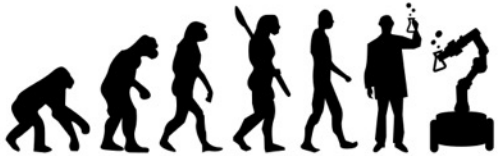
Catalysis Hub - Swiss CAT+



Algorithmique et automation en chimie; le Bon, la Brute et le Truand

CCCTA Villars, 3-4 octobre 2024
Journées Scientifiques

Pascal Miéville
Directeur exécutif West Hub

Centre de compétences
Chimie et Toxicologie
Analytiques

Project fundmentals

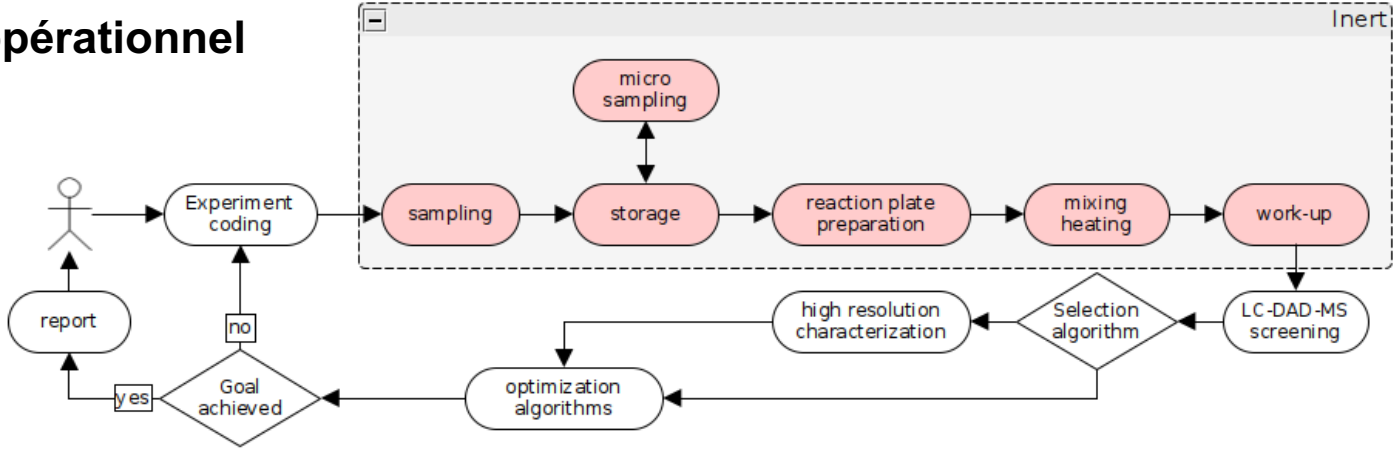
West Hub
homogeneous catalysis

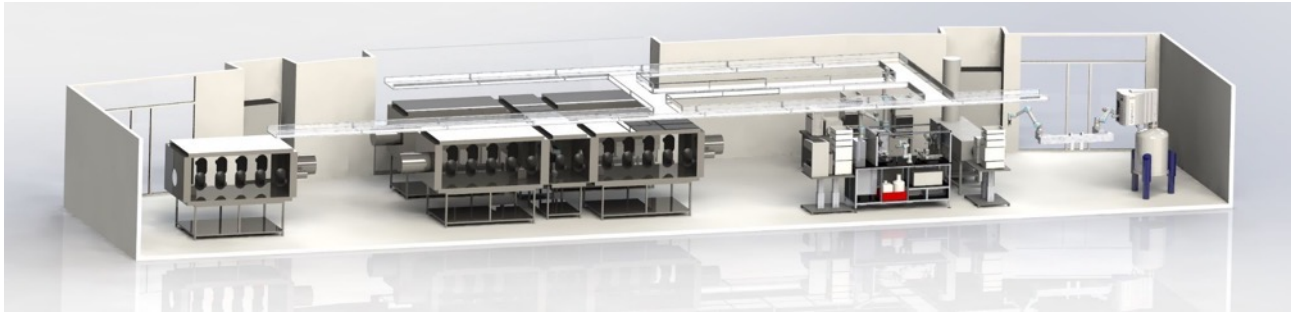
East Hub
heterogeneous catalysis

- National research infrastructure
- Optimization & discovery of catalysts
- Automated -> Autonomous
- Data driven
- Project initiated in 2016 with start Jan 2021

- Must be cutting-edge
- Call for project based service
  currently initiated
- Continuous development with dedicated research funding and collaborations
- 2021-2024 : 19.2 MCHF (21 USD)
- 2025-2028 : 4.05 MCHF secured
  + service revenue
  + looking for partnerships

Workflow opérationnel

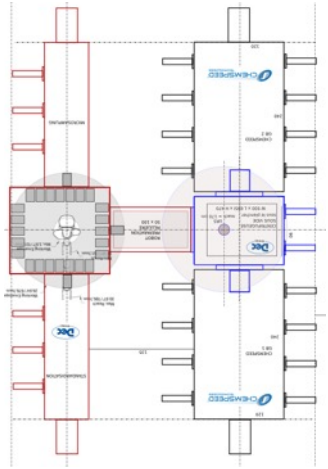




P. Laveille*, P. Miéville* et al, Chimia 2023, 77, 154, DOI: 10.2533/chimia.2023.154
P. Miéville, F. de Nanteuil, Modern Automated Organic Synthesis Comprehensive Organic Synthesis 3rd Edition, 2024, Elsevier

Workflow opérationel (dans la vraie vie)

La station de prélèvement de solides et de synthèse

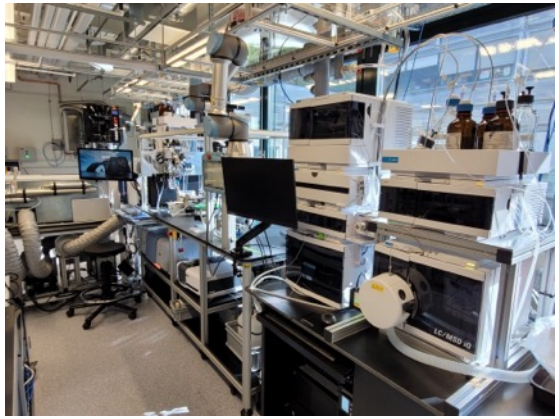


Workflow opérationel (dans la vraie vie)

