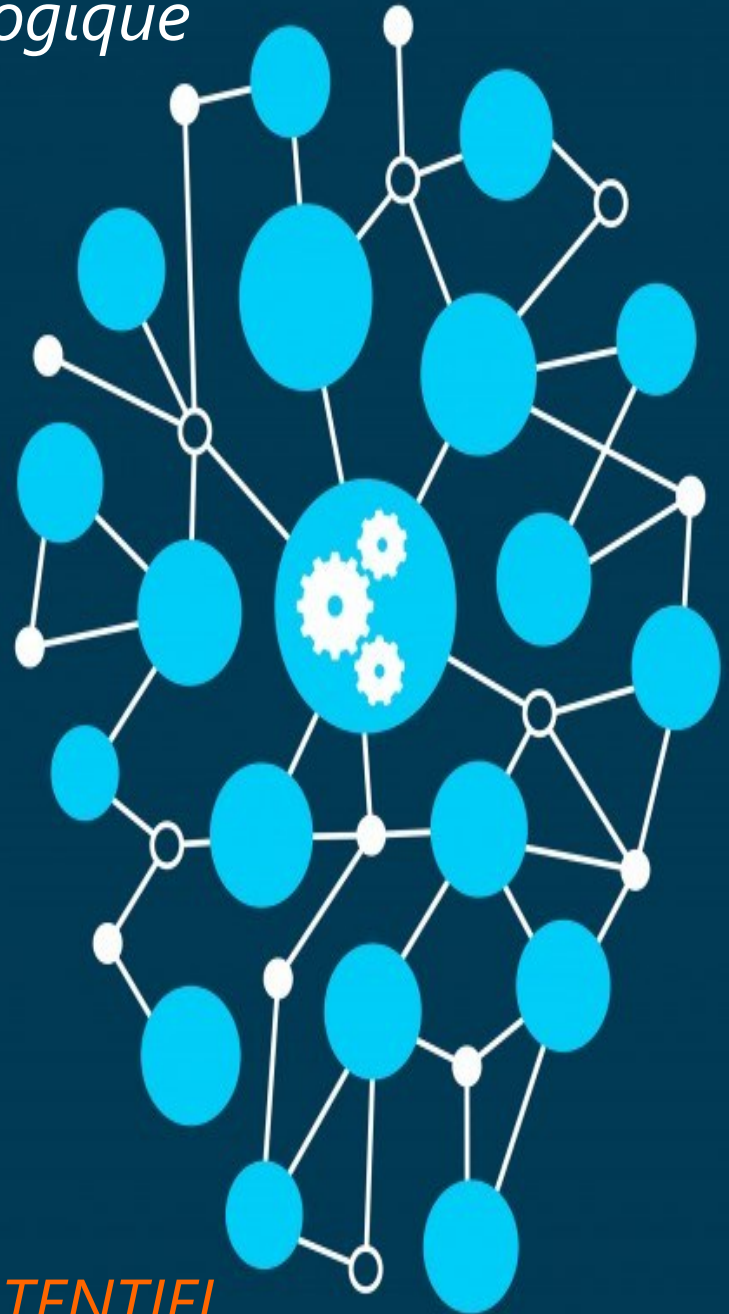


BULLETIN

Avril 2019

de *Veille Technologique*

MACHINE LEARNING



*COMPRENDRE LE POTENTIEL
ET LES DÉFIS À RELEVER*



Editorial



Depuis l'antiquité, l'humanité éprouve le besoin d'automatiser certaines tâches récurrentes et non moins fastidieuses du quotidien. Cette quête, accompagnée d'une obsession sans faille de créer une certaine forme d'intelligence, s'est traduite par l'invention au premier siècle du distributeur de vin de Héron d'Alexandrie jusqu'aux ordinateurs actuels les plus évolués. Dans l'intervalle, il y a eu les calculatrices de Wilhem Schickard en Allemagne et de Blaise Pascal en France sans oublier les emblématiques inventions de Léonard de Vinci qui fascinent encore aujourd'hui.

Plus proche de notre époque, en 1950, Alan Turing va poser les bases de ce qui deviendra l'Intelligence Artificielle et l'une de ses sous-branches, le Machine Learning (ou Apprentissage automatique en français), dans son article « Computing Machinery and Intelligence » (Mind, octobre 1950) dans lequel il décrit le « jeu de l'imitation (*imitation game en anglais*) » plus connu sous le nom de « test de Turing ».

Fondamentalement, le Machine Learning est une technologie qui, grâce aux données, permet aux ordinateurs d'apprendre sans avoir été programmés explicitement à cet effet. C'est donc une technologie qui permet d'exploiter pleinement le potentiel du Big Data où la volumétrie des données – *chaque jour, nous générons 2,5 trillions de données*¹ – rend obsolète les méthodes analytiques traditionnelles.

On assiste alors à des avancées improbables dans des domaines divers et variés tels que les jeux, la santé, l'automobile pour la conduite autonome, la finance pour l'analyse notamment des marchés boursiers, les Télécommunications/TIC pour optimiser le fonctionnement des systèmes et réseaux. On parlera plus globalement d'analyse prédictive dont la finalité serait d'améliorer les processus décisionnels.

C'est l'objectif que IBM chercherait à atteindre avec « Project Debater », un programme d'Intelligence Artificielle capable, non seulement de débattre avec les humains, mais aussi et surtout vu par ses concepteurs comme un outil d'aide à la prise de décision fondée sur des arguments impartiaux. De l'exécutif au simple citoyen en passant par les législateurs, les juges et les décideurs, tout le monde pourra dans l'avenir recourir à de tels outils pour gagner en efficacité dans leur prise de décision.

Cependant, même si ces outils d'aide à la décision sont déjà à l'œuvre dans nos usages quotidiens (reconnaissance vocale, reconnaissance faciale, détection de SPAM, recommandation de produits, etc.), sommes-nous préparés à un avenir où les décisions les plus importantes seraient prises sur la base de l'avis d'une machine ? Parce que, aussi intelligente soit-elle, une machine sera toujours limitée dans son apprentissage par la quantité et la qualité des données auxquelles elle aura accès.

Ainsi, en attendant la généralisation de tels procédés, il est du devoir de l'humanité de s'assurer que ces outils ne concourent pas à sa déliquescence. Il nous faut pour ce faire prendre les dispositions idoines afin de lever tous les obstacles d'ordre moral, éthique, technique voire réglementaire, administratif et organisationnel qui pourraient entraver le plein essor du Machine Learning.

J'appelle donc à des réflexions et à la création de cadres d'échanges sur la question auxquelles nous devrions tous prendre part avec le plus grand intérêt.

BILE Diéméléou

Directeur Général de l'ARTCI

¹<https://www.lebigdata.fr/definition-big-data>

Sommaire

Directeur de Publication:

M. BILE Diéméléou

Rédacteur en Chef:

M. KOUAKOU Guy-Michel

Equipe de rédaction:

M. ZEBOUA Patrick

M. YAO N'Guessan Kevin

Mlle LASME Mel Paule Renée

Editorial	2
Introduction	4
De l'intelligence artificielle au machine learning	5
Les domaines d'application du machine learning	8
Le machine learning, des défis à considérer	12
Le machine learning à l'épreuve du droit et de l'éthique	13
Le machine learning, quid de la normalisation ?	14
Le machine learning, catalyseur de développement pour l'Afrique?	15
Ce qu'il faut retenir	16
Webographie	18

Contacts

Marcory Anoumanbo, 18 BP
2203 Abidjan 18.

