



fr

Introduction à l'intelligence artificielle



Ce cours est ronéisé allez voir la roné000 © (vous inquiétez aps si y'a des modifs elles apparaitrons dans ma fiche.#ouais je sais je suis la meilleure)

HEY LES SANGS COMMENT ÇA VA. C'EST VOTRE TUTRICE DU SUD QUI VOUS PARLE. 🍷.

LE PETITES INFOS IMPORTANTES AVANT DE COMMENCER :

EN vert => RAJOUTS au fil de l'année le prof peut rajouter des notions. Elles apparaitrons en vert pour vous permettre de les repérer plus rapidement

Barrée en orange mandarine => les infos qui ne sont plus dites

I. Définition de l'IA

- IA = « ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. »
- définition du Larousse
- L'IA se **positionne à l'intersection de diverses disciplines** (les mathématiques, l'informatique et la santé) et n'est pas une intelligence humaine.
- Elle reproduit seulement un résultat similaire aux résultats qu'aurait produit une intelligence humaine.
- L'IA c'est donc une technique d'illusionniste qui reproduit non pas la pensée humaine mais son **résultat**.
- Grace aux technologies, l'IA se trouve dans un monde avec de plus en plus d'applications et donc de plus en plus d'algorithmes comme :

- ✓ Computer vision (vision par ordinateur) : les ordinateurs peuvent maintenant voir et analyser le contenu des images (reconnaissance faciale, classification d'objets).
- ✓ La reconnaissance vocale.
- ✓ Netflix avec ses recommandations en fonction de ce que vous avez l'habit qui analyse les séries que les clients regardent pour pouvoir leur proposer des nouvelles séries qui vont leur plaire.
- ✓ Les robots humanoïdes (**≠ IA**) : pour automatiser une tâche et utilisent des systèmes électroniques, mécaniques, informatiques et éventuellement des algorithmes d'IA.
- ✓ Alphago : algorithme plus fort que les humains qui reproduit les performances d'un joueur.
- ✓ Big Data : données massives numériques analysées, classées, triées par des algorithmes d'IA.

Donc l'IA regroupe énormément de domaines et d'applications variées.

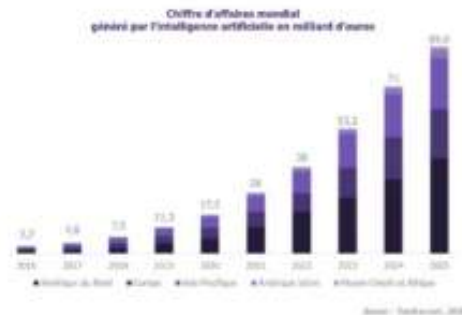


Le développement et le marché de l'IA

IA se développe **très rapidement** et a un potentiel énorme.

=> Sur le marché, on remarque que son chiffre d'affaires mondial = en **2015** inférieur à **1**

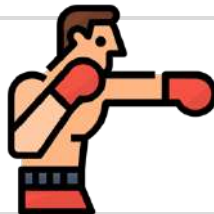
milliard d'euros et on estime qu'en **2025** il sera supérieur ou d'environ égal à **90 milliards d'euros**.



L'approche data Driven est-elle vraiment intelligente... ?

Il existe 2 types d'Intelligence artificielle :

L'IA FORTE = COMPARABLE À L'INTELLIGENCE HUMAINE



L'IA FAIBLE = Algorithmes = ensemble de règles ou d'instructions en vue d'obtenir un résultat

- ◆ Consciente de soi
- ◆ Compréhension de ses propres raisonnements
- ◆ Capable d'avoir des émotions, des intentions, de la créativité.
- ◆ Elle n'existe pas, on la voit **que dans les films** de science-fiction.
- ◆ Les scientifiques débattent aujourd'hui sur la possibilité d'arriver un jour à cette forme d'IA.

- ♣ Analyse et tri des données massives +++ (big data)
- ♣ Automatiser une tâche répétitive
- ♣ Formuler des recommandations après avoir examiné des milliers de décisions passées
- ♣ Elle **simule** le comportement/raisonnement humain pour obtenir le résultat mais **sans compréhension du processus** qui mène à ce résultat



Match time: IA vs Intelligence humaine!



