Fiche

Dès 1950, Alan Turing publie un article intitulé *Computing Machinery and Intelligence* dans lequel il se demande si les machines peuvent imiter l'homme. Mais la véritable recherche en intelligence artificielle (IA) n'est apparue qu'en 1956, lors d'une université d'été sur le thème des machines pensantes. Aujourd'hui, l'intelligence artificielle permet l'accomplissement de tâches et la résolution de problèmes jusqu'ici réservés aux humains, comme reconnaître et localiser des objets sur une image, conduire une voiture, jouer aux échecs, etc. Mais l'IA pose également certains problèmes éthiques.

Histoire, enjeux et débats

Dès 1725, B. Bouchon invente un système de ruban perforé pour automatiser un métier à tisser : les pièces du métier réagissent différemment selon qu'il y a un trou ou non sur le ruban. C'est l'un des premiers systèmes binaires ayant jamais existé. En 1728, son assistant remplace les rubans par des cartes qui resteront utilisées jusqu'aux années 1960. Ainsi, jusqu'au début du XX^{e} siècle, les machines traitant l'information sont limitées à une ou quelques tâches prédéterminées. Turing a été le premier à proposer le concept de machine universelle en 1936, matérialisée en 1943 par le premier ordinateur : ENIAC.

Alan Turing démontre en 1936 que certains problèmes mathématiques ne peuvent être résolus. Pour cela, il postule qu'il existe une machine programmable capable d'effectuer toutes sortes de calculs. Cette machine portant le nom de « machine de Turing » est la première affirmation qu'un appareil, à condition d'être programmé, peut faire toutes sortes de calculs et donc de tâches. Elle montre aussi qu'elle peut, en simulant l'activité mentale, devenir aussi « intelligente » que l'homme. Turing propose aussi l'idée d'universalité, puisque sa machine pourrait accomplir les tâches de n'importe quelle autre machine. Il est aussi connu pour avoir réussi à décrypter le code des machines Enigma, utilisées par les nazis au cours de la Seconde Guerre mondiale.

Dans les années 1950, les progrès de l'électronique ont permis de développer la première génération d'ordinateurs constitués de mémoire vive et d'un processeur. Puis, l'essor de la miniaturisation des composants a amené des processeurs de plus en plus puissants, qui permettent de produire des ordinateurs plus puissants, lesquels peuvent aider à concevoir des processeurs plus puissants, et ainsi de suite.

Aujourd'hui, les ordinateurs utilisent des fichiers de différentes natures : image, vidéo, texte, etc. Mais, comme cela a déjà été vu en SNT en Seconde, un ordinateur fonctionne uniquement avec le langage binaire (constitué de 0 et 1, chacun d'eux représentant 1 bit). Ainsi, chaque caractère est codé sur un octet (8 bits).

Un fichier .txt (écrit avec le bloc-note, par exemple) occupe autant d'octets qu'il contient de caractères. Mais un fichier .doc ou .docx contient, en plus, un en-tête descriptif avec la date, l'auteur, le logiciel utilisé, etc., et occupera donc une taille bien plus grande que le même texte écrit avec un bloc-note.

De même, un fichier image comportera aussi un en-tête dans lequel on pourra trouver la résolution, la taille, mais aussi, selon le format, l'appareil photographique utilisé et même la position GPS de l'image. Chaque pixel de l'image est codé sur un ou plusieurs octets selon la qualité. La taille des fichiers image peut donc atteindre plusieurs Go (Giga-octets) selon la définition (nombre de pixels de l'image). Un fichier vidéo constitué d'un nombre très important d'images peut atteindre plusieurs centaines de Go.

Les fichiers son ont également une taille variant selon leur qualité (la compression utilisée), comprise généralement entre 1 et 10 Mo (Méga-octets).

Il existe aussi des fichiers écrits dans des langages spécifiques (Python, Scratch, C, Sharp, etc.). Ces programmes permettent de faire le lien entre le programmeur et les instructions comprises par le(s) processeur(s). Certains langages de plus bas niveau, comme le langage assembleur, permettent au programmeur d'indiquer directement les instructions à chaque partie (mémoire, registre) du processeur.

