éduSCOL
Informer et accompagner les professionnels de l'éducation

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DE L'ÉDUCATION NATIONALE ET DE LA JEUNESSE

VOIE GÉNÉRALE

2DE | 1RE | TLE

Enseignement scientifique

ENSEIGNEMENT COMMUN

THÈME 3
SOUS-THÈME 3-5 : INTELLIGENCE ARTICIELLE

Mots-clés

Apprentissage automatique ; courbes de tendance ; big data

Références au programme

L'intelligence artificielle (IA) permet l'accomplissement de tâches et la résolution de problèmes jusqu'ici réservés aux humains : reconnaître et localiser les objets dans une image, conduire une voiture, traduire un texte, dialoguer... Un champ de l'intelligence artificielle ayant permis des applications spectaculaires est celui de l'apprentissage machine.

Savoirs

L'apprentissage machine (ou « apprentissage automatique ») utilise des programmes capables de s'entraîner à partir de données. Il exploite des méthodes mathématiques qui, à partir du repérage de tendances (corrélations, similarités) sur de très grandes quantités de données (big data), permet de faire des prédictions ou de prendre des décisions sur d'autres données.

La qualité et la représentativité des données d'entraînement sont essentielles pour la qualité des résultats. Les biais dans les données peuvent se retrouver amplifiés dans les résultats.

Savoir-faire

Utiliser une courbe de tendance (encore appelée courbe de régression) pour estimer une valeur inconnue à partir de données d'entraînement.

Analyser un exemple d'utilisation de l'intelligence artificielle : identifier la source des données utilisées et les corrélations exploitées.

Sur des exemples réels, reconnaître les possibles biais dans les données, les limites de la représentativité.

Expliquer pourquoi certains usages de l'IA peuvent poser des problèmes éthiques.

Histoire, enjeux, débats

La question de faire appel à une machine pour imiter l'Homme est ancienne. On peut, par exemple, citer les automates de Vaucanson au milieu du XVIIIe. C'est au XXe siècle que des travaux plus sérieux sont entrepris. En 1950, Alan Turing publie un article fondateur (*Computing Machinery and Intelligence*) où il pose la question de savoir si les machines peuvent imiter l'homme (*imitation game*).

L'Intelligence Artificielle, en tant que domaine de recherche, est née officiellement en 1956 lors de la conférence de Dartmouth, une école d'été sur le thème des machines pensantes organisée par quatre chercheurs américains John McCarthy, Marvin Minsky, Claude Shannon et Nathaniel Rochester. John McCarthy y propose l'expression « Intelligence Artificielle » pour désigner cette discipline émergente.

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

Si l'adjectif « artificielle » est à peu près compris par tout le monde sans équivoque, la signification du nom « intelligence » est loin d'être bien cernée : est-ce la capacité d'apprendre, d'utiliser ses connaissances, d'interagir avec son environnement, de planifier une succession de tâches… ?

Ainsi, le domaine de l'IA possède une délimitation variable en fonction des interlocuteurs, la frontière évoluant également au cours du temps.

On peut en reprendre la définition donnée par Yann Le Cun (lauréat du prix Turing 2018 avec Yoshua Bengio et Geoffrey Hinton pour leurs travaux sur l'apprentissage profond) : « On pourrait dire que 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 ».

L'IA vise donc à reproduire à l'aide de machines des activités humaines, qu'elles soient de l'ordre de la compréhension, de la perception ou de la décision ; des activités comme localiser un objet dans une image, jouer aux échecs, conduire une voiture, traduire des textes, conduire un dialogue, établir un diagnostic…

Elle désigne plus un projet loin d'être achevé, celui d'imiter l'intelligence biologique, qu'un ensemble de techniques bien délimitées. Néanmoins, depuis quelques années, l'IA est souvent confondue avec l'apprentissage machine, qui n'en est qu'un sous-domaine, mais qui connaît aujourd'hui un grand essor notamment en raison de la capacité qu'il confère aux machines de traiter massivement des données (*big data*).