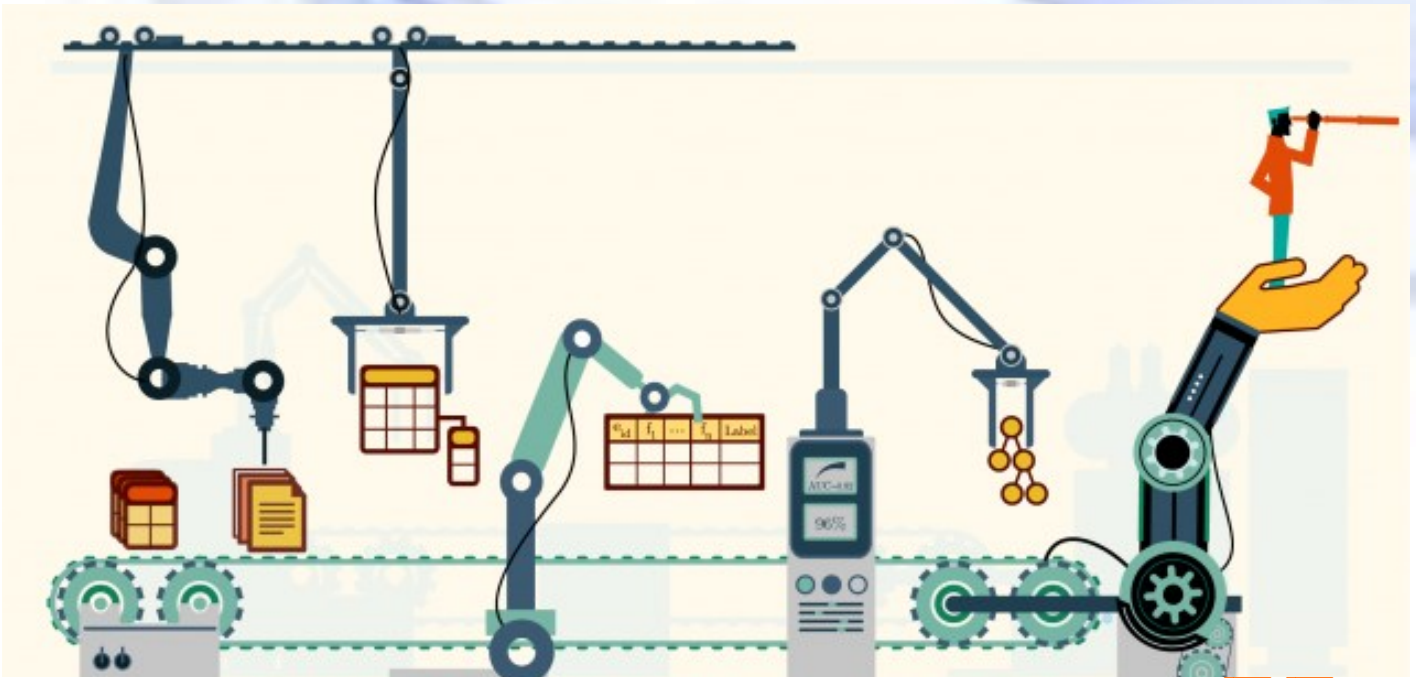
Tél : + 225 20 34 58 80

Fax : + 225 20 34 43 75

...Au lecteur

*Parce que votre avis compte,
nous serions heureux de
recevoir vos suggestions et
remarques, afin d'améliorer nos
prochaines publications, à :*

veille techno@artci.ci



INTRODUCTION



*S'*il est un domaine en plein développement actuellement, c'est bien celui de l'intelligence artificielle (IA).

De la reconnaissance faciale, aux assistants conversationnels, en passant par les véhicules autonomes et les systèmes de recommandations d'achats sur Internet, ces nouvelles applications envahissent notre quotidien. Ces applications sont rendues possibles par de nouvelles technologies ou approches dont la plus en vue ou importante est le machine Learning.

Le machine learning aborde la question de savoir comment construire des ordinateurs qui s'améliorent automatiquement grâce à l'expérience. C'est l'un des domaines techniques les plus en expansion, à la croisée de l'informatique et des statistiques, et au cœur de l'intelligence artificielle et de la science des données.

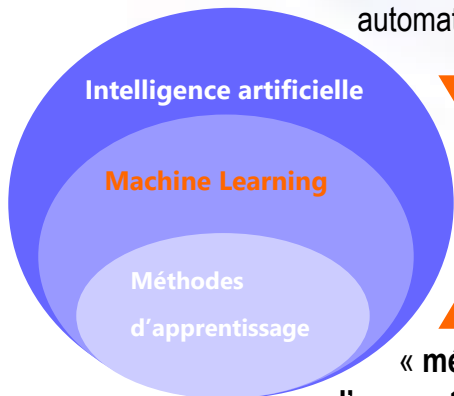
Les progrès récents en matière de machine learning ont été favorisés à la fois par le développement de nouveaux algorithmes d'apprentissage ainsi que par l'explosion actuelle de la disponibilité des données en ligne.

A l'heure où les premières applications du machine learning se développent sur le marché, il est important de pouvoir poser sur ces technologies un regard éclairé : comprendre ce dont il s'agit, quels en sont les possibilités et les défis...

Machine learning

DE L'INTELLIGENCE ARTIFICIELLE AU MACHINE LEARNING

Le **machine learning** ou « apprentissage automatique » en français constitue l'une des branches de l'Intelligence Artificielle (IA). Il se décline de plusieurs manières appelées « **méthodes d'apprentissage** ».



Resituons donc le machine learning par rapport au vaste domaine de l'Intelligence Artificielle et reprenons progressivement les différents termes y afférents.

1. Au départ : l'intelligence artificielle

Notion apparue dès les années 50, l'intelligence artificielle peut-être définie comme l'« ensemble des techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux » (Yann

LECUN). Il s'agit donc de techniques qui visent à reproduire, imiter, simuler l'intelligence, ou en tout cas les capacités que l'on peut associer à ce terme : capacité à percevoir son environnement, à interagir avec lui, à communiquer au moyen d'un langage, à mettre en œuvre un raisonnement et à résoudre des problèmes, à planifier des actions afin d'atteindre un but, à représenter la connaissance, à appliquer cette connaissance au moment opportun. C'est un domaine très vaste qui s'organise en plusieurs branches. (voir schéma ci-dessous)

Les évolutions en matière d'IA ont suivi différents courants. Dans les années 80, par exemple, les « systèmes experts » correspondaient à une approche basée sur la capacité à reproduire un raisonnement logique. Ces programmes permettaient de répondre à des questions ou problèmes dans un domaine très spécifique, où la machine sert à mettre en œuvre des règles logiques décrites par des experts humains et implémentées dans des bases de connaissance. Ce type d'approche est encore assez proche de l'informatique classique où il faut identifier à l'avance toutes les informations à fournir au système. Une autre approche consiste maintenant à s'inspirer du fonctionnement du cerveau, sa façon de modéliser et de s'adapter à l'environnement. Le machine learning relève de cette approche.

