

1ST GREEN ANALYTICAL CHEMISTRY WORKSHOP

5/6 FEBRUARY, 2024 - NOVOTEL PARIS CHARENTON



Artificial Intelligence (AI) at the service of Laboratories



#LaboratoireduFutur ...?



EXPERTSCIENCE

Yvon Gervaise

E.S.C. (ExpertScience – Consulting)

animateur GT SECF #LaboratoireduFutur

<http://www.expertscience.fr>



Thinking the Laboratory of the Future and Factory of the Future ...it's taking the measure of this mutation induced by A.I.

- The Mission of the GT #LaboratoireduFutur of the Society of Chemical Experts of France
 - enlighten you on the mysteries of A.I
 - make you aware of the importance of the exploitation and manipulation of data
 - become aware of the design and usefulness of a range of AI applications



CHIMIE, BIOCHIMIE ET INTELLIGENCE ARTIFICIELLE : UNE NOUVELLE ÈRE POUR LE LABORATOIRE DU FUTUR

Conférence animée par :

Yvon GERVAISE

Animateur du GT

#LaboratoireduFutur à la SECF -

Fondateur, ExpertScience-

Consulting



SECF
FORUM LABO PARIS

JEUDI 30 MARS
14H00



EXPERTSCIENCE

A.I. and Laboratory the advent of Artificial intelligence (A.I.)



Neural networks...

Machine learning algorithms...

Generative AI...

...at the heart of a new revolution
for the Laboratory ?



Artificial Intelligence (A.I.) ?...

Artificial intelligence (A.I.)...

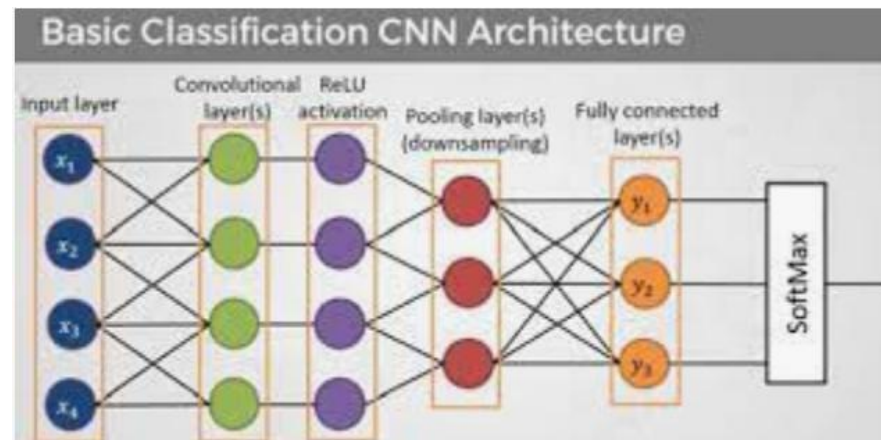
“it is the set of techniques allowing machines to accomplish tasks and. to solve problems normally reserved for humans and certain animals”

A.I. includes 3 categories :

1 - **Symbolic I.A.:** algorithms based on of rules (**Expert Systems**)

2 - **Connectionist AI:** the algorithms learning with **neural network**

3 - **Generative AI (GPT)** networks neurons with Transformer & variant.
Transformer



EXPERTSCIENCE

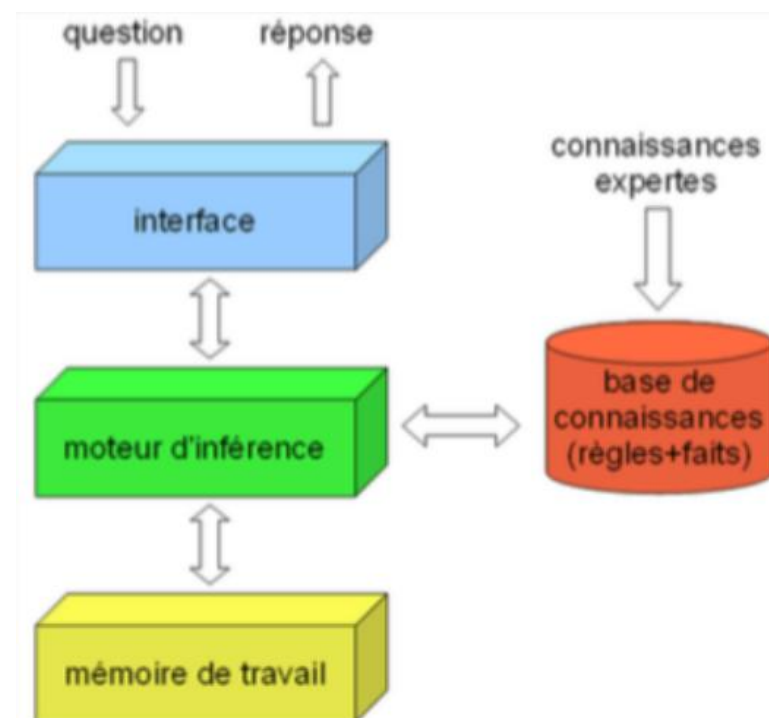


Symbolic A.I : Expert Systems...

Expert systems (S.E) :

- Encapsulate **expert knowledge** in the form of rules (if X then Y)
- Have an **inference** mechanism
- Use this knowledge to solve problems and **automate the processing of process**
- Have evolved according to different **models of representation** semantics, conceptual graph, tree decisions, computer languages

(Examples of S.E: prospector Mycin, **Dendral**)



The enigmatic **Meta-Dendral Expert System** for chemists...for a long time **1965 !** in **LABORATORIES!**

Allows constituents to be identified chemicals, Molecules of a material from its mass spectrum.

Use the rule: “if the spectrum of the molecule presents 2 peaks X1 and X2 such that $X1 - X2 = M + 28$...then the Molecule contains a ketone group.”

This S.E (**E.S.**) is the first algorithm **that models the knowledge** of an expert chemist analyst