L'utilisation du terme « intelligence artificielle » pour désigner ces techniques s'est répandu mais n'est pas neutre puisqu'il peut faire croire, à tort, que les machines accèdent aujourd'hui à une compréhension du monde. On pourra en effet noter que le mot *intelligence* vient du latin *intellegere* (« discerner, saisir, comprendre »), lui-même composé du préfixe inter- (« entre ») et du verbe *lĕgĕre* (« cueillir, choisir, lire »). Étymologiquement, l'intelligence consiste à faire un choix, une sélection.

Qu'est-ce que l'apprentissage machine ?

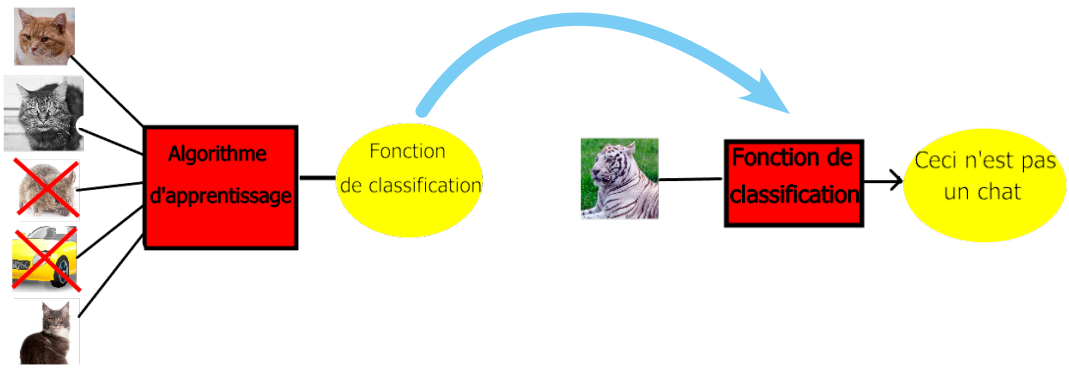
C'est une branche commune de l'IA et des statistiques. L'apprentissage machine, encore connu sous le nom d'apprentissage artificiel ou d'apprentissage automatique, consiste en des programmes capables d'ajuster leur comportement à des données dites d'entraînement. On fournit à une machine un très grand nombre de données. À partir de ces données, la machine s'entraîne (phase d'apprentissage ou d'entraînement) pour définir son comportement ultérieur (phase d'inférence).

Par exemple, si on voulait apprendre à une machine à reconnaître des images de chats sans avoir recours à de l'apprentissage machine (comme on l'aurait fait il y a 15 ans), on devrait écrire des procédures complexes pour détecter la présence des parties du corps caractéristiques d'un chat (tête, moustache, queue, etc.) et comment elles se situent les unes par rapport aux autres : c'est une tâche très difficile. Avec l'apprentissage machine, on fournit à l'algorithme d'apprentissage automatique un grand nombre d'images dont certaines contiennent des chats et d'autres n'en contiennent pas. L'algorithme détermine alors les paramètres qui permettront de distinguer les images de chats des autres : c'est la phase d'entraînement. Une fois la machine entraînée, on lui demande de décider si une nouvelle image proposée contient ou non un chat : c'est la partie d'inférence.On peut ici souligner une différence notable avec l'apprentissage humain : s'il faut à un jeune enfant moins d'une dizaine d'images pour apprendre à identifier un chat, il en faut plusieurs milliers à une machine.

On distingue communément trois modalités principales d'apprentissage : l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement.

L'apprentissage supervisé consiste à apprendre à une machine à catégoriser des données à partir d'un grand nombre de données préalablement étiquetées par l'Homme. Pendant la phase d'apprentissage, on calibre un algorithme en adaptant ses paramètres de catégorisation aux données fournies. On utilise ensuite cet algorithme pour catégoriser de nouvelles données.