Station de contrôle des reactions (screening)

Workflow opérationel (dans la vraie vie)

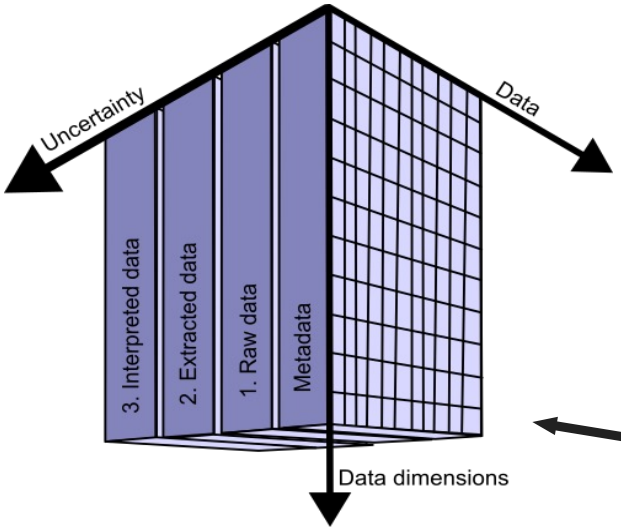
Station de caractérisation des molecules à haute résolution

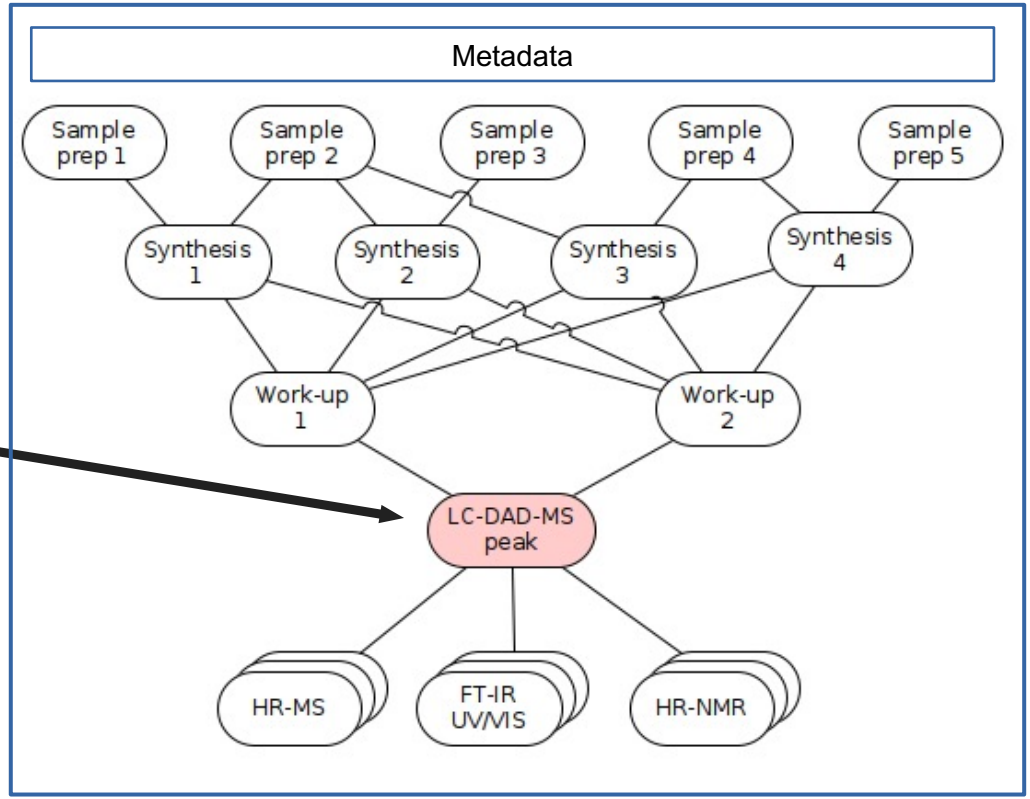


The 2D drone swarm sample transfer system

Vitesse de lecture 2x

Ontologie de données, le matryoshka file (ou le fichier poupée russe)





Une **représentation moléculaire** complète, adaptée au digital et contenant toutes les donnés préparatives et analytiques (format .zarr)

Algorithmes de chimie digitale utilisés et développés

Algorithmes d'optimisation

Optimisation bayesienne

BoFire

https://experimental-design.github.io/bofire/

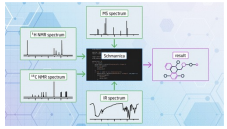
Algorithmes génératifs

Prof P. Schwaller and group
Prof K. Jorner ETHZ and group
IBM

- AI automatically synthesis protocols
- Chemical space exploration strategies

Elucidation automatique de structures

Journal of Chemical Information and Modeling 2021 *61* (2), 756-763
Chem. Sci., 2021,12, 15329-15338

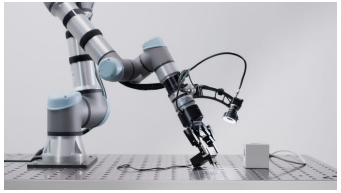


Laboratoire autonome (SDL)

- Dynamic workflow generation
- **Automatic sample selection for high-resolution analysis**
- Automated solubility analysis and prediction

