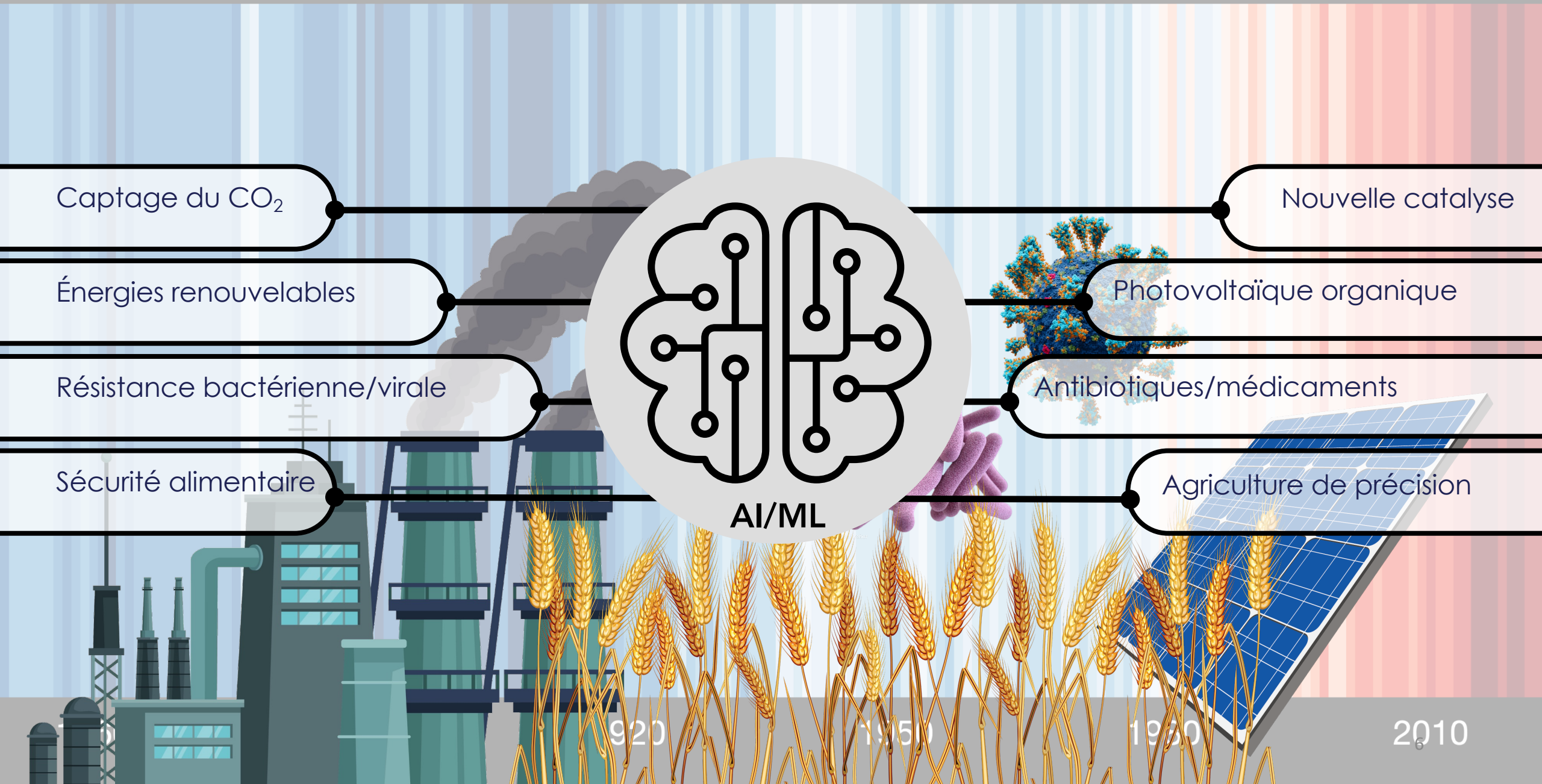
POURQUOI CE SUJET ME TIENT-IL À CŒUR ?

Global temperature change (1850-2022)

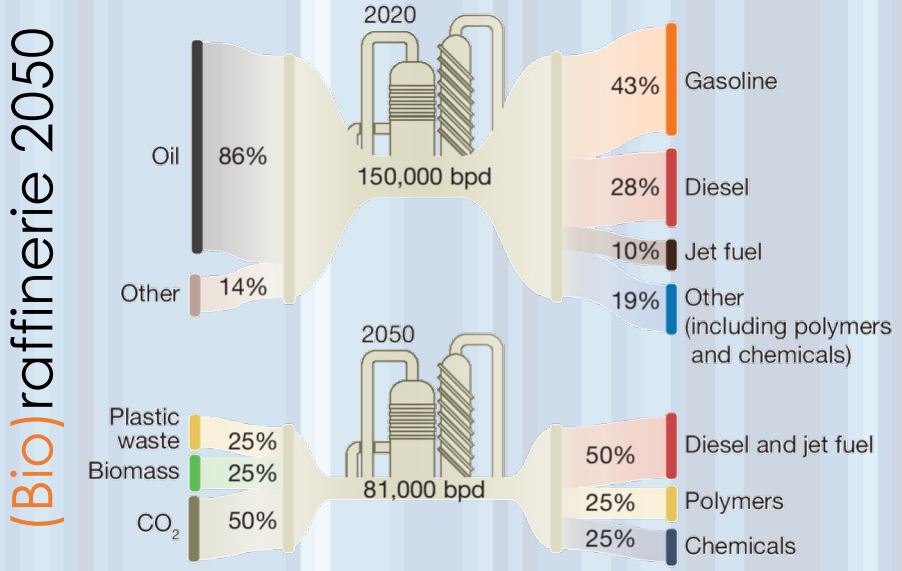


1860 1890 1920 1950 1980 2010

Global temperature change (1850-2022)

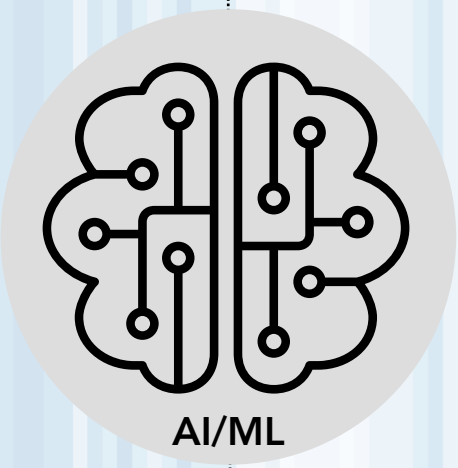


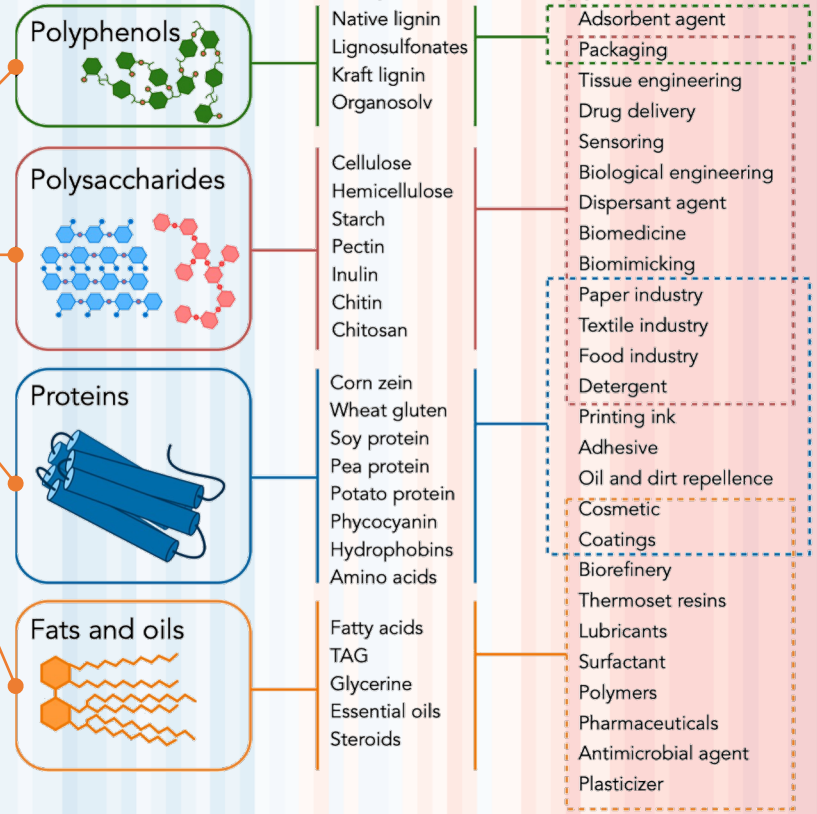
Nouvelle exploitation des matériaux issus des déchets



Nature 629, 295, **2024**

Déchets de biomasse
140 Gt/an

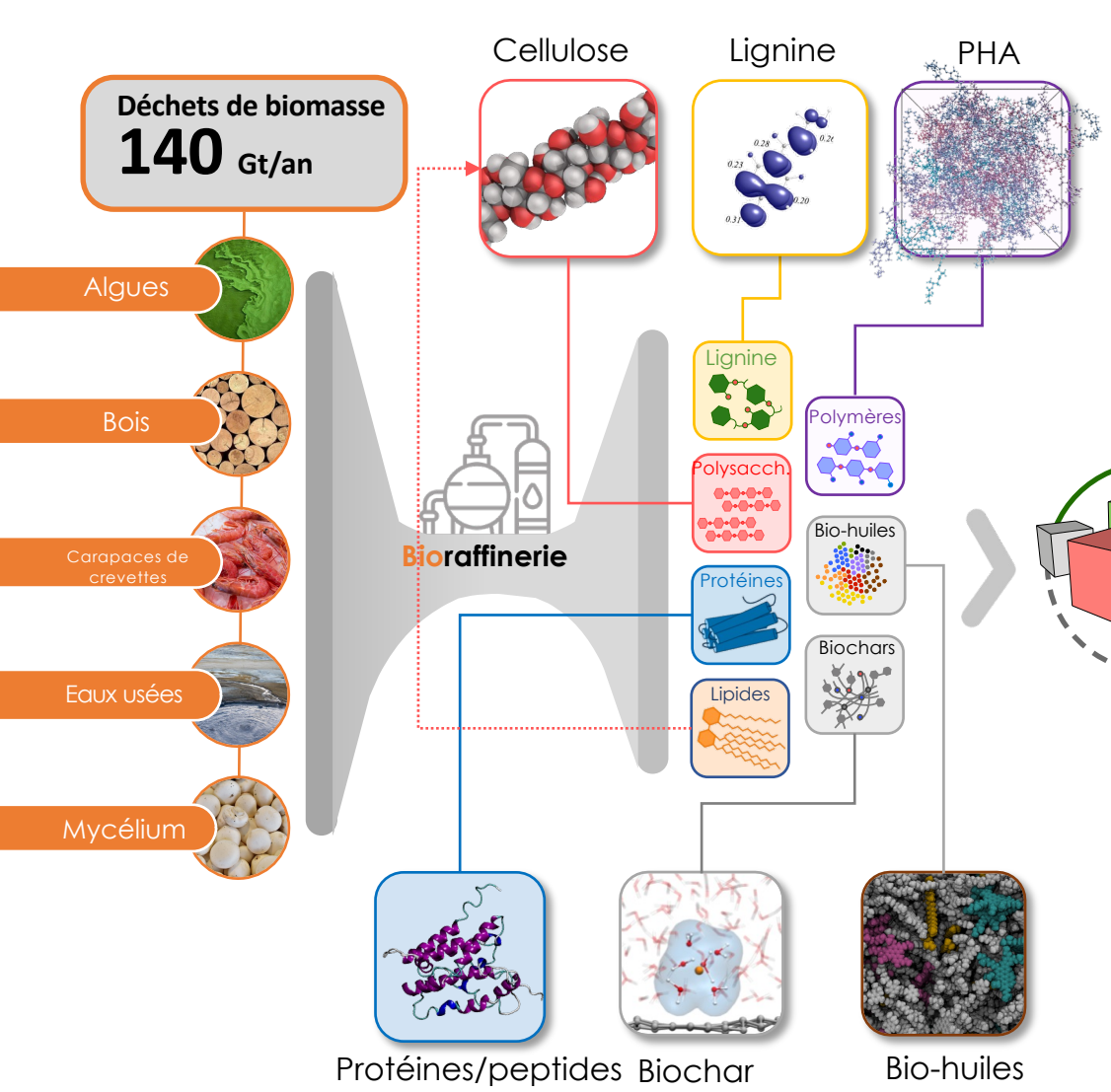


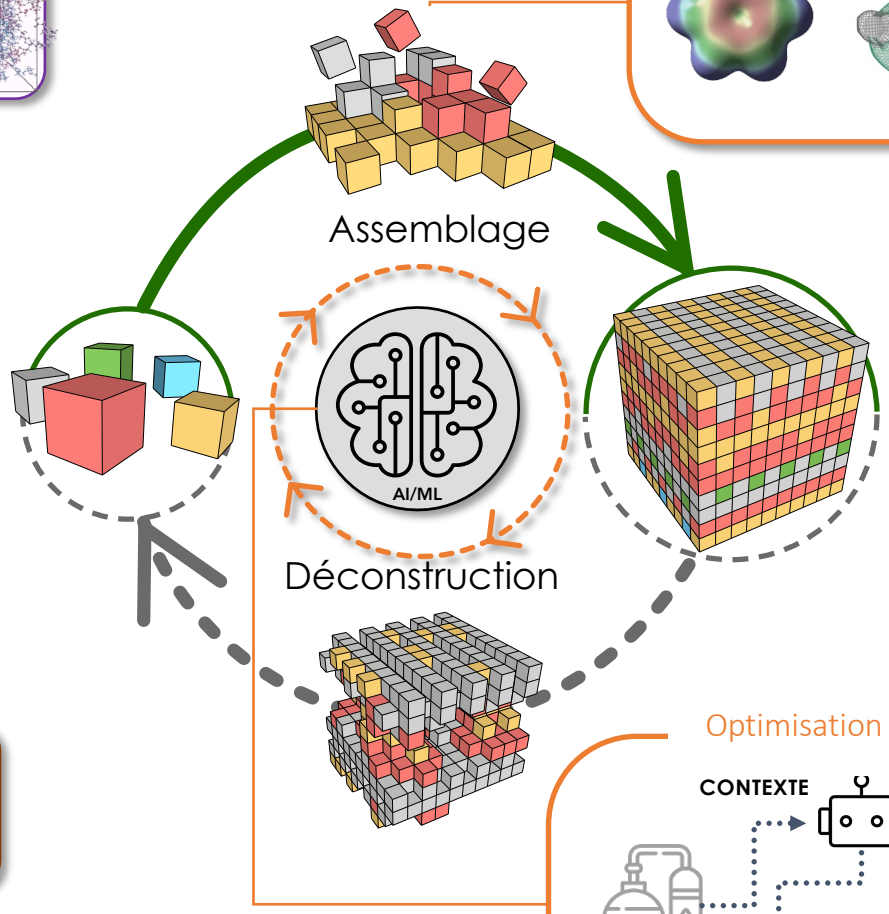


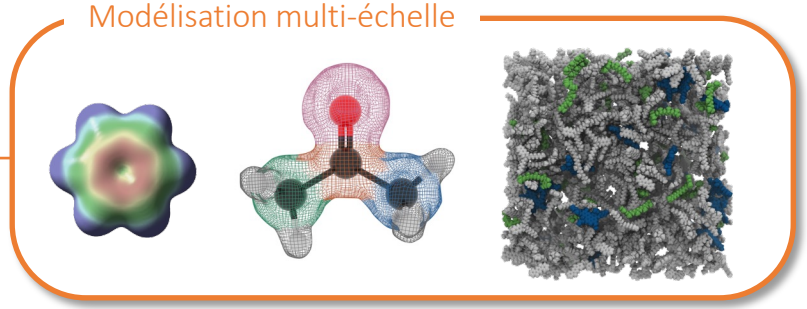
Bachs-Herrera, …, Martin-Martinez **iScience** 26, 106549, **2023**

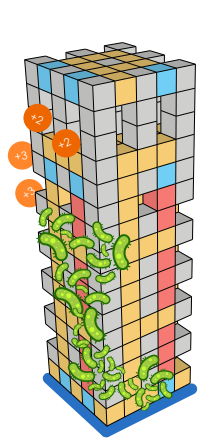
Valorisation des déchets en molécules et matériaux biosourcés

Matériaux circulaires biosourcés ➡ Vers une bioraffinerie quantique pilotée par l'IA

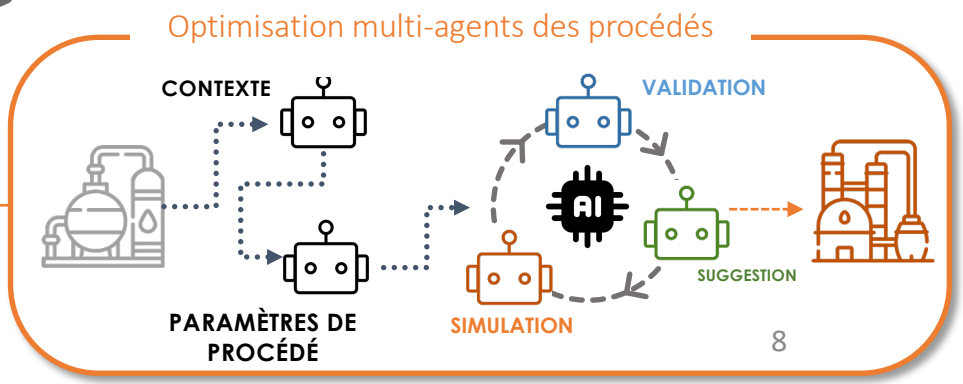












Bachs-Herrera et al., Martin-Martinez **iScience** 26, 106549, **2023**

AU PROGRAMME AUJOURD'HUI

AU PROGRAMME

01 Pourquoi les matériaux sont un goulot d'étranglement climatique

02 Le paradigme traditionnel (toujours utile) de la modélisation et de la simulation des matériaux

03 Un nouveau paradigme dans la découverte de matériaux

04 La boîte à outils de l'IA

05 Parler chimie aux ordinateurs

04 IA responsable

05 IA explicable en chimie

PARTICIPATION

Sondage en direct – Utilisez le chat

Selon vous, quel défi lié aux matériaux est le plus urgent pour le climat ?
Il n’y a pas de mauvaise réponse : c’est pour briser la glace

DES QUESTIONS ?

LE GOULOT D'ÉTRANGLEMENT DES MATÉRIAUX

Où l'IA intervient dans la chaîne des matériaux

1

Représenter

Encoder molécules/cristaux pour l'AA (empreintes, graphes OHE, texte)

2

Prédire

Prédiction de propriétés, modèles de substitution, **potentiels interatomiques par AA**

3

Générer

Conception inverse/générative de nouveaux candidats

4

Synthétiser

Laboratoires autonomes et auto-pilotés qui bouclent la boucle

5

Caractériser

Analyse assistée par IA des spectres, images et structures

Les matériaux, goulot d'étranglement climatique

Photovoltaïque

Rendement et stabilité limités par la chimie des matériaux

Batteries

Densité énergétique, durée de vie, dépendance aux minerais critiques

Carburants alternatifs

Catalyseurs et vecteurs pour l'H_2 vert et les e-carburants

Sorbants de CO_2

Matériaux de captage bon marché, durables et extensibles

Autres : infrastructures durables, sécurité alimentaire, découverte de médicaments, antibiotiques…

Le problème : le cycle de découverte est lent et coûteux.

Concevoir → synthétiser → caractériser → tester peut prendre des années par candidat.