En reprenant l'exemple de la reconnaissance d'un chat dans une image, on définit deux catégories « contient un chat » et « ne contient pas de chat ». Au cours de la phase d'apprentissage, le programme adapte au mieux ses paramètres pour identifier la présence d'un chat. Au cours de la phase d'inférence, le programme détermine si l'image qui lui est fournie en entrée contient ou non un chat.



Phase d'apprentissage Phase d'inférence

La fonction à apprendre n'est pas toujours une catégorisation, mais peut être l'estimation d'un nombre. Par exemple un programme d'apprentissage automatique pour faire des prévisions météorologiques à partir de données de capteurs pourra faire des prédictions de température, de pression, etc. On parle alors non plus de classification mais de régression.

Parmi les méthodes d'apprentissage supervisé, **l'apprentissage profond** (*deep learning*) est très utilisé depuis les années 2010. De manière simpliste, on peut définir l'apprentissage profond par l'enchaînement de plusieurs modules pour lesquels les sorties des uns fournissent les entrées des autres. On parle de réseaux de neurones artificiels, car ces modules s'inspirent directement des découvertes sur les neurones dans le cerveau. Le succès récent de l'apprentissage profond repose sur la très grande puissance de calcul des ordinateurs, qui leur permet de traiter un nombre considérable de données.

L'apprentissage non supervisé consiste à fournir à la machine un grand nombre de nombreuses données, non étiquetées. La machine y repère des régularités, des proximités, des corrélations pour construire elle-même son algorithme de classification.

L'apprentissage par renforcement consiste à induire le comportement d'un « agent » (par exemple un robot) évoluant dans un « environnement » (par exemple dans une pièce où il doit manipuler des objets) et qui apprend grâce à un système de « récompenses ». 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. Si on voulait entraîner un programme à jouer au jeu de go avec de l'apprentissage supervisé il faudrait lui donner comme données d'entraînement un grand nombre de parties de maîtres et il apprendrait à en reproduire les coups. En apprentissage par renforcement, le programme joue des parties contre lui-même et n'a pour unique information pour apprendre que les récompenses finales des victoires (ou les « punitions » des défaites).

L'apprentissage supervisé reste le cas de figure le plus courant, et par ailleurs les méthodes d'apprentissage non supervisé et par renforcement reposent souvent elles-mêmes sur des méthodes d'apprentissage supervisé.

La plupart des méthodes d'apprentissage machine s'appuient sur des outils mathématiques.

Exemples de méthodes d'apprentissage

Différentes méthodes sont ici présentées dans des contextes simples (nombre limité de données). L'objectif pédagogique est de faire comprendre les méthodes mathématiques utilisées dans un cadre beaucoup plus complexe (notamment quant au nombre de données utilisées) en intelligence artificielle.

Exemple 1 : les courbes d'ajustement

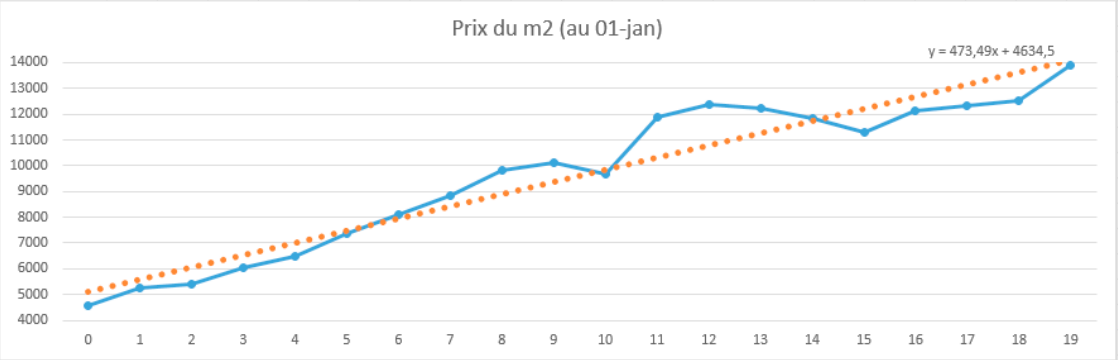
Cette méthode repose sur un traitement statistique des données : celles-ci, fournies lors de la phase d'apprentissage, peuvent être représentées par un nuage de points. Des méthodes statistiques permettent d'ajuster ce nuage par une courbe (droite, parabole, exponentielle).