Robotique optimisée

- Vision integration
- ML boosted motion control
- Soft robotics

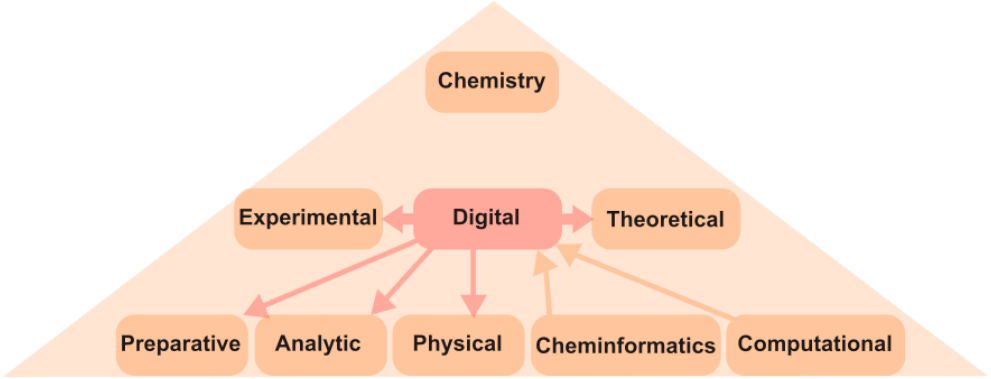


Algorithmique et automation : un couple indissociable

Mais multifacettes !

Le Bon

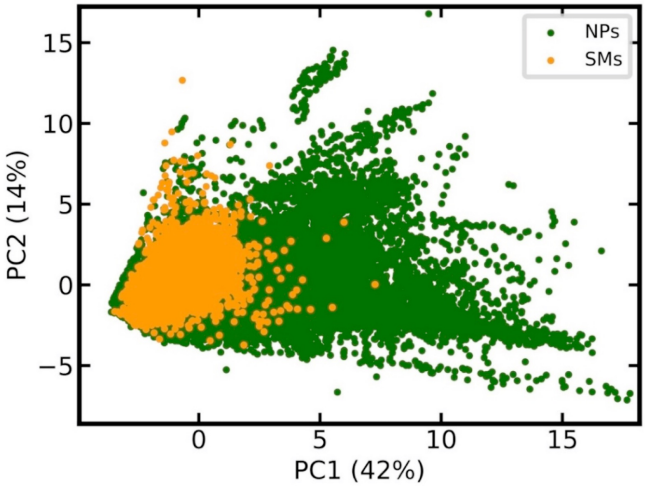
Le Bon : Systématiser l'acquisition de données en chimie



- Fusion des méthodes préparatives et analytiques
- Une systématisation du traitement des données
- Récupération de l'ensemble des données produites « positives et négatives »
- Plus d'apriori sur les réactions/conditions à tester
- Coût par data point nettement réduit

P. Miéville, F. de Nanteuil, Modern Automated Organic Synthesis Comprehensive Organic Synthesis 3rd Edition, 2024, Elsevier, /10.1016/B978-0-323-96025-0.00047-8

Le Bon : Accélérer l'exploration des espaces chimiques



Chen et al., Biomolecules 2019, 9(2), 43;
https://doi.org/10.3390/biom9020043

>10^{60} petites molecules potentiellement réalisables

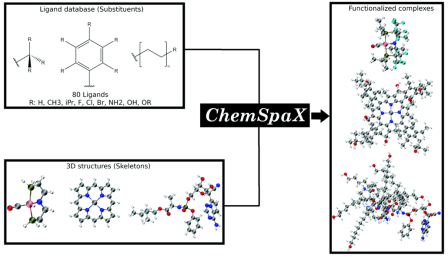
Stratégie d'exploration clusterisée

Stratégie d'exploration basée sur les similitutudes de structures

Stratégie d'exploration contrainte par la simplicité et les coûts

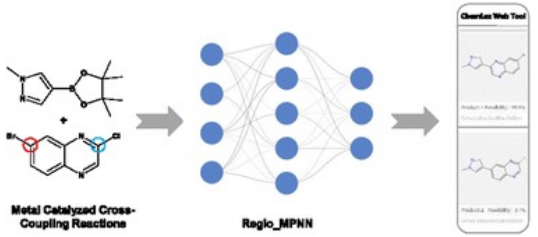
Le Bon : Accélérer l'exploration des espaces chimiques

ChemSpaX: exploration of chemical space by automated functionalization of molecular scaffold



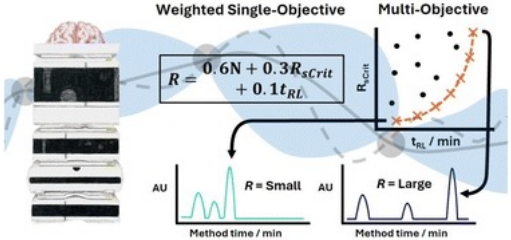
A. V. Kalikadien et al., Digital Discovery, 2022,**1**, 8-25 10.1039/d1dd00017a

Regio-MPNN: predicting regioselectivity for general metal-catalyzed cross-coupling …



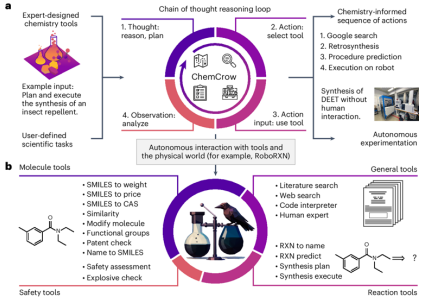
Baochen Li et al., Digital Discovery, 2024 10.1039/d4dd00244j

Operator-free HPLC automated method development guided by Bayesian optimization



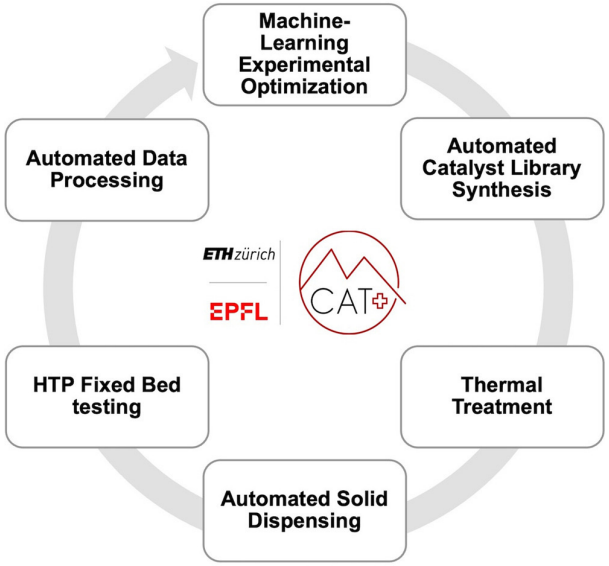
T. M. Dixon et al., Digital Discovery, 2024, **3**, 1591-1601 10.1039/D4DD00062E