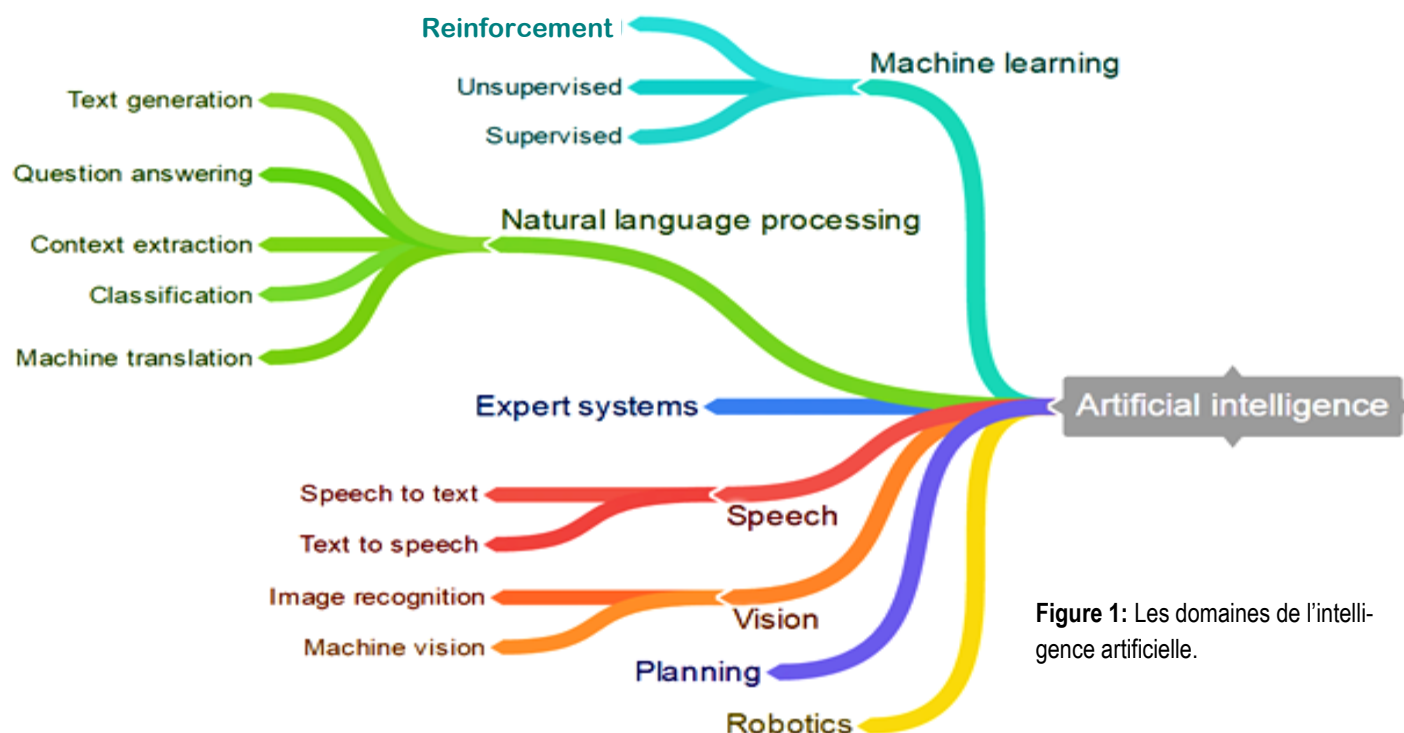


Figure 1: Les domaines de l'intelligence artificielle.

²Yann Le Cun est un chercheur français en intelligence artificielle et vision artificielle (robotique). Il est considéré comme l'un des inventeurs de l'apprentissage profond. Il reçoit le Prix Turing le 27 mars 2019. https://fr.wikipedia.org/wiki/Yann_Le_Cun

2. L'avènement du machine learning

Les premiers algorithmes ont été créés à la fin des années 1950. Le plus connu d'entre eux est Perceptron³, algorithme d'apprentissage supervisé (défini plus bas) inventé par Franck Rosenblatt. Entre les années 70 et 80, les efforts se sont portés sur des systèmes experts restreints à des domaines spécifiques comme l'analyse en chimie, le diagnostic médical, les systèmes experts de l'industrie, et les robots ou logiciels très spécialisés. L'émergence du machine learning en tant que discipline à part entière a eu lieu dans les années 90. Les premières applications du Machine Learning, tel qu'on le connaît aujourd'hui, remontent à cette période mais c'est depuis les années 2010 que le phénomène prend de l'ampleur.

Le machine learning, est une technologie d'intelligence artificielle permettant aux ordinateurs d'apprendre sans avoir été

programmés explicitement à cet effet. Il s'agit notamment de tâches comme le diagnostic des problèmes ou la prédiction des résultats dans les situations où l'utilisation des algorithmes classiques est problématique. En se nourrissant de nouvelles données, les algorithmes d'apprentissage automatique sont capables d'évoluer en permanence. « Par exemple, le machine learning peut permettre à un robot ayant la capacité de bouger ses membres mais ne sachant initialement rien de la coordination des mouvements permettant de marcher, d'apprendre à le faire. Le robot commencera par effectuer des mouvements aléatoires, puis, en sélectionnant et privilégiant les mouvements lui permettant d'avancer, mettra peu à peu en place une marche de plus en plus efficace.»⁴

Un problème de machine learning comporte plusieurs éléments spécifiques : des données, une tâche à accomplir, un algorithme d'apprentissage, une mesure des performances. Le traitement de ce problème peut être représenté comme suit :

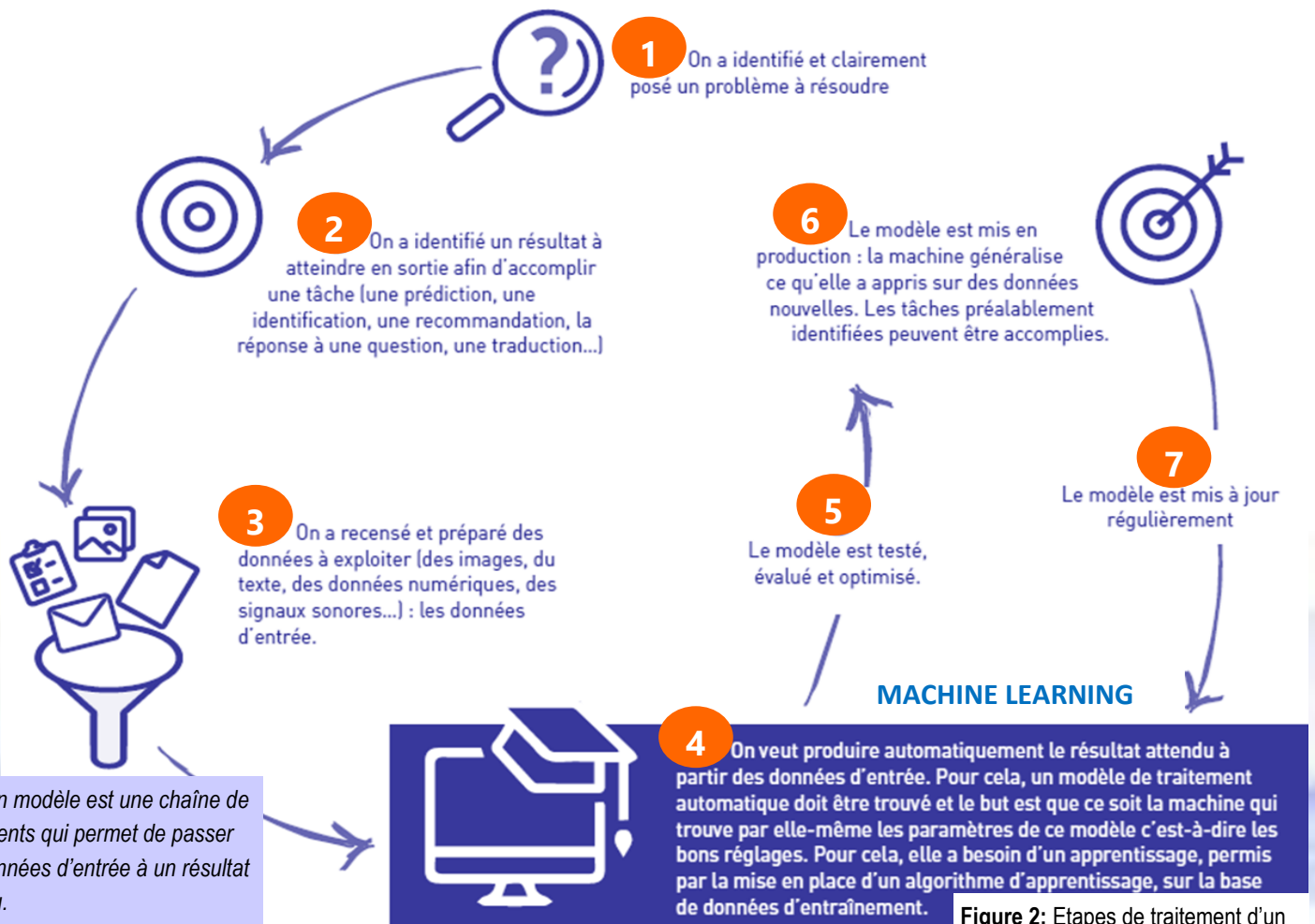


Figure 2: Etapes de traitement d'un problème avec le machine learning

³ Le perceptron est un algorithme d'apprentissage supervisé inventé en 1957 par Frank Rosenblatt¹ au laboratoire d'aéronautique de l'université Cornell.