Les paramètres de cette courbe sont déterminés à l'aide des données : c'est la phase d'entraînement. La connaissance de cette courbe permet alors de prédire des valeurs pour des points de mesure absents : c'est la phase d'inférence.

Exemple 1 : évolution du prix du mètre-carré dans le 6e arrondissement de Paris

Evolution du prix du mètre carré dans le 6ème arrondissement de Paris (année 0 en 2000)

Année (x_i)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Prix du m2 y_i (au 01-jan)	4562	5270	5400	6020	6460	7360	8090	8840	9830	10100	9690	11870	12400	12250	11820	11280	12150	12320	12530	13880



Les données antérieures à l'année 2020 constituent des données à partir desquelles on détermine l'équation d'une courbe approchant les données.

La forme du nuage de points laisse penser qu'il peut être ajusté par une droite.
On parle alors de régression linéaire.

La détermination de cette droite constitue à la phase d'apprentissage, et peut être réalisée par la méthode dite des moindres carrés qui donne directement son équation.

Il convient de préciser comment est définie cette droite d'ajustement : c'est celle qui passe par des valeurs aux points de mesures au plus près des vraies valeurs mesurées, dans un sens mathématiques bien précis, celui des moindre carrés. Si on note $y=ax+b$ l'équation de la droite, on veut minimiser la somme des carrés des différences entre les vraies valeurs y_i et leur prédictions par la droite x_i+b , soit trouver a et b qui minimisent $\sum_{i=1}^{n}(y_i-(ax_i+b))^2$. Il convient de remarquer que la plupart des algorithmes d'apprentissage consistent à déterminer les valeurs de paramètres (ici a et b) pour minimiser l'erreur effectuée par la fonction de classification ou de régression qui est en train d'être apprise : on parle d'**optimisation**.

Les valeurs de a et b peuvent être obtenues dans un tableur.

Avec l'équation de la droite obtenue, on peut ensuite inférer le prix du mètre-carré à des moments autres que ceux qui étaient initialement connus. Par exemple, pour l'année 2020, on peut estimer le prix au mètre-carré par $y \approx 473{,}49 \times 20 + 4634{,}5 \approx 14104€$.

On peut aussi estimer le prix d'une année précédente non disponible. Pour l'année 1999, on peut inférer un prix au mètre carré de 4161 €.

$$y \approx 473{,}49 \times (-1) + 4634{,}5 \approx 4161$$

Selon l'allure du nuage de points correspondant aux données, on peut envisager d'autres courbes d'ajustement (ajustement polynomial, exponentiel, logarithmique, etc.). Voir les activités proposées ci-dessous.

L'appliquette GeoGebra « Régression linéaire-Méthode des moindres carrés » permet de visualiser l'ajustement affine au sens des moindres carrés de différents nuages de points.

Exemple 2 : l'inférence bayésienne

« Le signalement par les usagers de messages considérés comme indésirables permet aux messageries électroniques de constituer une base conséquente et constamment alimentée à partir de laquelle le système peut apprendre à déterminer les caractéristiques des spams qui vont ensuite lui permettre de proposer de lui-même quels messages filtrer. »

Source : Extrait du document CNIL : Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Page 18

Le principe de l'inférence bayésienne et son application au filtrage de courriers indésirables sont détaillés dans une ressource spécifique présente sur la page éduscol dédiée à l'enseignement scientifique.

Exemple 3 : l'algorithme des plus proches voisins

Modèle simplifié

On suppose que les données d'apprentissage sont classifiées en deux catégories (M et B). On veut placer dans l'une ou l'autre de ces deux catégories une donnée nouvelle. Pour cela, on calcule sa distance à chacune des données déjà étiquetées et on lui attribue la catégorie de son plus proche voisin.