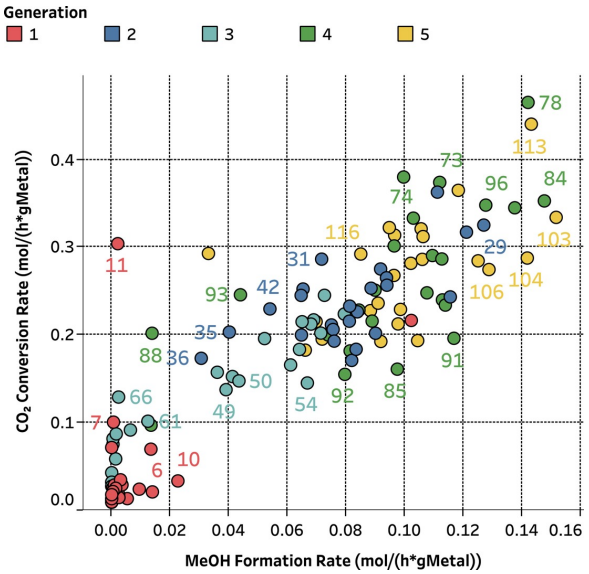
Augmenting large language models with chemistry tools



M. Bran, A., Cox, S., Schilter, O. *et al.* Augmenting large language models with chemistry tools. *Nat Mach Intell* **6**, 525–535 (2024). 10.1038/s42256-024-00832-8

Le Bon : Accélérer l'exploration des espaces chimiques et l'identification de molécules ou méthodes performantes





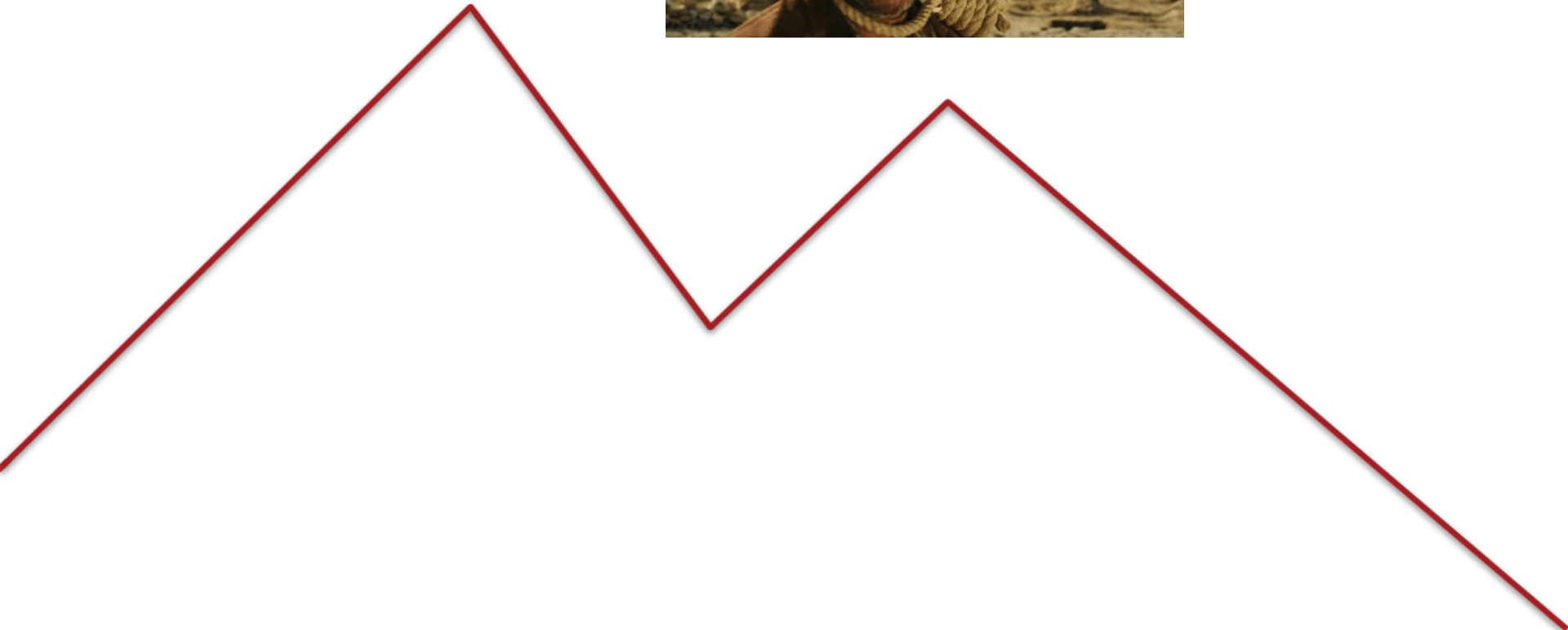
Automated and high-throughput synthesis and testing of **144 catalysts in 6 weeks**