EXPERTSCIENCE

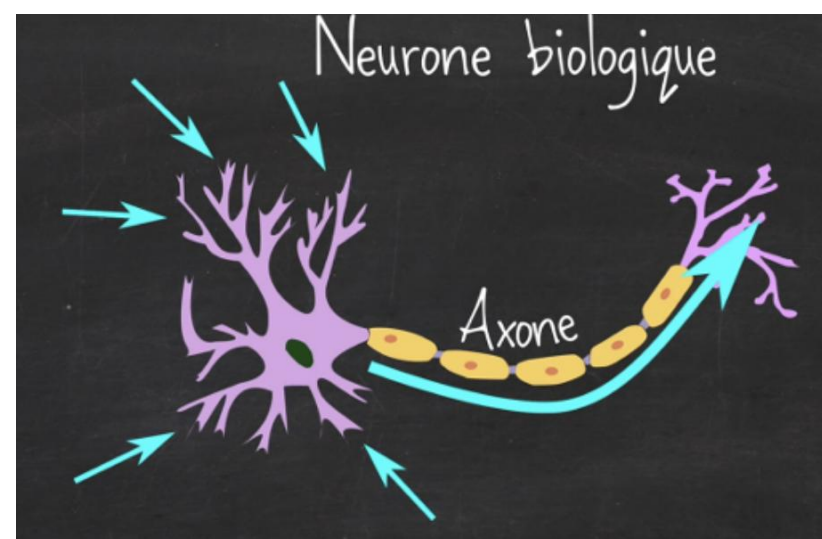
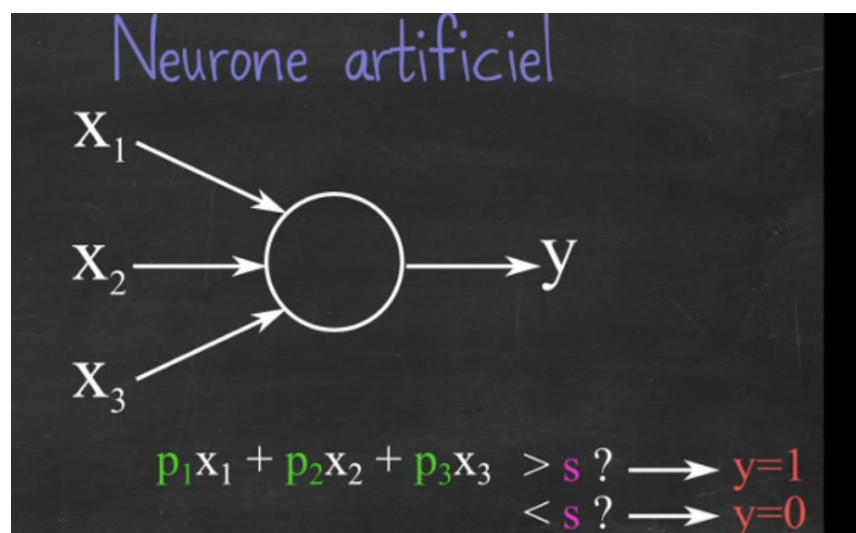
 ResearchGate

1. A Typical Meta-DENDRAL Rule ...



Connectionist AI: Machine Learning artificial neural network...

- Artificial neuron is a mathematical function that mimics the biological neuron with its hardware

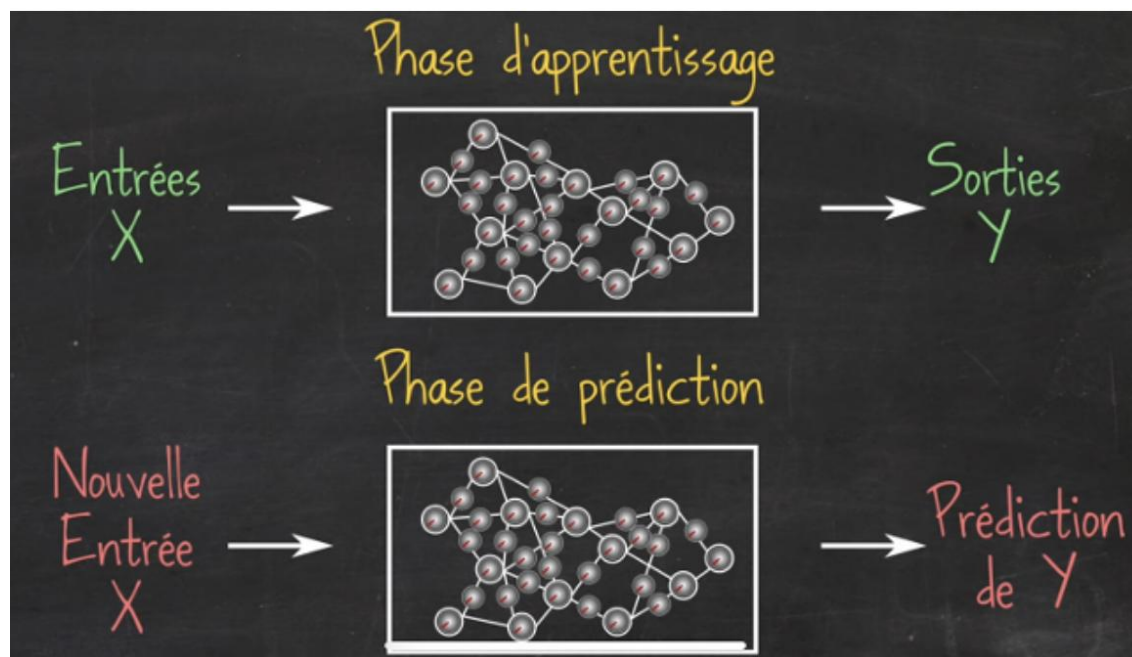
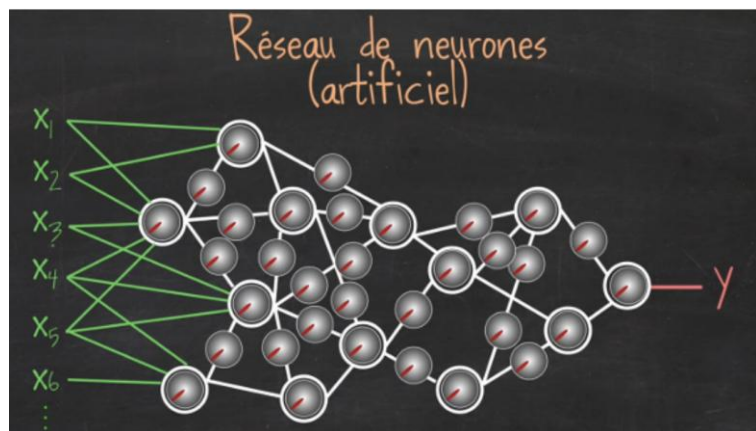


- ✓ Connectionist AI imitates the basic architecture of the brain
- ✓ Artificial neural networks can receive and transmit information



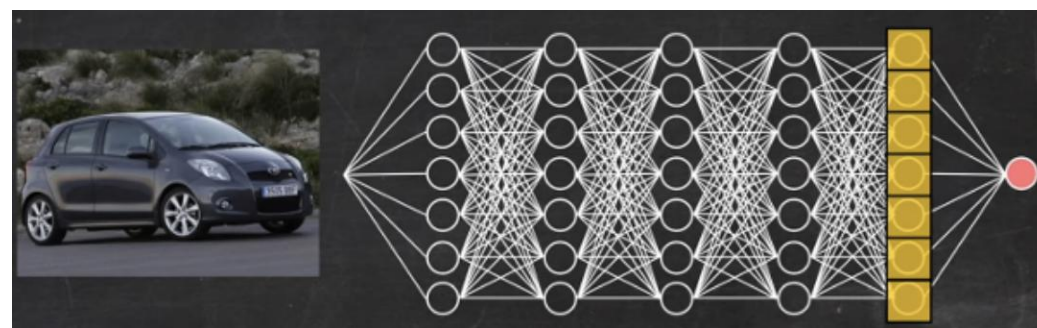
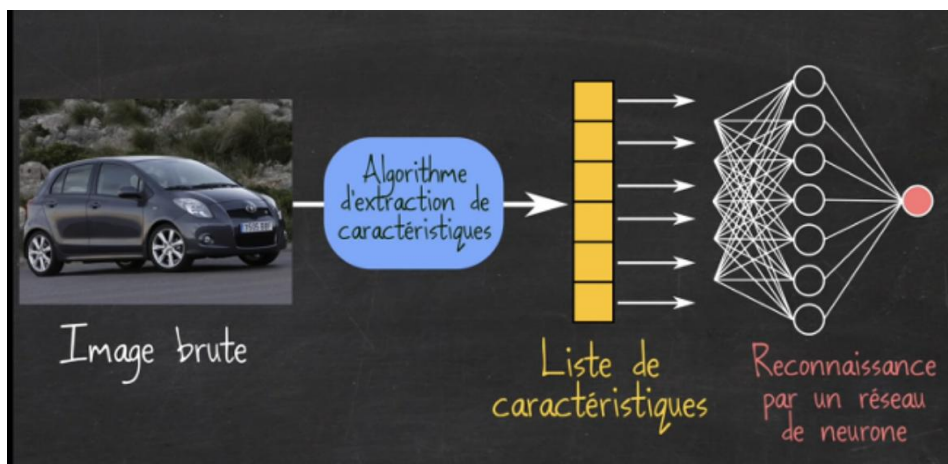
Connectionist AI : the machine learns then solves

- **Learning phase** feeds the algorithm with multiple characteristic examples
- **Prediction phase**
- **Test phase** Response/prediction quality



From machine learning to deep learning thanks to the multi-layer deep neural network

The machine discovers the characteristics itself and is then able to capture the relationship between entrance and exit.