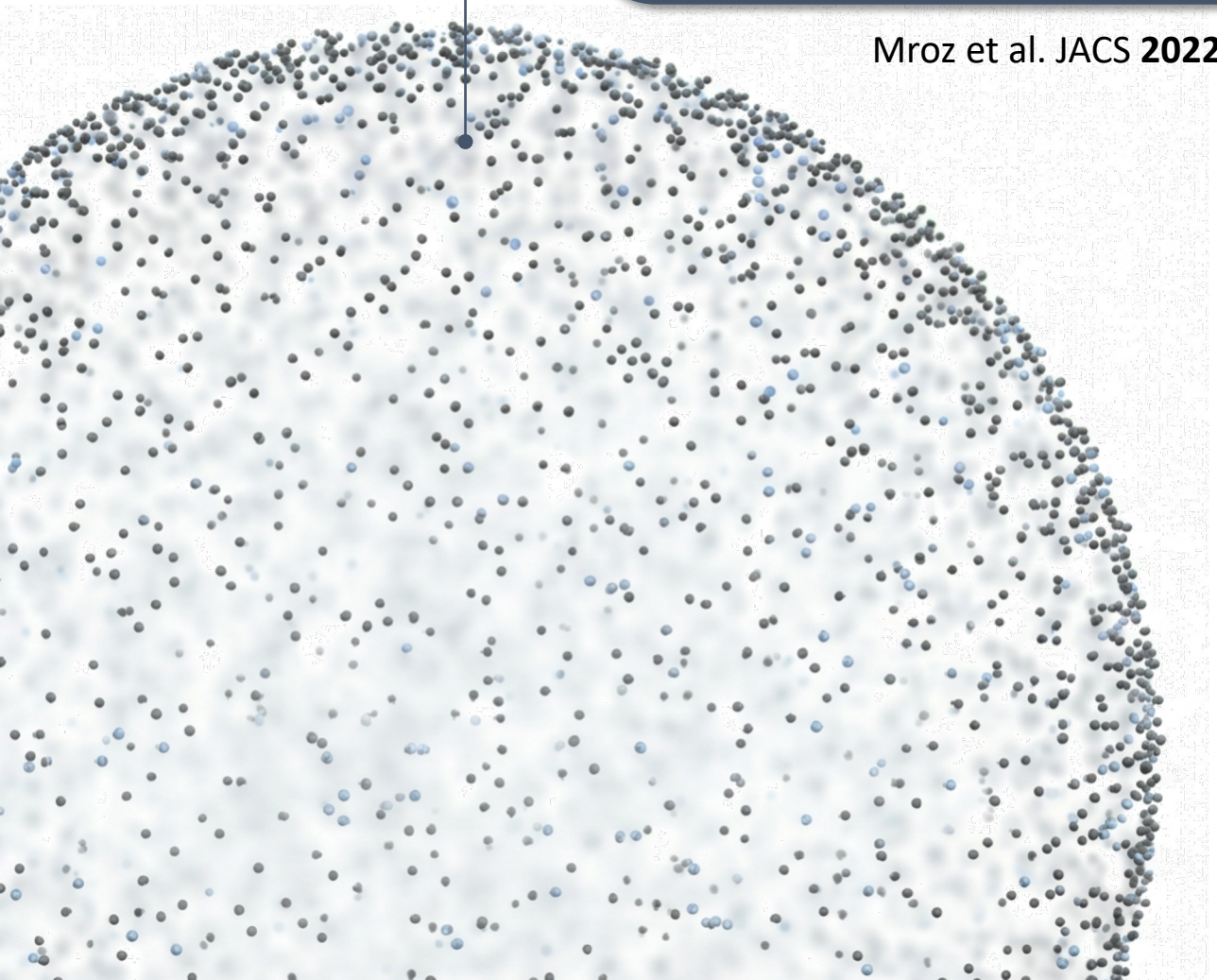
L'atténuation ET l'adaptation dépendent toutes deux de l'obtention des bons matériaux, plus rapidement et à moindre coût.

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

Le goulot d'étranglement de la découverte

10^{60} Petites molécules (de type médicament) pour la conception de matériaux

Mroz et al. JACS **2022**



L'espace chimique combinatoire est immense

VS

La validation expérimentale est lente et coûteuse

LE SAVIEZ-VOUS ?

Il existe plus de petites molécules organiques possibles que d'atomes dans l'univers observable.

PARTICIPATION

Quiz éclair – Vrai/Faux

1. **Il existe plus de petites molécules organiques possibles que d'atomes dans l'univers observable.**

   Vrai | Faux

2. **Un nouveau matériau arrive généralement sur le marché en quelques semaines.**

   Vrai | **Faux**

3. **L'IA intervient dans la chaîne aux étapes représenter, prédire, générer, synthétiser et caractériser.**

   Vrai | Faux

Messages clés

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. Les matériaux sont un goulot d'étranglement climatique : l'atténuation comme l'adaptation exigent de meilleurs matériaux, plus vite.

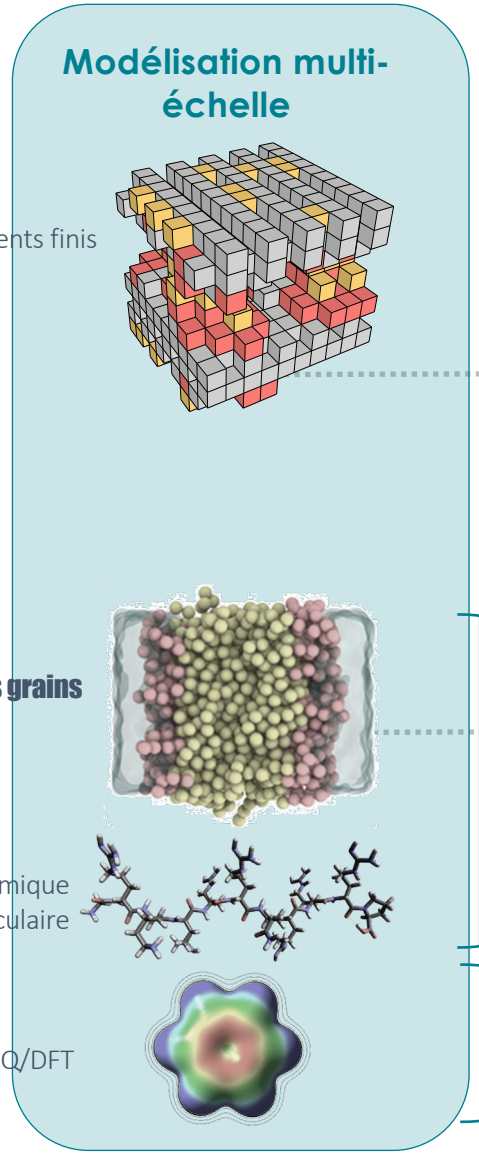
2. L'espace chimique ne peut pas être exploré par la force brute. L'IA le transforme en recherche guidée.

3. Les objectifs climatiques sont soumis à un calendrier fixe. Découvrir plus vite, c'est déployer plus vite.

DES QUESTIONS ?

MODÉLISATION ET SIMULATION DES MATÉRIAUX

Modélisation multi-échelle



Simulations de dynamique moléculaire

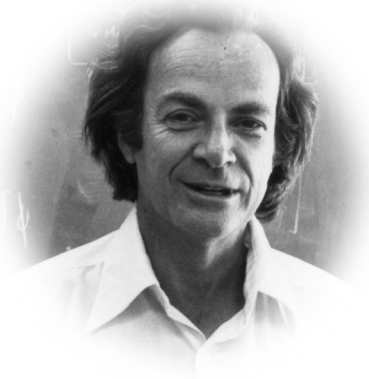


« Tous les modèles sont faux, mais certains sont utiles. »

– George E. P. Box



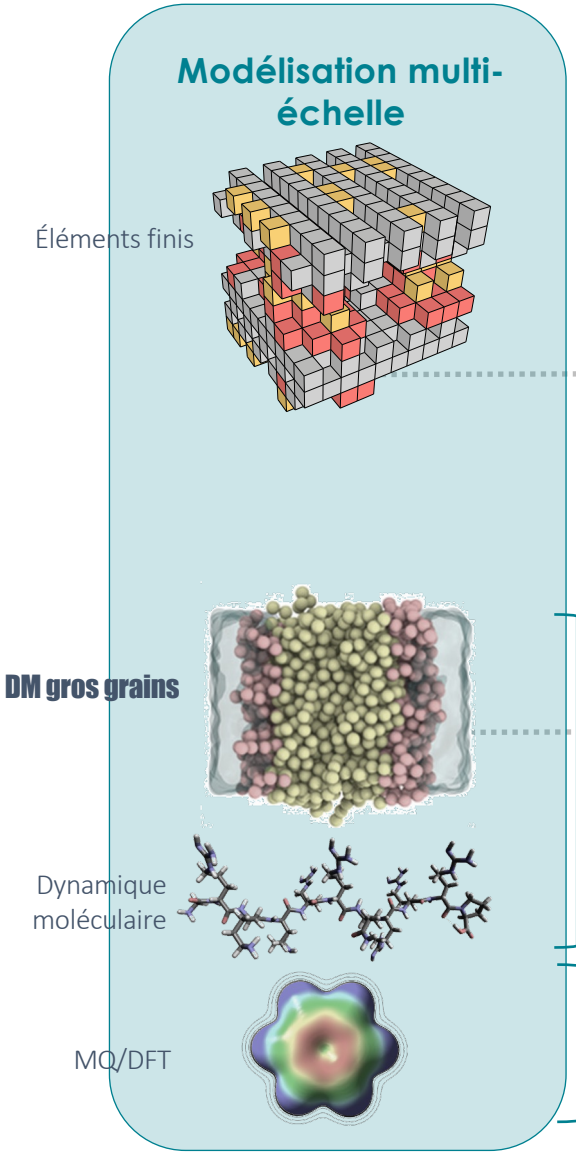
Théorie de la fonctionnelle



« Je crois pouvoir affirmer sans risque que personne ne comprend la mécanique quantique. »

– Richard Feynman

Modélisation multi-échelle



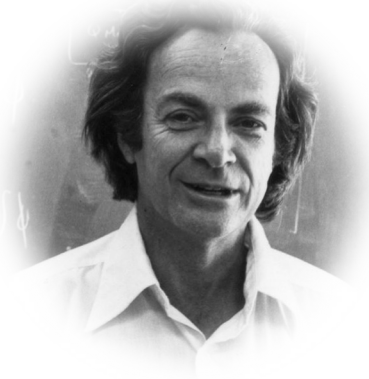
Simulations de dynamique moléculaire



« Tous les modèles sont faux, mais certains sont utiles. »

– George E. P. Box

Théorie de la fonctionnelle de la densité (DFT) – Mécanique quantique

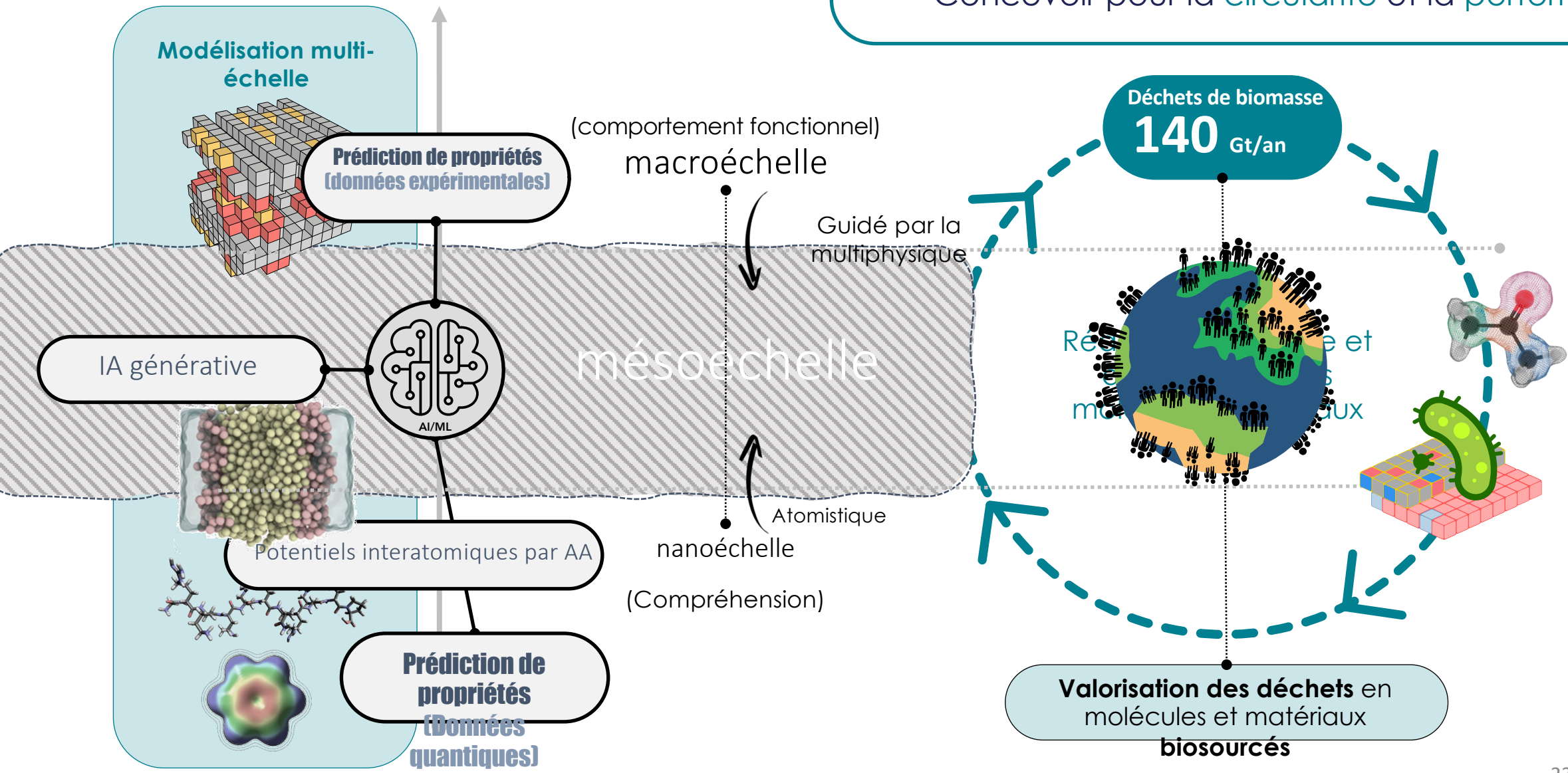


« Je crois pouvoir affirmer sans risque que personne ne comprend la mécanique quantique. »

– Richard Feynman

Modélisation multi-échelle

Concevoir pour la circularité et la performance



Logiciels de DFT, de DM et d'AA pour la chimie

DFT moléculaire



ORCA

DFT de l'état solide



siesta™

DM (tous atomes et gros grains)

LAMMPS

GROMACS
FAST. FLEXIBLE. FREE.



NAMD
Scalable Molecular Dynamics

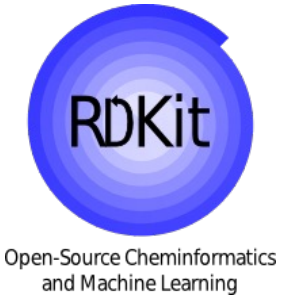


MD ANALYSIS



Intégration et chémoinformatique



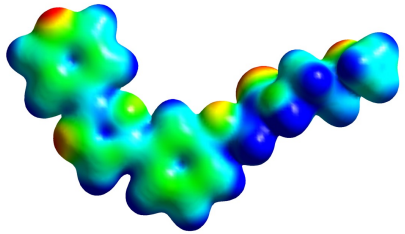


+ Atomic Simulation Environment (ASE)

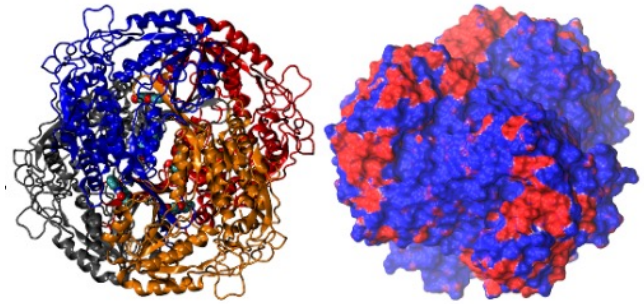
Bases de données au service de la découverte en chimie

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

Théorie de la fonctionnelle de la densité (DFT)



Dynamique moléculaire (DM)



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

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) DFT (quantique)

D) Éléments finis

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) **DFT (quantique)**

D) Éléments finis

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) DFT (quantique)

D) Éléments finis

Q2 **Pourquoi utilise-t-on la modélisation multi-échelle ?**

A) Une seule échelle suffit

B) Les phénomènes diffèrent selon l'échelle

C) Pour économiser des citations

D) C'est une obligation légale

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) **DFT (quantique)**

D) Éléments finis

Q2 **Pourquoi utilise-t-on la modélisation multi-échelle ?**

A) Une seule échelle suffit

B) **Les phénomènes diffèrent selon l'échelle**

C) Pour économiser des citations

D) C'est une obligation légale

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) **DFT (quantique)**

D) Éléments finis

Q2 **Pourquoi utilise-t-on la modélisation multi-échelle ?**

A) Une seule échelle suffit

B) **Les phénomènes diffèrent selon l'échelle**

C) Pour économiser des citations

D) C'est une obligation légale

Q3 **Par rapport à la DM tous atomes, la DM gros grains atteint…**

A) Des systèmes plus petits

B) Des systèmes plus grands et des durées plus longues

C) Une plus grande précision quantique

D) Aucune différence

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – choix multiple

Q1 **Quelle méthode est la plus précise, mais aussi la plus coûteuse ?**

A) DM classique

B) DM gros grains

C) **DFT (quantique)**

D) Éléments finis

Q2 **Pourquoi utilise-t-on la modélisation multi-échelle ?**

A) Une seule échelle suffit

B) **Les phénomènes diffèrent selon l'échelle**

C) Pour économiser des citations

D) C'est une obligation légale

Q3 **Par rapport à la DM tous atomes, la DM gros grains atteint…**

A) Des systèmes plus petits

B) **Des systèmes plus grands et des durées plus longues**

C) Une plus grande précision quantique

D) Aucune différence

MODÉLISATION ET SIMULATION

Messages clés – Le meilleur endroit pour échouer à moindre coût, c’est l’ordinateur.

POURQUOI C’EST IMPORTANT POUR LE CLIMAT

1. Aucune échelle ne suffit à elle seule : la modélisation multi-échelle est nécessaire.

2. Nous devons prédire la dégradation et la recyclabilité : la circularité, pas seulement la performance.

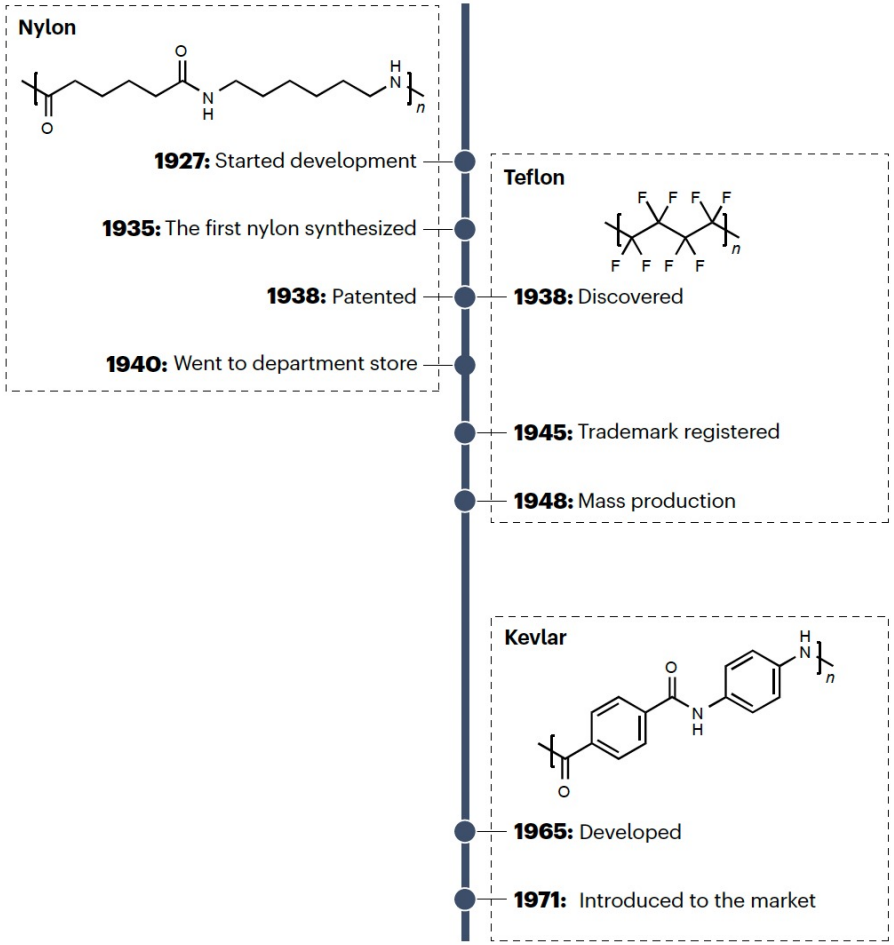
3. Les données sont le socle sur lequel l’IA se construit. De bonnes données font de bons modèles.

DES QUESTIONS ?

UN NOUVEAU PARADIGME DANS LA DÉCOUVERTE DE MOLÉCULES ET DE MATÉRIAUX

Matériaux lents, monde rapide

Les innovations en matière de polymères au cours des deux derniers siècles.



