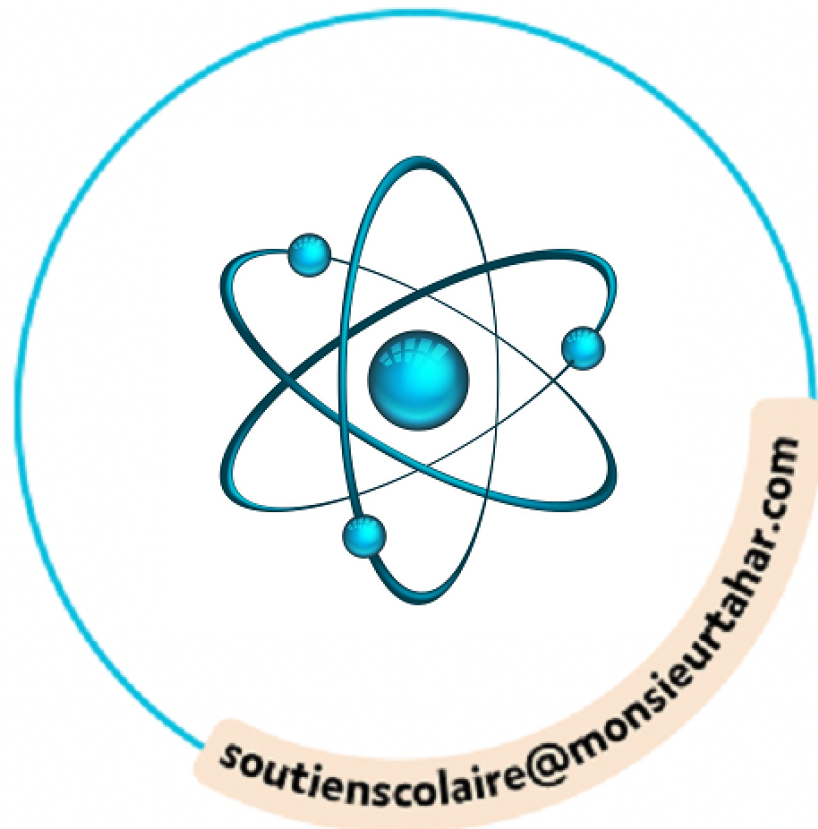
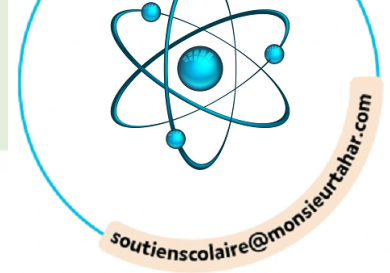


ENSEIGNEMENT SCIENTIFIQUE



PHYSIQUE - CHIMIE

CHAPITRE 7



1. Le traitement de l'information numérique

► Jusqu'au début du xx^e siècle, les machines automatisées sont limitées à une ou quelques applications particulières. Turing est le premier à imaginer une **machine universelle** capable de traiter n'importe quel type de données. Cette machine est matérialisée 10 ans plus tard avec les premiers ordinateurs, composés au minimum d'un processeur et d'une mémoire vive. L'informatique se répand ensuite largement avec l'**informatique embarquée**. > **Unité 1**

► Les ordinateurs manipulent divers types de données numérisées : images, sons, vidéos... Les programmes sont des informations stockées dans le même type de mémoire que les données. Un programme écrit dans un langage de programmation de haut niveau, comme Python, peut être traduit en instructions spécifiques à chaque type de processeur. Ainsi, un **fichier exécutable** sur une machine n'est pas obligatoirement exécutable sur une autre. > **Unité 2**

► Un programme complexe moderne peut comporter plusieurs centaines de millions de lignes de code, ce qui rend très probable la présence de **bugs** (ou bogues). Ces erreurs peuvent conduire un programme à avoir un comportement inattendu et entraîner des conséquences graves. > **Unité 3**

2. Intelligence artificielle et apprentissage machine

► L'**intelligence artificielle** (IA) regroupe des techniques permettant à des machines de réaliser des actions normalement associées à l'intelligence humaine. L'**apprentissage machine** (ou automatique) est une branche de l'IA qui utilise une grande quantité de données (big data) pour entraîner un programme à repérer des corrélations ou des similitudes. Le programme, une fois entraîné correctement, est capable de prédire ou de décider à partir de données nouvelles. > **Unité 4**

