

L'INTELLIGENCE ARTIFICIELLE, UN MOTEUR DANS LA RECHERCHE EN CHIMIE !

Éric Bausson

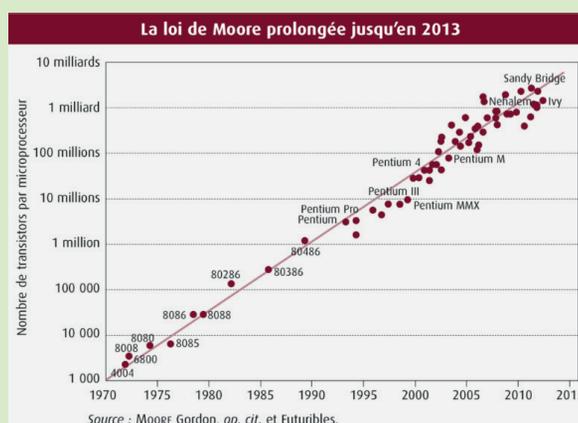
Parties des programmes de physique-chimie associées

- Programme d'enseignement scientifique de terminale, thème 3 : Une histoire du vivant, partie 5 : De la machine de Turing à l'intelligence artificielle
- Programme de la spécialité physique-chimie de terminale générale, partie « Constitution et transformations de la matière », 2.A : Suivre et modéliser l'évolution temporelle d'un système, siège d'une transformation chimique, 4. : Élaborer des stratégies en synthèse organique
- Programme de physique-chimie de terminale STL : partie « Chimie et développement durable » / Synthèses chimiques
- Programme de physique-chimie et mathématiques de première STL : partie « Transformation chimique de la matière » / Cinétique d'une réaction chimique

Mots-clés : intelligence artificielle – synthèse chimique – cinétique chimique

INTRODUCTION

L'intelligence artificielle est un des domaines de l'informatique qui prend une place de plus en plus importante dans notre quotidien (moteurs de recherche, itinéraires routiers...) mais aussi dans la recherche depuis quelques décennies. En doublant tous les dix-huit mois, la puissance des ordinateurs n'a cessé d'augmenter de façon très significative au fil des ans. Ceci a permis d'accroître les capacités de ces machines à exécuter des tâches normalement associées à l'être humain pour résoudre, entre autres, des problèmes complexes en analysant des données en nombre de plus en plus colossal. Depuis les années 1960, les chimistes utilisent ces machines capables de traiter des données et les applications liées à leur utilisation permettent d'optimiser la recherche et le développement de nouvelles molécules.



LA SÉRENDIPITÉ... AVANT ET APRÈS UTILISATION DE L'INTELLIGENCE ARTIFICIELLE

Avant l'essor de l'intelligence artificielle, la sérendipité – l'aptitude à faire par hasard une découverte inattendue et surtout à en saisir l'utilité – avait une place plus importante qu'aujourd'hui.

Dans le passé de la recherche en chimie, bon nombre de molécules comme la mauvéine (premier colorant industriel synthétique), ont été découvertes fortuitement, tout comme leur intérêt thérapeutique ou autres.

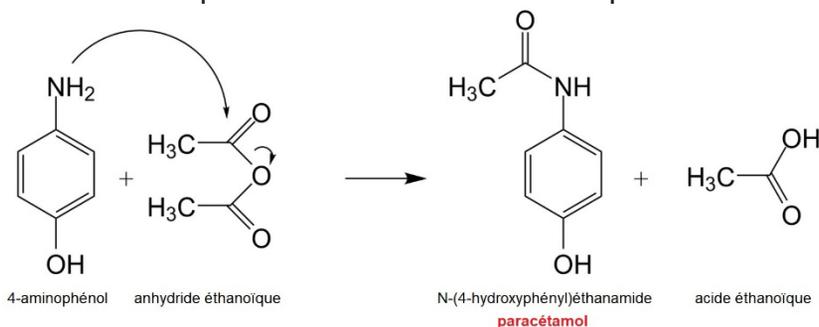
Mais sans observations effectuées et analysées par des chercheurs, elles seraient passées aux oubliettes !

Un des exemples les plus connus est celui de l'usage du paracétamol¹, actuellement le médicament le plus prescrit en France. Il s'agit d'un antalgique (anti-douleur) et d'un anti-pyrétique (anti-fièvre). La molécule a pour nom officiel « N-(4-hydroxyphényl)éthanamide », de formule brute $C_8H_9NO_2$.