Qu'est-ce que l'intelligence artificielle (IA) ?

La signification même d'*intelligence artificielle* est variable selon l'époque et les interlocuteurs. Yann Le Cun (prix Turing 2018) en énonce une définition : « l'IA est un ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux ». Le terme *intelligence artificielle* recouvre donc un ensemble de théories et de techniques qui traitent de problèmes dont la résolution fait appel à l'intelligence humaine.

La recherche en intelligence artificielle a permis de réaliser des progrès fulgurants en matière de robotique, de véhicules autonomes, de traitement de la parole, en compréhension du langage naturel. Quelques exemples remarquables : la description automatique du contenu d'une image par Google, les systèmes de reconnaissance faciale, le niveau de compétence des robots au Robotic Challenge de Darpa, la compréhension de la parole avec Siri d'Apple, Cortana de Microsoft ou OK Google.

Parmi ces nombreuses réalisations, beaucoup sont le fruit des progrès accomplis dans le domaine de l'apprentissage machine (apprentissage automatique).

Qu'est-ce que l'apprentissage machine (*machine learning*) ?

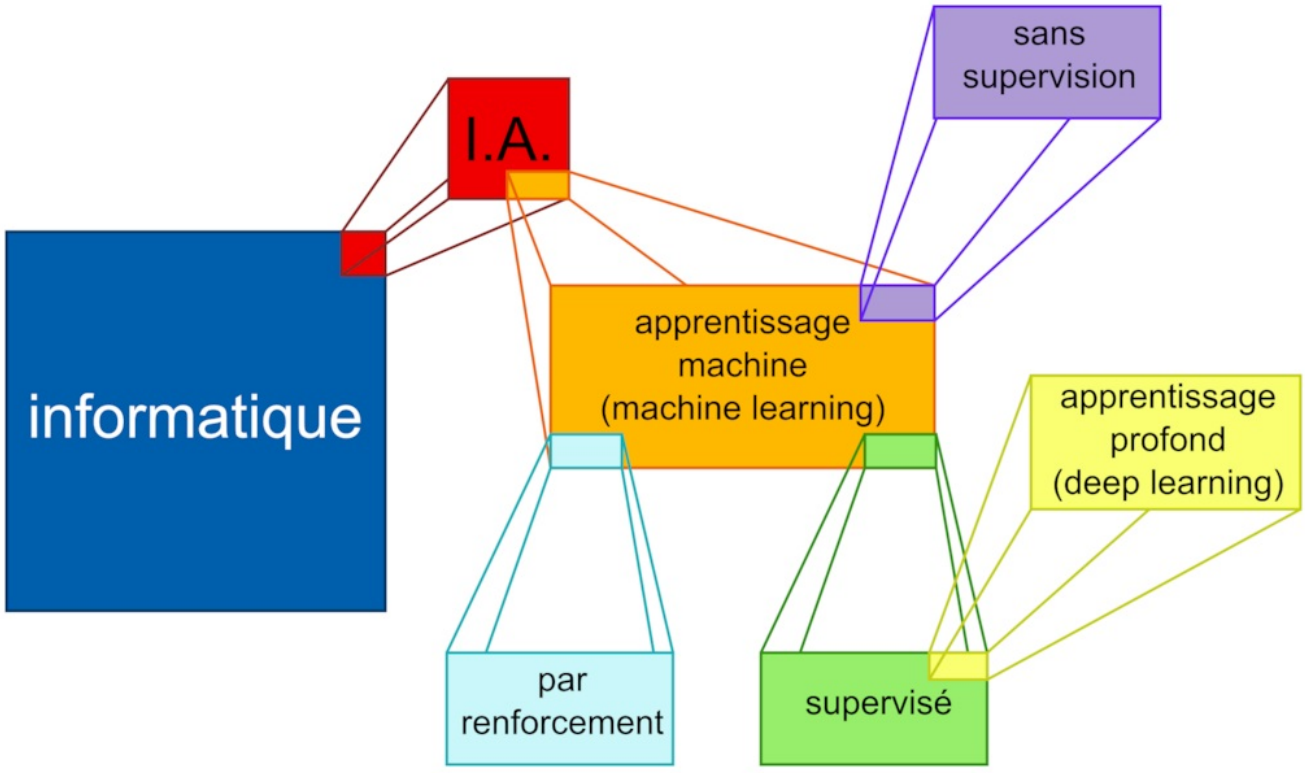
L'arrivée des centres de données (*data center*), lieux où sont stockés de multiples données, a permis une explosion de l'apprentissage machine. Il correspond à une branche commune de l'intelligence artificielle et des statistiques. Il consiste en des programmes capables