⁴ https://fr.wikipedia.org/wiki/Apprentissage_automatique

Machine learning

En une phrase, un ordinateur apprend à partir de données pour résoudre une tâche en faisant attention à mesurer les performances. S'il améliore les performances sur cette tâche, lorsqu'on lui fournit les données d'entraînement, **on dit alors qu'il apprend.**

Par exemple, dans un zoo, un visiteur distingue quatre zèbres. Le premier âgé de 10 ans a 100 rayures noires. Le deuxième âgé de 5 ans en a 50. Le troisième âgé de 3 ans en a 30 et le quatrième âgé d'1 an en a 11. En faisant une analyse, le visiteur constate que le nombre de rayures correspond à presque dix fois le nombre d'années. Il sait désormais corréler le nombre de rayures et l'âge d'un zèbre : c'est l'apprentissage.

Le machine learning traite selon un procédé similaire des cas plus complexes dans des domaines variés grâce à des algorithmes ou méthodes d'apprentissage.

3. Les méthodes d'apprentissage

La phase d'apprentissage repose sur un **algorithme d'apprentissage**. Mais, selon le résultat visé, les données dont on dispose au départ et le contexte, il existe plusieurs façons de faire apprendre à une machine. On distingue l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage par renforcement, l'apprentissage semi-supervisé et l'apprentissage par transfert. Les deux dernières méthodes ne seront pas abordées dans le présent document.

♦ **L'apprentissage supervisé : apprendre par un exemple**

Dans le cas de l'apprentissage supervisé, le système est guidé dans son apprentissage. On lui indique le type de résultat à atteindre en le nourrissant d'exemples. Pour cela, on lui fournit des données

d'entrée pour lesquelles le résultat est connu et communiqué au système. Le but est qu'il puisse ensuite généraliser ce qu'il a appris pour des données non connues.

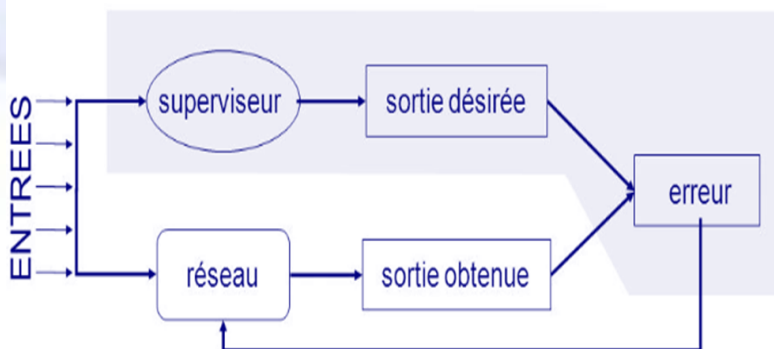


Figure 3 : Schéma de fonctionnement de l'apprentissage supervisé

Par exemple, si le système doit apprendre à reconnaître des feuilles de manguiers dans une image, on lui fournit des données étiquetées ou annotées en l'occurrence des images où la feuille est signalée et où le label « feuille de manguiers » est associé. Le jeu de données d'entraînement annotées permet au système de calculer ses erreurs en comparant ses résultats avec les résultats connus et ainsi d'ajuster le modèle pour progresser. Jim Sterne⁵ résume ainsi l'apprentissage supervisé : « Je connais la réponse, l'ordinateur non. Je lui donne de gros volumes de données étiquetées et de nombreux exemples. J'explique, il écoute. ». L'apprentissage supervisé est le concept derrière plusieurs applications : reconnaissance faciale dans la photographie, filtres anti-spam des messageries électroniques, maintenance prédictive, etc.

♦ **L'apprentissage non supervisé : apprendre en explorant**

Dans le cas de l'apprentissage non supervisé, on ne donne pas d'exemples de résultats attendus au système. Seules les données d'entrée sont fournies et le système doit apprendre, de façon autonome, la meilleure façon d'explorer les données. Il doit chercher à identifier dans le jeu de données une

⁵ Jim Sterne est l'auteur du livre Marketing digital Evolution Experience. Il est aussi cofondateur et président de la DAA (Digital Analytics Association)

Machine learning

façon de les structurer (trouver des modèles) ou encore à extraire des caractéristiques particulières.

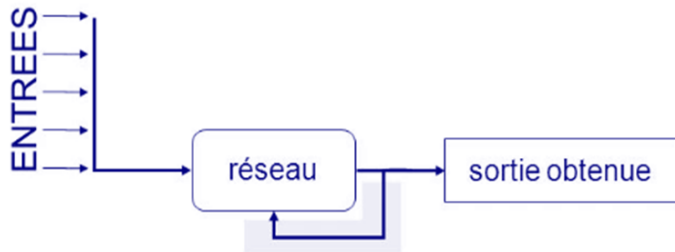


Figure 4 : Schéma de fonctionnement de l'apprentissage non-supervisé

Il y a toutefois un écueil à éviter avec l'apprentissage non supervisé : celui de supposer l'existence d'une causalité là où il n'y en a pas. L'apprentissage non supervisé peut être utilisé notamment dans le cadre de la détection d'anomalie ou de segmentation de marché etc.

- ♦ **L'apprentissage par renforcement : apprendre selon un principe de cause à effet**

Le système apprend cette fois par interaction avec l'environnement. A chaque instant, le système connaît l'état actuel et toutes les actions possibles. Il effectue une des actions et reçoit un signal de retour qui lui notifie son nouvel état et la récompense associée. Il n'y a alors pas de réponse absolue, seulement des récompenses. « Si la réponse est meilleure, on donne à l'ordinateur une récompense quantitative. Lorsqu'elle est moins bonne, la récompense devient négative », explique Jim Sterne. L'apprentissage par renforcement est très souvent utilisé pour la résolution de jeux notamment le jeu de Go où il

s'est fort bien illustré en 2017. AlphaZero, une variante d'AlphaGo⁶, n'a mis que quatre heures en partant des règles de base pour vaincre Stockfish, le meilleur programme de jeu du moment.

LES DOMAINES D'APPLICATION DU MACHINE LEARNING

Le machine learning est un domaine d'expertise assez vaste dont l'application est en plein essor dans de nombreux secteurs. Il est tellement répandu aujourd'hui que nous l'utilisons probablement des dizaines de fois par jour sans le savoir. Les différentes déclinaisons du machine learning ont changé la donne dans plusieurs secteurs de l'économie présentés ci-dessous.

1. Industrie – maintenance prédictive

En associant objets connectés et machine learning, il est possible de prévoir les pannes. On parle alors de maintenance prédictive. Dans le secteur industriel, les enjeux opérationnels et financiers sont énormes à ce sujet.