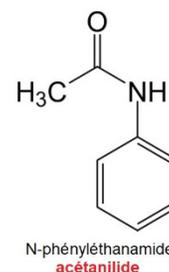


Figure 1 – Boîtes de paracétamol.
@ Maxppp – Luc Nobout.

En 1878, la synthèse du paracétamol fut réalisée pour la première fois par le chimiste américain Harmon Northrop Morse en réduisant du 4-nitrophénol en 4-aminophénol en présence d'étain dans l'acide éthanóique glacial, suivie d'une acylation par l'anhydride éthanóique pour conduire à l'amide désiré. Voici l'équation-bilan de cette dernière étape :



Les propriétés d'un dérivé du paracétamol ont été découvertes par hasard, en 1886, par Arnold Cahn et Paul Hepp, deux médecins strasbourgeois travaillant sur l'effet du naphthalène et de ses dérivés sur les parasitoses intestinales. À court de produits de départ, ils s'approvisionnent dans une pharmacie de la ville et n'observent pas l'effet antiparasitaire attendu, mais une action antipyrétique. En fait, le produit livré n'était pas du naphthalène mais de l'acétanilide [N-phényléthanamide].



Cependant, comme l'acétanilide est très toxique, de nombreuses recherches se consacrent alors à l'élaboration de dérivés mieux tolérés.

En 1948, une étude américaine démontre que l'acétanilide est en fait dégradé dans le corps humain (réaction *in vivo*) en paracétamol et que seule cette molécule est active contre la douleur.

Un peu moins de quatre-vingts ans après sa première synthèse, le paracétamol fut autorisé en 1955 sur le marché des médicaments aux États-Unis à condition qu'il soit utilisé à des doses thérapeutiques, c'est-à-dire 4 g au maximum par jour en 4 prises pour un adulte, et 60 mg/kg chez l'enfant toujours en 4 prises fractionnées, car au-dessus des doses, le médicament peut créer une intoxication du foie.

Mais la sérendipité réserve encore son lot de surprises...

Par exemple, en 2004 dans l'université de Manchester, deux scientifiques, Andre Geim et Konstantin Novoselov, ont isolé le graphène en utilisant du ruban adhésif pour prélever une fine couche de graphite². En réduisant celle-ci jusqu'à atteindre une couche d'un seul atome de carbone, ils isolèrent le graphène sans le chercher spécifiquement au départ. Cette découverte a depuis des répercussions très importantes dans des domaines très variés comme l'électronique, les matériaux composites et l'énergie, pour ne citer qu'eux ! Six ans plus tard, en 2010, donc très rapidement, ils reçurent le prix Nobel de Physique pour « leurs expériences révolutionnaires sur les matériaux bidimensionnels en graphène ».

... Il y en aura d'autres dans le futur !

1. Contraction de : para-acéthylaminophénol.

2. Matériau des mines de crayon.

QU'ENTEND-ON PAR « INTELLIGENCE ARTIFICIELLE » ?

Il n'y a pas de définition universelle de l'intelligence artificielle. Il en existe autant que d'experts dans ce domaine ! La définition de l'intelligence artificielle donnée par le dictionnaire *Le Robert*[®] permet de dessiner ses contours : « Ensemble des théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine (raisonnement, apprentissage...) ».