Exemple d'utilisation dans le dépistage automatique d'une tumeur maligne

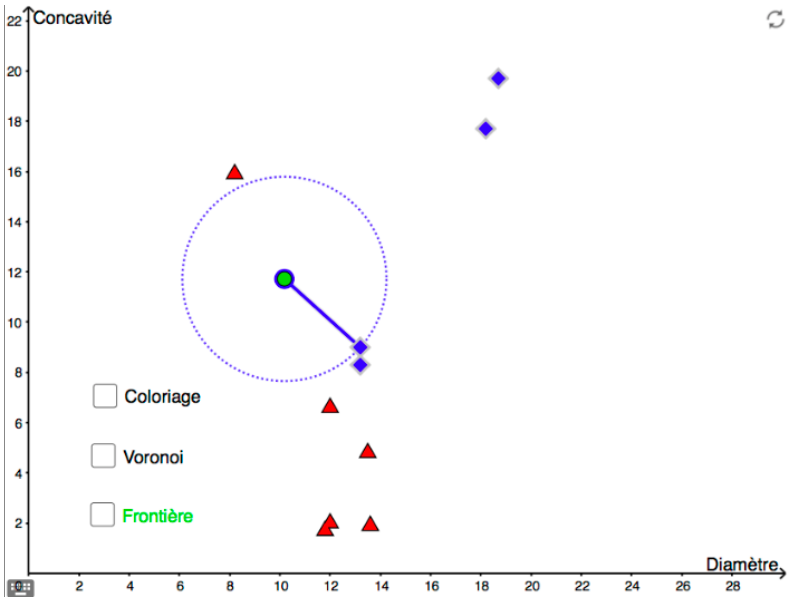
On se propose de répartir en deux catégories (Maligne, Bénigne) des tumeurs du sein dont on connaît un certain nombre de caractéristiques. Dans la réalité, un tel diagnostic automatique est effectué à partir de données d'apprentissage portant sur une cinquantaine de caractéristiques mesurées sur un échantillon d'un millier de tumeurs déjà étiquetées « Maligne » ou « Bénigne ». Dans un souci de simplification, on se restreint ici à 2 caractéristiques (diamètre et concavité) mesurées sur un panel de 10 patientes.

Données d'apprentissage :

Diamètre moyen (en mm)	Concavité moyenne	Catégorie
13,2	8,3	Bénigne
18,7	19,7	Bénigne
8,2	15,9	Maligne
13,2	9	Bénigne
13,5	4,8	Maligne
11,8	1,7	Maligne
13,6	1,9	Maligne
12	2	Maligne
18,2	17,7	Bénigne
12	6,6	Maligne

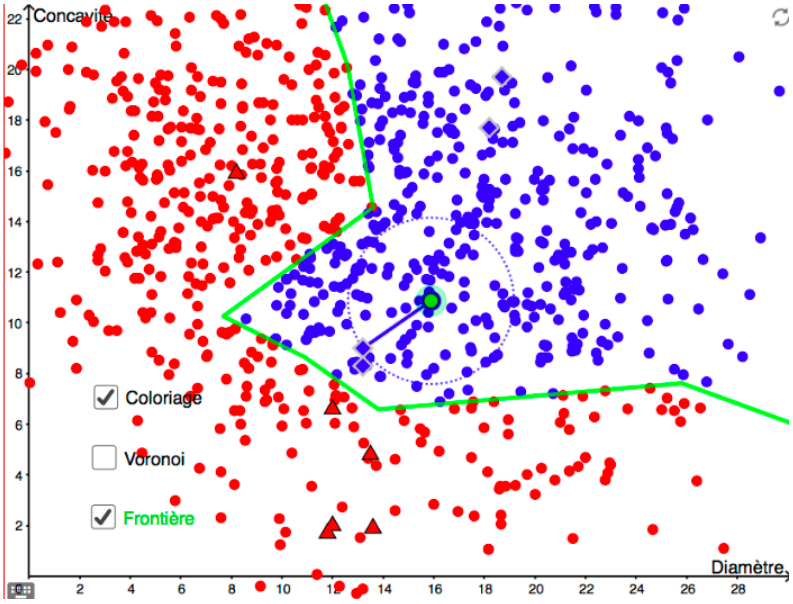
On considère une nouvelle tumeur de diamètre et de concavité connues et on cherche à savoir si elle est maligne ou bénigne.