 **Objectif du match : reconnaître un chat par rapport à un autre animal.**

IA (algorithme)	BB de 1 an
<p>Input (=entrée de l'algorithme) = des centaines de milliers d'images de chats.</p> <p>Output (=sortie, résultat) sera "ceci est un chat" si l'image montre un chat.</p>	<p>Input = quelques chats qu'il aura vu dans sa vie</p> <p>Output = Si on lui montre un chat, il le reconnaîtra. Il aura compris ce qu'est le concept du chat, que c'est un être vivant différent de lui.</p>

→ À la suite de ce match, on retient donc que le cerveau humain a une capacité d'apprentissage et de conceptualisation **supérieure** à l'algorithme d'IA. +++ (ndr encore heureux même si parfois...)

Le chatbot (comme les zommes au final #c'est pas du cours)

- Ils ne comprennent rien non plus à ce qu'on leur dit.
- Ils ne peuvent pas tenir une conversation
- Ils fonctionnent par **étapes** : on pose une **question**, l'algorithme analyse les sons, reconnaît la succession des mots et à partir de grosses bases de données, il ressort la réponse qui lui paraît la plus adaptée.
- → Ex. du robot Siri : s'il n'arrivait pas à répondre à la question, il était programmé pour raconter quelque chose de généraliste, à côté de la plaque ou une petite blague.

Le point faible de l'IA : l'adaptabilité +++

- L'IA a un **manque de flexibilité** car un algorithme d'IA est entraîné à résoudre une tâche très précise, dans laquelle il est très performant. Si l'on modifie la question ou le contexte dans laquelle la question a été posée, l'algorithme sera en **échec** et devra **complètement réapprendre**.
- Cette caractéristique du cerveau humain, c'est la **neuromodulation** +++ (=capacité du cerveau humain à adapter son apprentissage à des situations/contextes différents).
 - Pour l'IA faible, cette neuromodulation n'existe pas. C'est son grand point faible.

II. Approche « model driven » et « data driven »

Il existe deux grandes approches en IA :

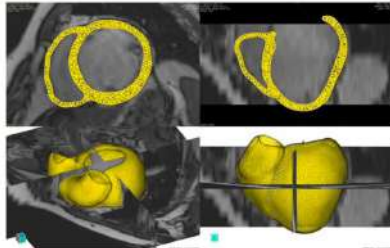
- L'approche guidée par le **modèle** = "model-driven" = "top to down"
- L'approche guidée par les **données** = "data-driven" = "Bottom up"



L'approche " model-driven "

Cette IA part d'un modèle qui peut être géométrique, mathématique, anatomique...

→ On part des **connaissances**, on construit notre **modèle** qui sont des **équations mathématiques** et ensuite on vérifie que ces équations s'appliquent à la réalité en les confrontant à des **données** et des **mesures** de la vraie vie.



A gauche nous avons l'exemple d'un modèle pour prédire l'évolution du patient : risque de développer une insuffisance cardiaque.

Grace a l'approche model driven on peut construire un **jumeau numérique** du patient basé sur des modèles statistiques et biophysiques du corps humain comme vu au-dessus.

On peut aussi développer des algorithmes sur ce jumeau numérique pour aider au diagnostic, au pronostic et à la thérapie : c'est la **médecine numérique**.

Pour devenir **+ efficace, robuste et explicable** il faut **combinaison** l'IA basée sur les données et l'IA basée sur un modèle.

L'approche data driven

Cette IA part des données pour construire un modèle statistique, elle « **donne du sens aux données** ». Elle est très utilisée puisqu'elle marche très bien pour certaines tâches et implique le producteur de données dès le départ en imposant des **conditions quantitatives et qualitatives** des données.

→ On part des données brutes du patient, on analyse ces données et ensuite on fait le modèle statistique qui sera utile pour le diagnostic

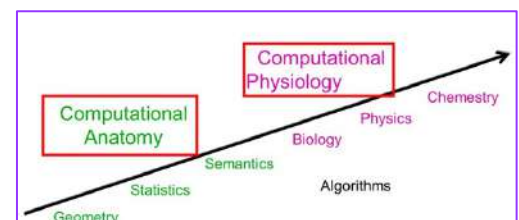
Comme on l'a dit, pour data-driven, on part **des données** qui peuvent être **quantitatives** et **qualitatives**. Il y a donc 4 caractéristiques, les 4 V = 4 conditions qui s'imposent pour le « data- driven » :

✓ **Vélocité** : vitesse à laquelle on va pouvoir capter les nouvelles données et les intégrer au modèle.

✓ **Volume** : quantité de données. Il en faut un assez gros pour pouvoir créer un modèle statistique issue du "data-driven".