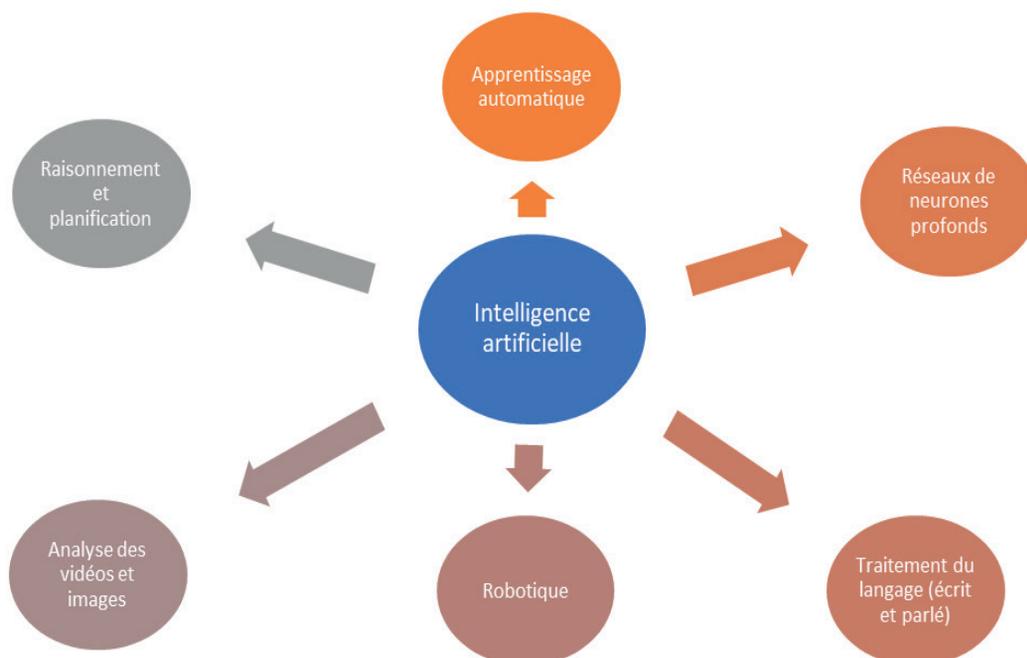


Figure 2 – Principales composantes de l'intelligence artificielle. - Source : É. Bausson.

Parmi les composantes de l'intelligence artificielle, voici celles très utilisées par les chimistes (Figure 3) :

- ▶ l'apprentissage automatique (*machine learning*) permet à un système d'apprendre de façon automatique et de s'améliorer avec l'expérience (traitement des données recueillies tout au long de ces décennies de recherches, recueil des erreurs, etc.) ;
- ▶ Les réseaux de neurones profonds (*deep learning*) sont des modèles informatiques inspirés du cerveau humain utilisant des algorithmes complexes.

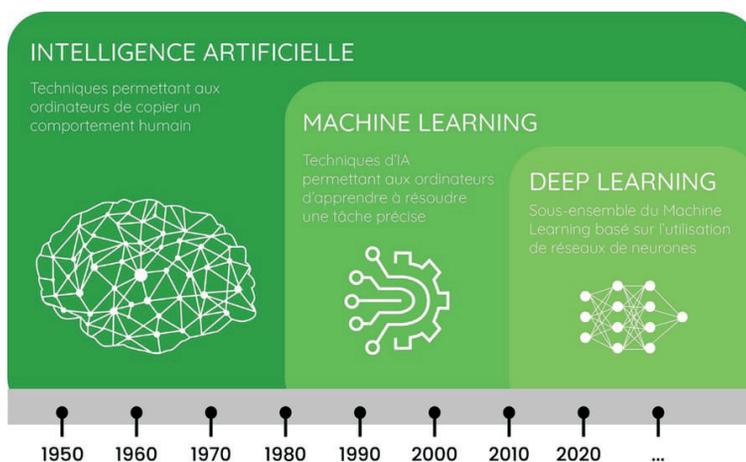
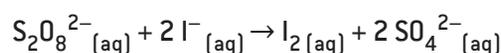


Figure 3 – L'intelligence artificielle et ses branches Machine Learning et Deep Learning.
Source : <https://www.dilepix.com/blog/difference-intelligence-artificielle-deep-learning-agriculture> ; © A-Yol-Dilepix.

En situation d'apprentissage, les élèves et étudiants pratiquent de plus en plus des activités en lien avec l'intelligence artificielle. Parmi celles-ci, nous pouvons citer l'utilisation des arbres de décision, les outils liés à la modélisation (régression linéaire, etc.).

Par exemple, lors de l'étude cinétique d'une réaction chimique, à partir de mesures effectuées au cours du temps comme celles de la conductivité, du pH et bien d'autres, il est possible de traiter les données avec un programme Python[®] pour déterminer la loi relative à l'évolution d'une concentration en fonction du temps.

Prenons l'exemple de la réaction d'oxydo-réduction suivante :