- Ability to solve complex problems “Deep learning” Yann Le Cun
- Requirements of :
 - Calculation capabilities (graphics processors)
 - Availability of large quantities of data for training the algorithm
- Neural network optimization (synaptic weights, activation thresholds, etc. gradient backpropagation)

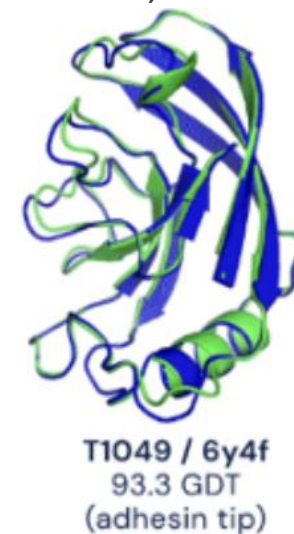
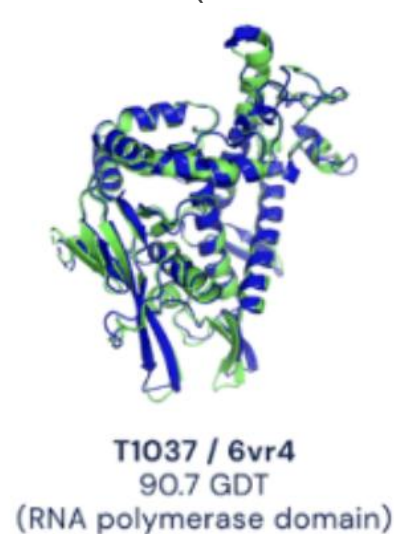
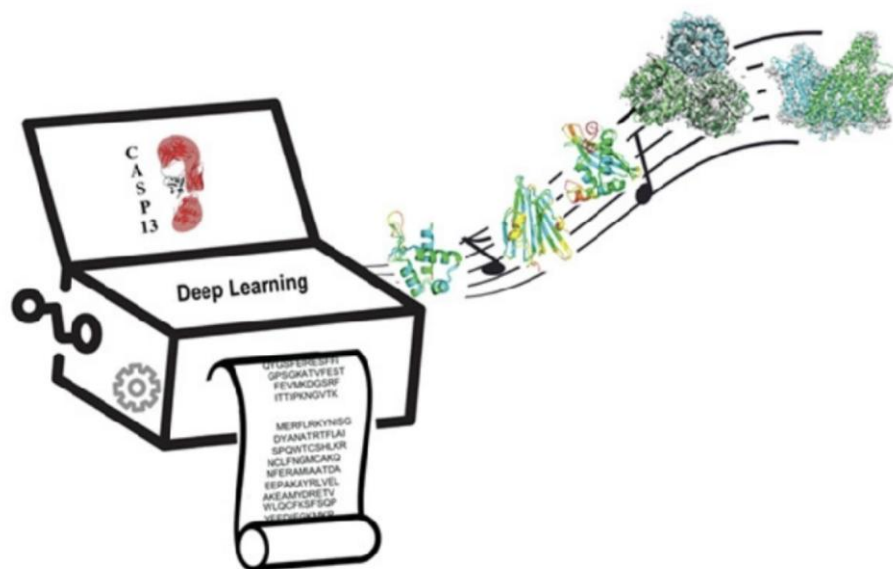


Deep learning at the heart of a disruption for Laboratories in biochemistry and molecular and structural biology

- **The problem:** Impossible to read the 3D structure in an amino acid sequence after folding of a protein

- **The solution:**

AlphaFold Deep Learning algorithm successfully wins the **CASP** in 2021 (Critical Assessment Proteins structure Prediction) with GDT (Global Distant Test) > 92



● Experimental result
● Computational prediction

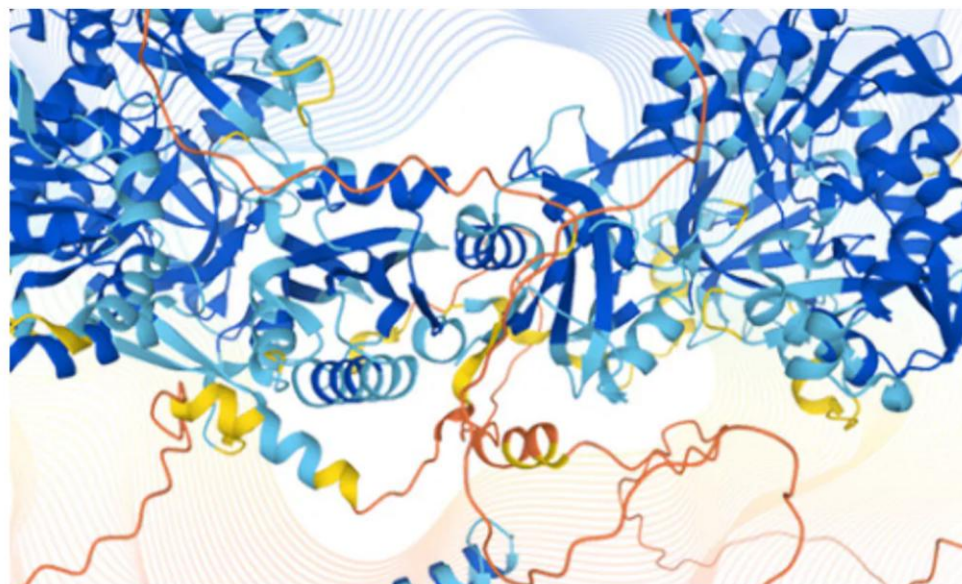


The A.I. a new challenge for Laboratories: the AlphaFold revolution

Google [DeepMind](#) in partnership with EMBL (European Molecular Biology Laboratory) has developed AlphaFold

From July 2021 to July 2022 the [AlphaFold database](#) of 3D protein structures has been multiplied by X 200

DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins



Protein structures representing the data obtained via AlphaFold. Source image: AlphaFold. Design credit: Karen Arnott/EMBL-EBI