L'évolution de la maintenance

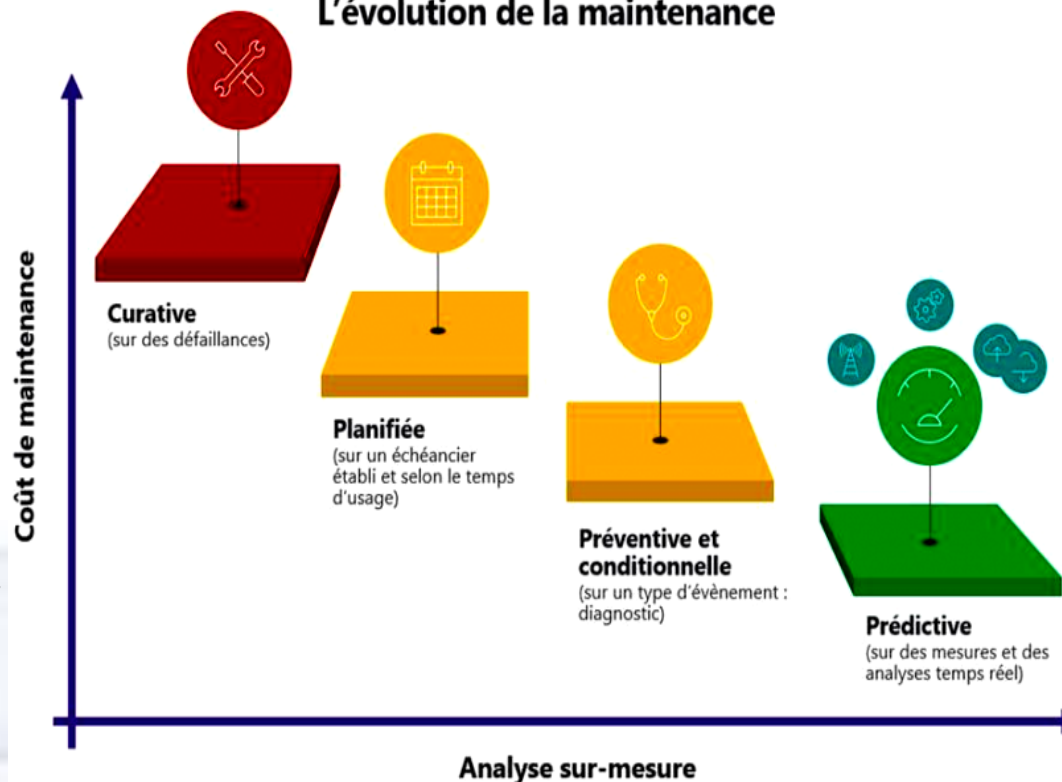


Figure 5 : Evolution de la maintenance

⁶ AlphaGO de Google DeepMind devint en octobre 2015, le premier programme à battre un joueur professionnel (le français Fan Hui). Il bat ensuite, en mars 2016, Lee Sedol, un des meilleurs joueurs mondiaux (9e dan professionnel), puis le 27 mai 2017

Machine learning

En effet, le machine learning permet d'ajuster le calendrier de maintenance en fonction des données collectées directement sur les équipements, pour éviter au maximum les pannes et le manque à gagner important qu'elles pourraient occasionner pour l'entreprise. .

Ce changement de paradigme constitue un véritable saut qualitatif dans le monde du service après-vente. En effet, dans les modèles actuels, la réparation de l'équipement n'intervient qu'une fois la panne constatée autrement dit la maintenance est curative ou corrective. Cependant, avec cette avancée, la maintenance sera prédictive et consistera à remplacer une pièce détachée en fonction de son taux d'usure statistique. En limitant voire en annihilant le risque de défaillance, l'entreprise allonge, toutes choses égales par ailleurs, la durée de vie de son équipement. D'après un sondage conduit en 2017 par Vanson Bourne, cabinet anglais spécialisé dans les études techniques, les trois-quarts des décideurs IT et des responsables de gestion des interventions estiment que les machines bénéficieront d'ici 2020 d'une meilleure prévention santé que les êtres humains ! Aujourd'hui, certaines machines intelligentes sont déjà en mesure de déclencher une demande d'intervention et, dans l'intervalle, de passer en mode dégradé.

2. Automobile

Dans le secteur automobile et particulièrement des véhicules autonomes, le machine learning facilite le travail des constructeurs qui se lancent dans le développement de véhicules autonomes. Il permet d'identifier en temps réel et d'éviter dans la majorité des cas, les éventuels obstacles (piétons, cyclistes etc.) qui se présentent sur une route. Cette technologie permet également de réaliser de la maintenance prédictive sur les voitures et autres camions afin de limiter au maximum les accidents.

3. Biens de consommation – marketing digital

Le machine learning permet aujourd'hui d'analyser finement de nombreuses variables extérieures propres aux clients en l'occurrence d'identifier les prospects les plus intéressants et ainsi optimiser les coûts, augmenter les volumes de vente d'un produit, personnaliser les offres en vue de satisfaire davantage les clients. Aussi, le machine learning dans le marketing est assez intéressant pour améliorer la logistique des magasins. Par exemple, en disséquant les données collectées en magasin sur le comportement des consommateurs, les grandes surfaces peuvent réorganiser leurs rayons pour booster leurs ventes et optimiser la gestion des stocks.

4. Finance

Dans la finance, cette technologie permet de débusquer les transactions frauduleuses en passant au crible les données bancaires des acheteurs, mais également de leur proposer des produits financiers ultra-personnalisés, en compilant des informations collectées via différents canaux.

Dans le monde, de grands groupes se sont déjà lancés dans des projets de transformation numérique en incluant le machine learning. C'est le cas de la BCE (Banque Centrale Européenne) qui prédit les évolutions des marchés financiers grâce à l'analyse de Tweets, du groupe d'assurance Covéa (MMA, Maaf, GMF) qui souhaite aller au-delà de l'approche statistique telle que maîtrisée depuis des années.

5. Agriculture

L'Agriculture fait face à de grands enjeux : produire en quantité, produire de la qualité et produire de façon durable avec des objectifs renforcés en termes de compétitivité économique, sociale et de réduction de l'impact environnemental. En effet, dès 2050, la

population mondiale devrait frôler les 10 milliards d'habitants (contre 7,6 milliards aujourd'hui)⁷. La



production alimentaire devra donc absorber cette croissance de population, alors que la quantité de terres arables disponibles est limitée et que le stress hydrique est quasi alarmant. Il est donc indispensable que l'agriculture se transforme, pour optimiser les ressources et éviter le gaspillage. Pour ces raisons, les agronomes et agriculteurs cherchent à disposer de méthodes et outils permettant de surveiller les facteurs de production, d'en prédire les évolutions et assurer une meilleure conduite des élevages. Grâce au machine learning, il s'offre de nouvelles possibilités pour répondre aux questions de suivi et de prévision à différentes échelles et pour des opérations difficilement automatisables jusqu'à présent. En effet, en production animale, cette avancée permet la prévision de comportements animaux et le contrôle des paramètres d'élevage (milieu, nourriture...) de sorte à anticiper et adapter les approvisionnements ainsi que les ressources humaines, éviter les pertes de bétail, garantir la santé animale et humaine. Du point de vue de la production végétale, elle permet la détection et la reconnaissance de plantes et des maladies, l'estimation des rendements, le comptage des fruits et la prédiction des caractéristiques du sol. Par conséquent, les agriculteurs peuvent customiser leurs techniques de culture en fonction de données météo et des caractéristiques singulières d'une

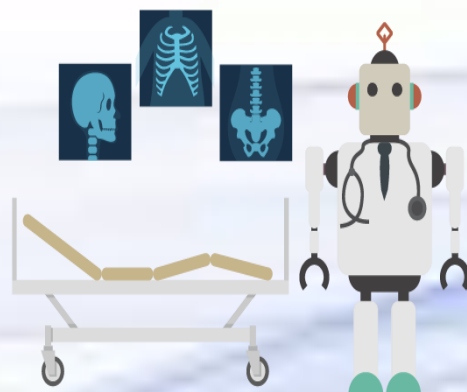
portion donnée de leur terrain.

6. Energie

A l'aide des capteurs posés sur les infrastructures et du machine learning qui permet de traiter les données collectées, les énergéticiens peuvent contrôler la production et la consommation de l'énergie. Prenons le cas de Google qui utilise le machine learning pour réduire la facture énergétique de ses datacenters. L'adaptation du niveau de refroidissement de ses centres de calcul par rapport aux besoins permet à l'entreprise d'atteindre un meilleur indicateur d'efficacité énergétique.

7. Santé

S'il existe bien un secteur dans lequel les humains profiteront directement et personnellement du machine learning, c'est bien celui de la santé avec l'amélioration des diagnostics et des chirurgies. Pour obtenir une illustration de cette technologie à l'œuvre, il suffit d'observer l'imagerie médicale, domaine dans lequel la vitesse de traitement et l'efficacité ont transformé le processus. Par exemple, Google vient de développer un algorithme de machine learning pour identifier les tumeurs cancéreuses à partir d'examens mammographiques. On peut aussi évoquer la plateforme PReDicT, développée par l'université d'Oxford, qui permet de diagnostiquer et de soigner les maladies du cerveau grâce à l'analyse prédictive. Un autre cas d'utilisation est celui des robots chirurgiens.



⁷ Le monde, « 10 milliards d'humains et alors ? » https://www.lemonde.fr/idees/article/2017/12/07/10-milliards-d-humains-et-alors_5226163_3232.html

On peut notamment citer Da Vinci, le robot chirurgien le plus avancé au monde spécialisé dans les opérations oculaires complexes, ou encore Heartlander, un mini-robot capable d'entrer dans le thorax d'un patient pour assister les chirurgiens humains pendant une opération du cœur.