Significant improvement of the catalyst's performances **in 5 generations**

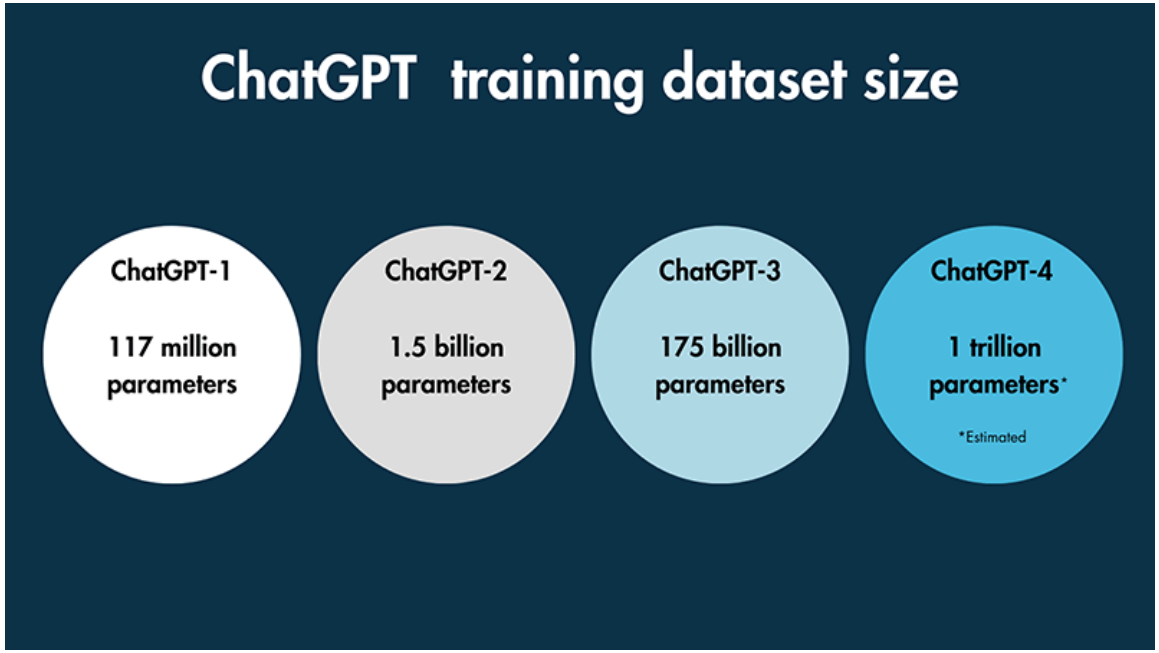
Similar to **ca. 10 years of catalytic development**