✓ **Variété** : On peut utiliser différents types de données.

✓ **Véracité** : fiabilité des données (en lien avec la qualité).



+++A bien retenir +++

♥ le **model-driven** part des **connaissances** (physiques, biologiques...)

♥ le **data-driven** part de **l'observation des données massives** (en médecine, elles proviennent du patient).



III. Machin learning et deep learning



Il y a une articulation entre « Intelligence Artificielle », Machin Learning (apprentissage automatique) et Deep Learning (apprentissage profond). Cependant, il ne faut pas les confondre. (*Il faut voir ces notions comme des poupées russes.*)

La **1ère poupée** (la + grande) est l'**IA**. L'IA est un concept ancien qui émerge dans les années 1950 quand Alan Turing, un mathématicien Anglais, se demande si une machine est capable de penser comme un humain et va écrire un article intitulé « Computing Machinery and Intelligence ».

L'IA imite les fonctions cognitives humaines complexes : raisonnement, langage, perception visuelle ou auditive, déplacement d'un robot dans un milieu complexe...

Machin Learning



Elle correspond à la **2ème poupée** russe et se développe dans les années 1980/90.

Ici, l'ordinateur **déduit les règles** uniquement par l'analyse des données qu'on lui fournit. C'est une **approche « data-driven »** (à partir de données, on crée un modèle statistique).

Les Machine Learning se font sur des outils statistiques qui modélisent les données et les classent.

C'est une **méthode statistique**. L'algorithme **apprend** à partir des données reçues et **améliore** sa performance. Il n'est pas explicitement programmé pour résoudre des tâches : il y a **peu de programmation**.



Illustration du Machin Learning avec le cancer du sein :



◇ Le cancer du sein se caractérise par différents paramètres (taille, ganglions, distance des métastases ...). Chacune de ces caractéristiques participe au pronostic (plus la patiente a ces caractéristiques, moins elle a de chances de guérir).

La question : « Comment prédire à partir de la taille de la tumeur et du nombre de ganglions axillaires, le risque de récurrence de la patiente ? »

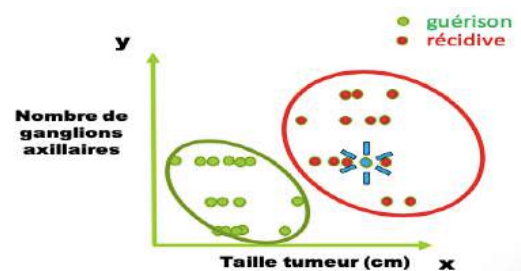
Petite explication d'anatomie : les ganglions axillaires c'est les ganglions des aisselles. (osf)

On utilise un graphique avec en abscisse : la taille de la tumeur et ordonnée : le nombre de ganglions axillaires de la patiente.

On positionne sur ce graphique les patientes en fonction de la taille de la tumeur et du nombre de ganglions. On suit les patientes donc on va pouvoir dire si elles vont guérir ou récidiver.

◇ On remarque que les patientes qui guérissent ont des tumeurs plus petites et moins de ganglions que les patientes qui font des récurrences. On peut faire un regroupement des patientes qui guérissent (en bas) et qui récidivent (en haut).

◇ On va prendre une **patiente n°1** avec un cancer du sein et la positionner sur le graphique (curseur) en fonction de la taille de la tumeur et du nombre de ganglions. On remarque que cette patiente se trouve dans le cercle « récurrence ». → Cette patiente appartient donc au groupe de patientes à hauts risques de récurrences. Là c'est « facile » (scuze nous), on n'a pas besoin de machin learning.



Le tutorat niçois est gratuit en plus d'être trop bien. Toute reproduction ou vente est interdite.



◇ Maintenant, on prend une **patiente n°2** mais cette fois-ci, elle ne se situe dans aucune des 2 zones (guérison ; récidive).

C'est là que rentrent en jeu les algorithmes statistiques pour essayer de classer cette patiente.

Pour cela, on utilise un outil statistique assez simple : le « **K nearest neighbors algorithm** » (= méthode des K plus proches voisins) .

En gros cette méthode regarde la distance qu'il y a entre le point de la patiente et les points des catégories identifiées « guérison » ou « récidive » .

Si l'on trace un cercle autour du point du curseur, on remarque qu'il y a 4 patientes qui guérissent et 1 qui fait une récidive.

→ La patiente a donc plus de probabilité de guérir que de récidiver.