de modifier leur comportement lorsqu'ils sont confrontés à des données.
En fournissant à une machine un très grand nombre de données, on permet à celle-ci de s'entraîner (phase d'apprentissage ou d'entraînement) pour définir son comportement ultérieur (phase d'inférence). La machine exploite des méthodes mathématiques qui, à partir du repérage de tendances (corrélations, similarités) sur de très grandes quantités de données (*Big data*), permettent de faire des prédictions ou de prendre des décisions.
Il y a trois catégories d'apprentissage machine :

• l'apprentissage avec supervision : les opérateurs donnent à l'ordinateur des exemples d'entrées et les sorties souhaitées, et l'ordinateur recherche les solutions par modélisation prédictive. Il existe de nombreux algorithmes qui entrent dans cette catégorie : méthode du *k* plus proche voisin, régression linéaire, arbres décisionnels, etc. Ceux-ci sont par exemple utilisés pour effectuer les prévisions météorologiques. Le *deep learning* (apprentissage profond) est l'une des méthodes d'apprentissage supervisé : les sorties de chaque module servent d'entrée aux suivants. On parle alors de « réseaux de neurones artificiels ».

• l'apprentissage sans supervision : l'algorithme est laissé à lui-même, alors qu'on lui fournit un grand nombre de données non étiquetées. La machine y repère des corrélations pour construire elle-même son algorithme de classification. L'algorithme de reconnaissance faciale de Facebook qui identifie les personnes sur les photos est un exemple d'apprentissage machine sans supervision.