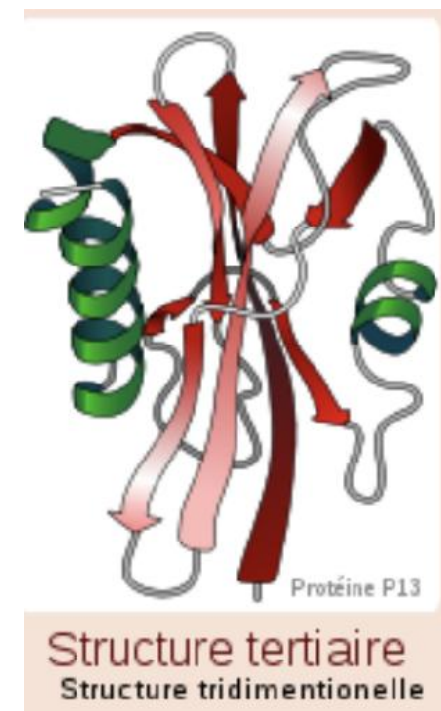
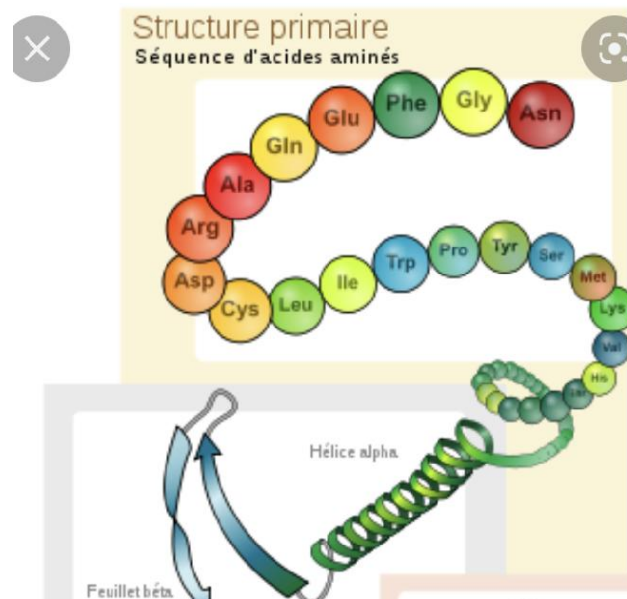
New approach & Performance of AI: the AlphaFold revolution

The acquired knowledge of the structure in space of these molecules which are proteins provide information on their function in living organisms

° This result can be appreciated as an approach combining:

- **training data:** a corpus of data from sequencing amino acids and some 3D structures established by the techniques *X-ray crystallography* and *cryo-microscopy electronic*

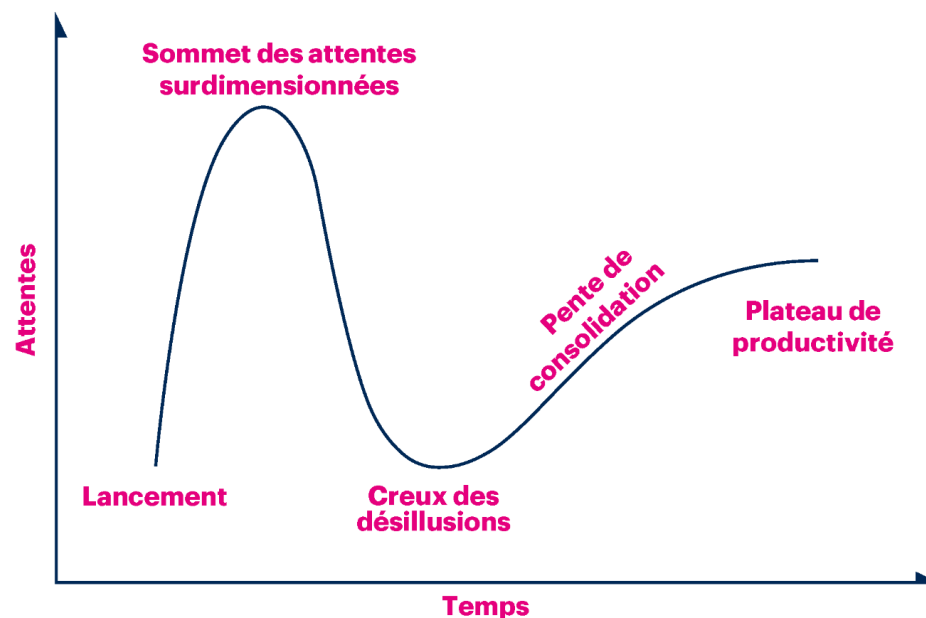
- **the capacity of the AI (AlphaFold):** to decipher the link uniting primary/tertiary structure



ARTIFICIAL INTELLIGENCE CHALLENGES FOR LABORATORIES IN THE COMING YEARS...?

Its integration as a solution to optimize the operation of a growing number of industrial processes and R&D

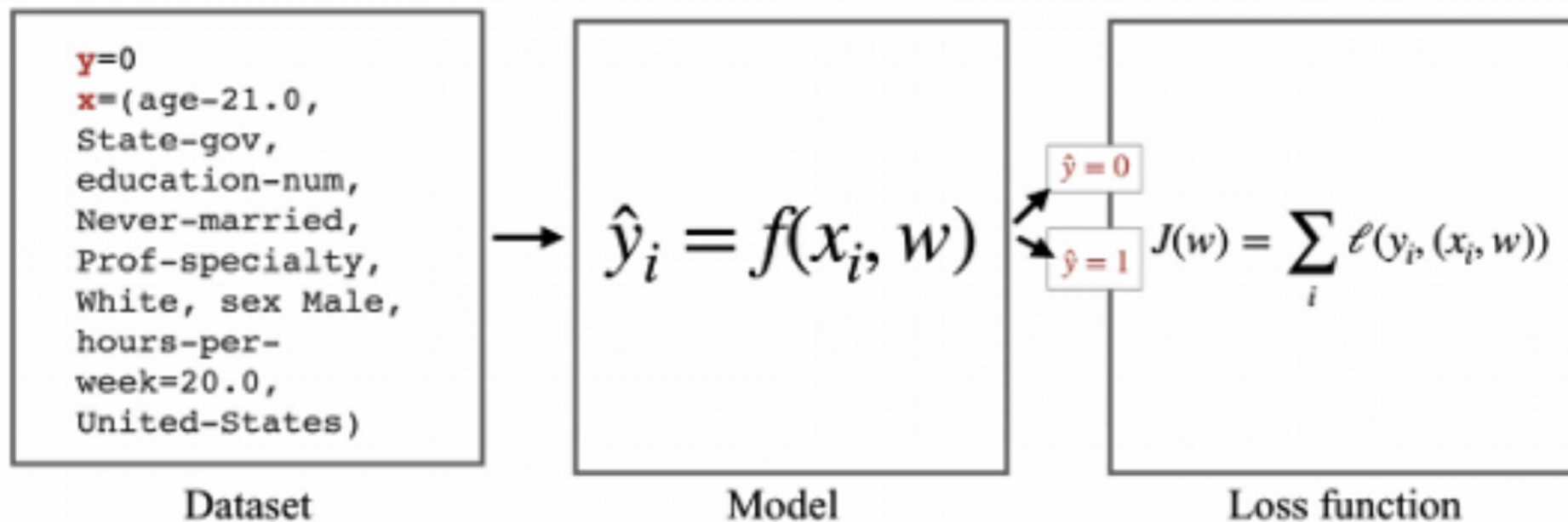
courbe de Gartner
(hype cycle)



A.I. technologies will be installed discreetly in the gaps where they actually create value



The A.I. Matrix of the laboratory of the future: the Ingredients...and the principle :