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

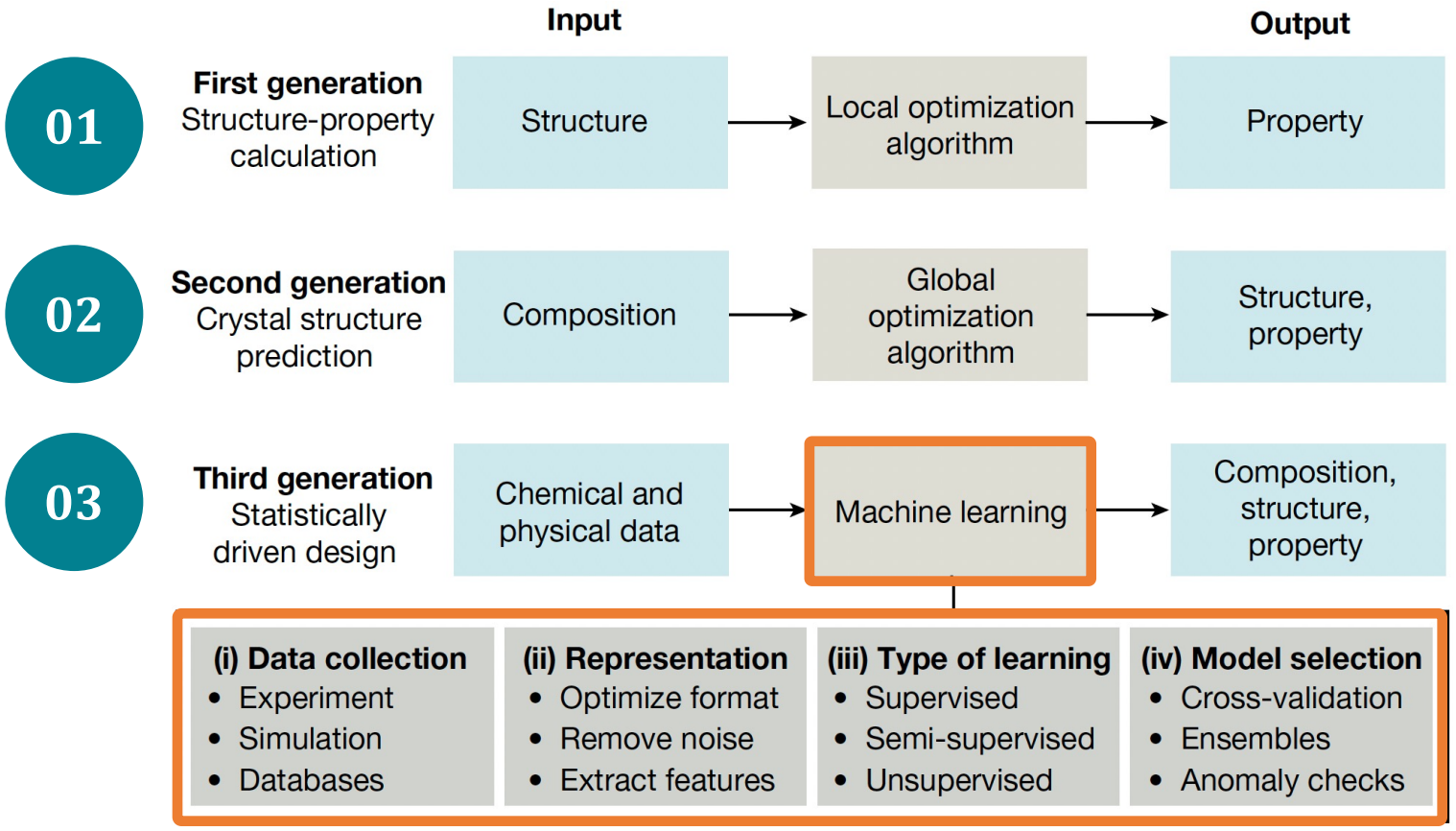
2011: Launch of Materials Genome Initiative

2020: 100-year anniversary of modern polymer science

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

De l'essai-erreur à la conception guidée par les données :

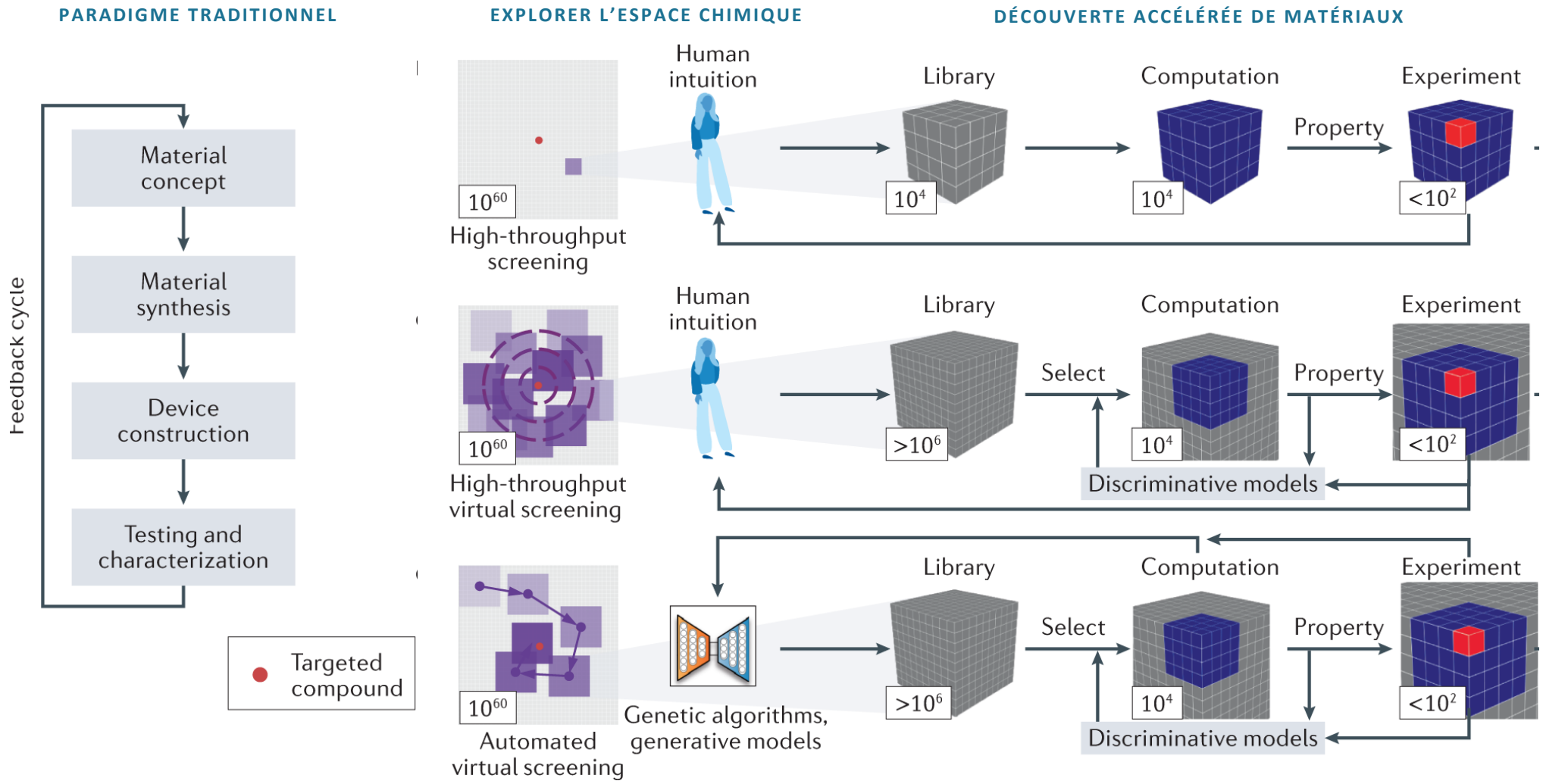
évolution du flux de travail de recherche en chimie computationnelle



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

Nouveau paradigme dans la découverte de matériaux

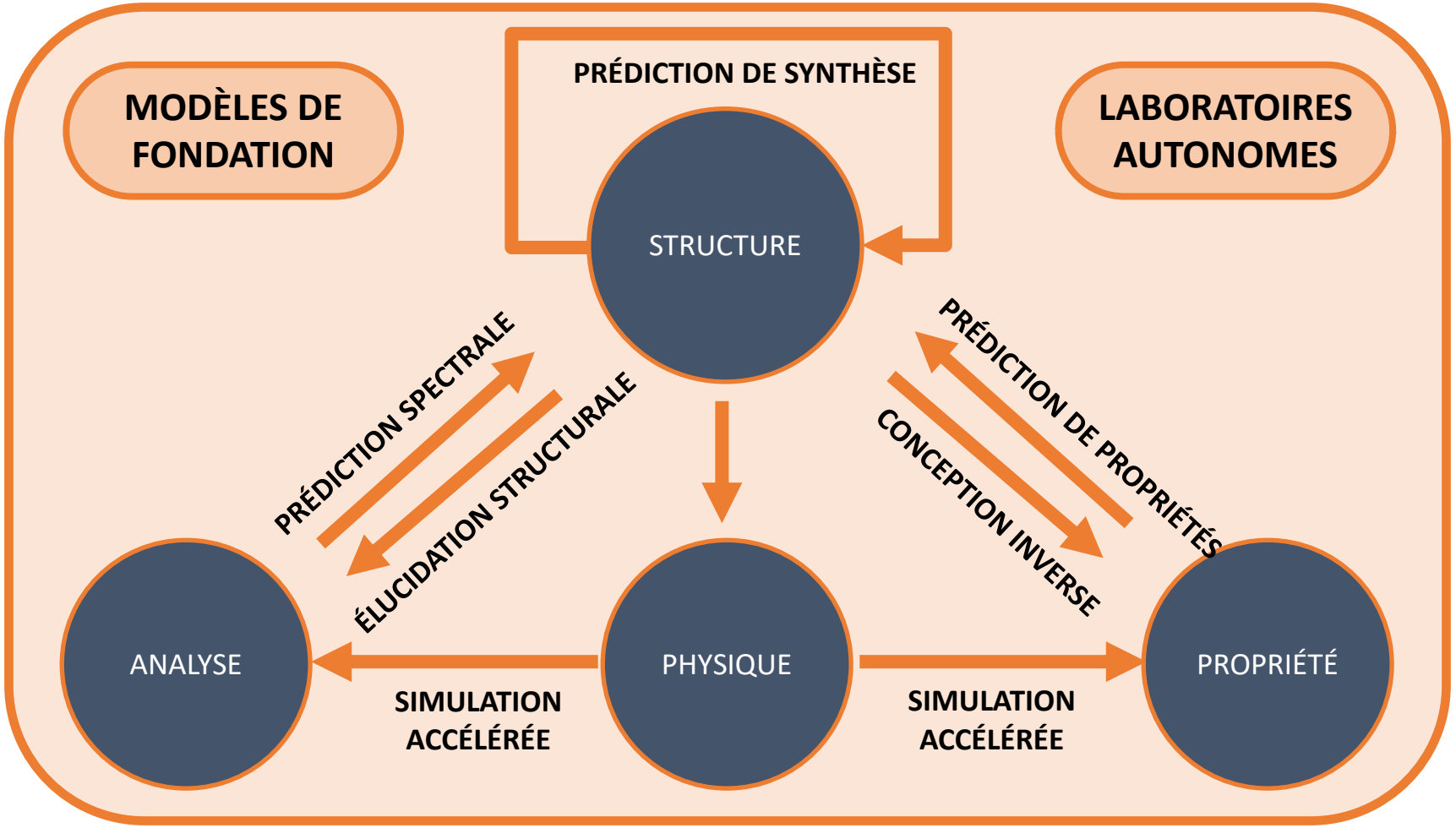
NOUVEAU PARADIGME



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

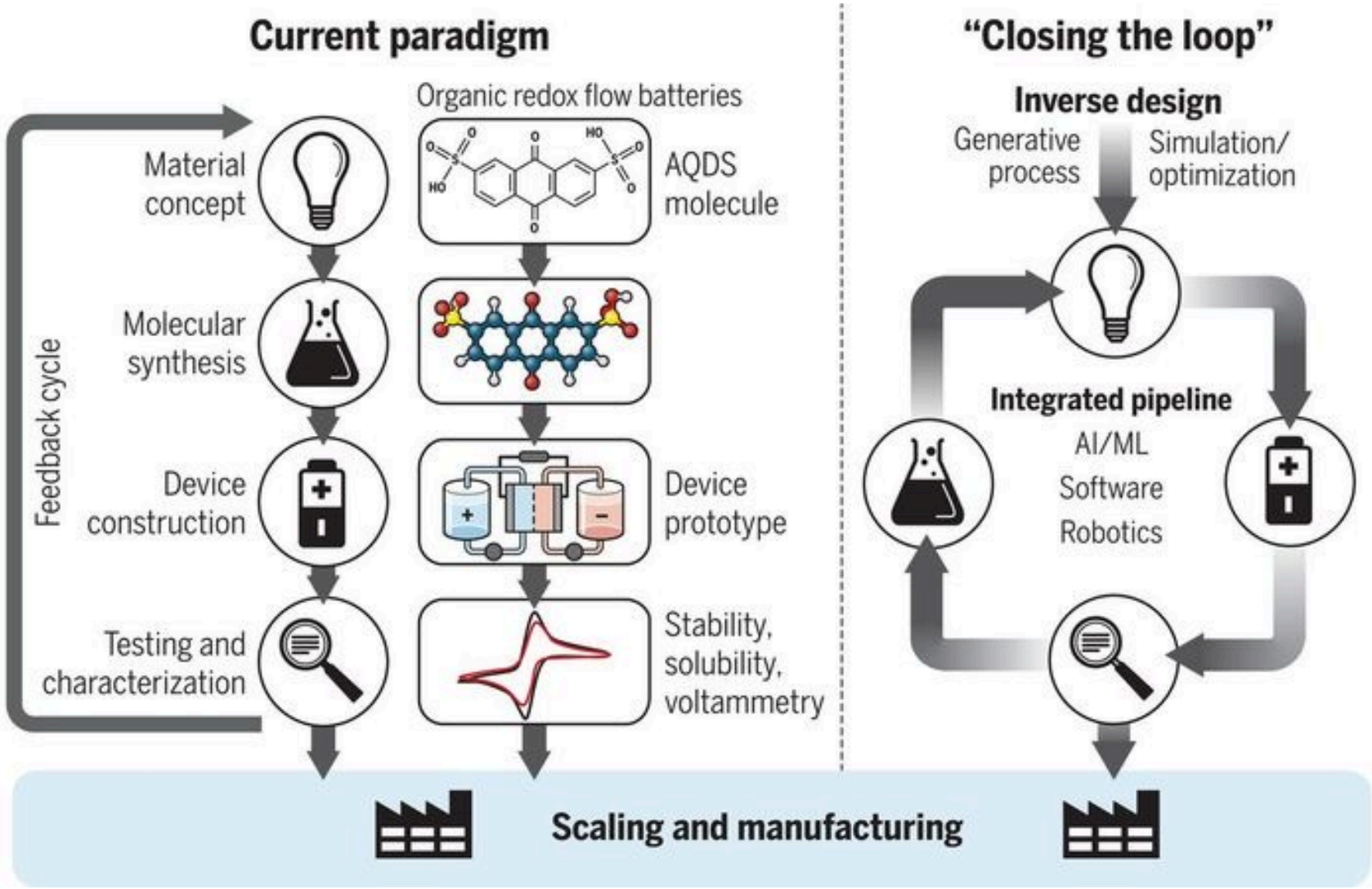
L'apprentissage automatique pour la chimie et les matériaux

NOUVEAU PARADIGME



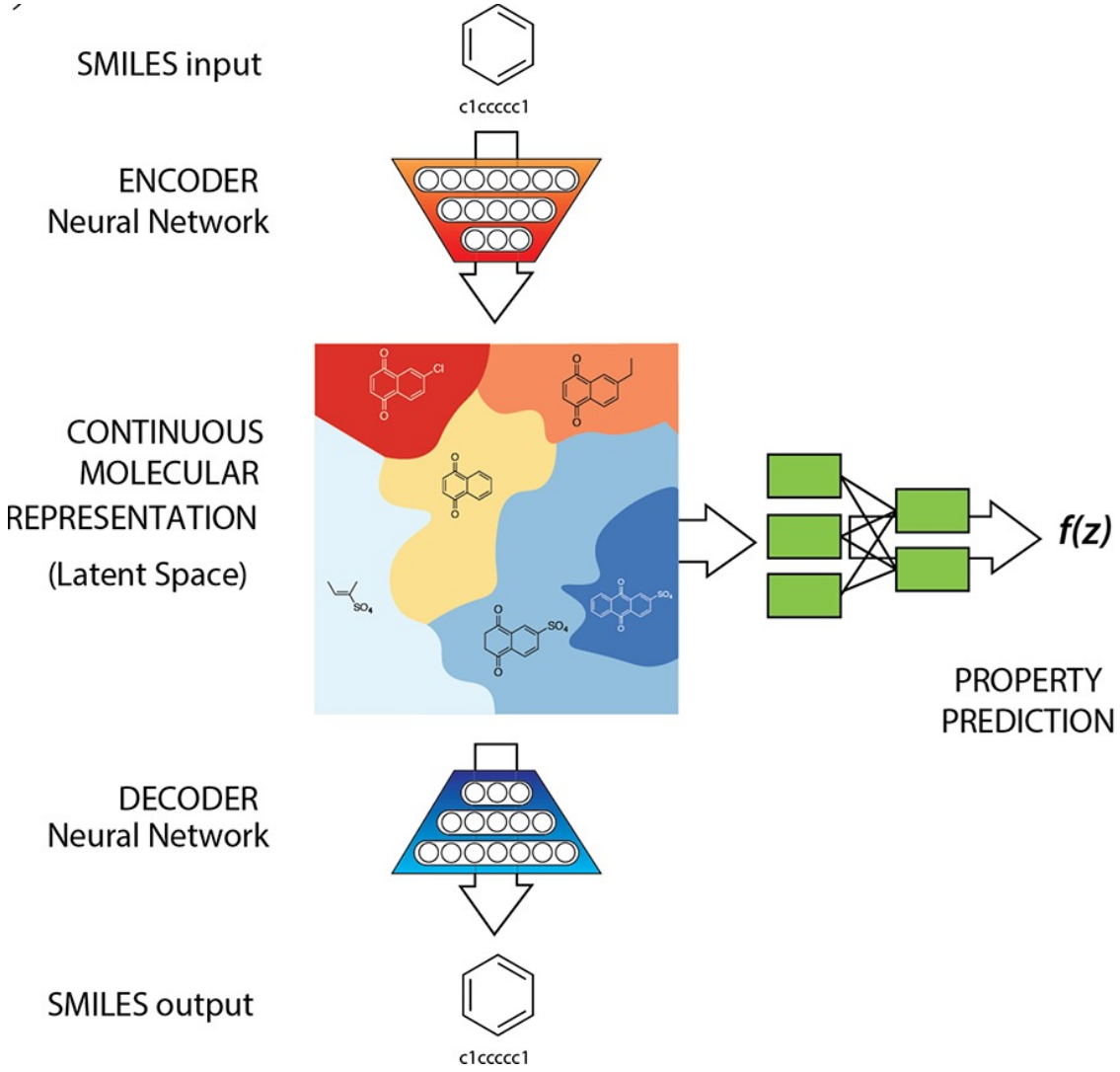
Cheng, Ser, Skreta, et al., Aspuru-Guzik, Faraday Discuss. 256, 10–60 (2025).

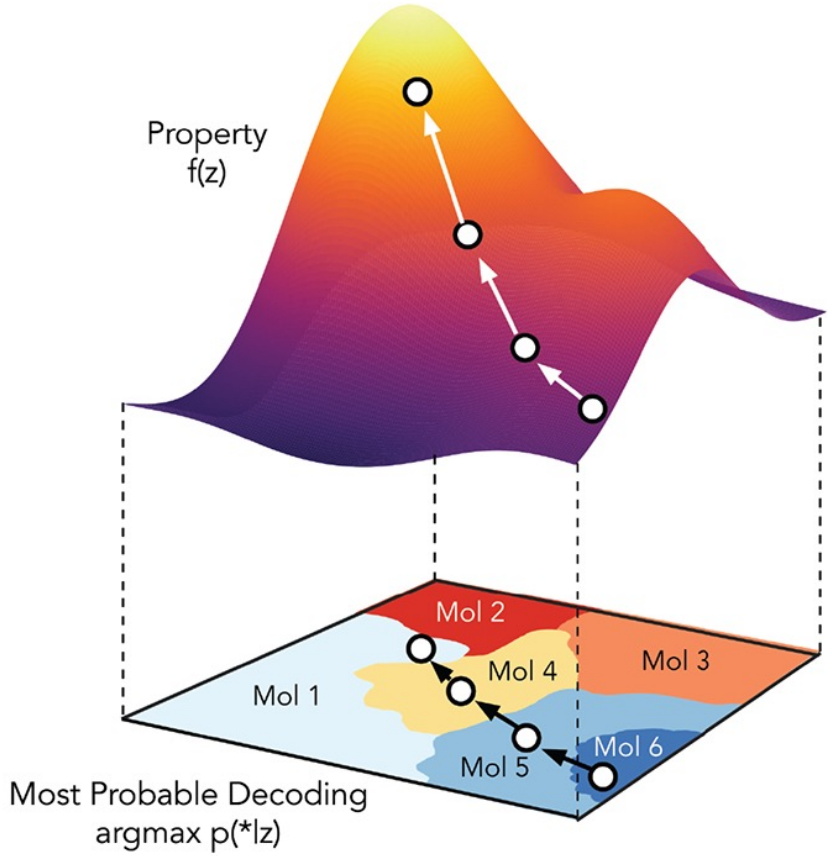
Conception directe ou inverse : l'exemple des batteries à flux redox



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

VAE chimique : transformer les molécules en une carte à optimiser





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

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – Vrai/Faux

1. **En conception inverse, on spécifie la propriété souhaitée et le modèle propose une structure.**

   Vrai | Faux

2. **Un VAE fournit un espace latent continu dans lequel on peut optimiser.**

   Vrai | Faux

3. **La « troisième génération » de la conception de matériaux est l'essai-erreur manuel.**

   Vrai | **Faux**

NOUVEAU PARADIGME

Messages clés – Concevoir selon un cahier des charges au lieu d'y arriver par hasard.