► L'IA soulève de nombreuses questions éthiques. L'apprentissage machine utilise une grande quantité de données personnelles pour lesquelles un consentement explicite doit être obtenu. Par ailleurs, la qualité des données d'entraînement est essentielle puisque des biais en entrée peuvent se retrouver amplifiés en sortie. Le développement des voitures autonomes soulève d'autres questions éthiques, comme la responsabilité engagée lors d'un accident. > **Unités 5 et 6**

3. L'inférence bayésienne dans l'apprentissage automatique

► L'**inférence bayésienne** permet d'estimer la plausibilité d'une hypothèse, étant donné des observations concrètes. Elle est très utilisée en apprentissage automatique, notamment en vue de prononcer un diagnostic (médical, détection de spam...). Lorsqu'on applique une procédure de décision, nécessairement imparfaite, il apparaît des erreurs qu'on appelle des « faux », et qu'on qualifie de positif ou négatif selon le résultat de la détection. La construction d'un tableau de contingence permet de calculer la fréquence de faux positifs et de faux négatifs, qui caractérise la qualité de la procédure de décision.

► La performance d'un test diagnostique est de plus caractérisée à l'aide de la sensibilité, qui mesure la probabilité qu'un test soit positif lorsqu'il doit l'être, et de la spécificité, qui mesure la probabilité qu'un test soit négatif lorsqu'il doit l'être. Ces grandeurs permettent notamment de calculer la probabilité d'avoir une maladie si le test est positif : c'est la valeur prédictive positive dont le résultat dépend de la **prévalence** de la maladie. > **Unité 7**

Les mots-clés du chapitre

• **Machine universelle** : Machine, imaginée par Turing, capable de simuler n'importe quelle machine réalisant un calcul. La configuration nécessaire fait alors partie des données d'entrée.

• **Informatique embarquée** : Appareils de la vie courante dont le fonctionnement repose sur un ou plusieurs programmes informatiques.

• **Fichier exécutable** : Fichier qui contient le programme sous une forme qui peut être traduite en instructions spécifiques à chaque type de processeur.

• **Bug** : Erreur dans un programme entraînant un comportement inattendu lors de l'utilisation du programme.

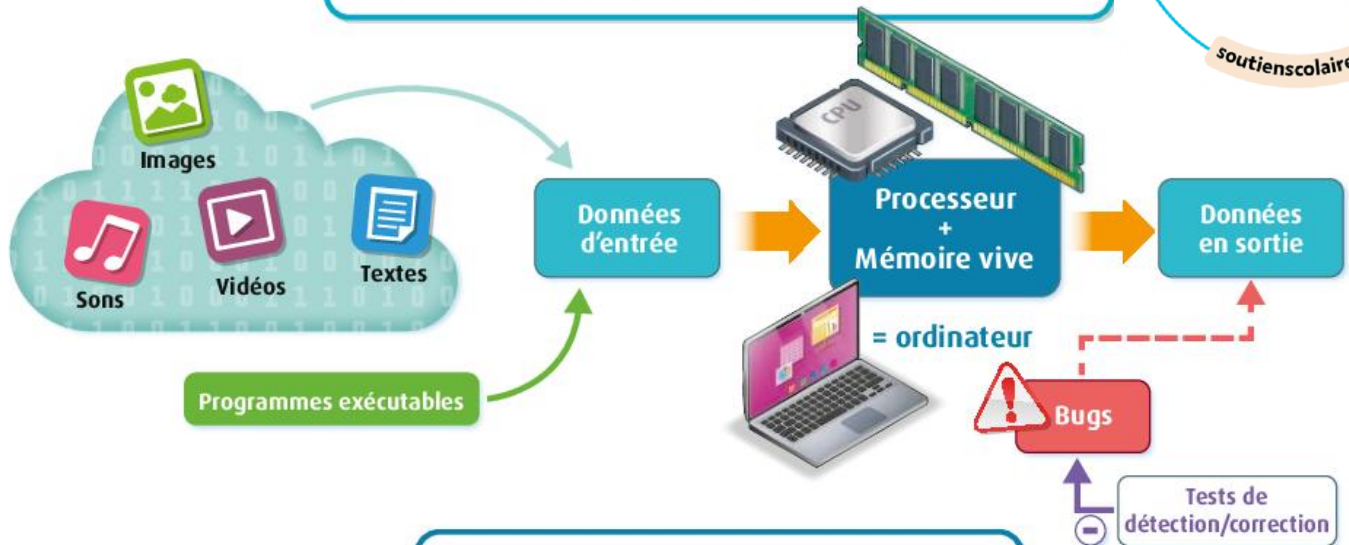
• **Intelligence artificielle** : Ensemble de théories et de techniques qui traite de problèmes dont la résolution fait normalement appel à l'intelligence humaine.