Un suivi spectrophotométrique de celle-ci permet, après calculs, de déterminer la concentration des réactifs et produits chimiques au cours du temps à partir de la mesure de l'absorbance au cours du temps. Tout cela peut se faire en utilisant le programme Python® ci-dessous (Figure 4), ayant pour but de vérifier que la cinétique de cette réaction d'oxydo-réduction est bien d'ordre 1 par rapport aux ions $S_2O_8^{2-}$.

```
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats as sc

t = np.array([0,120,0, 180,0, 240,0, 300,0, 360,0, 420,0, 480,0, 540,0, 600,0, 660,0, 720,0, 780,0, 840,0, 900,0, 960,0, 1020,0, 1080,0, 1140,0, 1200,0, 1260,0, 1320,0])
A = np.array([0,0.3069, 0.4514, 0.5904, 0.7182, 0.833, 0.9418, 1.0385, 1.1308, 1.2151, 1.2885, 1.3599, 1.4244, 1.4811, 1.533, 1.5791, 1.6202, 1.6587, 1.6943, 1.7248, 1.756, 1.7814,])
concI2 = np.array(A / 435)
cpero = np.array(5.E-3 - concI2)

lndepero = np.log(cpero) # nouvelle grandeur

plt.plot(t,lndepero,'+')

#modélisation de la droite sc.Linregress calcul le coef directeur
droite = sc.linregress(t,lndepero)
coefficient=droite.slope
origine = droite.intercept
lndeperomodel=coefficient*t+origine
print('ln [S2O8^2-] = ',coefficient, 'x t + ', origine)
plt.plot(t,lndeperomodel,color='r')

#calcul d'erreur facultatif
erreur = droite.stder
erreur1 = round(erreur,6)
print("coefficient directeur k = ", coefficient, "+/- ",erreur1)

#configurer l'aspect du graphique
plt.title('ln [S2O8^2-] en fonction du temps') #code qui permet les indices et exposants
plt.xlabel("temps en seconde")
plt.ylabel("ln [S2O8^2-]")

plt.grid()

plt.show()
```

Figure 4 – Modélisation Python® de la réaction d'oxydo-réduction suivante : $S_2O_8^{2-}(aq) + 2 I^-(aq) \rightarrow I_2(aq) + 2 SO_4^{2-}(aq)$. Source : É. Bausson.

Après exécution de ce programme, les résultats obtenus (Figure 5) après modélisation permettent de valider le fait qu'il s'agit bien d'une loi de vitesse d'ordre 1 par rapport aux ions $S_2O_8^{2-}$:

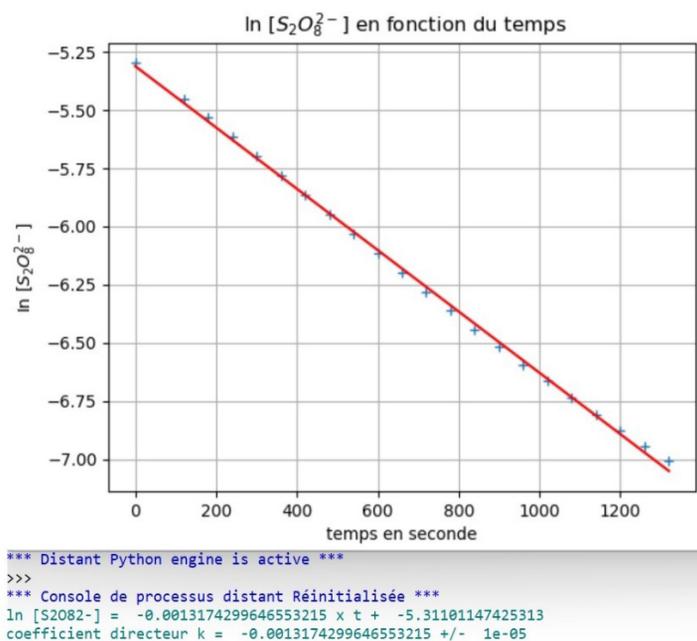


Figure 5 – Résultat de la modélisation, loi de vitesse d'ordre 1. Source : É. Bausson.

QUELS ÉTAIENT LES PREMIERS PAS DE L'INTELLIGENCE ARTIFICIELLE EN CHIMIE ?

La multiplication du nombre de données et la croissance des ressources en informatique ont permis très tôt d'utiliser l'intelligence artificielle comme outil de prédiction en chimie, laissant ainsi moins de place à la sérendipité.

Vers 1960, avec seulement 512 ko de mémoire vive et 8 Mo de stockage, la première utilisation a été faite en spectrométrie de masse, permettant de déterminer la masse et la structure des composés chimiques présents dans un échantillon. Cela a permis une analyse plus rapide et plus précise des données issues des mesures effectuées.

Focus sur la spectrométrie de masse

Tout est expliqué très clairement dans une [vidéo](#) sur le site [Mediachimie](#). Si nécessaire, vous pouvez utiliser l'intelligence artificielle pour avoir les sous-titres en français !