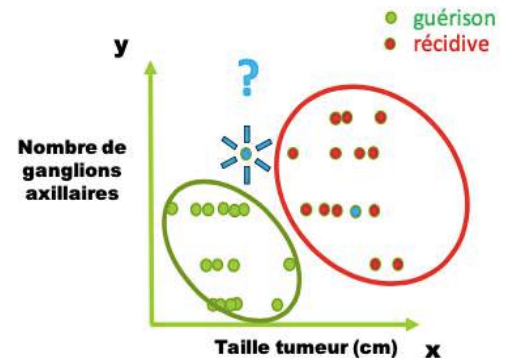
◇ **La caractéristique du machine learning, c'est que ces outils statistiques simples vont s'améliorer avec le nombre de données +++.**

Plus il y a de données, plus l'algorithme réussit à classer les patientes et donc prédit mieux la possibilité de guérison ou de récidive (pour cet exemple).

◇ **L'autre caractéristique du machine learning, c'est qu'ils vont pouvoir intégrer différentes caractéristiques de la patiente.**

Ici, on en a pris que 2. En réalité, le risque de récidiver dépend de beaucoup plus de facteurs...

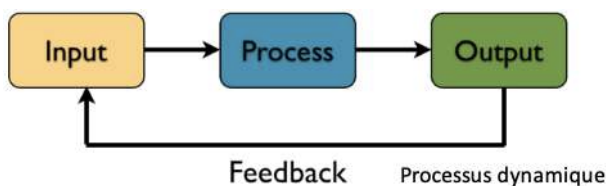
◇ **Donc plus l'algorithme a de données en Input, mieux il va classer et meilleur sera l'Output pour une prédiction précise**



La notion de Feedback

▫ Pour que l'algorithme apprenne, on va lui dire s'il a bien classé les nouvelles données. C'est un processus dynamique.

En effet, les **données** arrivent et on a la **classification, l'output**, puis le **feedback**. Si l'algorithme a mal fait le travail, il va de lui-même modifier certains paramètres pour améliorer sa prédiction.



- 1) phase d'apprentissage pour créer le modèle
- 2) Phase d'évaluation du modèle

▫ Il y a donc 2 phases dans le développement des algorithmes du machine learning :

1. Phase **d'apprentissage** (pour créer le modèle)
2. Phase **d'amélioration** (on vérifie qu'il ait bien classé et s'il y a erreur il se corrige lui-même)

Il existe différentes approches de machine learning :

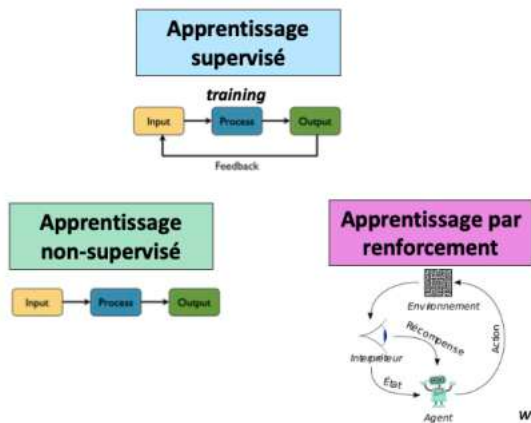
✓ **L'apprentissage supervisé** : on apporte des données à l'algorithme et on donne le feedback ce qui va lui permettre de résoudre des problèmes de classification ou de régression.

Le tutorat niçois est gratuit en plus d'être trop bien. Toute reproduction ou vente est interdite.

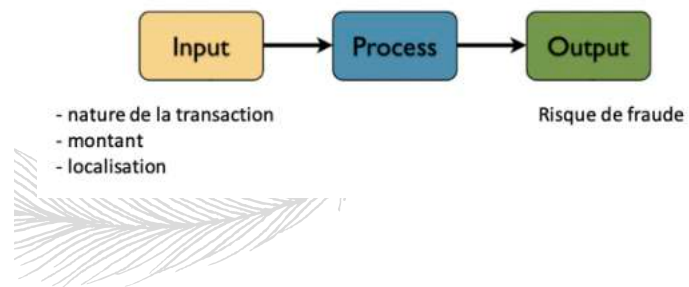


✓ **L'apprentissage non-supervisé** : On apporte à l'algorithme des données mais sans feedback.