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. Nous vivons une génération de conception de matériaux guidée par l'AA.

2. Conception inverse = optimiser dans un espace latent appris (espace de conception continu).

3. Il existe deux catalyseurs transversaux : les modèles de fondation et les laboratoires autonomes.

DES QUESTIONS ?

LA BOÎTE À OUTILS DE L'IA

Une boîte à outils de l'IA

Six choses que l'AA fait réellement tout au long de la chaîne des matériaux

1 Prédiction de propriétés

PRÉDIRE

Les modèles supervisés prédisent une propriété y (p. ex. la bande interdite) à partir d'une structure x, en complément ou en remplacement de simulations coûteuses.

2 Conception générative

GÉNÉRER

Les modèles non supervisés/génératifs produisent de nouveaux candidats x', conditionnés par des propriétés cibles : la voie de la conception inverse.

3 Laboratoires autonomes

SYNTHÉTISER

Des laboratoires autonomes planifient et réalisent les expériences : rétrosynthèse, prédiction de réactions, optimisation par apprentissage actif.

4 Aide à la caractérisation

CARACTÉRISER

L'AA interprète les mesures : structure cristalline par MET, coordination par XAS, symétrie par diffraction électronique.

5 Gestion des systèmes

DÉPLOYER

L'AA gère les appareils et les réseaux : prédiction de la durée de vie des batteries, prévision de la charge, optimisation du pilotage des réseaux intelligents.

6 Accélérer la théorie

SIMULER

L'AA rend réalisables des simulations insolubles : potentiels et fonctionnelles moins coûteux pour des échelles plus grandes.

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

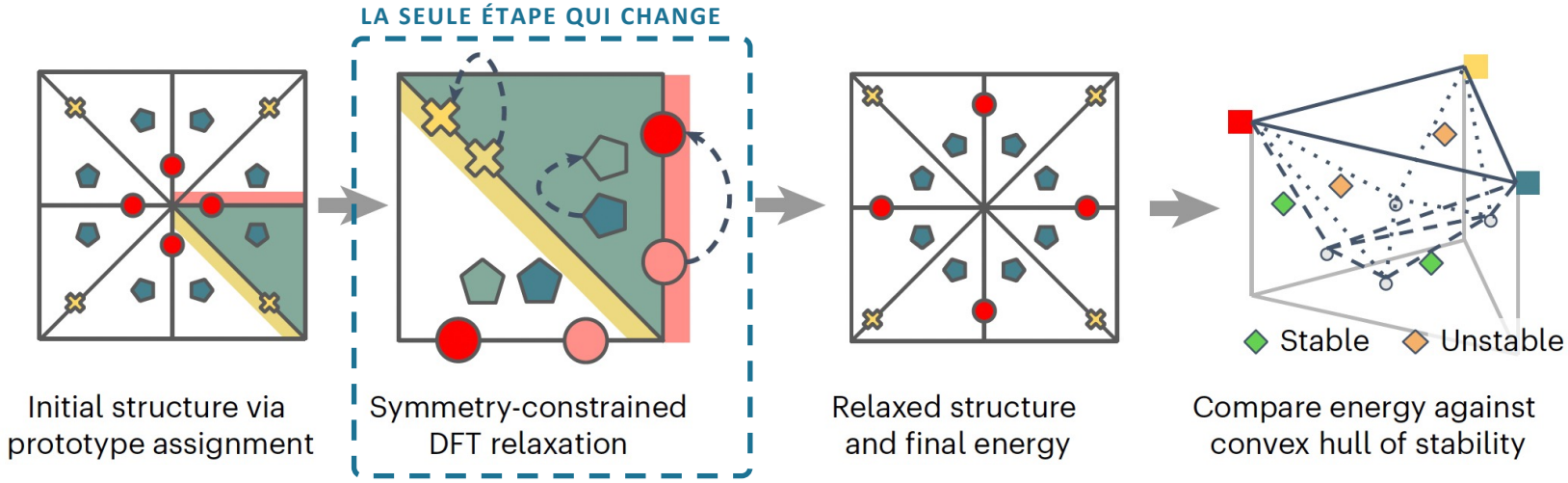
1

Prédiction de propriétés

PRÉDIRE

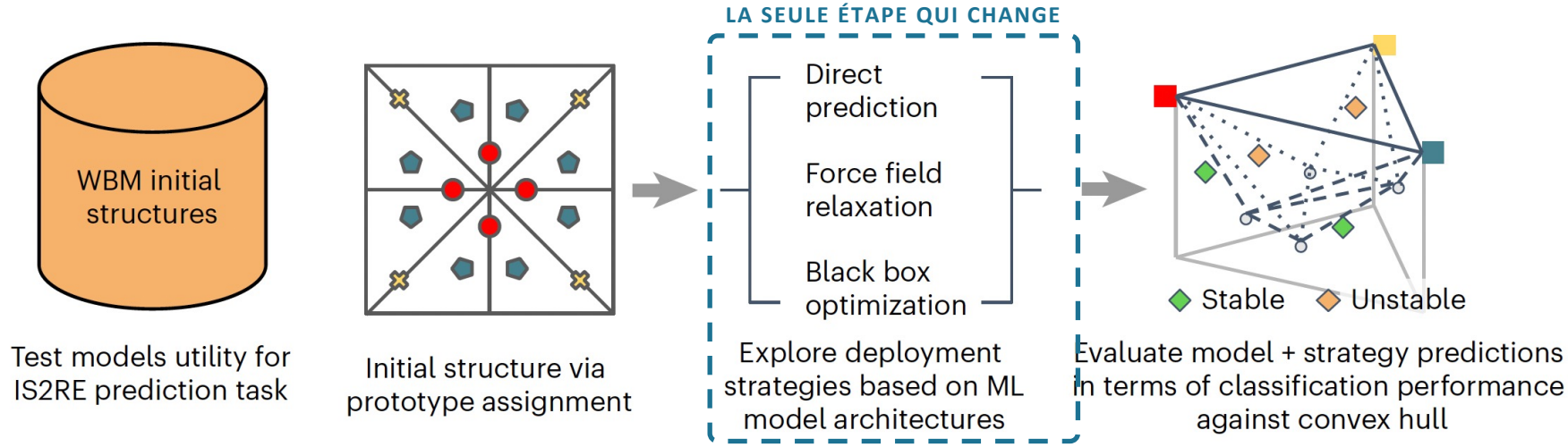
Prédire les propriétés DFT avec un substitut AA

CLASSIQUE



MATBENCH DISCOVERY

IS2RE = de la structure initiale à l'énergie relaxée (Initial Structure to Relaxed Energy)



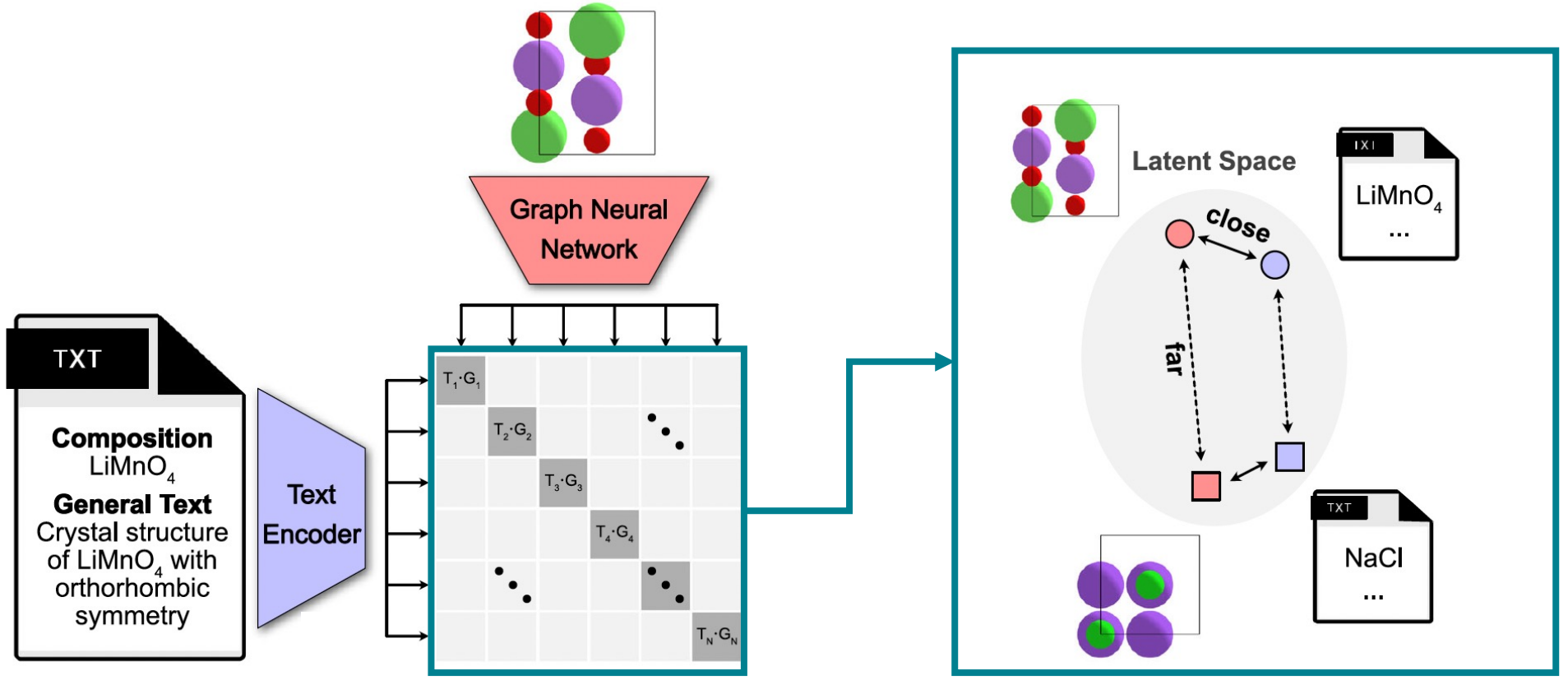
Adapté de Riebesell et al., « A framework to evaluate machine learning crystal stability predictions », Nat. Mach. Intell. 7, 836–847 (2025).

2

Conception générative/inverse

GÉNÉRER

Chemeleon : transformer une requête textuelle en cristal, en trois étapes



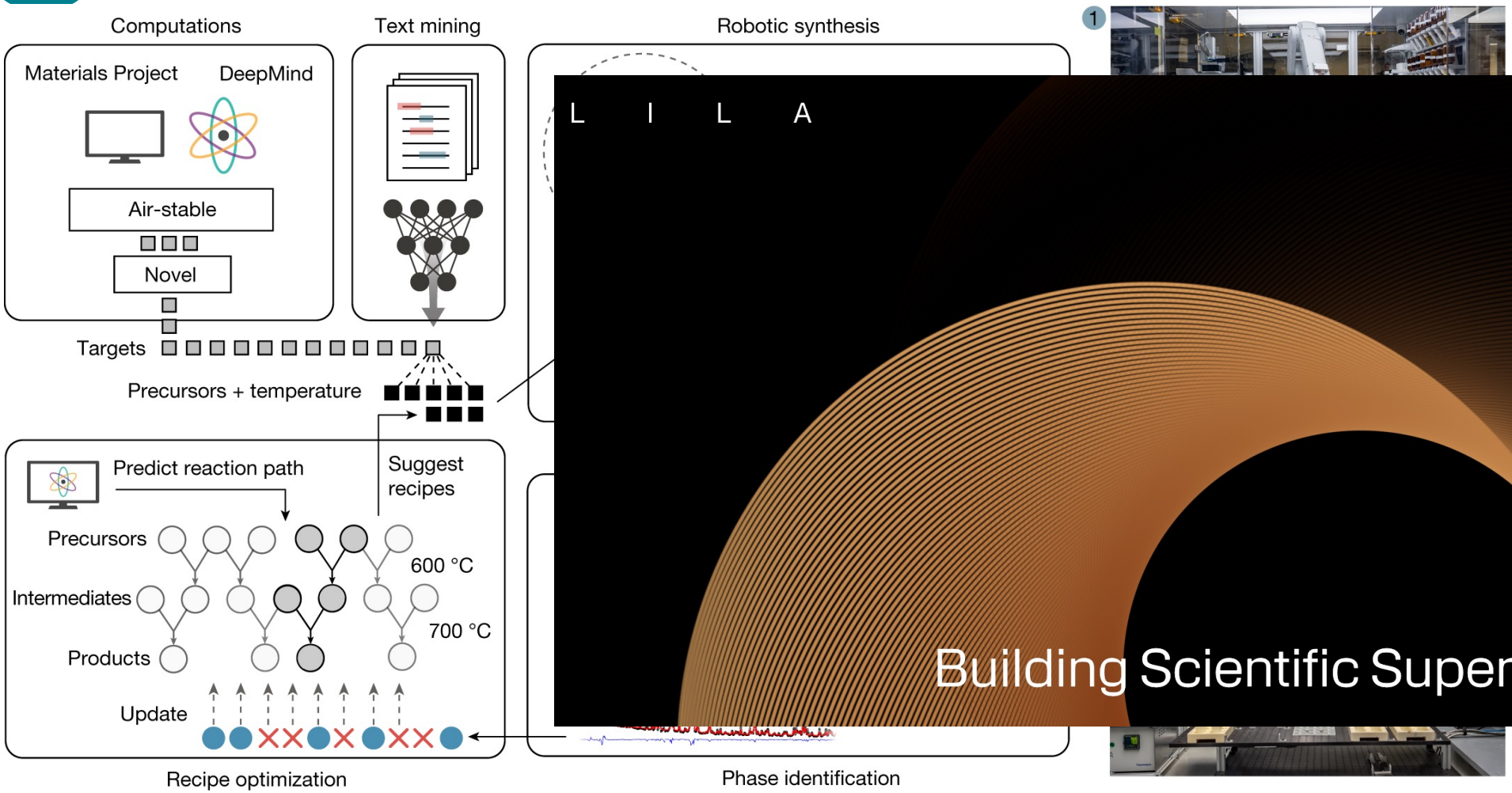
Adapté de Park, Onwuli & Walsh, Nat. Commun. 16 (2025)

3

Laboratoires autonomes

SYNTHÉTISER

Un laboratoire autonome : l'A-Lab



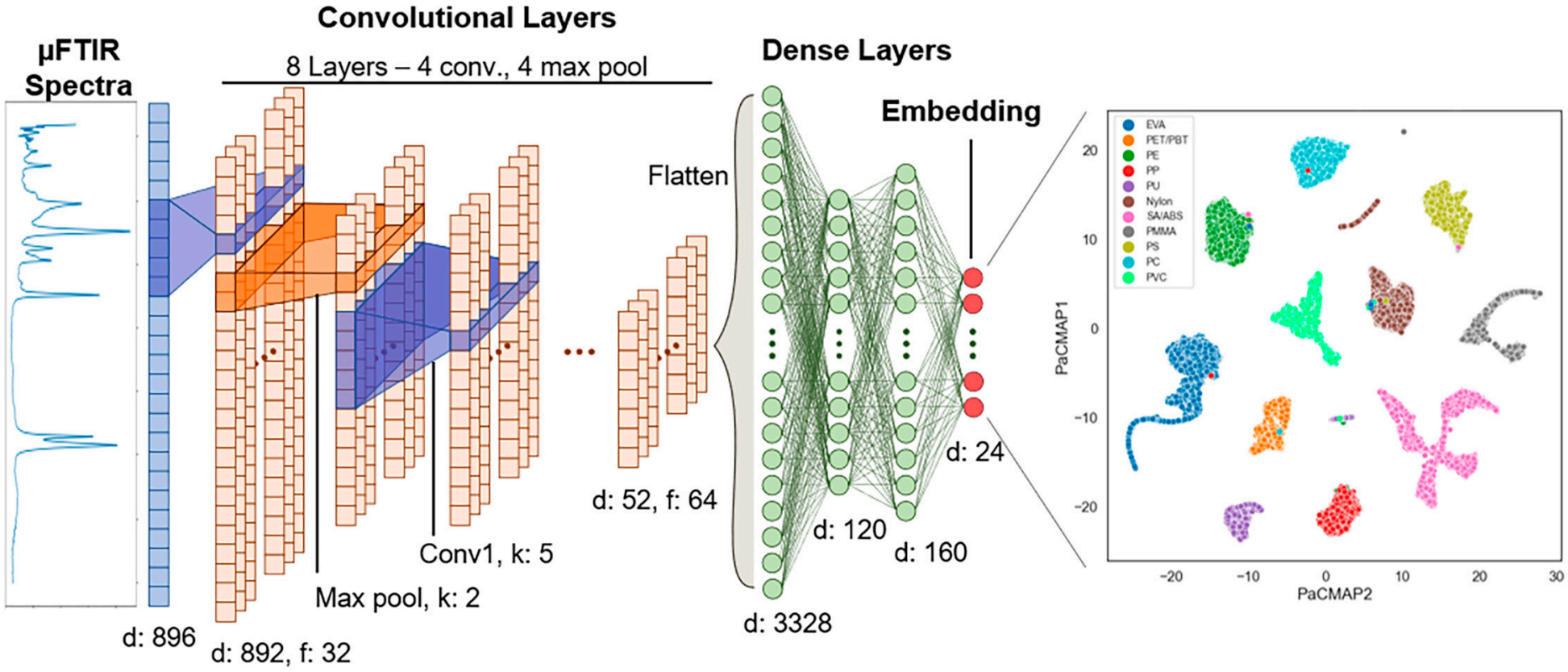
Adapté de Szymanski et al., Nature 624, 86 (2023)

4

Aide à la caractérisation

CARACTÉRISATION

Apprentissage de similarité appliqué à l'imagerie spectrale µFTIR des microplastiques



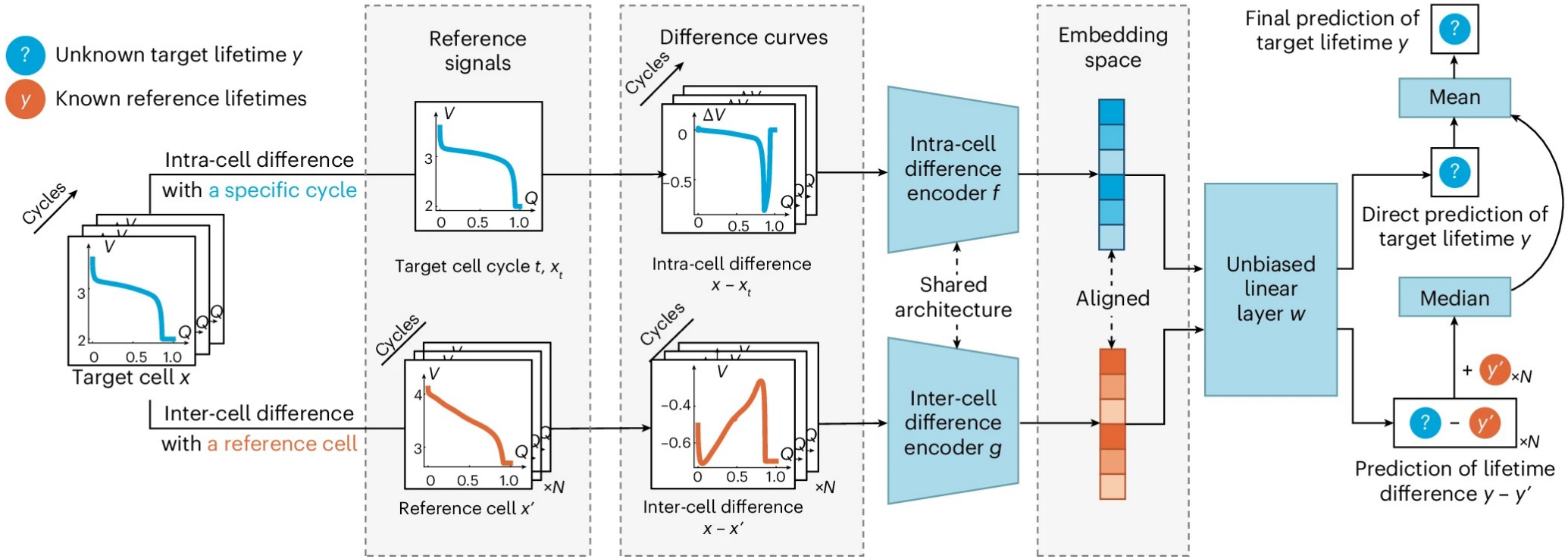
Smolen, Moore, Perez & Wooley, PNAS 122(42), e2509745122 (2025)