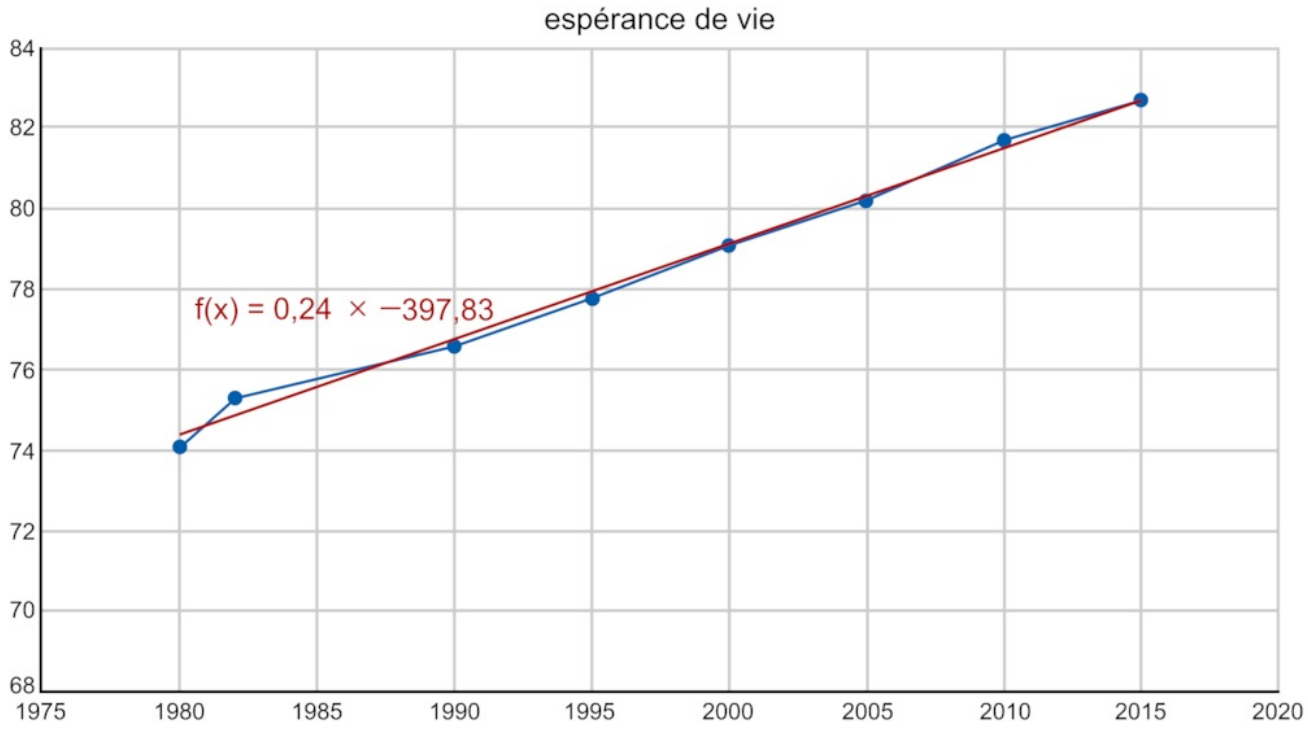
• l'apprentissage par renforcement : un programme informatique interagit avec un environnement dynamique dans lequel il doit atteindre un but. Le programme-apprenti reçoit des récompenses ou des punitions pendant qu'il navigue dans l'espace du problème et qu'il apprend à identifier le comportement le plus efficace. Le programmeur fixe les règles qui déterminent si l'intelligence artificielle sera punie ou récompensée : à chaque décision prise par le système, les règles préétablies permettent de modifier les décisions suivantes selon qu'il s'agit d'une bonne décision (récompense) ou d'une mauvaise décision (punition). Le système évolue ainsi en maximisant les performances, donc les bonnes décisions. C'est ce type d'apprentissage qui a permis au programme Alpha Zero de Google de battre le champion de jeu de Go, Lee Sedol, en 2016.



La courbe d'ajustement

La méthode de la courbe d'ajustement repose sur un traitement statistique des données : lors de la phase d'apprentissage, on fournit un nuage de points à la machine, et celle-ci sort une courbe de type droite, parabole ou exponentielle.
Le choix du type de courbe est déterminé avec les données, lors de la phase d'entraînement. La connaissance de cette courbe permet alors de prédire les valeurs : c'est la phase d'inférence.
Exemple : Espérance de vie à la naissance en France

Année de naissance	1980	1985	1990	1995	2000	2005	2010	2015
Espérance de vie (en années)	74,1	75,3	76,6	77,8	79,1	80,2	81,7	82,7



La forme allongée du nuage de points permet un ajustement par une droite : on parle alors de régression linéaire. La détermination de cette droite constitue la phase d'apprentissage, et peut être réalisée par une méthode mathématique (la méthode des moindres carrés). L'algorithme d'apprentissage consiste à déterminer la droite qui passe au plus près des valeurs des points mesurés, minimisant ainsi l'erreur d'approximation : c'est l'optimisation.
Avec l'équation de la droite, on peut ensuite prévoir l'espérance de vie pour d'autres années. Par exemple, pour l'année 2040, on peut estimer une espérance de vie de $y = 0{,}2385 \times 2040 - 397{,}83 = 88{,}7$ ans.