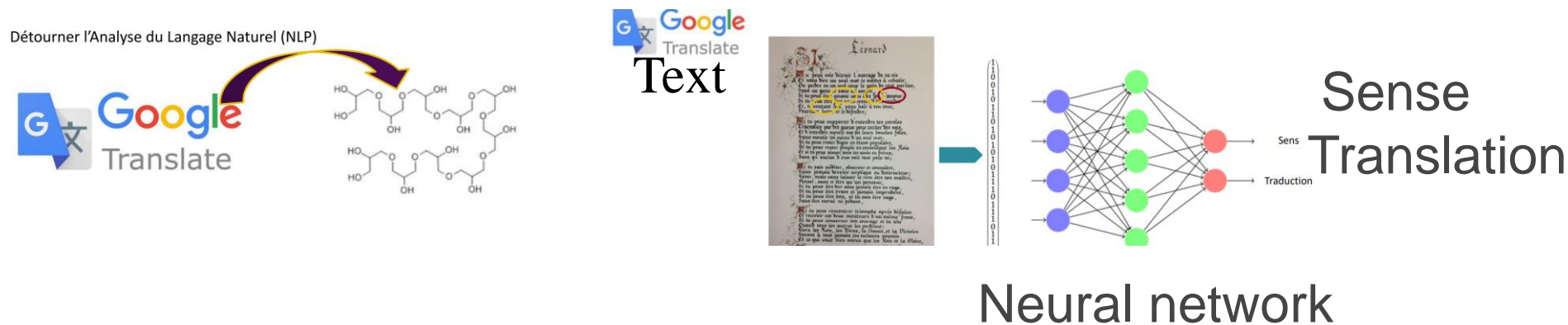
Datasets – Model – Cost Function

- **Parametric Models :**
 - decision tree
 - regression
 - neural network

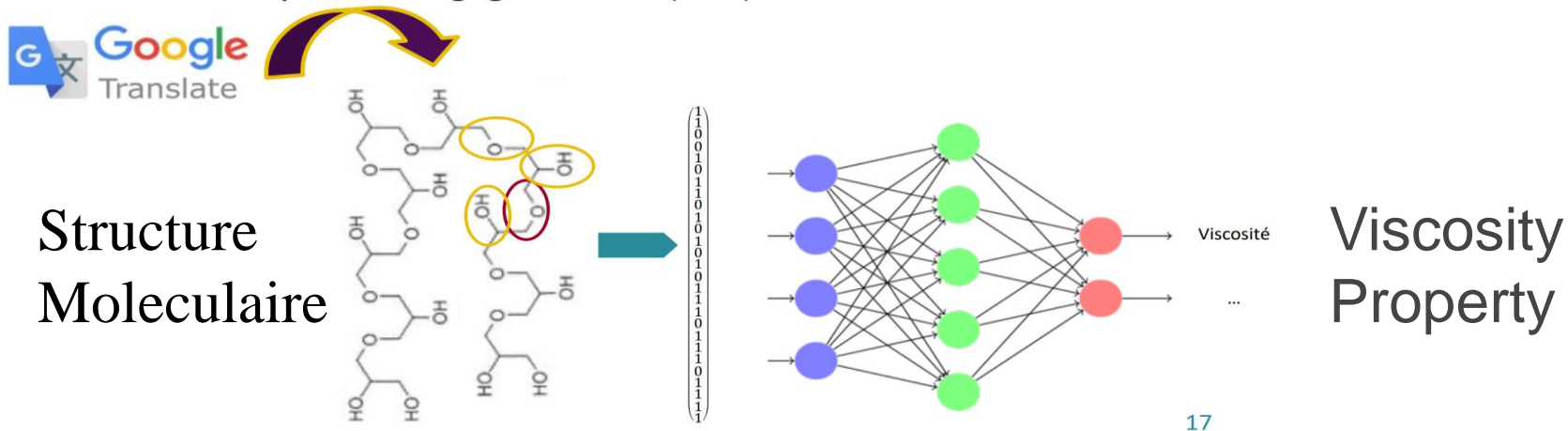


Neural Network: AI and Prediction Property Chemical

Diversion natural language analysis



Détourner l'Analyse du Langage Naturel (NLP)



Chemistry, Biochemistry and Artificial Intelligence...?

- 1- **Chemistry** and **Biochemistry** are scientific disciplines which study: composition, structure, material properties as well as reactions chemicals that take place between them
- 2- There are different types of data commonly used in chemistry, biochemistry
- 3- A.I. with **ML** (**Machine – Learning**) methods makes it possible to new exploration of this data to valorize it

Artificial Intelligence (AI) can be used to:

- > Classify molecules
- > Predict Physical, Biochemical Properties
- > Discover new Molecules, Materials



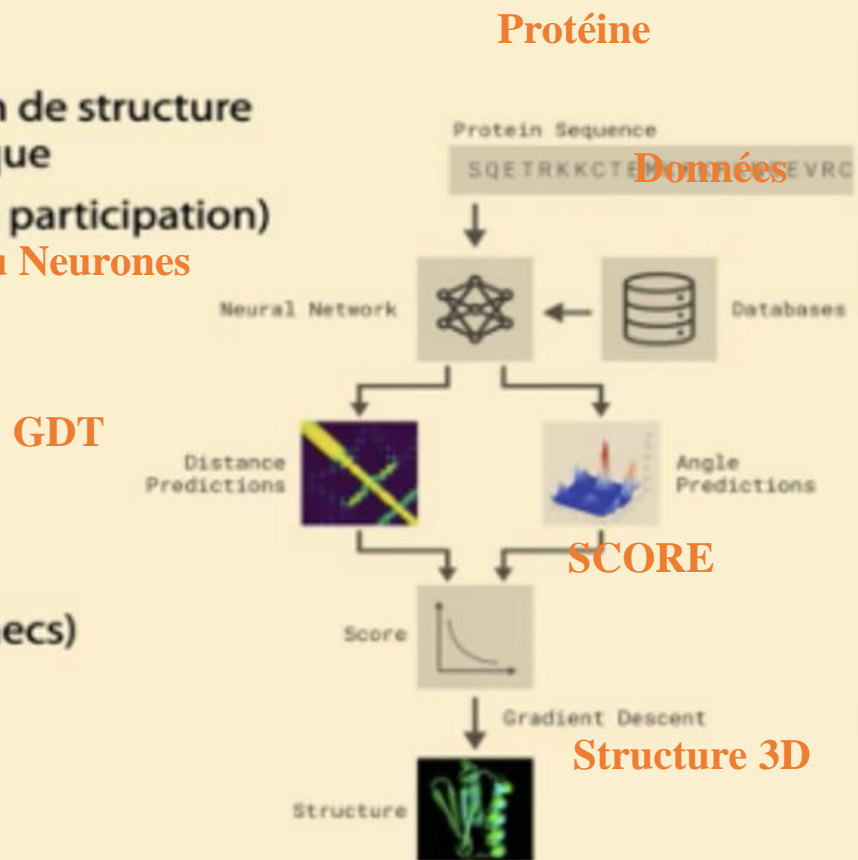
A.I. and exploitation of laboratory data ...Key to How Alpha Fold Works

Un tournant en biochimie

- ★ Compétition CASP 13 de prédiction de structure (repliement) de protéines au Mexique
- ★ Un nouveau programme (première participation) a remporté la première place **Réseau Neurones**
- ★ AlphaFold gage 25 / 43 tests
- ★ L'équipe n° 2 a remporté 3 tests