8. Pharmacie

Le machine learning a investi le terrain de la pharmacie et des découvertes moléculaires. Ainsi, en croisant des molécules qui, sans le machine learning, n'auraient jamais été associées entre elles, les scientifiques trouvent par sérendipité⁸ des médicaments dont la découverte était peu probable vu l'état actuel des connaissances.

Les essais cliniques constituent un autre grand domaine. Toute innovation médicale doit être validée par des essais thérapeutiques longs et coûteux. Leem, syndicat des entreprises du médicament, évalue à 1 milliard d'euros la production d'une nouvelle molécule. Or, aujourd'hui tous les industriels pharmaceutiques évoquent leurs difficultés à trouver des volontaires satisfaisants cliniquement les critères de sélection de plus en plus complexes et s'engageant sur la durée du protocole. Souvent, ils abandonnent les essais cliniques générant alors des retards de commercialisation des produits et des dépassements de coûts colossaux. Dans un tel contexte, le machine learning constitue un moyen efficace d'identifier des profils ad hoc.

Une fois les médicaments fabriqués, le machine learning est un puissant outil pour repérer où sont les patients. Expédier des médicaments coûte cher surtout lorsqu'ils exigent des conditions très spécifiques de transfert (température, rapidité, mode de conservation). Aussi, collecter et traiter les données sur les types de pathologies présentent dans le monde permet d'anticiper la demande des produits et de les distribuer de façon ciblée.

Le machine learning est aussi une aide précieuse à la prescription médicale en vue d'éviter des erreurs. Les algorithmes peuvent révéler l'incompatibilité de deux médicaments prescrits par un médecin, la dangerosité d'une molécule en cas de pathologie chez un patient, ou le dépassement d'une posologie.

Enfin, on peut imaginer que le machine learning permette dans un futur plus ou moins proche une pharmacologie personnalisée, c'est à dire des médicaments adaptés à la génétique des patients.

9. Secteur public

Dans le secteur public, peut-être plus encore que dans le privé, la prise de décision est un processus complexe, multifactoriel, nécessitant de croiser de nombreuses sources de données différentes, avec potentiellement un impact sur plusieurs aspects de la vie des citoyens. Pour l'Etat, comme pour la plupart des entités publiques, c'est une vraie nécessité que de pouvoir regrouper, croiser et visualiser l'information en temps réel, afin d'améliorer la prise de décision. Qu'il s'agisse de mesurer l'impact d'une politique, de simuler ou de prédire les conséquences d'un choix, voire d'identifier de nouveaux axes stratégiques, l'exploitation des données tient une place centrale. Pour y parvenir, celui-ci dispose d'une richesse, en général encore peu exploitée : ses données. Grâce aux techniques d'analyse prédictive, de simulation et d'exploration visuelle, de machine learning, celles-ci peuvent être utilisées pour nourrir la prise de décision.

10. Médias

L'omniprésence d'Internet dans nos vies et l'explosion des réseaux sociaux ont considérablement changé la manière dont les contenus médiatiques sont consommés et donc publiés et traités. Afin de pérenniser leur métier et leurs organisations, les médias doivent continuellement produire efficacement, organiser

⁸ La **sérendipité** est le fait de réaliser une découverte scientifique ou une invention technique de façon inattendue à la suite d'un concours de circonstances fortuites et très souvent dans le cadre d'une recherche concernant un autre sujet.

automatiquement puis diffuser leurs contenus, et ce, en un tour de main. Pour ce faire, il est nécessaire qu'ils puissent s'appuyer sur un système de production performant et intelligent. Un enjeu de taille lorsque l'on connaît la dispersion des équipes de production, l'hétérogénéité des systèmes dans lesquels ils se trouvent et le flux de données dans lequel ils baignent. Avec le machine learning, les médias peuvent classifier les données, leur ajouter automatiquement une structure et automatiser leurs opérations de traitement. Ce qui entraîne un gain de temps considérable mais aussi une optimisation des coûts.

Le machine learning permet aussi d'assurer la modération des commentaires sous les articles de sorte à faire gagner du temps aux journalistes qui en ont la charge et ouvrir un plus grand nombre d'articles aux avis des lecteurs. C'est le cas du New York Times ou The Economist, des organes de presse qui utilisent un algorithme nommé « Perspective » qui détecte le niveau de toxicité d'un article sur une échelle de 1 à 100, en se basant sur le champ lexical issu des archives de commentaires déjà modérés. Cette notation a permis aux modérateurs de gagner du temps et d'intervenir en priorité sur les messages posant problème. Il existe aussi par ailleurs des robots journalistes qui savent donner la météo, rendre compte de centaines de rencontres sportives le week-end et afficher les résultats boursiers.

LE MACHINE LEARNING, DES DÉFIS À CONSIDÉRER

Le machine learning est piloté par les données. C'est ce qui en fait sa force mais également ce qui constitue sa faiblesse. Il consomme de très grands ensembles de données de sorte qu'il est souvent nécessaire de disposer de plusieurs jeux de données :

- ♦ des données pour l'apprentissage : celles qui vont permettre d'entraîner le modèle par l'algorithme d'apprentissage ;
- ♦ des données pour le test : les données gardées de côté, encore inconnues du système, qui vont permettre de mesurer l'erreur du « modèle ».

Plus le problème à résoudre est compliqué, plus le nombre de données nécessaires a tendance à être grand. Il peut atteindre des dizaines de milliers de données d'entrée. Ces données doivent pouvoir couvrir des variétés de situations et nécessitent donc parfois un recours à des sources multiples, pas toujours faciles d'accès. Ce qui soulève la question de la disponibilité de la donnée.

Si les données ne sont pas toujours à portée de main, il faut en plus s'assurer en amont de leur fiabilité, et en aval, de l'explicabilité et l'interprétabilité des résultats. Pourtant, le machine learning peut constituer « une véritable boîte noire » du fait des nombreuses combinaisons qui sont opérées sur les données en entrée avec des niveaux de complexités et d'abstraction différents. Dans ce cas, il devient difficile d'assurer la véracité des résultats, le cas échéant, de comprendre les biais et améliorer les performances du système.

Il est à noter aussi un manque de compétences dans ce secteur. Car, jamais les machines ne pourront tout faire d'elles-mêmes surtout que l'apprentissage automatique, pour peu que l'on prenne le temps d'identifier et rectifier les erreurs, permet d'améliorer les performances. Par conséquent l'exploitation des systèmes incluant le machine learning n'est pas toujours aisée.

Etant donné la multiplicité des sources, une question également à prendre en compte est celle de la sécurité et de la vigilance autour des systèmes d'informations. Il est crucial d'assurer un niveau de sécurité accrue d'un point de vue matériel, humain et

règlementaire (loi sur la protection des données à caractère personnel) aux systèmes d'informations pour gérer ces gros volumes de données et éviter les attaques de nature à corrompre les données.

Par ailleurs, la problématique de la responsabilité reste un enjeu majeur. Par exemple, lorsqu'un véhicule autonome intégrant l'apprentissage automatique ne reconnaît pas un panneau et que survient un incident, il est nécessaire de savoir à qui incombe réellement la responsabilité en vue de garantir la réparation du dommage.

Enfin, il n'existe pas encore de module de certification universelle des dispositifs de machine learning. Toutefois des travaux de normalisation sont en cours.

LE MACHINE LEARNING, A L'ÉPREUVE DU DROIT ET DE L'ÉTHIQUE

Alors que partout dans le monde des règlements sur la protection des données à caractère personnel sont adoptés, l'avènement des nouveaux outils algorithmiques du machine learning, soulève de nouveaux risques en termes de traitement des données. Toutefois, il n'existe aucun encadrement juridique spécifique en matière de machine learning.