Méthode du plus proche voisin
On cherche parmi les données (triangles rouges pour les tumeurs malignes et losanges bleus pour les tumeurs bénignes) celle qui est la plus proche de la nouvelle donnée à catégoriser (point vert). On lui attribue la catégorie de son plus proche voisin.



Par exemple, on considère une nouvelle tumeur ayant un diamètre de 10 mm et une concavité de 12. L'algorithme la range dans le groupe du point le plus proche. Ainsi, on estime ici que la tumeur est bénigne. Cette méthode est appelée méthode du plus proche voisin.

L'appliquette GeoGebra « Algorithme des plus proches voisins » permet de déplacer le point vert pour observer différents cas de figure. Le bouton « Coloriage » permet de garder la trace des données ainsi catégorisées et de délimiter le plan entre deux régions « Maligne » et « Bénigne ». Le bouton « Frontière » trace la ligne de partage.



Cette méthode peut être généralisée à k plus proches voisins. Il suffit de considérer les k voisins les plus proches, d'identifier la catégorie à laquelle appartient la majorité de ces voisins et d'affecter au nouveau point cette catégorie. L'exemple avec k = 5 est également proposé algorithme des 5 plus proches voisins.

On remarquera que pour la méthode des plus proches voisins il n'y a pas vraiment de phase d'apprentissage : comme l'inférence se fait directement à partir des données d'entraînement, la phase d'apprentissage ne fait pas plus que garder en mémoire l'ensemble des données d'entraînement, elle n'estime pas de paramètres qui en font la synthèse (comme tout à l'heure les paramètres a et b de la droite de régression).

L'importance du choix des données

C'est à partir des données que les algorithmes s'entrainent et calibrent leurs algorithmes. Afin d'éviter les biais induits par ces données, on doit être attentif à trois aspects :

- la qualité des données : celles-ci doivent être exactes et non périmées ; la corruption de données peut être due à un problème technique (dysfonctionnement d'un capteur de mesure) comme à une erreur, une négligence, voire une malveillance humaine ;
- la quantité de données : des données en grand nombre (plusieurs milliers) sont nécessaires au bon fonctionnement de la phase d'apprentissage ;
- la représentativité des données : elles doivent être les plus exhaustives possibles.

Reprenons l'exemple de la reconnaissance d'images.

Chacun comprend aisément que si les images sont mal étiquetées, cela entrainera des erreurs dans la classification. Trop peu de données dans la phase d'entrainement ne permettra pas à la machine d'adapter correctement ses paramètres.

Cependant, une grande quantité de données de qualité ne suffit pas pour assurer le bon fonctionnement de l'algorithme. N'utiliser que des images contenant des chats roux lors de la phase d'apprentissage conduira à la machine à intégrer dans ses paramètres de la catégorie « contient un chat » la couleur rousse et une image contenant un chat noir ne sera pas classée dans la bonne catégorie.

Il est erroné et dangereux d'attribuer à l'IA un caractère objectif au motif que les résultats obtenus sont attribués à une machine et non à un être humain dont on peut toujours suspecter la subjectivité.

Le respect de la quantité et de la représentativité des données est par ailleurs essentiel pour une bonne utilisation de l'IA dans la prise de décision.

Des résultats discriminatoires fournis par des algorithmes d'IA peuvent s'expliquer en partie par des données biaisées. Par exemple, un algorithme de classement de CV ayant appris sur la base des recrutements antérieurs d'une société risque, en l'automatisant, de renforcer le sous-recrutement des femmes à des postes à responsabilité.