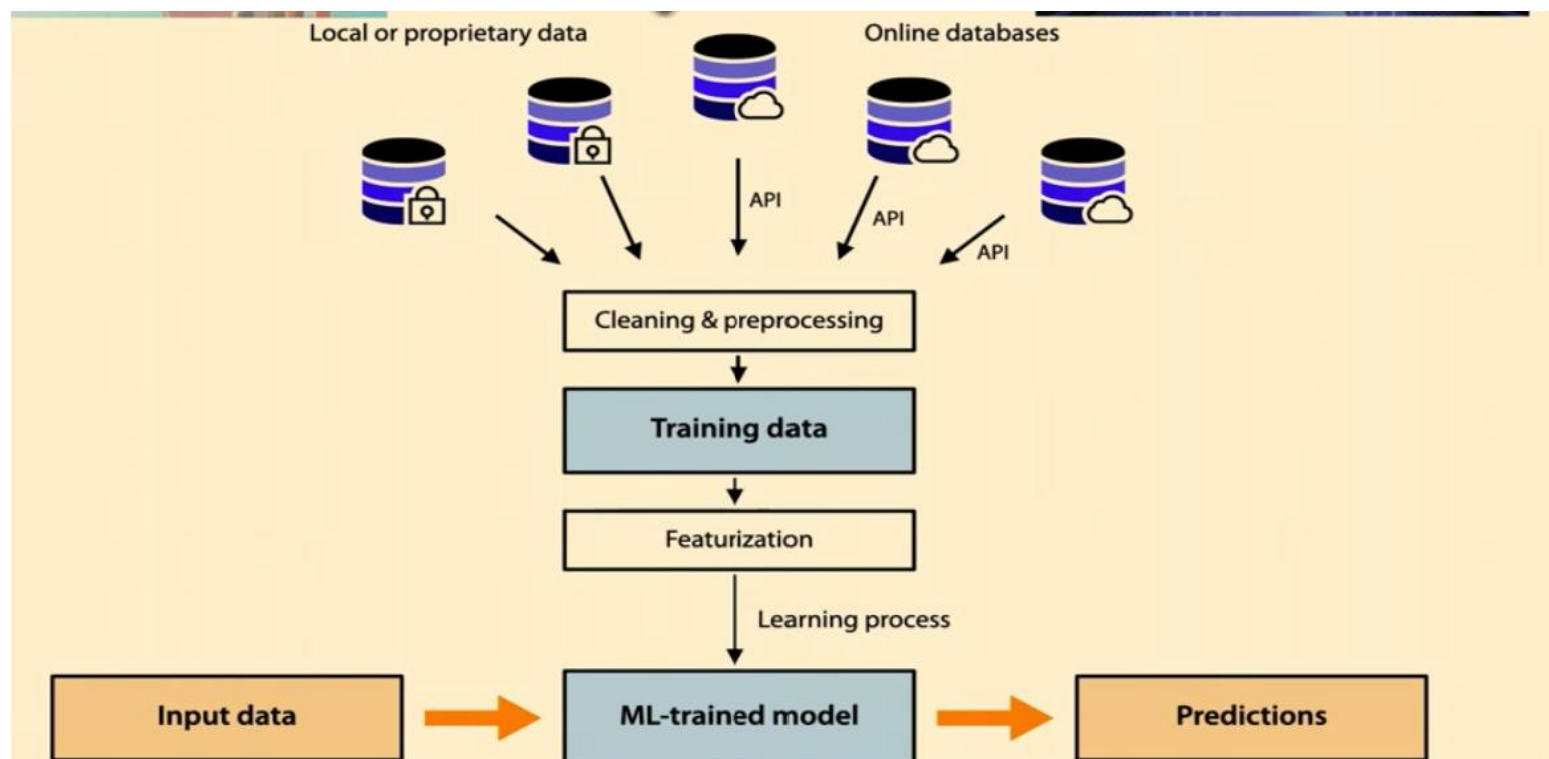
Dans de nombreux domaines

- ★ 1997 : Deep Blue bat Kasparov (échecs)
- ★ 2011 : Watson gagne à *Jeopardy!*
- ★ 2016 : AlphaGo bat Lee Sedol (go)
- ★ 2022 : ChatGPT



Laboratory and AI: training data, learning and exploitation Machine Learning Model...

Databases Laboratories



Learning
Training
of the Model



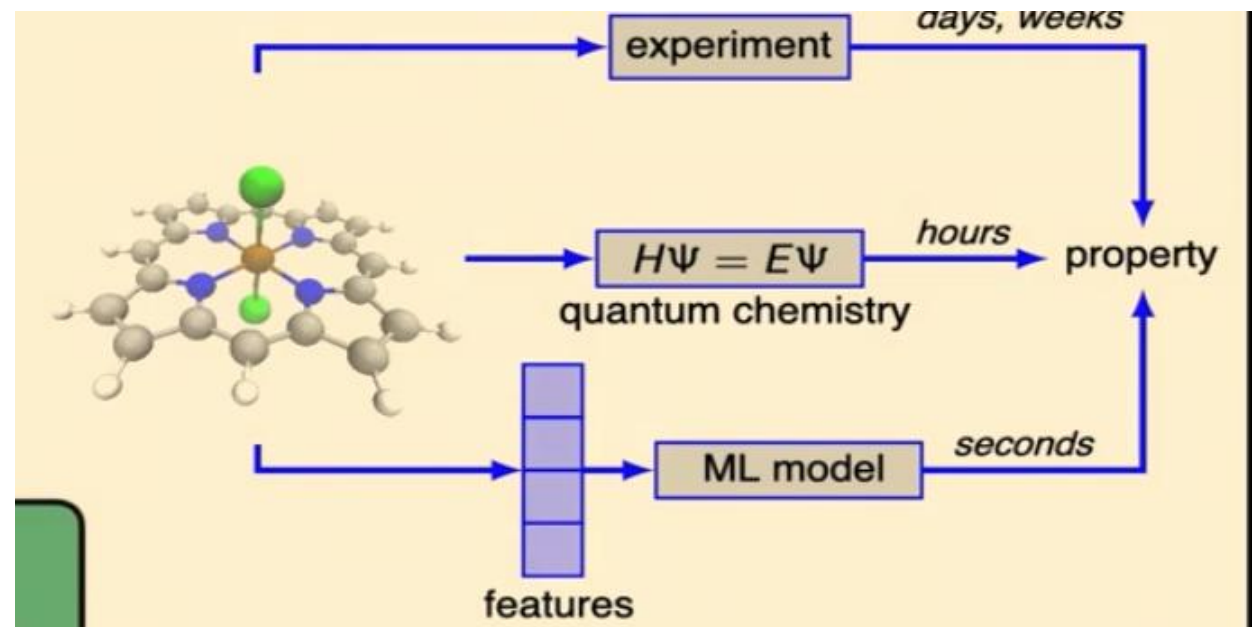
EXPERTSCIENCE

Use of the Model: solubility prediction, enthalpy etc.

Chemistry & AI (ML supervised learning)... Performance in the prediction of structure/property relationships...