✓ **L'apprentissage par renforcement** : Consiste pour un agent autonome (ex : robot), un algorithme d'être plongé dans un environnement complexe et de prendre ses décisions en fonction de son état dans l'environnement. S'il prend une mauvaise décision, il aura un retour négatif et vice-versa. La notion de Feedback est donc très importante. Le robot apprend à avoir un comportement décisionnel optimal dans un environnement donné grâce au Feedback. C'est donc un processus dynamique.

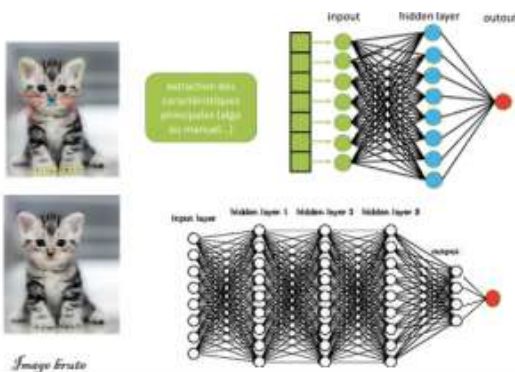


Petit exemple de la détection de la fraude avec CB :



Deep Learning

- ❖ Il s'agit d'un apprentissage « en profondeur » constituant la **3ème** poupée russe. C'est une **catégorie de machin learning** +++.
 - ❖ Il utilise les **big data** et nécessite de **fortes puissances de calcul**. Il est très lié à l'explosion des **réseaux de neurones** (type d'algorithme) en 2012 : performance ++
 - ❖ Les réseaux de neurones sont des **fonctions mathématiques** qui ont la particularité d'être très nombreuses avec de multiples connexions entre elles et une organisation en différentes couches. Elles s'inspirent de l'organisation des neurones.
- ⇒ **L'intérêt de ces réseaux de neurones, c'est qu'ils peuvent gérer les Big Data, d'avoir beaucoup d'informations en entrée.**



→ Si on prend une image de 400 pixels sur 600, il y aura 240 000 pixels (600x400), ce qui correspond à 240 000 informations, entrées du modèle de machin learning.

→ Si on utilise un algorithme standard de machine learning, à un moment il sature, on ne peut pas faire de classification d'images avec, contrairement au réseau de neurones.

- ❖ Cette révolution des algorithmes d'apprentissage profond pour faire de la classification d'image **date de 2012**. Les réseaux de neurones existaient avant mais leurs performances ont été prouvées lors d'un challenge de reconnaissances d'images qui a lieu tous les ans. Le professeur montre sur son diapo les résultats d'algorithmes à ces challenges.
- ❖ **Plus le pourcentage est bas, meilleur est le résultat. En 2012, il y a l'algorithme supervision, un réseau de neurones profonds** qui fait 16% au test et écrase la concurrence.
- ❖ Sur les réseaux de neurones, il y a un nom à connaître, c'est un **français, Yann Le Cun**. C'est un des pionniers de ces réseaux. Il y travaillait depuis les 1990's quand personne ne s'y intéressait et était un des seuls à en créer avant 2012.



- ♣ Les raisons du succès des réseaux de neurones :
 - ✓ la sophistication des algorithmes : plus il y a de couches, plus l'analyse de l'image est fine.
 - ✓ la puissance de calcul nécessaire : grâce aux GPU.
 - ✓ Disponibilité des Big Data :

Pour l'apprentissage de ces réseaux de neurones, il faut des milliers d'images. Cela a été possible grâce à internet, où des grandes bases de données ont été mise à disposition des chercheurs.

- ♣ Ces réseaux sont aujourd'hui très utilisés dans la :
 - ✓ Reconnaissance faciale
 - ✓ Reconnaissance vocale
 - ✓ Reconnaissance et suivis d'objets
 - ✓ Classification des objets (véhicules autonomes).

++les points clés du Deep Learning++

- Il va « dans la profondeur » des données
- Il permet de faire des tâches spécifiques, limitées, bien définies : « narrow AI »
- Les approches sont purement basées sur les données : il faut de grandes bases de données annotées, on a un risque de biais, des problèmes éthiques, etc...
- Il existe un effet de « **Black Box** », c'est-à-dire que l'on a du mal à expliquer le résultat du réseau de neurones car l'algorithme est capable de très bien classer les images mais sans comprendre comment il fait cette classification