En effet, l'application du machine learning bouleverse la définition même de la donnée personnelle et conduit ainsi à remettre en question les modalités d'application des dispositions légales applicables en la matière. Pour rappel, une donnée personnelle est définie par son pouvoir identifiant direct ou indirect. Elle peut être soit un nom, un prénom, un numéro de téléphone, un matricule ou une adresse IP etc. Pourtant, le machine learning, par définition implique le croisement de nombreuses données brutes, hétérogènes, individuellement insignifiantes pour aboutir à l'extraction de modèles

et d'informations significatives. Autrement dit, lorsque les informations identifiantes ne sont pas immédiatement disponibles, l'intersection d'un nombre suffisant de catégories auxquelles un individu est supposé appartenir peut faire émerger un sous-ensemble plus étroit, jusqu'à finalement pouvoir sélectionner une seule et unique personne.

Si ces algorithmes sont capables de détecter les ténues des corrélations entre les données produites par les individus, les profiler et les identifier, ils posent alors des problèmes d'interprétation et d'application des dispositions légales en vue de la protection des données à caractère personnel.

Par ailleurs, en plus des questions d'identification et de profilage des individus, bien souvent les algorithmes du machine learning s'intègrent à la robotique, et amènent, de par l'autonomie décisionnelle qu'ils favorisent, de véritables problématiques liées à la responsabilité du robot. En règle générale, pour que la responsabilité d'un individu soit engagée, il doit être impliqué dans le dommage et y jouer un rôle. Cependant, dans le cas d'espèce, c'est certainement la responsabilité du fabricant qui sera engagée.

Pour anticiper ces régimes de responsabilité relatifs au machine learning qui peuvent s'avérer compliqués à mettre en pratique, il est toujours possible de prévoir contractuellement un aménagement de la responsabilité afin de déroger aux principes de droit commun. Par exemple, le Parlement européen a adopté une résolution 2015/2103(INL) contenant des recommandations relatives aux règles de droit civil applicables en matière de robotique. Celle-ci stipule que « le futur instrument législatif devrait prévoir l'application d'une règle de responsabilité stricte pour les dommages causés par des robots "intelligents", en exigeant une simple preuve d'un lien de causalité entre le comportement dommageable du robot et le dommage subi par la partie lésée. Il convient de

créer un régime d'assurance robotique reposant éventuellement sur l'obligation du fabricant de contracter une police d'assurance pour les robots autonomes qu'il fabrique. Ce régime d'assurance devrait être complété par un fonds afin de garantir un dédommagement, y compris en l'absence de couverture. ».

Au regard des subtilités issues de l'utilisation du machine learning, il conviendrait de s'assurer que les systèmes fonctionnant à partir d'algorithmes d'apprentissage automatique respectent les dispositions générales en la matière:

- ◆ Informer les individus sur les traitements qui seront faits de leurs données collectées ;
- ◆ Obtenir leur consentement ;
- ◆ Respecter les finalités définies pour le traitement ;
- ◆ Garantir l'exercice des droits des individus concernés ;
- ◆ Mettre en place de mesures de sécurité techniques et organisationnelles appropriées ;
- ◆ Respecter les principes de *Privacy by Design*⁹, *Privacy by Default*¹⁰.

Enfin, l'émergence de nouvelles capacités technologiques en l'occurrence le machine learning impose un exercice d'interprétation des règles en vigueur. Ainsi, au risque de constater le dépassement et l'obsolescence de la notion traditionnelle de donnée à caractère personnel, il conviendrait d'y opposer une interprétation plus large. La donnée à caractère personnel ne doit plus être en effet conçue comme une propriété singulière, attachée à l'individu, mais comme une variable qui peut être obtenue à partir de traitement.

LE MACHINE LEARNING, QUID DE LA NORMALISATION ?

Considérant l'émergence du machine learning, un groupe spécialisé sur le machine learning (FG-ML5G) pour les réseaux futurs incluant la 5G a été mis en place par la commission d'études 13 de l'UIT-T lors de sa réunion de juin 2017. L'objectif de ce groupe est de mener une analyse sur le machine learning pour les réseaux futurs afin d'identifier les lacunes et les problèmes pertinents dans les activités de normalisation liées à ce sujet. Cette analyse comprend un aperçu des activités connexes menées par d'autres organismes de normalisation et groupes. En outre, il inclut des aspects techniques tels que des cas d'utilisation, des exigences éventuelles, des architectures et autres.

Plus précisément, les objectifs sont entre autres de:

- ◆ Faciliter l'adoption de machine learning dans les réseaux futurs, notamment l'architecture, les interfaces, les cas d'utilisation, les protocoles, les algorithmes, les formats de données, l'interopérabilité, les performances, l'évaluation, la sécurité et la protection des informations personnelles;
- ◆ Etudier, examiner et analyser les technologies, plates-formes, directives et normes existantes en matière de machine learning dans les réseaux futurs;
- ◆ Reconnaître et mettre en évidence les différentes perspectives d'avenir des réseaux et des systèmes informatiques impliquant le machine learning;
- ◆ Identifier les aspects permettant une utilisation sûre et sécurisée des frameworks;
- ◆ Examiner et étudier comment former, adapter, compresser et échanger des algorithmes machine learning dans les réseaux futurs et

9 Privacy by Design, ou La protection de la vie privée dès la conception en français, est une approche de l'ingénierie des systèmes qui prend en compte la vie privée tout au long du processus.

10 Default by design, ou en français C'est le corollaire du Privacy by design, qui implique que dès la conception du service ou du produit, le plus haut niveau de protection de la vie privée soit mis en place et applicable par défaut.

comment faire interagir différents algorithmes les uns avec les autres.

- ◆ Identifier les problèmes rencontrés dans les activités de normalisation pour le machine learning dans les communications;
- ◆ Produire une analyse des lacunes du machine learning afin d'identifier le champ d'application pertinent des recommandations de l'UIT-T sur ces sujets et d'élaborer une feuille de route pour le machine learning;
- ◆ Établir des liaisons et des relations avec d'autres organisations susceptibles de contribuer aux activités de normalisation pour le machine learning.

Les différentes questions discutées par le FG-ML5G constituent le point de départ afin de relever les défis soulever par le machine learning. Pour ne pas rester en marge de ces discussions internationales, l'Afrique et particulièrement la Côte d'Ivoire doit y mandater leurs délégués et assurer la prise en compte des recommandations aux niveaux régional et national.

En marge des activités du FG-ML5G, le machine learning suscite de plus en plus d'intérêt sur la scène internationale. Par exemple, en Novembre 2018, a eu lieu en France la première édition de la Conférence internationale sur l'apprentissage automatique pour la mise en réseau (MLN). L'objectif de la conférence est de fournir un forum aux scientifiques, ingénieurs et chercheurs pour discuter et échanger de nouvelles idées, résultats, expériences et travaux en cours sur tous les aspects de l'apprentissage automatique et de la mise en réseau. Chaque année, les participants aux MLN apprécient et bénéficient d'échanges multidisciplinaires sur les défis inhérents au machine learning.

Il faut noter aussi, l'existence de la Société internationale du machine learning une organisation à but non lucratif dont l'objectif est d'encourager la

recherche sur l'apprentissage automatique et l'activité phare, la présentation d'une conférence annuelle dénommée la conférence internationale sur l'apprentissage automatique (ICML). Elle est la principale conférence académique internationale sur l'apprentissage automatique. La dernière édition de cette conférence a eu lieu en 2018 à Stockholm (Suède).

Avec elle, il y a également la « Neural Information Processing Systems (NeurIPS, anciennement NIPS) », une conférence scientifique en intelligence artificielle et neurosciences computationnelles qui se tient chaque année. Les actes de conférence sont publiés sous forme de livre par le MIT Press et Morgan Kaufmann Publishers sous le titre « Advances in Neural Information Processing Systems »

Ces deux conférences constituent les deux principales plateformes de discussion à fort impact dans la recherche sur l'apprentissage automatique et l'intelligence artificielle.

LE MACHINE LEARNING, CATALYSEUR DE DEVELOPPEMENT POUR L'AFRIQUE ?

L'apprentissage automatique est déjà utilisé pour aider à résoudre les problèmes de développement avec des résultats prometteurs comme l'attestent les initiatives privées présentées dans notre document.