3 possibilities:

- Lab Experience
- Theoretical Chemistry
- Machine Learning



The 3rd way if you have enough **data** to train a **Model** accelerates structure/property prediction
Performance = Weeks / Hours / + Seconds

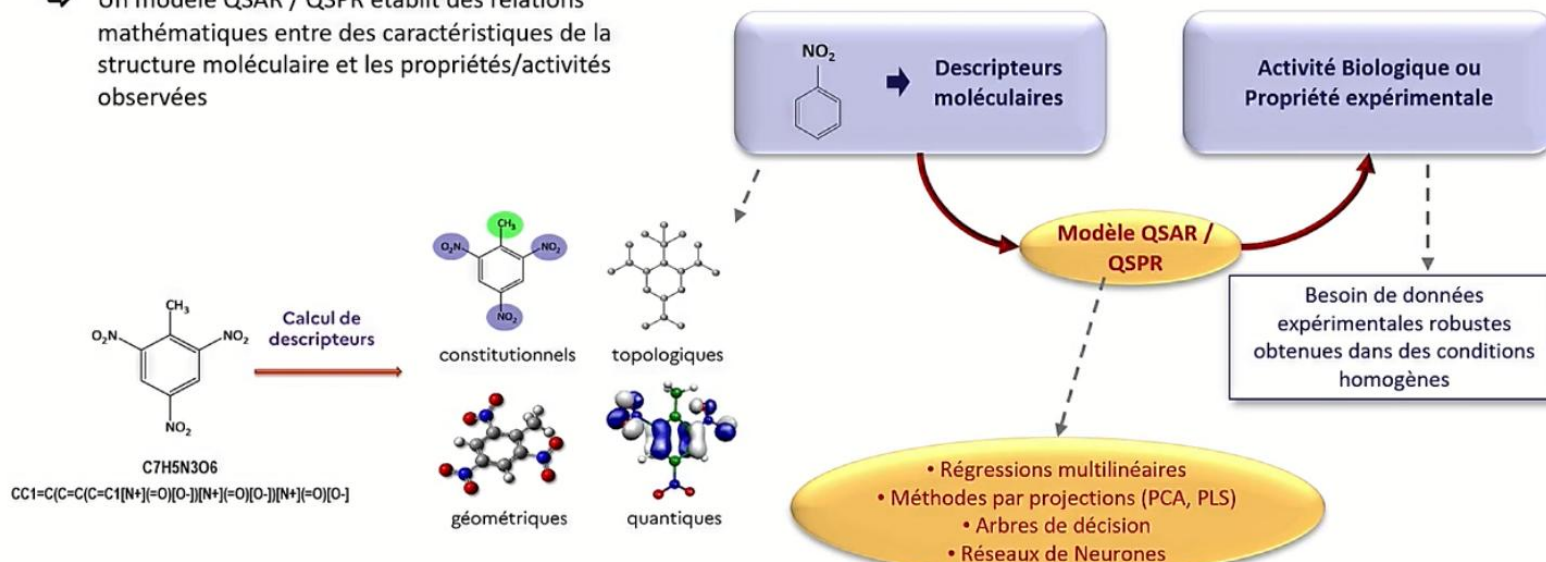


AI and New Methodological Approaches for control of Chemical Risk: alternative methods (QSAR / QSPR) uses of neural networks, supervised and unsupervised AI Machine Learning



Quantitative Structure-Activity/Property Relationships (QSAR/QSPR)

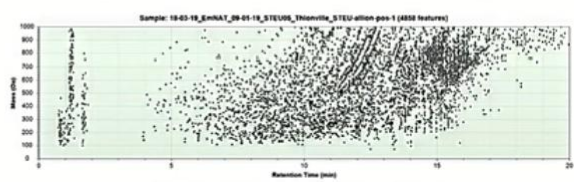
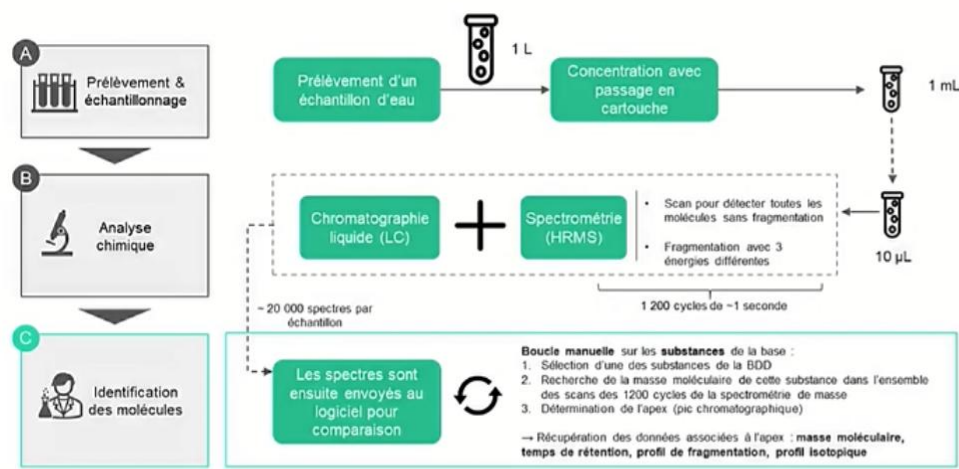
⇒ Un modèle QSAR / QSPR établit des relations mathématiques entre des caractéristiques de la structure moléculaire et les propriétés/activités observées



IDENTIFICATION of Contaminants...thousands of entities !

from the human loop... to the application of A.I. & Machine Learning > REX from INERIS

Épaulé l'Innovat | montrer le risque pour un développement durable



Limites des outils constructeurs

- 👎 Critères de décision et bases de données fermées et limitées
- 👎 Fiabilité et reproductibilité limitées
- 👎 Validation entièrement à la main de l'expert (vérification ligne par ligne) → long et fastidieux

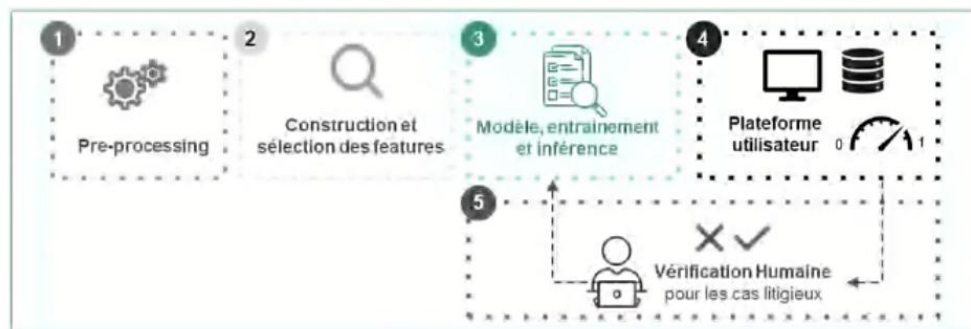
A.I. APPROACH and HRMS Analysis of Contaminants...

... construction of a **MODEL** from identification characteristics particular **in the spectra**

Projet « Empreintes environnementales »
(AMI IA#2 2020 & France Relance 2021)



SIAPARTNERS

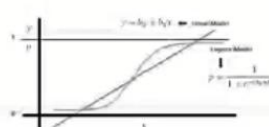
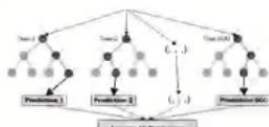


Bases de données

- Base interne Ineris
400 molécules
- Base constructeur
1200 molécules
- Base externe : MassBank
35000 molécules
(4000 avec même équipement que Ineris)

Différentes approches testées :