L'inférence bayésienne

L'inférence bayésienne est une méthode de calcul de probabilités de causes à partir des probabilités de leurs effets. Elle est utilisée en apprentissage automatique pour modéliser des relations au sein de systèmes complexes, notamment en vue de prononcer un diagnostic (médical, industriel, détection de spam, etc.). Cela permet de détecter une anomalie a` partir d'un test imparfait.
Le filtre de détection des spams est un exemple d'inférence bayésienne. Il ne connaît pas à l'avance les probabilités d'apparition de ces e-mails indésirables, c'est pourquoi il lui faut une phase d'apprentissage pour les évaluer. Cet apprentissage est réalisé à partir de l'observation du comportement des utilisateurs, qui doivent indiquer manuellement si un message est un spam ou non. Pour chaque mot de chaque message « appris », le filtre ajuste les probabilités de rencontrer ce mot dans un spam ou dans un courrier légitime et le stocke dans sa base de données.

L'importance du choix des données

Les algorithmes s'entraînent et calibrent leurs algorithmes à partir d'un grand nombre de données. La qualité et la représentativité de ces données d'entraînement sont essentielles pour obtenir des résultats fiables : en effet, les biais dans les données se retrouvent souvent amplifiés dans les résultats.
D'autre part, une grande quantité de données de qualité ne suffit pas pour assurer le bon fonctionnement de l'algorithme. Il faut aussi veiller à ce que les données soient représentatives.

La programmation

Un programme peut comporter plusieurs centaines de millions de lignes de code, ce qui rend très probable la présence d'erreurs appelées bogues (ou bugs). Ces erreurs peuvent conduire un programme à se comporter de manière inattendue et avoir de graves conséquences.
On considère la fonction informatique dont le code est le suivant :

```
def f(x):
    if x<=-1:
        return x*x-1
    elif -1<=x and x<=1:
        return 1-x*x
    else:
        return x*x-1
```

Pour tester le programme, on peut proposer le jeu de données suivant :

• Pour $x = -2$, la première condition est vérifiée, la fonction renvoie la valeur 3.

• Pour $x = 0$, la première condition n'étant pas vérifiée mais la deuxième si, la fonction renvoie la valeur 1.

• Pour $x = 3$, les deux premières conditions n'étant pas vérifiées, la fonction renvoie la valeur 8.

On peut modifier le programme pour qu'il ne teste qu'une seule condition :

```
def f(x):
    if -1<=x and x<=1:
        return 1-x*x
    else:
        return x*x-1
```

Zoom sur...

L'éthique et l'intelligence artificielle

Les intelligences artificielles (IA) se font de plus en plus présentes dans les voitures autonomes, les algorithmes de recommandation, les drones, etc. En effet, en s'appuyant sur de très grandes quantités de données, les algorithmes sont capables d'effectuer un nombre important de tâches : reconnaître et analyser des voix ou des images, accepter ou refuser des offres bancaires de prêt, conduire un véhicule, etc. Les questions éthiques posées par l'intelligence artificielle sont nombreuses et continueront à croître avec le développement de nouveaux algorithmes : peut-elle être utilisée à des fins critiquables comme la surveillance civile, le marketing ciblé ? En cas de défaillance de ma voiture autonome, celle-ci devra-t-elle choisir entre tuer deux enfants ou bien trois personnes âgées, si ceux-ci venaient à croiser simultanément sa trajectoire ? L'IA doit-elle uniformiser ou non les contenus mis en avant sur les réseaux sociaux ?

L'intelligence artificielle qui s'appuie sur les réseaux de neurones artificiels (*deep learning*) est particulièrement critiquée : on lui reproche d'être opaque, de ne pas laisser voir le raisonnement qui permet aux algorithmes d'arriver au résultat final. Il est alors difficile d'avoir confiance en des décisions qui ne sont pas mesurables.

Des limitations existent déjà : au niveau européen, le règlement général sur la protection des données (RGPD) tente de restreindre la collecte de données personnelles par des entités juridiques, pourtant nécessaires lors de l'entraînement de programmes d'IA. Un comité chargé d'« aborder de manière globale les enjeux éthiques du numérique et de l'intelligence artificielle » a vu le jour fin 2019.

En France, c'est la CNIL qui s'occupe de la protection des données, mais l'une des difficultés est que, en grande majorité, les sites sont hébergés hors de France...