5

Gestion des systèmes

DÉPLOYER

Prédire précocement la durée de vie en cycles des batteries

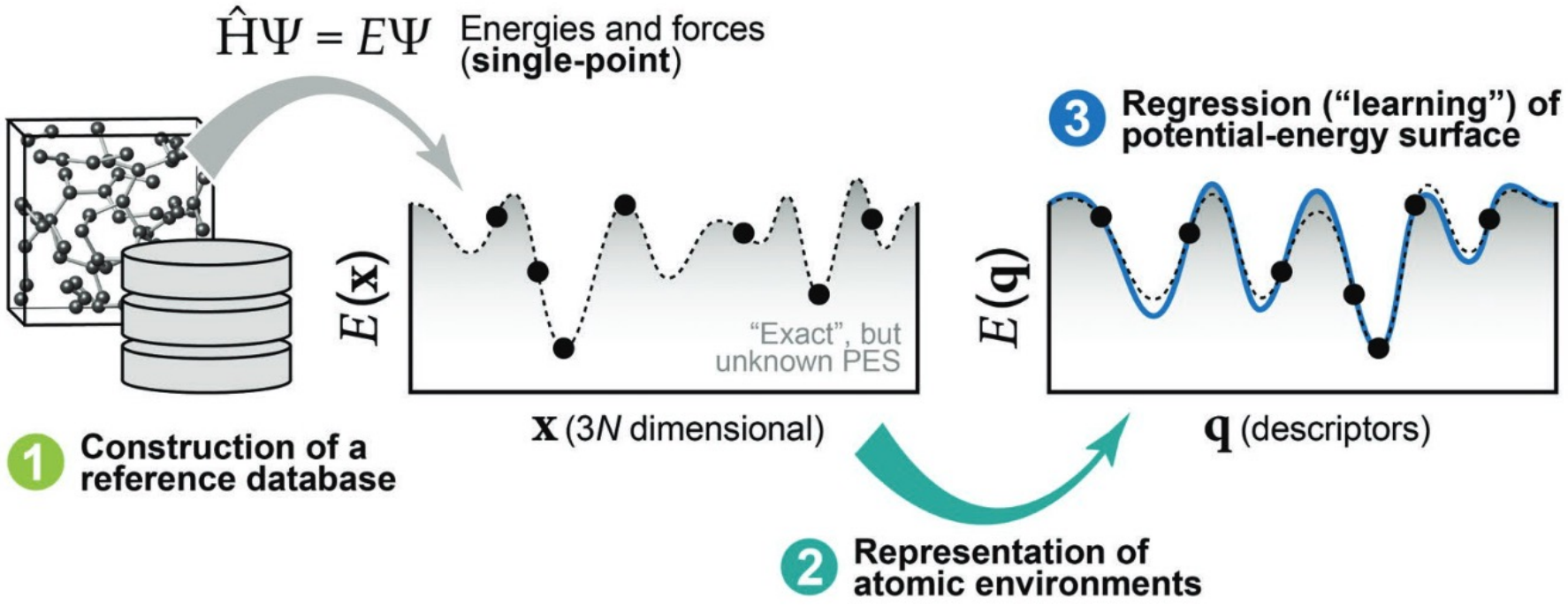


BatLiNet compare la cellule cible à une cellule de « référence » entièrement cyclée dont la durée de vie est déjà connue

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

Comment fonctionne un potentiel interatomique par apprentissage automatique

Une même recette pour tous : environnement atomique local → énergie, apprise à partir de la DFT



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

Les modèles MLIP en un coup d'œil

Modèles représentatifs par famille : idée clé, usage typique et source

Modèle	Famille	Idée clé	Application typique	Référence
HDNNP	Descripteur + RN	fonctions de symétrie → réseau de neurones	solides massifs, eau	*Behler & Parrinello 2007*
GAP	Descripteur + noyau	SOAP + processus gaussien	Si, C, amorphes ; économe en données	*Bartók et al. 2010*
DeePMD	Descripteur appris + RN	potentiel profond de bout en bout	DM à grande échelle (eau, alliages)	*Zhang et al. 2018*
SNAP	Développement linéaire	composantes du bispectre (linéaire)	métaux (Ta, W)	*Thompson et al. 2015*
MTP	Développement linéaire	tenseurs de moments	alliages ; apprentissage actif	*Shapeev 2016*
ACE	Développement par ordre de corps	développement en amas atomiques	général ; rapide + précis	*Drautz 2019*
SchNet	GNN invariant	convolutions à filtre continu	molécules (MD17, QM9)	*Schütt et al. 2018*
NequIP	GNN équivariant	messages E(3)-équivariants	molécules + matériaux, économe en données	*Batzner et al. 2022*
Allegro	GNN équivariant local	équivariant strictement local	DM à très grande échelle	*Musaelian et al. 2023*
MACE	Équivariant + ordre de corps élevé	messages d'ordre supérieur par couche	DM précise et rapide	*Batatia et al. 2022*
M3GNet	GNN universel	potentiel en graphe entraîné sur MP	relaxation/criblage de cristaux	*Chen & Ong 2022*
CHGNet	Universel (sensible à la charge)	GNN informé par la charge	Li-ion, oxydes, magnétisme	*Deng et al. 2023*
MACE-MP-0	Fondation (équivariant)	modèle de fondation couvrant le tableau périodique	DM prête à l'emploi pour toutes les chimies	*Batatia et al. 2024*

SOAP = recouvrement lisse des positions atomiques (smooth overlap of atomic positions) · E(3) = rotations, translations et réflexions en 3D · MP = Materials Project · GNN = réseau de neurones en graphe.

Potentiels interatomiques par apprentissage automatique, dans l'ordre à peu près chronologique

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

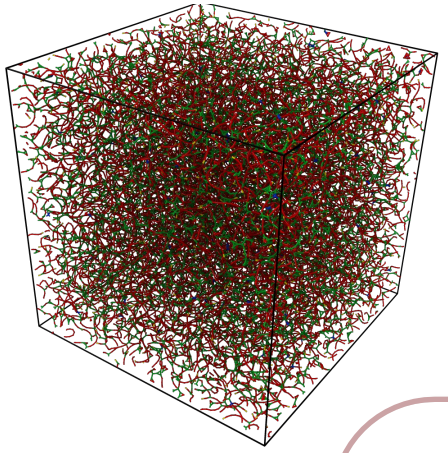
13. Merchant, A. et al. Scaling deep learning for materials discovery (GNoME). Nature 624, 80–85 (2023). 10.1038/s41586-023-06735-9

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

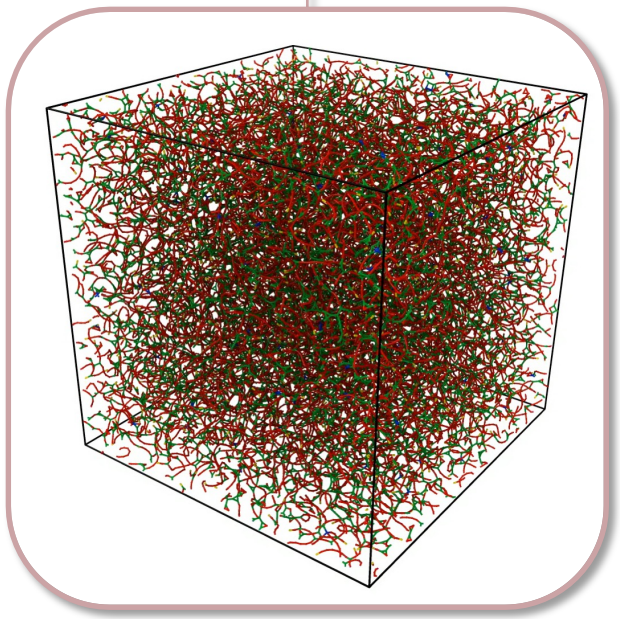
Modéliser le biochar avec des MLIP (GAP20)

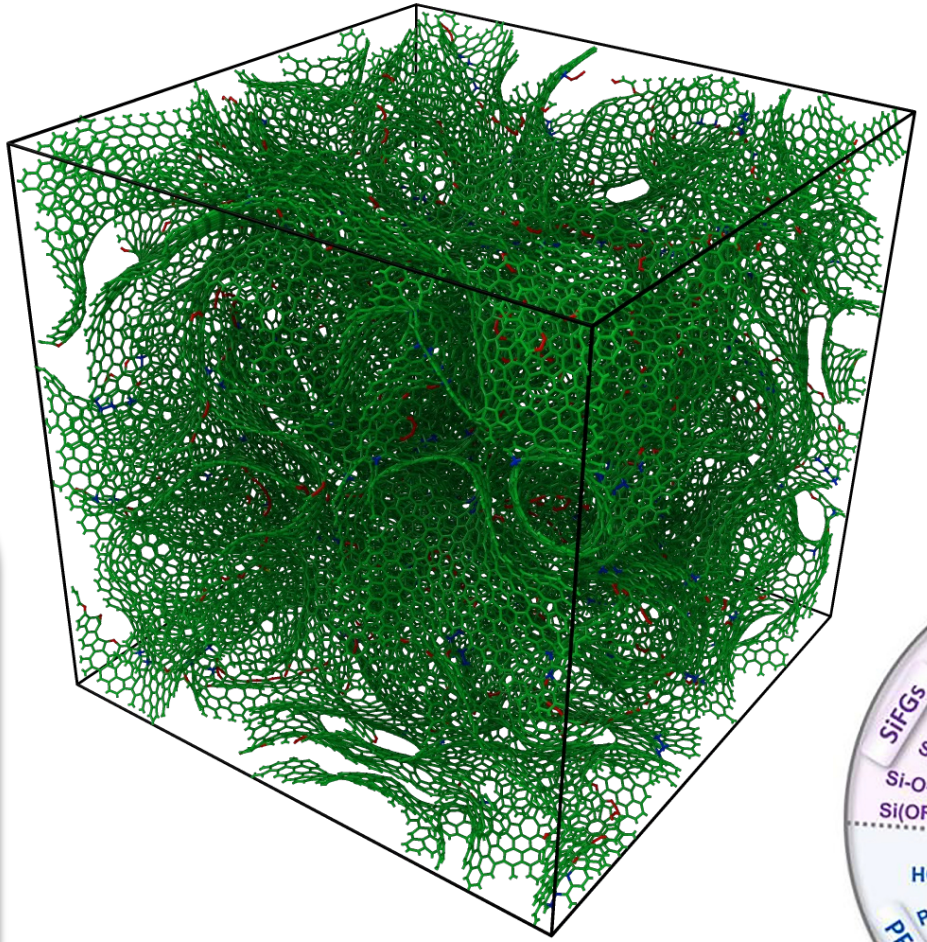
Échelle atomique

Caractérisation structurale :
Porosité, structure atomique, hybridation du carbone, ordre structural

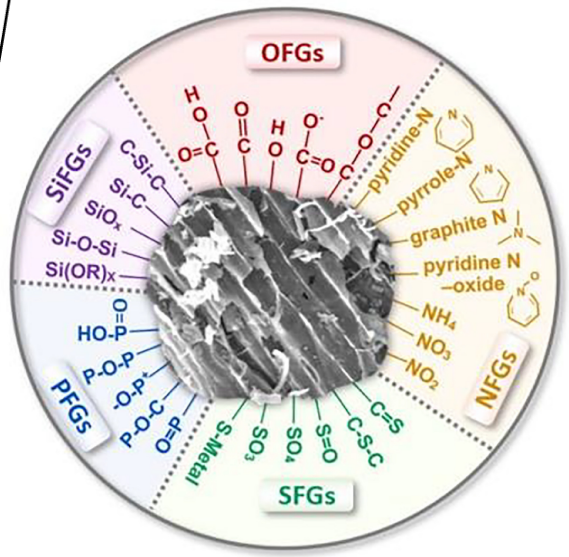


Simulation de DM





Biochar Modèle



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

PARTICIPATION

CLÔTURE DE SECTION · QUIZ

Quiz éclair – Vrai/Faux

1 **Un potentiel interatomique par AA vise la précision DFT à bien moindre coût.**

Vrai | Faux

2 **Un modèle à 100 % de précision à l'entraînement est toujours le plus fiable sur de nouvelles données.**

Vrai | **Faux**

3 **Les candidats générés par l'IA nécessitent toujours une validation par DFT/expérimentale.**

Vrai | Faux

POURQUOI C’EST IMPORTANT POUR LE CLIMAT

Une seule boîte à outils pour tous les matériaux d’énergie propre.

- Les potentiels par AA offrent une précision proche de la DFT pour une fraction de l’énergie et du temps : on peut cribler des millions de candidats.
- Les laboratoires autonomes transforment rapidement les meilleures prédictions en échantillons validés.
- La validation évite de gaspiller un temps de laboratoire précieux sur des faux positifs.

BOÎTE À OUTILS IA

Messages clés

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. Les laboratoires autonomes transforment rapidement les meilleures prédictions en échantillons validés.

2. Les MLIP offrent la précision de la DFT à la vitesse de la DM : la case « accélérer la simulation » devient réalité.

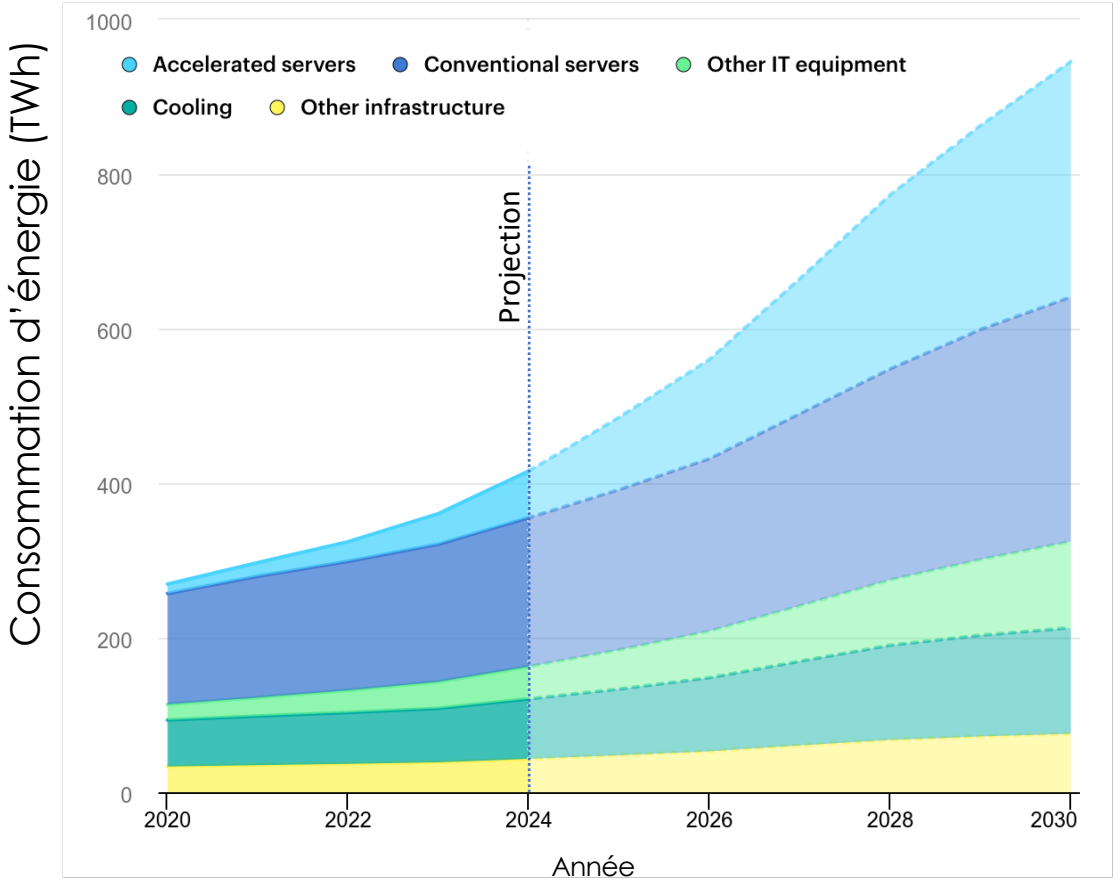
3. Générer est facile ; valider est difficile

DES QUESTIONS ?

PARLER CHIMIE AUX ORDINATEURS

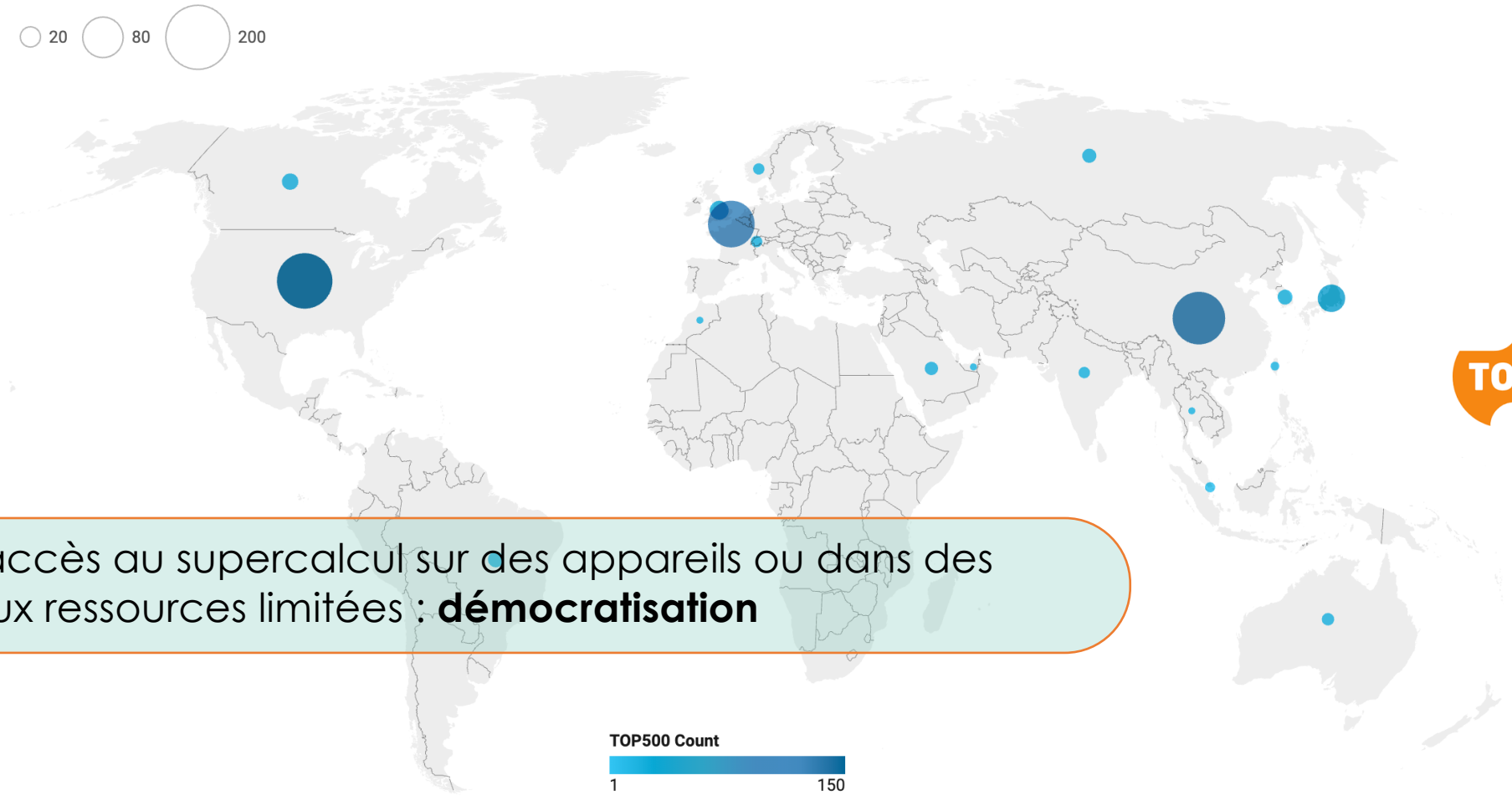
Coût énergétique croissant de l'IA et des centres de données

Besoin d'une informatique plus verte



Source : Agence internationale de l'énergie (AIE)

Où se trouve le supercalcul ?