De la même manière, la sous-représentation d'une population ethnique dans un ensemble de données peut s'avérer problématique. Tel a été le cas dans des bases de données de reconnaissance faciale, composées majoritairement de personnes de couleur blanche. L'algorithme reconnaissait aisément les membres de cette catégorie, alors qu'elle se trompait beaucoup plus fréquemment dans la reconnaissance de personnes de couleur, insuffisamment représentées dans les bases de données.

Dans quels domaines et avec quelles précautions peut-on utiliser l'intelligence artificielle ?

L'IA occupe une place de plus en plus importante dans de nombreux domaines par l'assistance qu'elle peut apporter à la prise de décision : santé (médecine personnalisée, tests de dépistage et de diagnostic), commerce (recommandation de produits), activités sociales (accompagnement des personnes, proposition de nouveaux amis sur les réseaux sociaux), communications (traduction automatique de la langue, assistant vocal, détection de spams), transports (voitures autonomes), finance (décisions d'investissements), justice, sécurité, détection de fraudes éducation (personnalisation des contenus), industrie (détection de défauts, robotique), et même création artistique.

Le document de la CNIL « Comment permettre à l'Homme de garder la main ? » mentionné dans la bibliographie développe en page 22 l'utilisation de l'IA dans chacun de ces secteurs.

Les questions éthiques posées par l'utilisation de l'IA sont nombreuses et continueront à croître parallèlement au développement de nouveaux algorithmes. Au-delà du choix des données déjà développé, se posent des questions relatives à l'usage de l'IA : Comme toute technologie cette dernière peut être utilisée à des fins criticables : surveillance civile (scores attribués aux citoyens chinois), marketing ciblé (recherche de personnes influençables sur les réseaux sociaux). En rendant possible la réalisation par la machine de tâches jusqu'alors réservées à l'homme elle peut être source de nouvelles inégalités sociales sur le marché de l'emploi, et dérange lorsqu'elle touche des domaines à haute valeur de responsabilité (armes autonomes, diagnostic médical). Elle peut aussi avoir des effets pervers inattendus (uniformisation voire détérioration des contenus mis en avant sur les réseaux sociaux, etc).

Ces usages font aussi émerger de nombreuses questions liées au droit. Le règlement général sur la protection des données (RGPD, article 14.2.g) tente de limiter la collection de données personnelles par des entités juridiques (entreprises, collectivités) lors de l'entraînement de programmes d'IA. En outre, il précise que le résultat fourni par un algorithme doit être explicable. Or si les algorithmes d'apprentissage sont très performants, leur fonctionnement notamment, dans le cas de l'apprentissage profond, reste actuellement très opaque.

Propositions d'activités

Exemples de recherche et d'utilisation, à la calculatrice ou au tableur, d'une courbe de tendance

Activité 1 : ajustement affine

Le tableau ci-dessous contient des données sur le diamètre moyen (en cm) d'un arbre (le pin Douglas) en fonction de son âge (en année) :

Âge	20	25	30	35	42	50
Diamètre moyen	14,3	17,8	21,0	24,2	29,6	35,3

Représenter le nuage de points. Déterminer (par exemple au tableur) une courbe d'ajustement et l'utiliser pour d'estimer le diamètre moyen d'un pin Douglas de 40 ans et de 60 ans.

Activité 2 : un exemple d'ajustement non linéaire

Le glacier d'Aletsch, classé à l'UNESCO, est le plus grand glacier des Alpes ; situé dans le sud de la Suisse, il alimente la vallée du Rhône.

Pour étudier le recul de ce glacier au fil des années, une première mesure a été effectuée en 1900 : ce glacier mesurait alors 25,6 km.