Ramirez et al., Chem Catalysis 4, 100888, February 15, 2024

La Brute

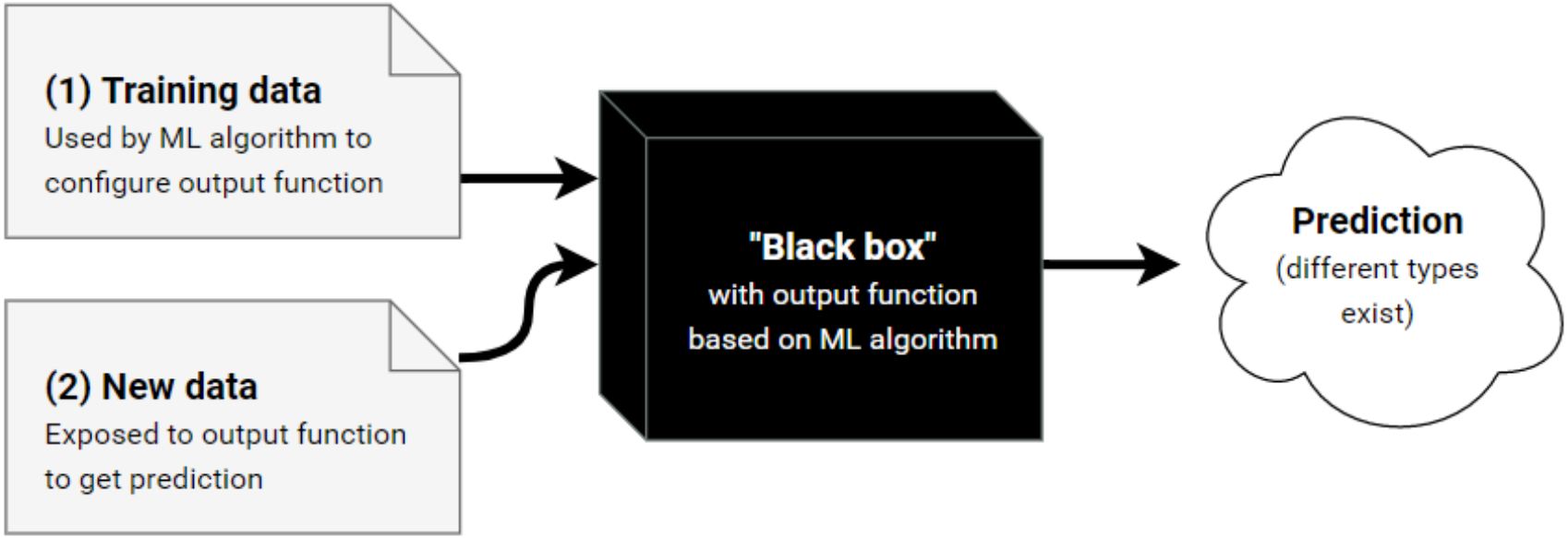


La Brute : il faut énormément de données pour entraîner des modèles génératifs

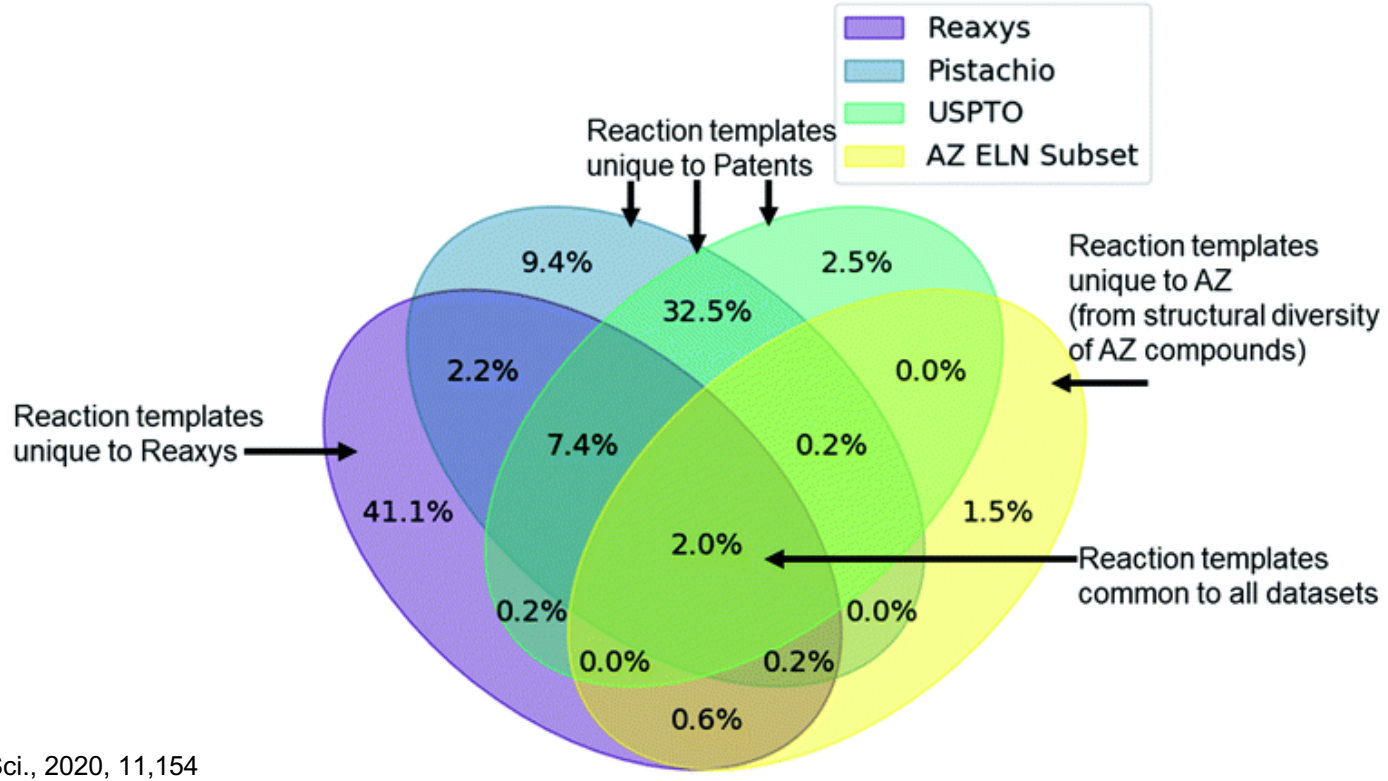


Sources: datacamp and Wired

La Brute : L'Importance de la qualité des données (en plus de la quantité)
ou le principe cochonneries dedans – cochonneries dehors (Garbage in – garbage out principle)

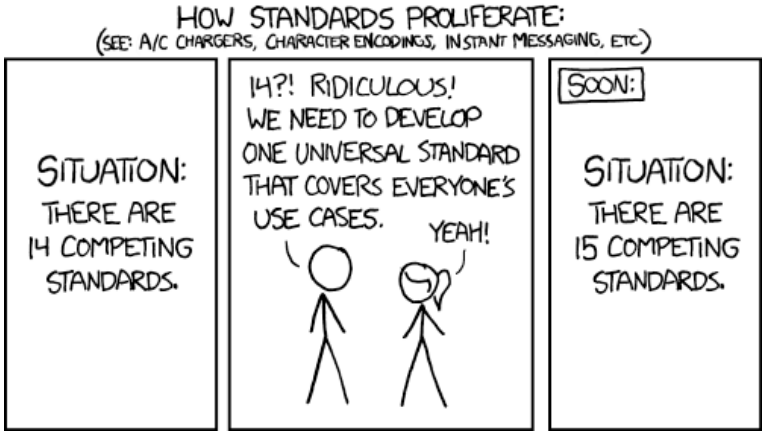


La Brute : A l'heure actuelle la chimie est très pauvre en données structurées



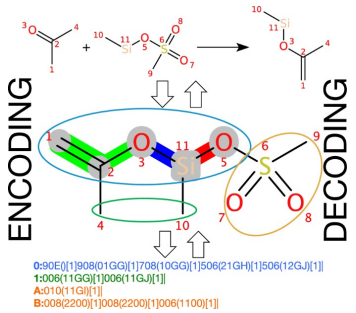
Chem. Sci., 2020, 11,154

La Brute : la production de toutes ces données structurée et partageables est compliquée

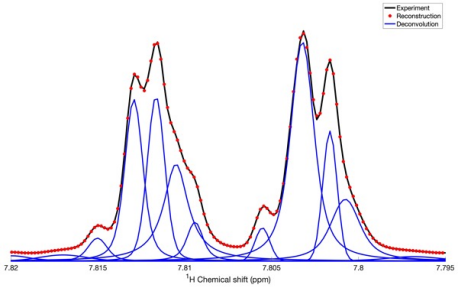


xkcd.org

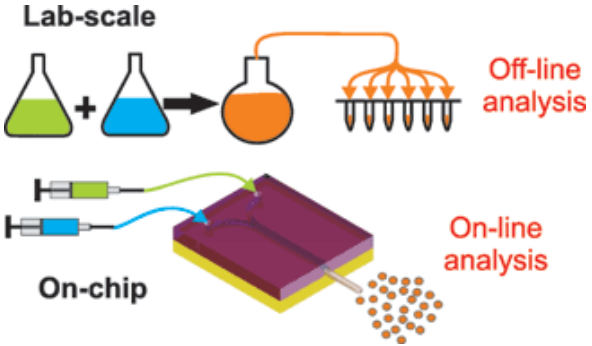
Le Brute : l'automatisation ne règle pas tout, loin de là