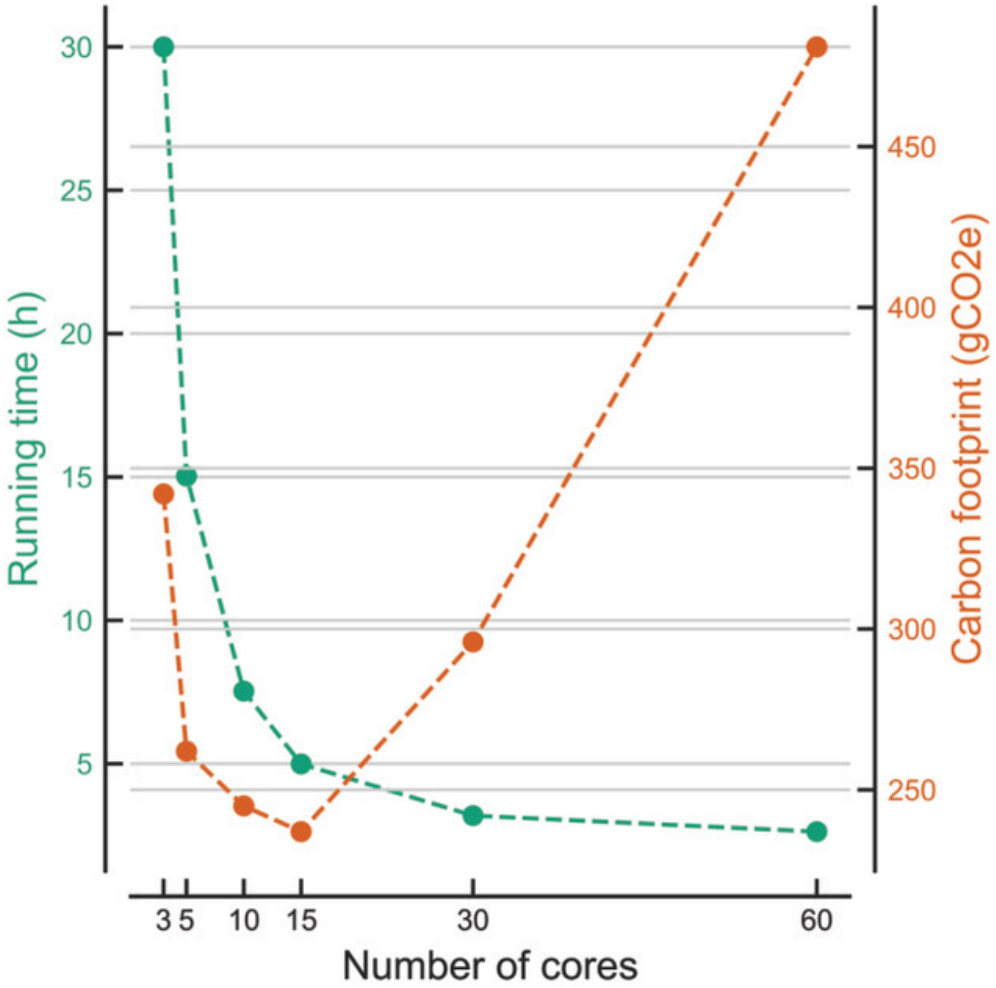


Améliorer l'accès au supercalcul sur des appareils ou dans des contextes aux ressources limitées : **démocratisation**

Source : top500.org

Algorithmes verts

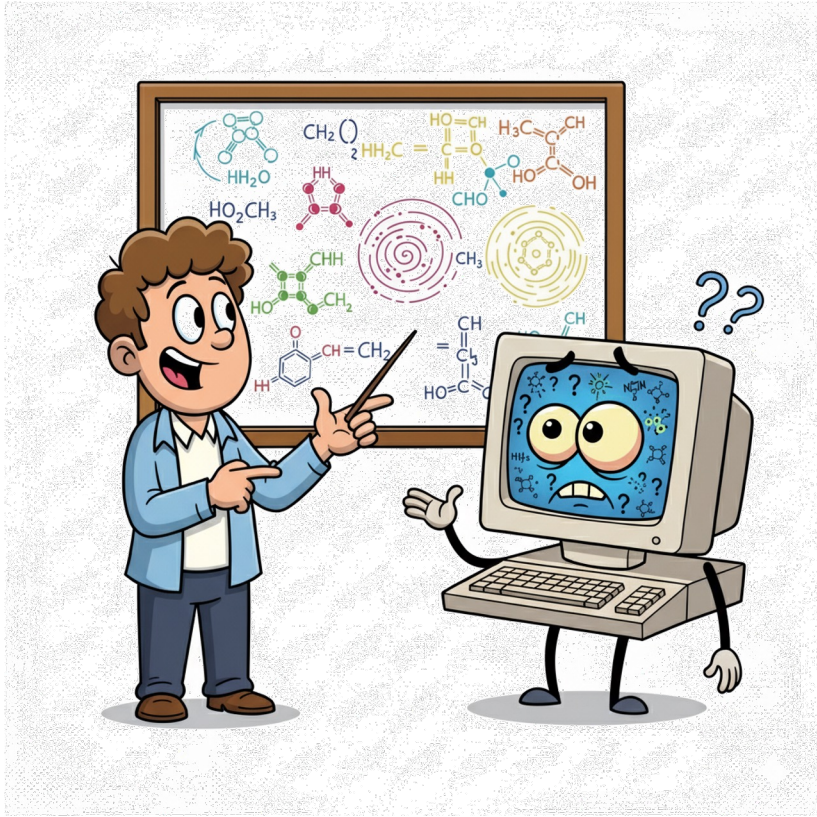
PARLER CHIMIE





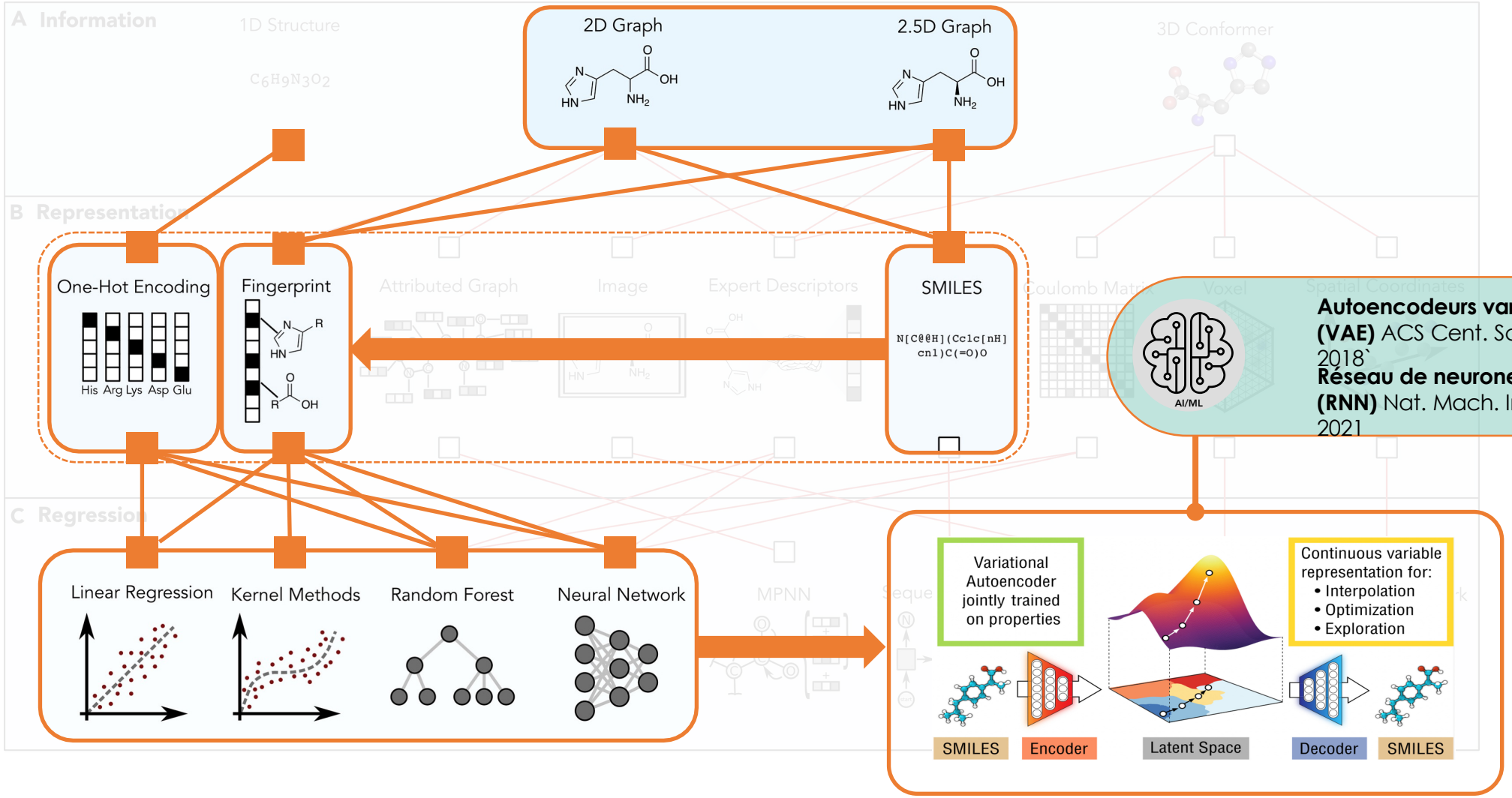
Lannelongue et al., Advanced Science, 8, 2021

Un pas en arrière : comment parler chimie aux ordinateurs ?



L'apprentissage automatique pour la chimie exige de **vectoriser les molécules**

L'AA pour la chimie exige de vectoriser les molécules



Autoencodeurs variationnels (VAE) ACS Cent. Sci., 4 , 268, 2018`

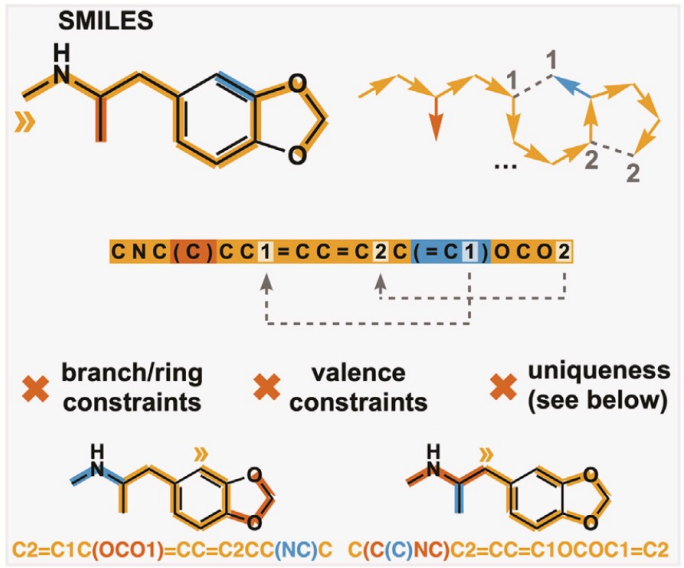
Réseau de neurones récurrent (RNN) Nat. Mach. Intell., 3 , 759, 2021

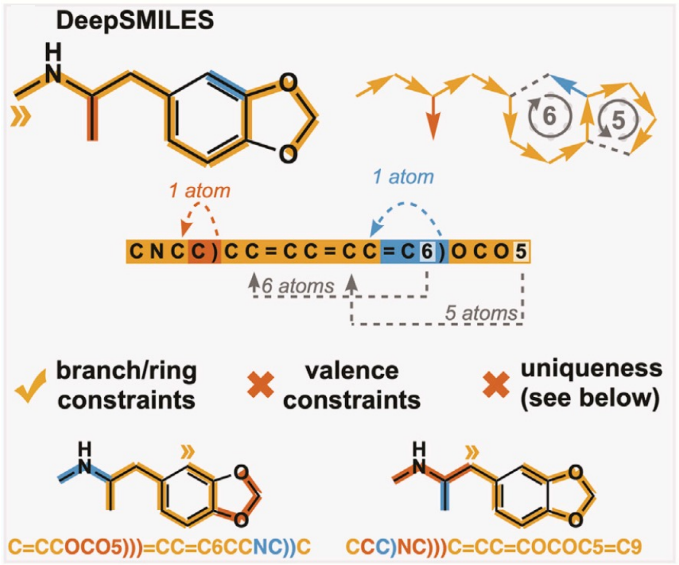
Pattanaik and Coley, **Chem 6, 1204-1207, 2020**

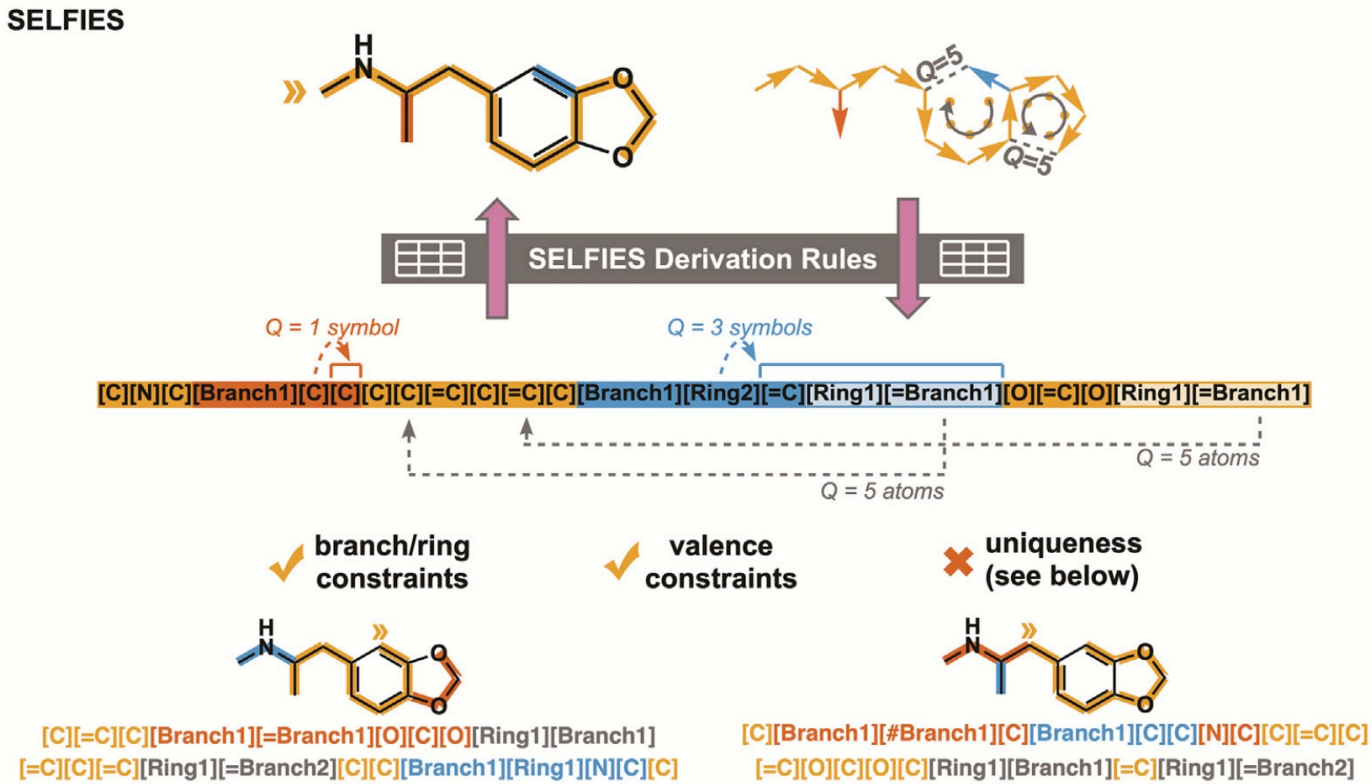
Skinnider et al. **Nat. Mach. Intell.**, 3 , 759, **2021**

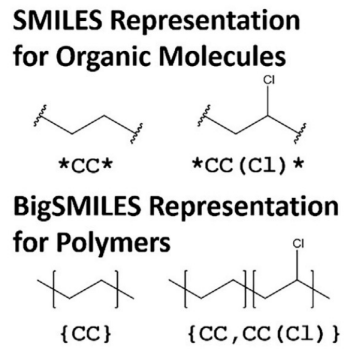
Gomez-Bombarelli et al. **ACS Cent Sci.** 4, 268–76, **2018**

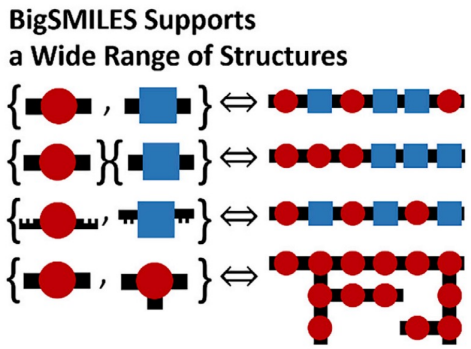
Encodage des données lisible par machine





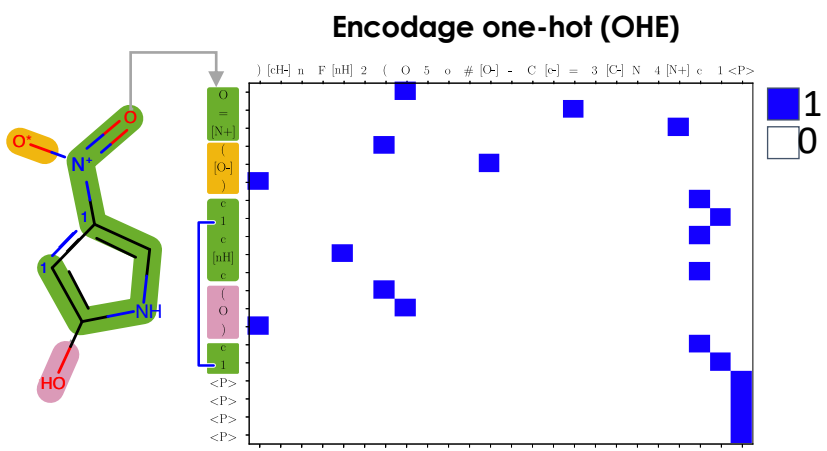


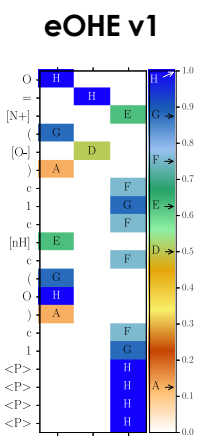




Krenn et al. Patterns 3, 10, 100588, 2022

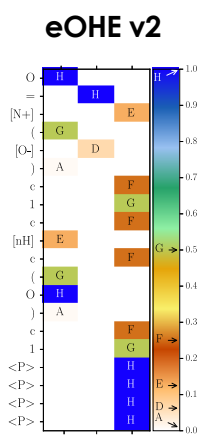
Encodages one-hot intégrés (eOHE)