• **Apprentissage machine** : Branche de l'IA qui utilise des programmes repérant des corrélations ou des similitudes sur un très grand nombre de données et peuvent, après cet entraînement, prédire des résultats ou prendre des décisions sur d'autres données.

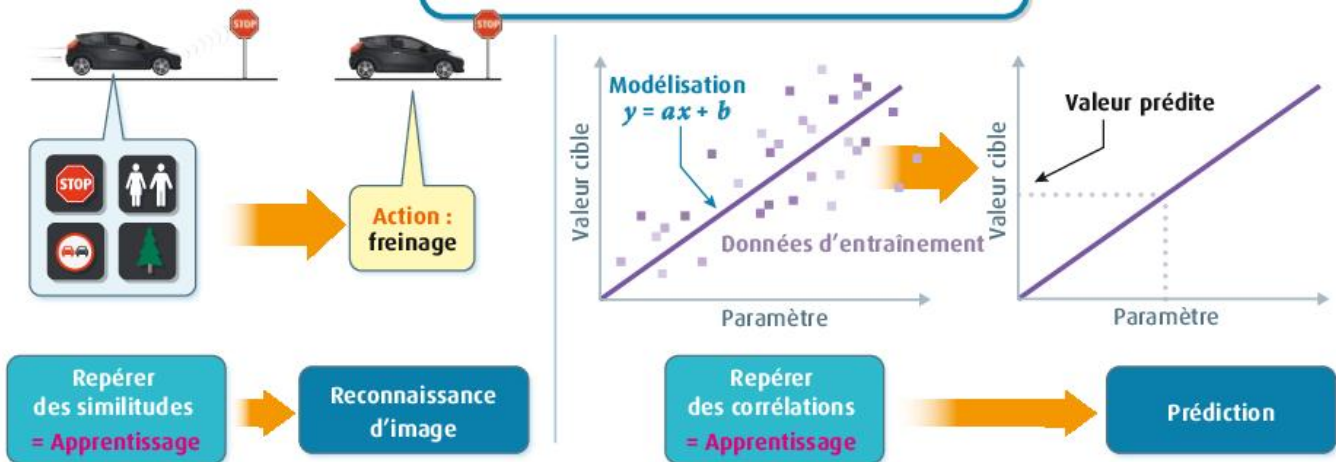
• **Inférence bayésienne** : Méthode de calcul de probabilités de causes à partir des probabilités de leurs effets.

• **Prévalence** : Proportion de personnes malades au sein d'une population.

Le traitement de l'information numérique



L'IA et l'apprentissage machine



La quantité et qualité des données d'entrée détermine la qualité des données en sortie ➡ Risque de biais

L'inférence bayésienne

► Méthode de calcul de la probabilité des causes à partir de la probabilité de leurs effets

► Tableau de contingence :

	Caractère présent	Caractère absent	
Test positif	Vrai positif (VP)	Faux positif (FP)	$f(FP) = \frac{FP}{VP + FP}$
Test négatif	Faux négatif (FN)	Vrai négatif (VN)	$f(FN) = \frac{FN}{FN + VN}$
Sensibilité = $\frac{VP}{VP + FN}$		Spécificité = $\frac{VN}{FP + VN}$	

➡ Caractérisation de la qualité d'une procédure de décision dans un test diagnostique