En voici les grandes lignes : les molécules sont portées à l'état gazeux, celles-ci et leurs fragments sont ensuite ionisés (le plus souvent avec une seule charge positive). Tous ces ions sont ensuite déviés par des champs électriques et magnétiques dans le vide, pour éviter tout choc avec d'autres molécules. Leurs trajectoires varient en fonction du quotient masse/charge, soit m/z . Le détecteur permet de déterminer le taux m/z et l'abondance relative de chaque ion formé avant d'être dévié. En utilisant des algorithmes d'apprentissage automatique, l'intelligence artificielle peut comparer les spectres de masse de ces composés avec des bibliothèques de spectres connus pour les identifier avec une grande précision.

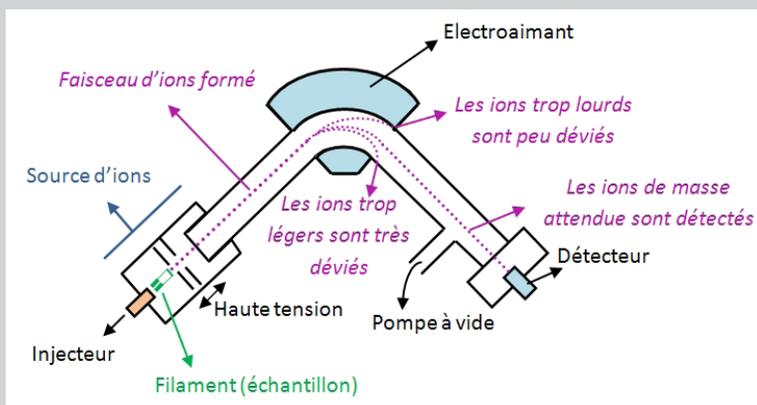


Figure 6 – Schéma de principe de la spectrométrie de masse. Source : <http://acces.ens-lyon.fr/acces/thematiques/limites/Temps/datation-isotopique/comprendre/le-spectrometre-de-masse> ; © ENS-Lyon.

Le spectre de masse obtenu avec le paracétamol, de masse molaire moléculaire $151,163 \text{ g}\cdot\text{mol}^{-1}$ est présenté en Figure 7.

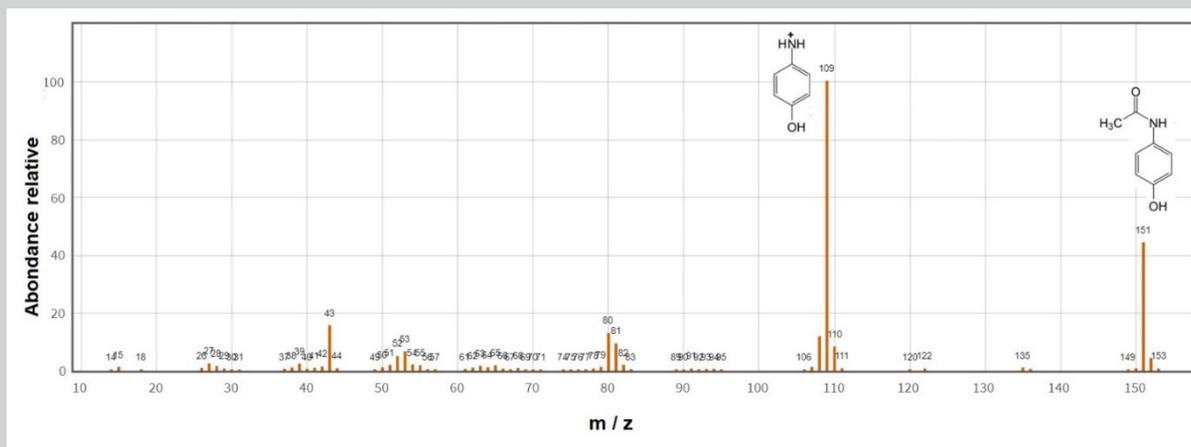


Figure 7 – Spectre de masse du paracétamol. Source : <https://webbook.nist.gov/cgi/cbook.cgi?ID=C103902&Mask=200>.

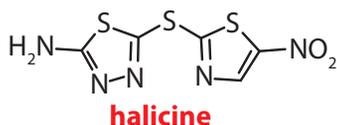
Parmi tous les pics présents, nous retrouvons le pic parent (ou pic moléculaire) pour lequel m/z vaut 151 et celui du fragment le plus abondant, après perte du groupement $\text{CO}-\text{CH}_3$, pour lequel m/z vaut 109.