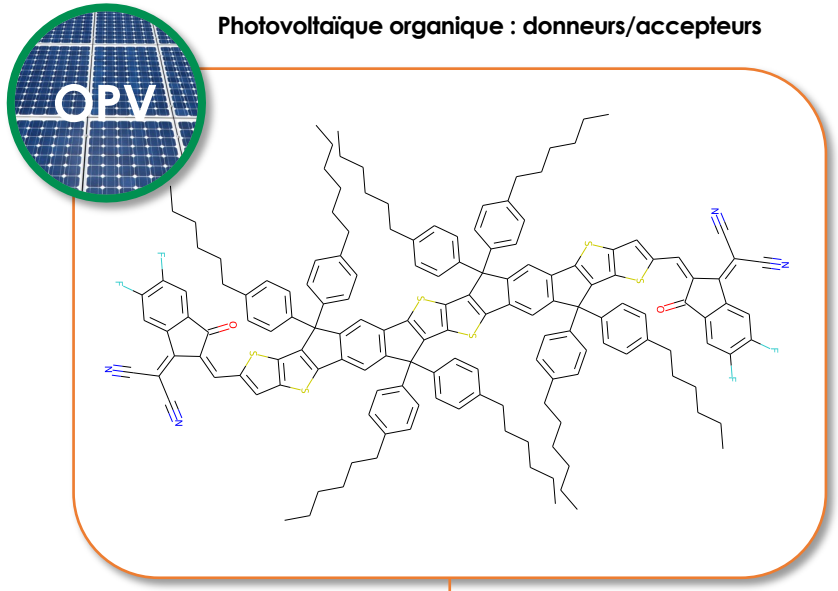
Réduction linéaire
$q = 8$

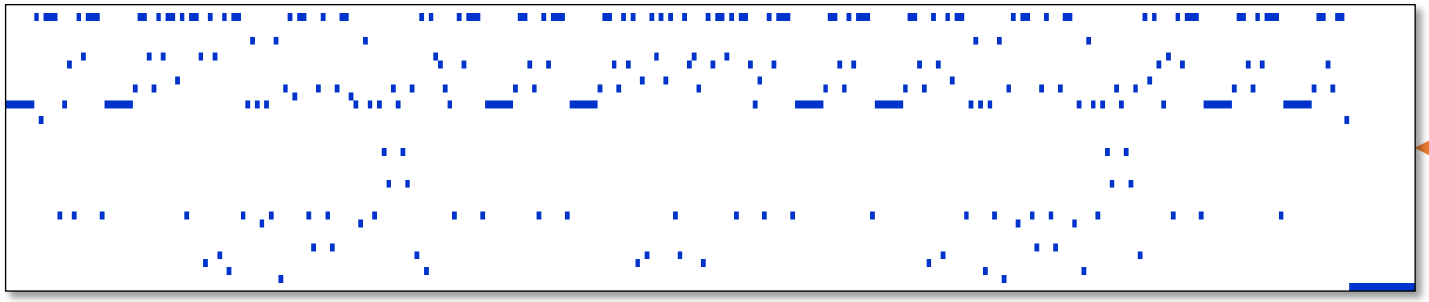
$$v_1(r,q) = \frac{r+1}{q}$$



Réduction par puissances de 2
$q = 8$

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

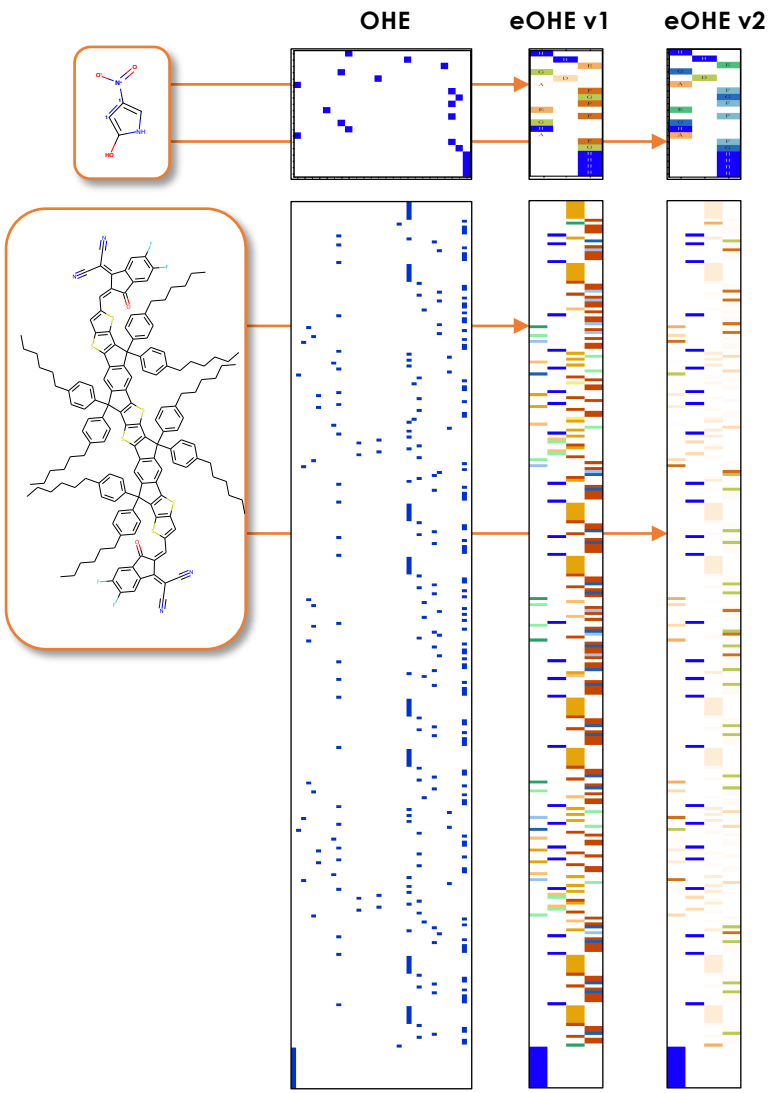


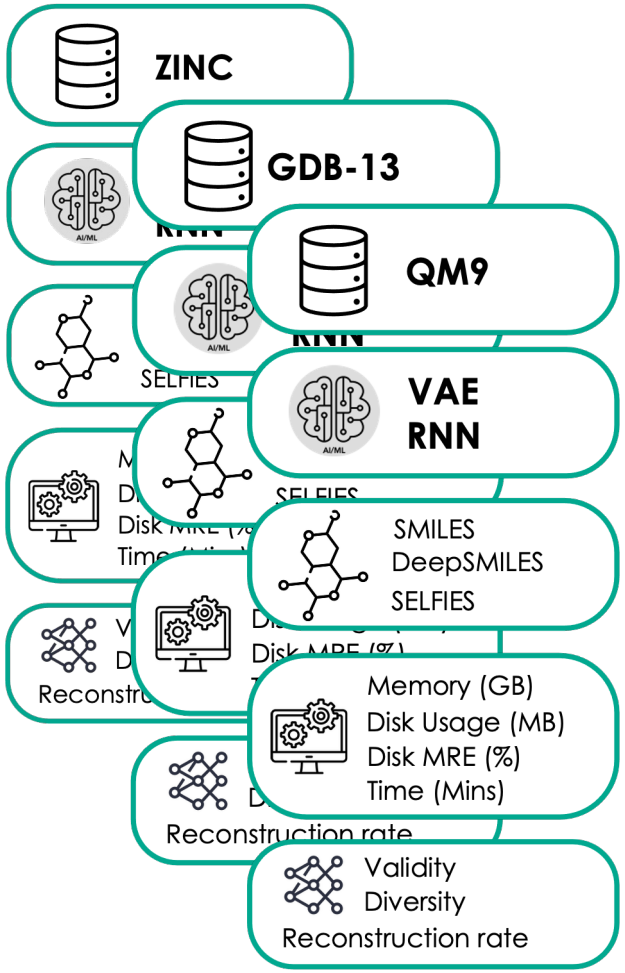


CCCCCCc1ccc(C2(c3ccc(CCCCCC)cc3)c3cc4c(cc3-c3sc5cc(C=C6C(=O)c7cc(F)c(F)cc7C6=C(C#N)C#N)sc5c32)C(c2ccc(CCCCCC)cc2)(c2ccc(CCCCCC)cc2)c2c-4sc3c4c(sc23)-c2cc3c(cc2C4(c2ccc(CCCCCC)cc2)c2ccc(CCCCCC)cc2)-c2sc4cc(C=C5C(=O)c6cc(F)c(F)cc6C5=C(C#N)C#N)sc4c2C3(c2ccc(CCCCCC) cc2)c2ccc(CCCCCC)cc2)cc1

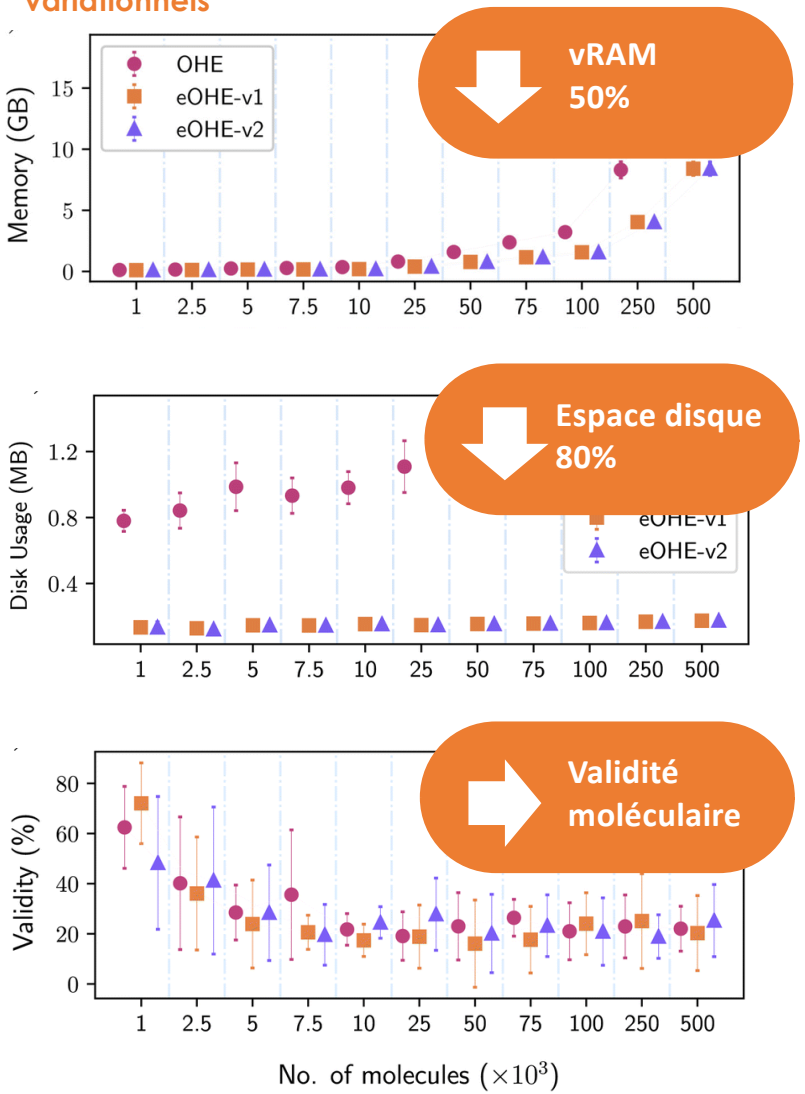
Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez,. **Digital Discovery**, 4, 776-789, **2025**

Codification lisible par machine plus efficace





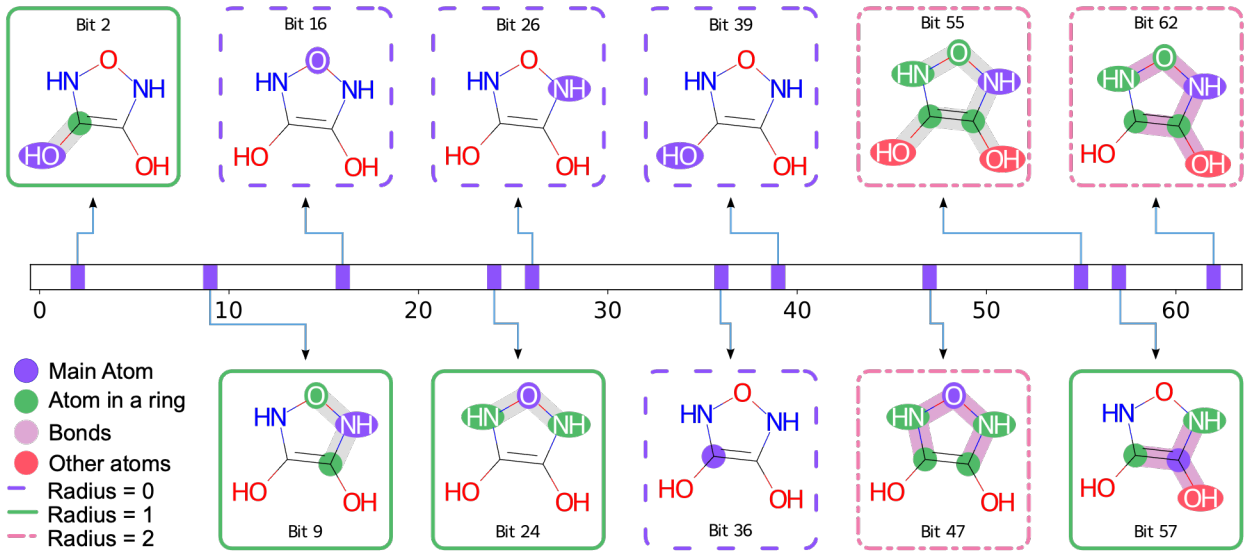
Performances dans les autoencodeurs variationnels



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

Empreintes de Morgan intégrées (eMP)

Une empreinte de Morgan est un vecteur binaire de taille fixe qui encode la présence (1) ou l'absence (0) de sous-structures dans une molécule.

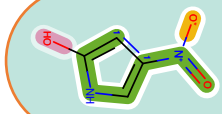




QM9 : 134 000+ molécules
NFA : 50 000 molécules
RedDB : 15 000 molécules



Forêt aléatoire (RF), Perceptron multicouche (MLP), Régresseur KNN (KNR), régresseur à gradient boosting (GBR), réseau de neurones profond (DNN)



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

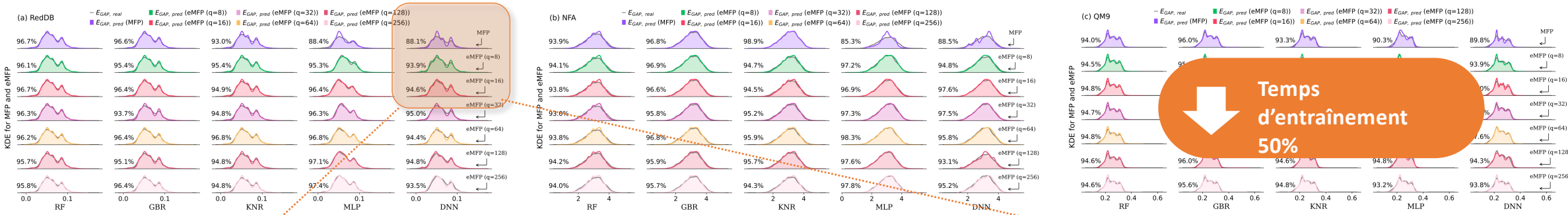


150 essais pour optimiser 15 750 hyperparamètres

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

Écart énergétique avec les empreintes de Morgan intégrées (eMP)

PARLER CHIMIE



Temps d'entraînement 50%

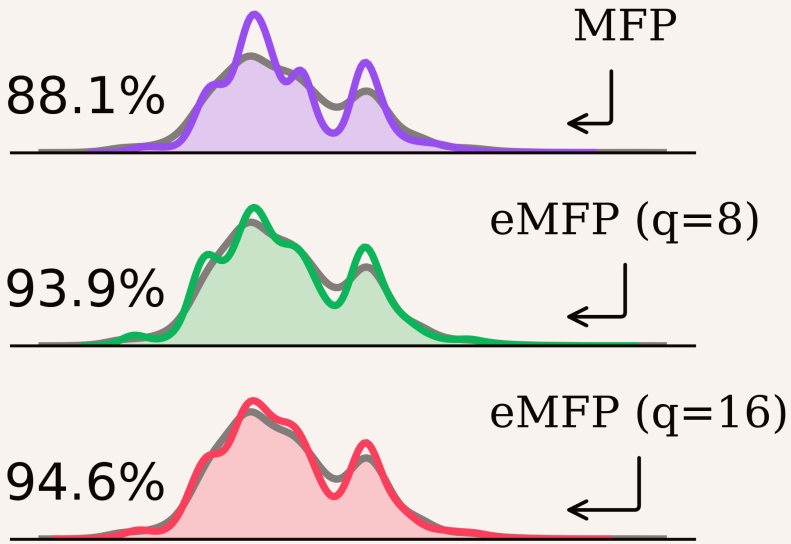
Base RedDB

– $E_{GAP,\ real}$

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

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

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



Nuñez-Andrade, Vidal-Daza, Gomez-Bombarelli, Ryan, Martin-Martinez, *en cours d'évaluation,* **2026**

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

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES

B) SELFIES

C) InChI

D) Aucune de celles-ci

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES

B) SELFIES

C) InChI

D) Aucune de celles-ci

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES

B) **SELFIES**

C) InChI

D) Aucune de celles-ci

Q2 **eOHE/eMFP améliorent principalement…**

A) La précision uniquement

B) Mémoire/disque/temps à précision égale

C) Le nombre de publications

D) Rien

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES
B) SELFIES
C) InChI
D) Aucune de celles-ci

Q2 **eOHE/eMFP améliorent principalement…**

A) La précision uniquement
B) Mémoire/disque/temps à précision égale
C) Le nombre de publications
D) Rien

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES
B) SELFIES
C) InChI
D) Aucune de celles-ci

Q2 **eOHE/eMFP améliorent principalement…**

A) La précision uniquement
B) Mémoire/disque/temps à précision égale
C) Le nombre de publications
D) Rien

Q3 **Pourquoi le coût énergétique de l'IA est-il important ?**

A) Il ne l'est pas
B) Durabilité et démocratisation
C) Le calcul est gratuit
D) Uniquement pour les jeux vidéo

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Quelle représentation garantit une molécule valide pour chaque chaîne ?**

A) SMILES
B) SELFIES
C) InChI
D) Aucune de celles-ci

Q2 **eOHE/eMFP améliorent principalement…**

A) La précision uniquement
B) Mémoire/disque/temps à précision égale
C) Le nombre de publications
D) Rien

Q3 **Pourquoi le coût énergétique de l'IA est-il important ?**

A) Il ne l'est pas
B) Durabilité et démocratisation
C) Le calcul est gratuit
D) Uniquement pour les jeux vidéo

PARLER CHIMIE

Messages clés – La représentation est une décision de durabilité.

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. La conversion en caractéristiques est une décision de conception de premier plan, pas une étape de prétraitement accessoire.

2. eOHE/eMFP réduisent la mémoire, le disque et le temps à précision égale efficacité = calcul plus vert et plus démocratique.

3. Une informatique verte et accessible fait partie d'une accélération responsable.

DES QUESTIONS ?

DÉVELOPPEMENT DE MATÉRIAUX GUIDÉ PAR LES LLM

Développement de matériaux guidé par les LLM : une nouvelle couche dans la pile

Au-delà de la prédiction de valeurs : des modèles qui lisent, raisonnent, planifient et utilisent des outils

Le changement : l'AA spécialisé associe une structure à une propriété. Un grand modèle de langage y ajoute une interface en langage naturel, de vastes connaissances préalables, une planification étape par étape et la capacité d'appeler des outils externes, **ce qui transforme un prédicteur en collaborateur de recherche**.

1

Interface en langage naturel

On l'interroge, on l'adapte avec quelques exemples ; performant quand les données sont rares, là où l'AA classique peine.

2

Connaissances du monde

Préentraîné sur la littérature scientifique : il apporte des connaissances chimiques préalables et du contexte, pas seulement un ajustement de courbes.

3

Planification et raisonnement

Décomposer un objectif (« trouver une cathode Na-ion stable ») en étapes ordonnées et vérifiables.

4

Utilisation d'outils

Appeler des simulateurs, des bases de données et des robots, pour fonder les réponses sur de vrais calculs au lieu de deviner.

… mais prudence. Les LLM sont fluides mais peu fiables. Ils inventent des faits, citent mal leurs sources et peuvent se tromper avec assurance. Tout dépend de **l'ancrage dans des outils et de la validation**

Ce que les LLM font réellement dans le développement de matériaux

1 Lire la littérature

Transformer des milliers d'articles en bases de données structurées et interrogeables (compositions, MOF, dopants).

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

2 Prédire des propriétés

Répondre à des questions de chimie en langage naturel ; compétitif face à l'AA, performant quand les données sont rares.

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

3 Générer et concevoir à l'inverse

Écrire molécules et cristaux sous forme de texte ; génération orientée par les propriétés.