J Cheminform **12**, 72 (2020).
10.1186/s13321-020-00476-x



MR, 4, 19–26, 2023
https://doi.org/10.5194/mr-4-19-2023

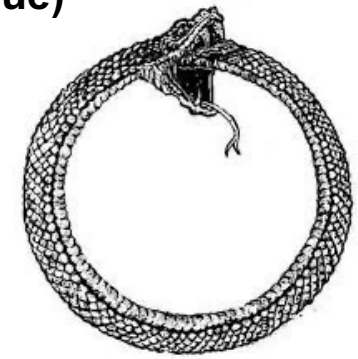


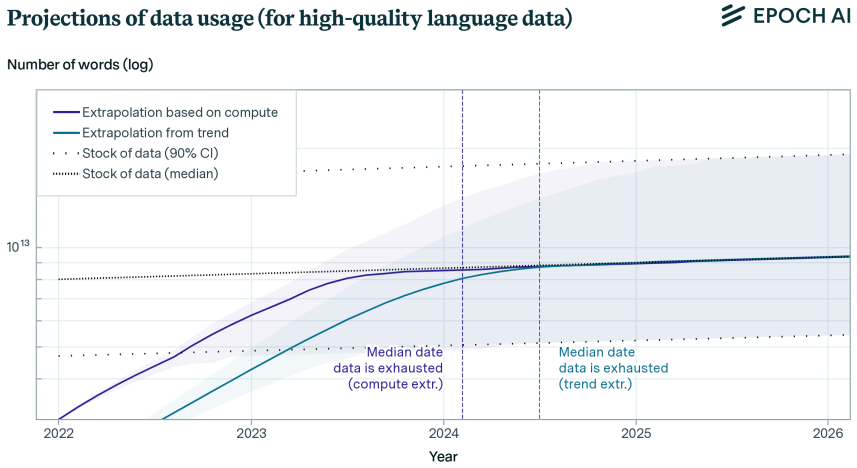
Lab Chip, 2006,6, 329-344

Le Truand

Le Truand : le problème Ouroboros (le serpent qui se mord la queue)

ou ne pas utiliser des données générées par IA pour entraîner des modèles IA





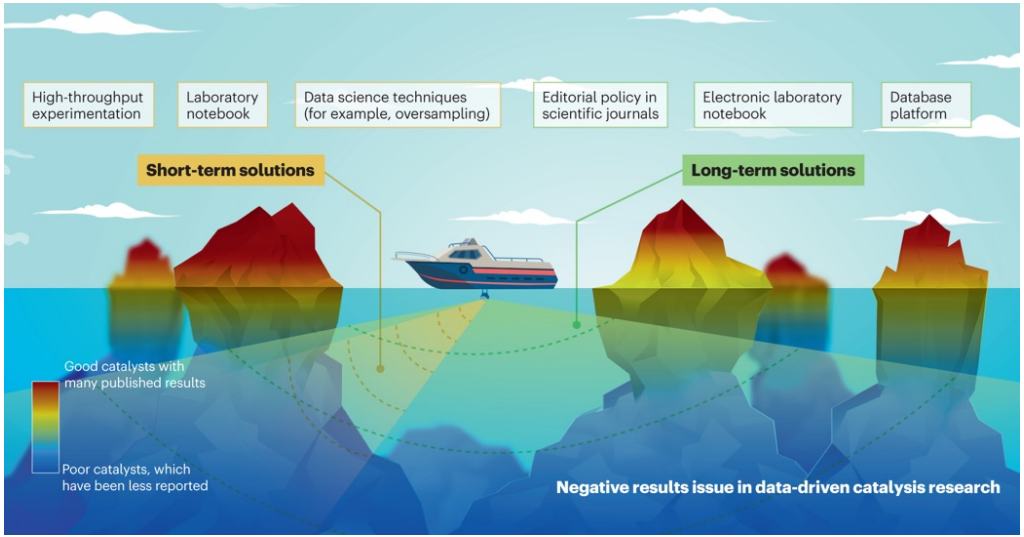
Pablo Villalobos, et al. 'Will We Run out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning'. *ArXiv [Cs.LG]*, 2022. arXiv.

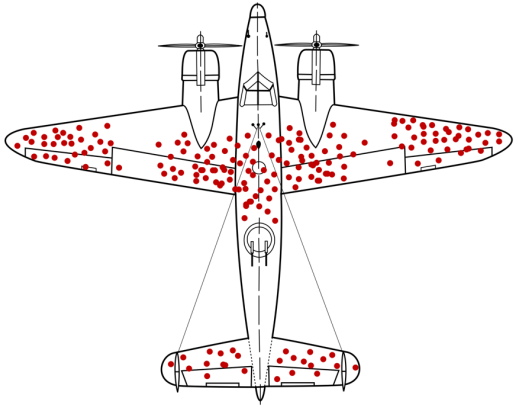
2026 – l'entraînement des principaux ChatBots aura consommé l'intégralité du **web actuel basé sur des documents générés par des humains**.

Une grande quantité de nouveaux document est maintenant crée par des IA.

Problème équivalent en chimie => différencier et tracer les sources de données

Le Truand : L'importance des données négatives pour entrainer les modèles





Nat Catal **6**, 108–111 (2023). 10.1038/s41929-023-00920-9
PLoS Med 2(8): e124, 10.1371/journal.pmed.0020124
Wald, Abraham. (1943). *A Method of Estimating Plane Vulnerability Based on Damage of Survivors*. Statistical Research Group, Columbia University. *CRC 432*

Le Truand : Les hallucinations et sur-interprétations non filtrées

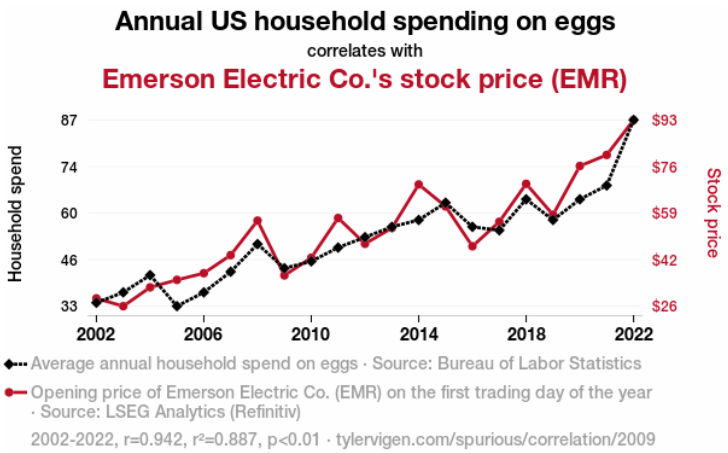
Hallucinations

https://medium.com/@gcentulani/understanding-hallucination-in-llms-causes-consequences-and-mitigation-strategies-b5e1d0268069