UN EXEMPLE DE DÉCOUVERTE D'UN ANTIBIOTIQUE EN UTILISANT L'INTELLIGENCE ARTIFICIELLE

Cette molécule a été nommée « halicine », en hommage à l'ordinateur HAL 9000 du film « 2001, l'Odyssée de l'espace » de Stanley Kubrick. L'intelligence artificielle a permis de rechercher par modélisation informatique les molécules chimiques qui seraient à même de s'attaquer à certaines bactéries, en faisant examiner des bibliothèques de composés chimiques. Dans une bibliothèque de six mille composés chimiques, le modèle a tourné en prenant comme référence les antibiotiques capables d'éliminer la bactérie *Escherichia Coli*. Il a recherché ceux qui avaient les caractéristiques recherchées en croisant toutes les données.

L'algorithme a trouvé un composé, déjà connu mais inutilisé comme antibiotique, à la structure différente des antibiotiques existants, et prédit qu'il serait efficace contre de nombreuses bactéries.

La structure de cette molécule est la suivante :



Après avoir effectué des tests, les chercheurs du M.I.T. ont validé la prédiction de l'intelligence artificielle.

Sans l'aide apportée par l'intelligence artificielle, cette découverte aurait sans nul doute pu être faite par l'Homme, mais quand ?

Actuellement, 10^{20} molécules sont en catalogue virtuel, pas encore synthétisées mais pouvant l'être à l'avenir. Les chimistes estiment que 10^{63} molécules peuvent avoir un intérêt pharmaceutique alors que « seulement » 10^{12} sont brevetées actuellement dans tous les secteurs de la chimie. L'intelligence artificielle est donc cruciale pour sélectionner à partir des propriétés recherchées un nombre limité de molécules pour réduire la durée de recherche.

CONCLUSION

Et si nous demandions à l'intelligence artificielle de se prononcer ? Voici la question posée à l'application ChatGPT le 19/04/2023 :

Quelles sont les perspectives et les limites de l'intelligence artificielle en chimie ?

À cette date, sa réponse est listée ci-dessous, sans modifier un seul mot (sans aucun doute, celle-ci devrait varier au fil du temps !).

L'intelligence artificielle (IA) offre des perspectives intéressantes pour la chimie, mais il existe également des limites à considérer.

Perspectives de l'IA en chimie :

- ▶ La prédiction de la réactivité chimique : l'IA peut aider à prédire la réactivité chimique des molécules, ce qui est important pour la conception de nouveaux médicaments et de nouveaux matériaux.
- ▶ La recherche de nouveaux matériaux : l'IA peut aider à trouver de nouveaux matériaux plus performants, tels que des catalyseurs plus efficaces ou des matériaux pour le stockage de l'énergie.
- ▶ L'optimisation des processus de synthèse : l'IA peut aider à optimiser les processus de synthèse chimique, ce qui peut réduire les coûts et augmenter l'efficacité.
- ▶ La découverte de nouveaux médicaments : l'IA peut aider à identifier des composés potentiellement actifs en pharmacologie, accélérant ainsi le processus de découverte de médicaments.

Limites de l'IA en chimie :

- ▶ La qualité des données : les prévisions de l'IA dépendent de la qualité des données utilisées pour l'entraînement du modèle. Si les données sont insuffisantes ou inexactes, les prévisions de l'IA peuvent être également inexactes.
- ▶ Le coût et la complexité de l'IA : l'utilisation de l'IA peut être coûteuse et nécessite une expertise pour la mise en œuvre et l'entretien.
- ▶ Les modèles d'IA ne peuvent pas remplacer complètement l'expertise humaine. Bien que l'IA puisse aider à accélérer et à améliorer certains aspects de la recherche en chimie, elle ne peut pas remplacer complètement l'expertise humaine dans la compréhension et l'interprétation des résultats.

En conclusion, l'intelligence artificielle a le potentiel de révolutionner la chimie, comme dans d'autres secteurs d'activités, mais elle n'est pas encore une solution complète pour tous les problèmes en chimie. Les chercheurs doivent comprendre les limites et les perspectives de l'intelligence artificielle pour l'utiliser efficacement dans leurs travaux de recherche, sans oublier de poser les garde-fous qui sont censés accompagner sa progression fulgurante.

SOURCES PRINCIPALES

D'après le colloque « Chimie et intelligence artificielle » - février 2023

Éric Bausson est professeur de physique-chimie

Comité éditorial : Danièle Olivier, Jean-Claude Bernier, Grégory Syoen