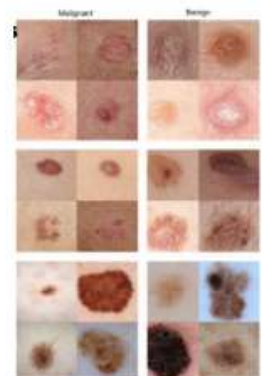
IA > Machine Learning > Deep Learning

IV. IA & médecine : quelles applications ? → l'exemple des omics

- ♣ On assiste à une digitalisation des informations et à l'explosion des données médicales. L'IA appliquée aux données médicales permet :
 - D'automatiser des tâches médicales répétitives = imitation du résultat produit par le médecin
 - D'intégrer et d'analyser un flux de données trop complexes pour le cerveau humain = augmentation des capacités cognitives des médecins pour améliorer le soin
 - D'identifier de nouveaux facteurs de risque / biomarqueurs / Nouvelles pistes de recherche

1. En dermatologie :

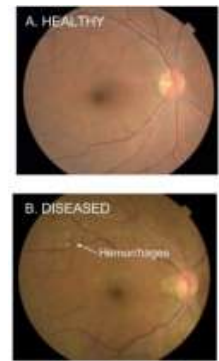
- En 2017 une publication sort dans le magazine Nature montre ce dont l'IA est capable. Dans l'optique de classer des photographies de lésions cutanées une équipe de chercheurs de Stanford utilise des réseaux de neurones artificiels profonds (multi-couches) pour savoir si la lésion était bénigne ou maligne (cancer).
- L'apprentissage est plutôt long, il faut environ 1,3 millions d'images puis encore 130 000 autres pour les ajustements finaux.
- Ces réseaux de neurones peuvent reconnaître des lésions de 2000 pathologies et ont donc les performances de dermatologues confirmés.





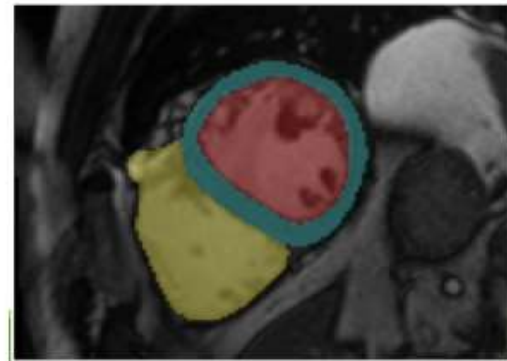
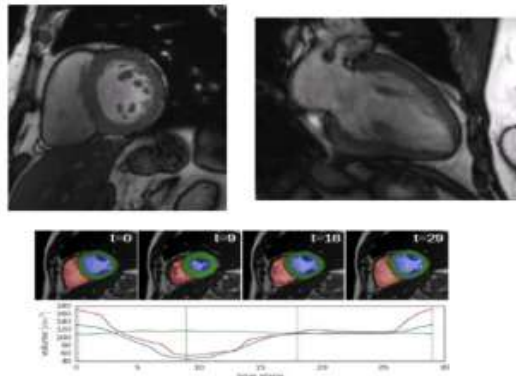
2. En ophtalmologie :

- En 2016, une équipe de Google a développé et fait valider un algorithme pour faire la classe des rétines et reconnaître les rétines normales et les rétinopathies diabétiques.
- Encore une fois l'apprentissage est plutôt long, il a fallu environ 130 000 patients puis encore 10 000 patients autres pour la validation.
- En avril 2018, le FDA a approuvé l'utilisation du logiciel en routine.



3. En segmentation anatomique :

Segmentation manuelle :	Segmentation automatique :
→ Chronophage ventricule gauche : 30min et ventricule droit : 50 min → Tâche peu gratifiante et stimulante → Manque de reproductibilité	→ Temps : 10 sec en tout → Deep Learning : 0.97 (score de corrélation) → Gain d'efficacité+++



4. L'essor des « omics » en médecine

- ∞ Les **omics** sont les Big Data biologiques produites par les nouvelles technologies nécessitant de l'IA pour en extraire des connaissances, les analyser et leur donner du sens.
- ∞ Les sciences omics :
 - ✓ L'analyse du **génom**e = **génomique**
 - ✓ L'analyse du **transcriptome** = **transcriptomique**
 - ✓ L'analyse du **protéome** = **protéomique**
 - ✓ L'analyse du **métabolisme** = **métabolomique**
- ∞ Elles produisent **des milliers de données biologiques (Big data) ce qui permet de faire de la médecine personnalisée** en proposant le traitement le plus adapté au patient. Elles ont besoin de l'IA pour en extraire des connaissances et leur donner du « sens » puisqu'elles travaillent sur les données brutes, sans sélection : dizaines de milliers de métabolites.