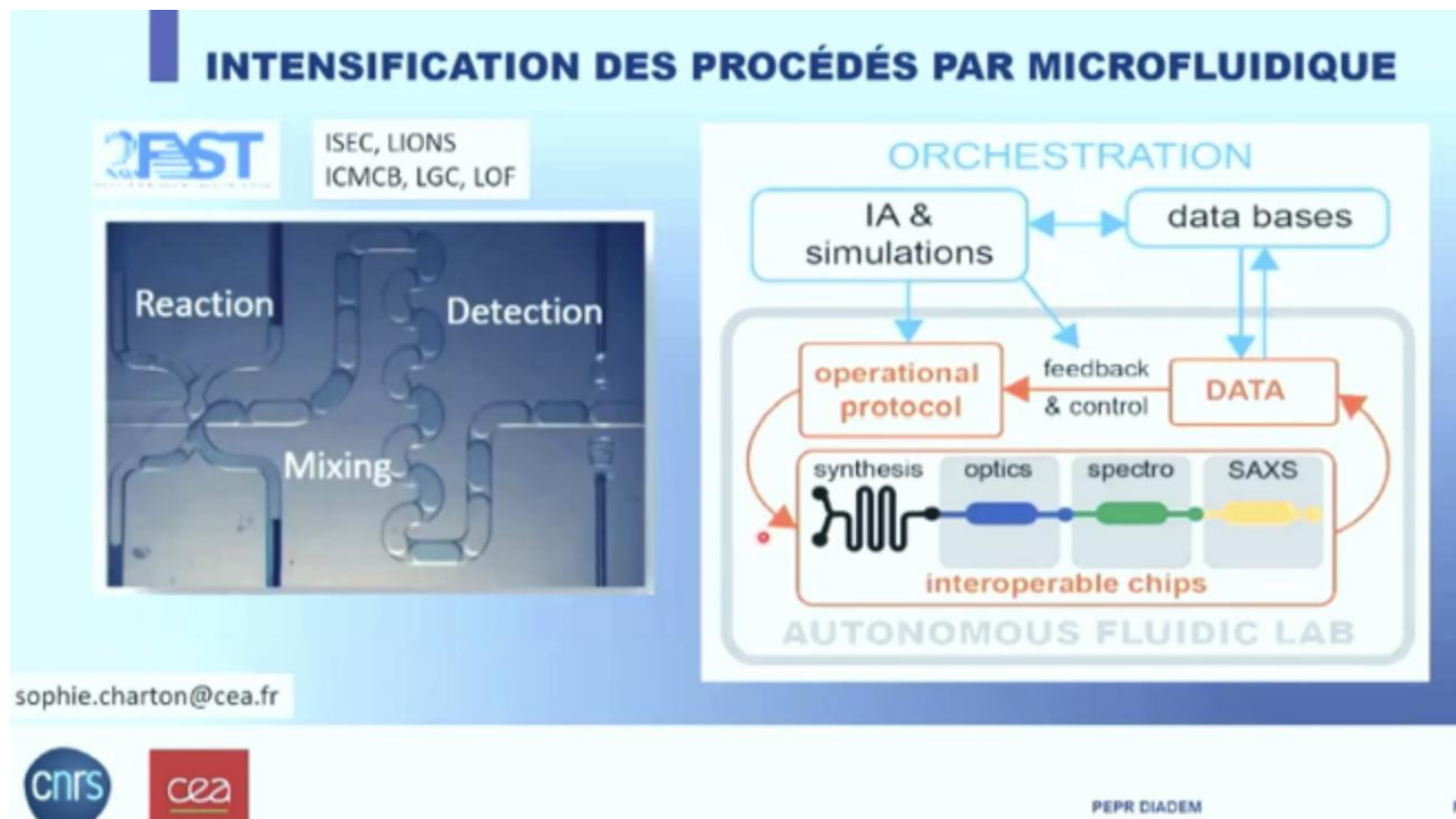
- MLP Classifier
- Decision Tree
- Random Forest
- KNN Classifier
- SVM Classifier
- Logistic Regression
- Bayes Classifier



GAINS: +> extended database
+> work + exhaustive & + fast

@Guillaume Fayet
INERIS

A.I. & Process Intensification by Microfluidics... orchestrated laboratory : synthesis, characterization, modeling, databases and A.I.



The Integration of A.I. in Laboratories...

...Revolutionises sample processing

- 1- the example of **PhénoMatrix**: AI embedded in an automated system of sample processing in **Microbiology**
- 2- an image processing algorithm (**Deep Learning**) works from a panel 5000 images provided by the laboratory learns and analyzes the images implementing the best strategies for detecting relevant elements for classify images
- 3- Thanks to advanced artificial intelligence (AI), the software **pre-evaluates and pre-automatically sorts culture plates** allowing laboratories to **microbiology to read, interpret and separate bacterial cultures**



EXPERTSCIENCE

Robotic Process Automation



Generative A.I. in Laboratories... a new deal for the profession

1- **Generative AI** = exploits AI techniques to generate content (texts, images, sound and even code)

2- 2000 **Language Model (LLM)** key concept is formulated
= model capable of predicting the next word in a sentence
= vector representation of words with coordinates in **semantic space**

3- 2022 arrival of ChatGPT (**GPT**=**G**enerative **P**retained **T**ransformer)
= Algorithm capable of learning contextualized word representations
= **Probabilistic statistical models** that determine the arrangement, distribution of words based on massive data ingested during their training

>> **Bing GPT4** — Bibliographie
-- solver d'équation

Microsoft openAI , Anthropic Amazon , Google



EXPERTSCIENCE



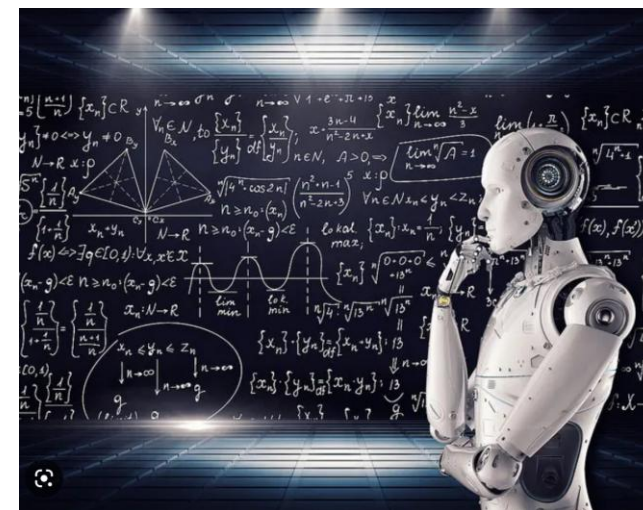
The A.I. : Matrix of the Laboratory of the Future?

Artificial intelligence:

An opportunity for LABORATORIES !

a Collaboration...Between the **Laboratorians**
and **A.I. SPECIALISTS**

Advent of augmented LabTechnicians?



“The power of Machine – Learning Algorithms
opens a new era for the **laboratory of the future!** ”





#LaboratoireduFutur To go further...

video replays of the entire #conferences of 2 symposiums >
and a link below >

TEST YOURSELF!!! LINK below

<https://webtv.insa-rouen.fr/videos/laboratoire-du-futur-2-le-defi-de-lintelligence-artificielle/>



<https://www.mediachimie.org/ressource/chimie-et-intelligence-artificielle-colloque-f%C3%A9vrier-2023>

A.I. and Laboratory ? = The A.I. Response !!!

Voici une réponse que j'ai obtenue à l'aide du nouveau Bing- GPT4 (Copilot)

LIEN : <https://sl.bing.net/kEK4X1Fu6zA>



Merci ! Thank You! let us keep in touch...

Yvon Gervaise  <https://www.linkedin.com/in/yvon-gervaise-22a1b035/>

 <https://twitter.com/expertscience>

Mail : expertscience.ygervaise@protonmail.com

Web: <http://www.expertscience.fr>

Tél : 06 60 67 91 50