Le perroquet stochastique

10.1145/3442188.3445922

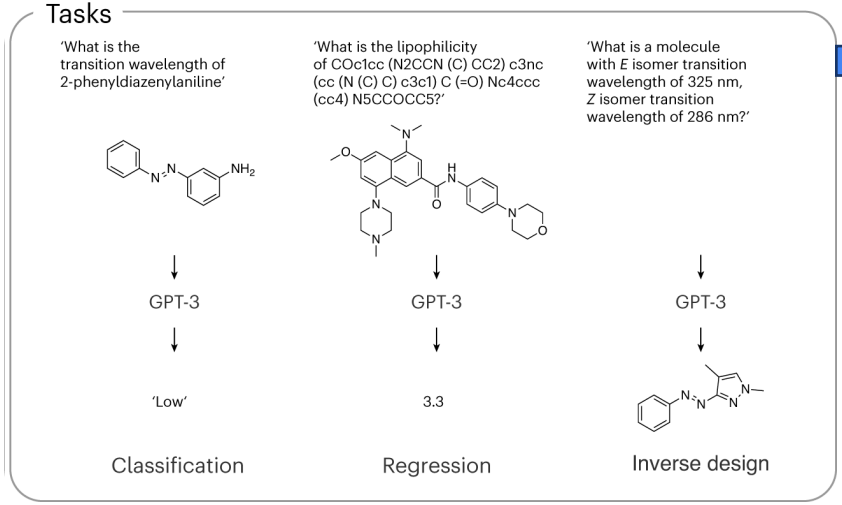
Les correlations absurdes



Impact massif sur des systèmes automatisés

Le Truand : pourquoi est-il aussi méchant ?

LLM appliqués à la chimie



Jablonka, K.M., Schwaller, P., Ortega-Guerrero, A. *et al.* Leveraging large language models for predictive chemistry. *Nat Mach Intell* **6**, 161–169 (2024). https://doi.org/10.1038/s42256-023-00788-1

Generation automatique du protocole

M. Bran, A., Cox, S., Schilter, O. *et al.* Augmenting large language models with chemistry tools. *Nat Mach Intell* **6**, 525–535 (2024). 10.1038/s42256-024-00832-8

Compilation pour application dans un setup spécifique

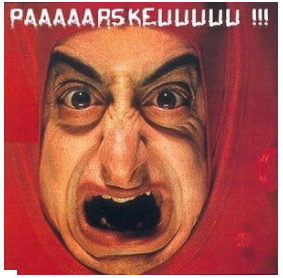
Swiss Cat+ team, *in preparation*

Le Truand : pourquoi est-il aussi méchant ?

ChemGPT

Remote automated lab

Prompt 1 : fabrique une molecule pour éliminer ma belle-mère



Prompt 2 : fabrique un nouveau désherbant efficace

Conclusion

L'automation et l'algorithmique en chimie forment un couple indissociable.

- Elles se nourissent mutuellement.
- Comme toute technologie, elle n'est ni blanche, ni noire, elle est surtout mutlifacette.
- Elles apportent certainement de nouvelles possibilités (le Bon)
- Elles sont très groumandes en ressources et posent de varies questions de durabilité (la Brute)
- Elles peuvent générer des résultats absurdes, mais est-elle vraiment plus mauvaise que les humains ? (Le Truand)

AFCAD

L’Association Francophone de Chimie Accélérée et Digitalisée

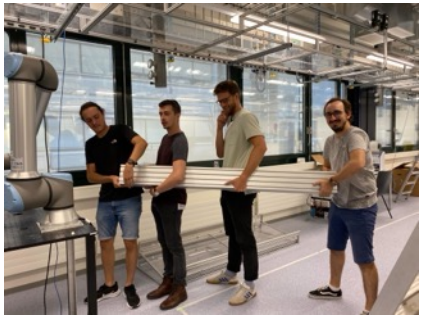
Fondée le 2 octobre 2024 !

Basée à l’EPFL, déjà plus de 60 members de Belgique, France et Suisse.

Académiques et industriels.

N’hésitez pas à me contacter en cas d’intérêt.

Acknowledgments



Active members funded by ETH-RAT
Dr. P. Miéville (executive Director chemist)
Dr. S. di Leone (scientist chemist)
Dr. N. V. T. Nguyen (scientist chemist)
H. Francelet (admin, 50%)
J.-C. Cousty (IT developer)
E. Mariano (robotics engineer)
E. Clerc (chem technician)
C. Cheyrouse (lab technician)
P. Gaumann (chem technician)

Active members funded by other funds
K. Villat (microtechnics engineer, 50%)
H. Zolonowski (microtechnics engineer, 50%)
C. Gachet-Aquillion (chem technician, 30%)
M. K. Villat (microtechnics engineer, 50%)
C. Monheo dos Anjos (apprentice lab technician 70%)
M. Lowenberg (apprentice automation technician 70%)

Students
Y. Coderey (robotics master student)
L. Choudhury (chem master student)
H. Greuilich (chem master student)
E. Sölvi Oskarson (chem master student)
M. Huewei Liu (robotics bachelor student)
M. Tanguy Cavagna (computer sciences bachelor student)
M. Alec Schmidt (computer sciences bachelor student)
N. Bui (chem master student)

Special collaborations
Prof. M. Kunze - automation - HEIG-VD
Prof. J. Hugues - robotics – EPFL
Prof P. Schwaller – digital chem - EPFL
Dr. R. Williams - Soft Director – Agilent
Dr. Santi Domingez – CEO – MestreLab
Dr. Till Khun – VP – Bruker Switzerland
Dr. A. Habou-Amdan - VP - Chemspeed

Centre de compétences Chimie et Toxicologie Analytiques

Featured
Clint Eastwood
Ellie Wallash
Lee Van Cleef
Sergio Leone
Enio Moricone

ETH-RAT
EPFL
Innosuisse
canton de vaud