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

4 Planifier la synthèse

Proposer des voies de synthèse et des recettes ; rétrosynthèse et conception de protocoles.

p. ex. **ChemCrow 2024**

5 Raisonner avec des outils

Doter le modèle d'outils experts (RDKit, calculateurs, recherche) pour qu'il cesse de deviner.

p. ex. **ChemCrow – 18 outils**

6 Mener des expériences de manière autonome

Boucles planifier-agir-observer qui pilotent instruments et robots pour tester des hypothèses.

p. ex. **Coscientist 2023**

Systèmes LLM marquants pour la chimie et les matériaux

Systèmes représentatifs : ce qu'ils font, l'approche et la source

Système	Fonction	LLM/approche	Référence
Coscientist	conçoit, planifie et mène des expériences (p. ex. couplages croisés au Pd)	GPT-4 + web/docs/code + robotique	*Boiko et al., Nature 2023*
ChemCrow	synthèse organique, conception de médicaments et de matériaux ; a réalisé de vraies synthèses	GPT-4 + 18 outils experts	*Bran et al., Nat. Mach. Intell. 2024*
LLM de chimie prédictive	prédiction et classification de propriétés à partir de texte ; peu de données	GPT-3 ajusté ; en contexte	*Jablonka et al., Nat. Mach. Intell. 2024*
Extraction structurée	lit des articles → enregistrements structurés (dopants, MOF, compositions)	GPT-3/Llama-2 ajustés	*Dagdelen et al., Nat. Commun. 2024*
LLM cristal-texte	génère des cristaux inorganiques stables écrits sous forme de texte	LLaMA-2 70B ajusté	*Gruver et al., ICLR 2024*
Chemeleon	génération de structures cristallines guidée par le texte	contrastif texte–cristal + diffusion	*Park et al., Nat. Commun. 2025*

Références

LLM et agents autonomes pour la chimie et les matériaux

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

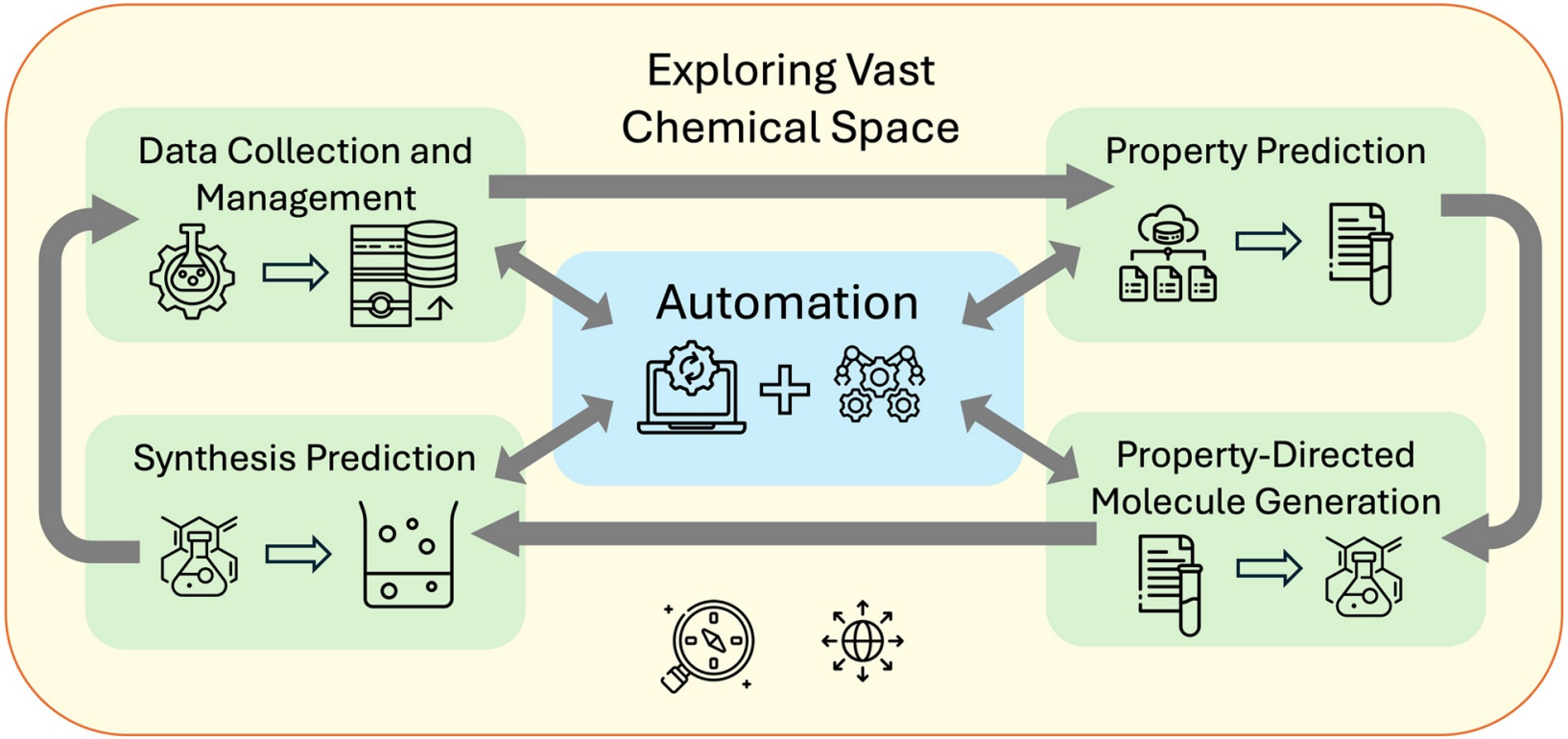
5. Gruver, N.; Sriram, A.; Madotto, A.; Wilson, A. G.; Zitnick, C. L.; Ulissi, Z. Fine-Tuned Language Models Generate Stable Inorganic Materials as Text. ICLR 2024. arXiv:2402.04379

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

Que peuvent réellement faire les LLM et les agents en chimie ?

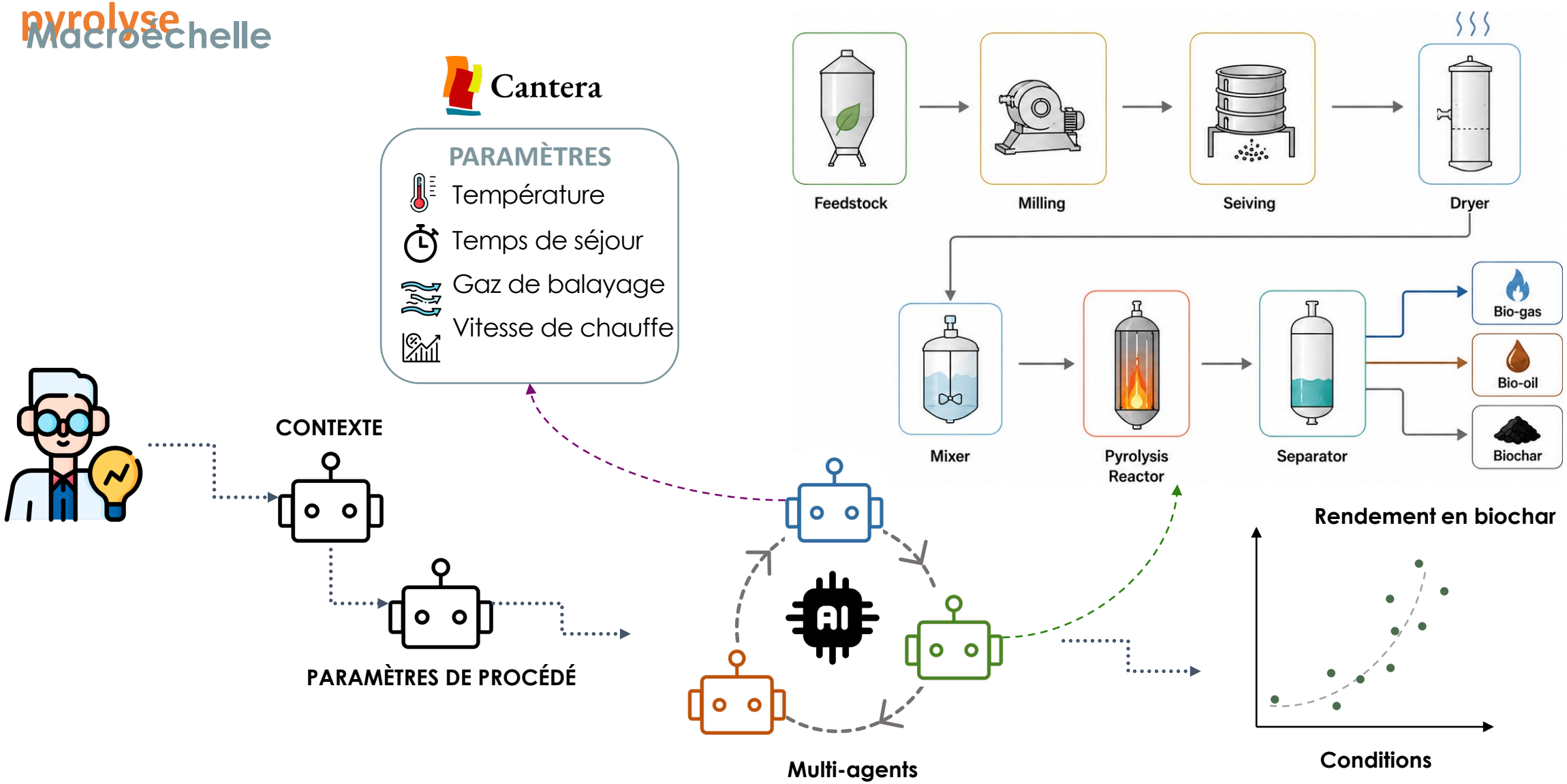


Butler, Davies, Cartwright, Isayev, Walsh, Nature 559, 547–555 (2018).

Modéliser le biochar – ingénierie du procédé de pyrolyse

DÉV. GUIDÉ PAR LES LLM

Macroéchelle



PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions

B) Planifie et agit au moyen d'outils, en boucle

C) N'utilise jamais d'outils

D) Fonctionne uniquement hors ligne

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions

B) Planifie et agit au moyen d'outils, en boucle

C) N'utilise jamais d'outils

D) Fonctionne uniquement hors ligne

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions

B) Planifie et agit au moyen d'outils, en boucle

C) N'utilise jamais d'outils

D) Fonctionne uniquement hors ligne

Q2 **Doter un LLM d'outils (RDKit, recherche, simulateurs)…**

A) Le rend moins performant

B) L'ancre et améliore sa fiabilité

C) Supprime toutes les erreurs

D) Est impossible

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions

B) Planifie et agit au moyen d'outils, en boucle

C) N'utilise jamais d'outils

D) Fonctionne uniquement hors ligne

Q2 **Doter un LLM d'outils (RDKit, recherche, simulateurs)…**

A) Le rend moins performant

B) L'ancre et améliore sa fiabilité

C) Supprime toutes les erreurs

D) Est impossible

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions

B) Planifie et agit au moyen d'outils, en boucle

C) N'utilise jamais d'outils

D) Fonctionne uniquement hors ligne

Q2 **Doter un LLM d'outils (RDKit, recherche, simulateurs)…**

A) Le rend moins performant

B) L'ancre et améliore sa fiabilité

C) Supprime toutes les erreurs

D) Est impossible

Q3 **Les LLM sont particulièrement utiles quand…**

A) Les données sont massives

B) Les données sont rares (peu de données)

C) Jamais

D) Uniquement pour les images

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (choix multiple)

Choix multiple

Q1 **Un « agent » LLM (par opposition à un assistant) est un système qui…**

A) Se contente de répondre aux questions
B) Planifie et agit au moyen d'outils, en boucle
C) N'utilise jamais d'outils
D) Fonctionne uniquement hors ligne

Q2 **Doter un LLM d'outils (RDKit, recherche, simulateurs)…**

A) Le rend moins performant
B) L'ancre et améliore sa fiabilité
C) Supprime toutes les erreurs
D) Est impossible

Q3 **Les LLM sont particulièrement utiles quand…**

A) Les données sont massives
B) Les données sont rares (peu de données)
C) Jamais
D) Uniquement pour les images

Messages clés – Libérer le savoir prisonnier des textes.

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. Les LLM ajoutent une couche langage + raisonnement + utilisation d'outils à l'AA spécialisé.

2. Transformer des décennies de littérature en données exploitables ; planifier des synthèses ; orchestrer des outils.

3. Garder un humain responsable et consigner chaque action. La traçabilité et la supervision rendent l'autonomie digne de confiance (et reproductible).

DES QUESTIONS ?

IA RESPONSABLE



Une IA responsable pour les matériaux et la chimie

Des outils puissants impliquent des responsabilités. La même accélération est à double tranchant

1 Intégrité et reproductibilité

Les fuites de données et la course aux classements gonflent les résultats.
Documenter suffisamment pour que d'autres puissent les reproduire.

2 Des prédictions fiables

Indiquer l'incertitude, respecter le domaine d'applicabilité et valider en aval (DFT → expérience).

3 Qualité des données et biais

Les jeux de données privilégient les systèmes bien étudiés et ayant fonctionné.
Données inexactes, résultats inexacts

4 Ouverture et FAIR

Partager données, code et poids des modèles
Faciles à trouver, accessibles, interopérables, réutilisables (FAIR).

5 Sécurité et double usage

Des outils de conception puissants peuvent être détournés pour nuire.
Contrôle d'accès et supervision humaine.

6 Durabilité et calcul

L'entraînement et l'inférence consomment de l'énergie.
Dimensionner les modèles au juste besoin et réutiliser des modèles préentraînés.

Biais des données

Défis

- Les croyances implicites et les biais issus de contextes occidentaux peuvent être normalisés dans l'IA, la rendant moins responsable à l'échelle mondiale.
- Il faut adopter une perspective mondiale dans les hypothèses fondamentales de l'IA : données d'entraînement, performances des modèles et **explicabilité** des systèmes
- Poser très tôt les questions difficiles sur les impacts positifs et négatifs, plutôt que de recourir à des réparations et des correctifs a posteriori.

Actions

- Être robuste au multilinguisme et à la diversité des niveaux de littératie
- Personnaliser les expériences grâce au modèle d'identité partagée
- **L'explicabilité** et la transparence doivent avoir du sens pour les nouveaux utilisateurs de la technologie
- Traiter de manière proactive l'équité face aux déséquilibres des données

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

Références

Une IA responsable et digne de confiance au service de la science

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

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

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

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

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

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

PARTICIPATION

CLÔTURE DE SECTION · QUIZ (vrai/faux)

Vrai ou faux ?

1. **Les jeux de données de chimie publiés sont biaisés, en partie parce que les expériences ratées sont rarement rapportées.**

   Vrai | Faux

2. **Un modèle de conception de médicaments peut être détourné pour concevoir des toxines.**

   Vrai | Faux

3. **FAIR signifie Findable, Accessible, Interoperable, Reusable (faciles à trouver, accessibles, interopérables, réutilisables).**

   Vrai | Faux

Messages clés – L'accélération n'est utile que si elle est fiable et partagée.

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. Les données sont biaisées : biais de publication, biais humain, présupposés propres au contexte occidental.

2. Les outils génératifs/agentiques puissants sont intrinsèquement à double usage.

3. La reproductibilité, l'ouverture (FAIR) et le coût de calcul font partie intégrante d'un travail bien fait.

DES QUESTIONS ?

IA EXPLICABLE

IA explicable : de la prédiction à la compréhension

En science, on veut généralement le « pourquoi », pas seulement le chiffre

Le fossé : un modèle peut prédire une propriété avec précision sans rien vous apprendre. L'explicabilité cherche à savoir ce que le modèle a appris, s'il l'a appris pour de bonnes raisons et quelle nouvelle science il laisse entrevoir.

1 Compréhension

Transformer un modèle ajusté en règles de conception et en mécanismes. Retrouver la physique/la chimie derrière les données.

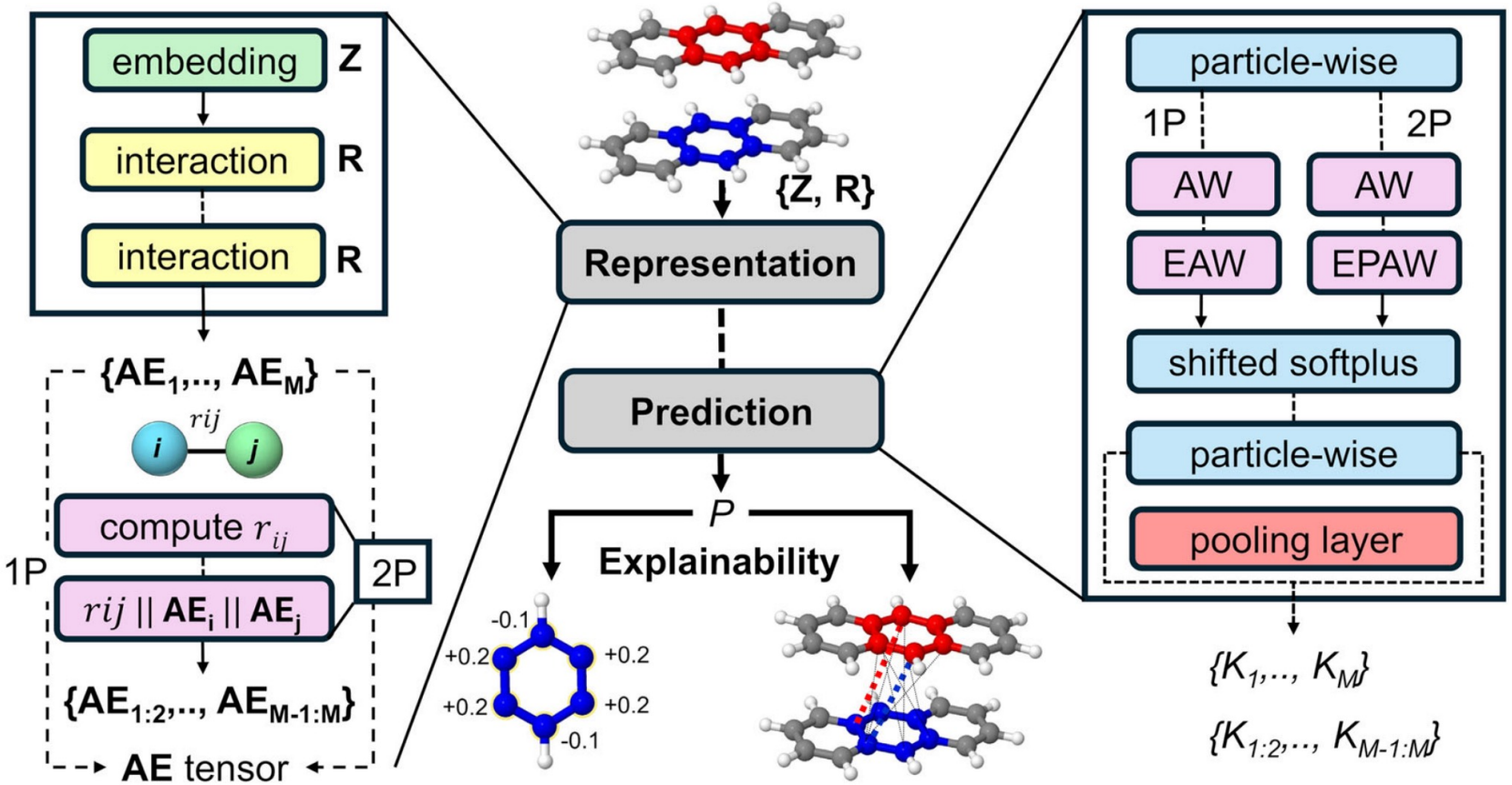
2 Confiance et validation

Vérifier que le modèle prédit pour de bonnes raisons, et non à cause de fausses corrélations ou d'artefacts du jeu de données.

3 Découverte

L'interprétation fait émerger de nouvelles relations structure–propriété et des hypothèses vérifiables.

IA chimique 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

Références

IA explicable et interprétable pour la science et les matériaux

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

Messages clés – Transformer des prédictions rapides en règles de conception fiables.

POURQUOI C'EST IMPORTANT POUR LE CLIMAT

1. En science, on veut le pourquoi, pas seulement le chiffre.

2. Les méthodes interprétables démasquent les modèles qui ont raison pour de mauvaises raisons.

3. Traiter chaque explication comme une hypothèse à vérifier par la physique et l'expérience.

DES QUESTIONS ?

RÉFLEXIONS 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.



Concevoir pour la circularité et la performance

Concevoir pour la performance

AI/ML

Dépôts en ligne : boîte à outils open source (matériaux et IA)

Un kit de démarrage à partager. Les outils de notre groupe sont mis en évidence ; vérifier que les liens sont à jour avant diffusion.

Domaine	Outil	Utilité	Dépôt
Données et outils	**pymatgen**	Analyse des matériaux, E/S et API Materials Project	*github.com/materialsproject/pymatgen*
Données et outils	**matminer**	Descripteurs/caractéristiques prêts à l'emploi pour l'AA	*github.com/hackingmaterials/matminer*
Données et outils	**RDKit**	Chémoinformatique : molécules, empreintes, descripteurs	*github.com/rdkit/rdkit*
Données et outils	**DeepChem**	Boîte à outils d'AA pour la chimie : jeux de données, modèles, bancs d'essai	*github.com/deepchem/deepchem*
Données et outils	**ASE**	Atomic Simulation Environment : construire/exécuter/analyser	*gitlab.com/ase/ase*
Représentations	**SELFIES**	Représentation moléculaire sous forme de chaîne toujours valide	*github.com/aspuru-guzik-group/selfies*
Représentations	**eOHE**	Encodages one-hot intégrés (entrées efficaces)	*github.com/MMLabCodes/eOHE*
Représentations	**eMFP**	Empreintes de Morgan intégrées (compactes, plus rapides)	*github.com/MMLabCodes/eMFP*
Génératif	**chemical_vae**	Conception moléculaire dans un espace latent continu	*github.com/aspuru-guzik-group/chemical_vae*
Génératif	**MatterGen**	Modèle génératif de diffusion pour cristaux inorganiques	*github.com/microsoft/mattergen*
Potentiels par AA	**MACE**	Potentiels interatomiques par AA (+ modèle de fondation MACE-MP-0)	*github.com/ACEsuit/mace*
Potentiels par AA	**CHGNet**	Potentiel universel par réseau de neurones informé par la charge	*github.com/CederGroupHub/chgnet*
Potentiels par AA	**MatGL (M3GNet)**	Potentiels universels en graphe + modèles de propriétés	*github.com/materialsvirtuallab/matgl*
Autonomie	**Ax / BoTorch**	Optimisation bayésienne pour laboratoires autonomes	*github.com/facebook/Ax*
Agents LLM	**ChemCrow**	Agent LLM de chimie relié à des outils experts	*github.com/ur-whitelab/chemcrow-public*
Explicabilité	**SHAP**	Explications par attribution de caractéristiques pour tout modèle	*github.com/shap/shap*
Explicabilité	**SISSO**	Découverte de descripteurs/formules interprétables	*github.com/rouyang2017/SISSO*
Informatique verte	**CodeCarbon**	Suivre le coût carbone de vos calculs	*github.com/mlco2/codecarbon*

Lignes en surbrillance = dépôts de notre groupe (github.com/MMLabCodes). Liste représentative, non exhaustive.

Lectures complémentaires

GOULOT DES MATÉRIAUX

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

MODÉLISATION ET SIMULATION

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

NOUVEAU PARADIGME

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

BOÎTE À OUTILS 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

PARLER CHIMIE

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

Merci. Des questions pour terminer ?