Des relevés ont ensuite été effectués tous les 20 ans : le recul du glacier est mesuré par rapport à la position où se trouvait initialement le pied du glacier en 1900.
Les mesures successives ont été relevées dans le tableau le tableau ci-dessous. On note t la durée, en années, écoulée depuis 1900, et r le recul correspondant, mesuré en kilomètres.

Année de mesure	1900	1920	1940	1960	1980	2000
Durée écoulée depuis 1900	0	20	40	60	80	100
Recul (en km)	0	0,3	0,6	1	1,6	2,3

1. Représenter le nuage de points.
2. Chercher un ajustement linéaire de ce nuage.
3. Selon ce modèle, quel serait le recul du glacier en 2020 ?
   Selon ce modèle, en quelle année le glacier Aletsch aura-t-il disparu ?
4. Répondre aux mêmes questions en utilisant un ajustement exponentiel.

Textes pouvant servir de supports à une analyse documentaire

Diagnostic médical automatisé

« Dans le domaine médical où la qualité de la prise de décision peut être plus facilement évaluée (ou, du moins, quantifiée), on peut logiquement se demander quelle marge d'autonomie resterait au médecin face à la recommandation (en termes de diagnostic et de solution thérapeutique à privilégier) qui serait fournie par un système d'« aide » à la décision extrêmement performant. On annonce en effet que l'intelligence artificielle serait supérieure à l'homme pour le diagnostic de certains cancers ou pour l'analyse de radiographies. Dans le cas où ces annonces s'avéreraient exactes, il pourrait donc devenir hasardeux pour un médecin d'établir un diagnostic ou de faire un choix thérapeutique autre que celui recommandé par la machine, laquelle deviendrait dès lors le décideur effectif. Dans ce cas, se pose alors la question de la responsabilité. Celle-ci doit-elle être reportée sur la machine elle-même, qu'il s'agirait alors de doter d'une personnalité juridique ? Sur ses concepteurs ? Doit-elle être encore assumée par le médecin ? Mais alors, si cela peut certes sembler résoudre le problème juridique, cela n'aboutit-il quand même pas à une déresponsabilisation de fait, au développement d'un sentiment d'irresponsabilité ? »

Source : Extrait du document CNIL : Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Page 27

Le dilemme du tramway en lien avec le développement des véhicules autonomes

« Le fameux dilemme du tramway est très souvent évoqué à l'occasion de réflexions portant sur ce problème. On sait que ce dilemme met en scène un tramway sans freins dévalant une pente ; le tramway arrive devant un embranchement ; selon qu'il s'engage sur l'une ou l'autre des deux voies, il tuera une seule personne ou bien plusieurs. Dès lors, quelle devrait être la conduite d'une personne ayant la possibilité de manœuvrer l'aiguillage et donc de choisir, pour ainsi dire, l'un des deux scénarios possibles ? L'intérêt de cette expérience de pensée est qu'elle peut donner lieu à toute une gamme de variations : qu'en est-il si la personne seule attachée à l'une des deux voies se trouve être un proche parent ? Si les personnes sur l'autre voie se trouvent être 5 ou bien 100 ? ».

Source : Extrait du document CNIL - Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Page 41

On voit aisément comment ce dilemme peut être adapté à l'utilisation de voitures autonomes. Selon quels principes une voiture placée dans une situation de dilemme éthique de ce type devrait-elle « choisir » de se comporter ? »

Pour éclairer la réflexion sur la voiture autonome, on peut consulter :
https://lejournal.cnrs.fr/billets/voitures-autonomes-par-quoi-serez-vous-choques

Bibliographie et sitographie

- Article de Yann Le Cun Les enjeux de la recherche en intelligence artificielle
- Article de Nicolas Rougier L'intelligence artificielle, mythes et réalités
- Turing, A.M. (1950). Computing machinery and intelligence. Mind, 59, 433-460
- Document de la CNIL Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle décembre 2017
- Conférence inaugurale de Stéphane Mallat au Collège de France le 11 janvier 2018
- Tangente Hors-série N°68 : intelligence artificielle – l'alliance des mathématiques et de la technologie, septembre 2019.