Comme nous le savons, l'Afrique est un continent à fort potentiel mais dont le développement tarde certainement parce qu'il n'est pas maîtrisé. Pourtant, dans de nombreux pays africains, les données, éléments décisionnels, sont disponibles sans que leur plein potentiel ne soit utilisé. Le machine learning arrive donc à point nommé car il pourrait

¹¹ MIT Press est une maison d'édition universitaire américaine affiliée au Massachusetts Institute of Technology à Cambridge, Massachusetts.

¹² Morgan Kaufmann Publishers est un éditeur basé à Burlington, Massachusetts (San Francisco, Californie jusqu'en 2008) et spécialisé dans les contenus informatiques et d'ingénierie.

résoudre la sous-exploitation des données. En effet, cette innovation technologique représente une opportunité car elle est un moyen pour aider à comprendre autrement le paysage africain et prendre des décisions plus éclairées.

C'est l'objectif qui a été visé lors le sommet « Données pour le développement en Afrique » qui a eu lieu à Nairobi en juin 2017 et qui a encouragé les initiatives déjà engagées par certains pays. C'est le cas du Kenya avec le développement d'un réseau intergouvernemental sur les données ouvertes pour l'agriculture et la nutrition. Ce réseau favorise un écosystème multipartenaire inclusif afin d'améliorer les capacités des petits exploitants agricoles à utiliser les données pour améliorer la productivité, augmenter la participation de la jeunesse dans les commerces agricoles et renforcer les capacités des départements statistiques du ministère de l'agriculture. Ce réseau est appuyé par un centre de données, né de la collaboration entre l'Université Strathmore du Kenya et Oracle en vue de fournir des services de cloud et des capacités d'analytique de données pour l'écosystème des données kenyan. Il comprend également un laboratoire de données qui sert d'espace pour l'idéation et la co-création de solutions fondées sur les données aux problèmes concrets du secteur public. Un autre exemple fort intéressant est celui de la Sierra Leone qui a lancé une initiative de données ouvertes et massives pour les activités fiscales et économiques, financée en partie par la Banque mondiale. Le gouvernement a développé une feuille de route sur les données pour les objectifs de développement durable adaptée au contexte national afin d'atteindre et d'assurer le suivi des progrès vers les objectifs mondiaux, en collaboration avec le Partenariat mondial pour les données sur le développement durable.

Eu égard à ce qui précède, force est de reconnaître que si les gouvernements africains sont capables de capter, d'agrèger, de structurer et de donner du sens

aux données, les politiques et stratégies nationales s'en verraient enrichies. Alors, on s'orienterait résolument vers un développement harmonieux et continu. Le développement de l'Afrique par le machine learning n'est donc pas une gageure. Mais pour atteindre cet objectif et tirer parti du machine learning, il faut :

- ◆ Inscrire le développement par les données et les TIC au cœur de la stratégie de développement globale.
- ◆ Investir pour améliorer les méthodes de travail, collecter des données de meilleure qualité, renforcer les capacités matérielles et humaines afin d'utiliser et analyser ces données à l'aide du machine learning ou tout autre technologie corollaire de manière à avoir un impact réel sur le développement.
- ◆ Encourager les entreprises privées à inclure le machine learning ou autre technologie innovante relatif au traitement des données dans leur plan de développement.

CE QU'IL FAUT RETENIR

Comme pour toute innovation de rupture, le développement du Machine Learning en particulier, et de l'intelligence artificielle en général, appellent encore bien des évolutions dans différents domaines. Parce que les technologies de l'intelligence artificielle peuvent susciter questions et craintes, leur adoption reposera en grande partie sur le bon suivi des systèmes.

Le machine learning ouvre les horizons pour de nouveaux usages ou des gains de performances par rapport aux méthodes actuelles. Mais dans tous les domaines déjà impactés, de nombreuses questions demeurent et constituent autant d'axes d'amélioration, notamment en matière de :

- ◆ Structuration, mutualisation et accès aux base

¹³ Une gageure : action, projet peu croyable ressemblant à un pari hasardeux. <https://www.larousse.fr/dictionnaires/francais/>

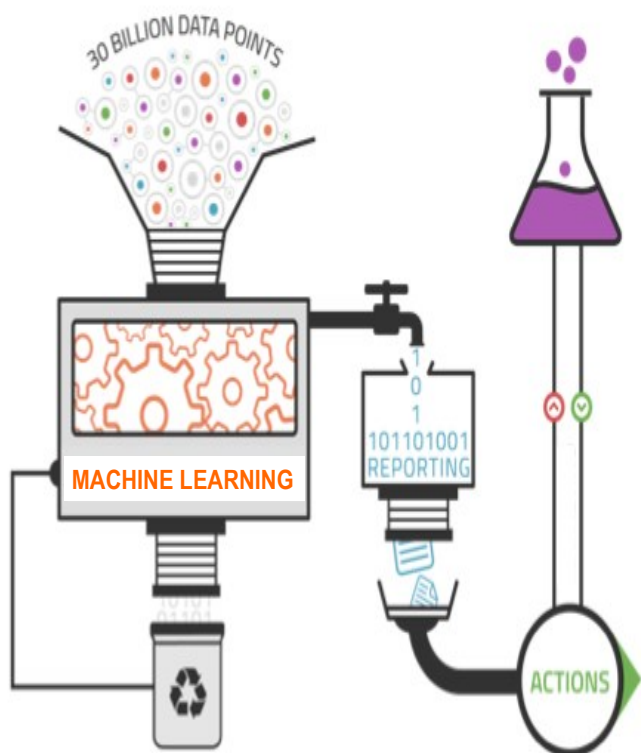
Machine learning

de données d'apprentissage ;

- ◆ D'adaptation des systèmes d'information ;
- ◆ De réglementation et de régulation ;
- ◆ D'évolution des composants matériels ;
- ◆ D'évaluation voire de certification des dispositifs.
- ◆ Etc.

WEBOGRAPHIE

1. <https://www.journaldunet.com/economie/expert/69528/le-machine-learning-au-c-ur-de-l-e-sante-de-demain.shtml>
2. <http://www.data4sdgs.org/news/les-pays-dafrique-mencent-la-revolution-des-donnees>
3. <https://www.info-afrique.com/les-donnees-au-service-du-developpement/>
4. <https://tel.archives-ouvertes.fr/tel-01282893/document>
5. <https://slideplayer.fr/slide/491441/>
6. https://www.lemonde.fr/idees/article/2017/12/07/10-milliards-d-humains-et-alors_5226163_3232.html
7. <https://www.larousse.fr/dictionnaires/francais/gageure/35786>
8. <https://www.wizaly.fr/blog/post/machine-learning-quelles-applications>
9. <https://www.insurancespeaker-wavestone.com/2016/07/machine-learning-application-domaine-banque-assurance/>
10. <http://www.topcom.fr/zoom/le-machine-learning-comme-levier-doptimisation-de-la-production-et-de-la-gestion-pour-les-medias-et-entreprises-du-divertissement/>
11. <https://www.frenchweb.fr/10-choses-que-vous-devez-absolument-savoir-sur-le-machine-learning/302284>
12. <https://www.quantmetry.com/single-post/2015/10/28/Une-petite-histoire-du-Machine-Learning>
13. <https://blog.atinternet.com/fr/machine-learning-et-marketing-les-conseils-de-jim-sterne/>
14. <https://www.maddyness.com/2018/07/20/comment-se-structure-le-secteur-du-machine-learning/>
15. <https://www.journaldunet.com/solutions/reseau-social-d-entreprise/1191979-machine-learning-12-secteurs/>
16. <https://itsocial.fr/innovation/intelligence-artificielle/6-defis-machine-learning-production/>
17. <https://medium.com/@mfouquet/m%C3%A9dias-et-intelligence-artificielle-la-mod%C3%A9ration-automatique-est-elle-pour-bient%C3%B4t-cace0efccec>
18. <https://www.fabernovel.com/insights/tech/lia-pour-les-nuls>
19. <http://ni.committees.comsoc.org/event/1st-international-conference-on-machine-learning-for-networking-mln-2018/>
20. https://en.wikipedia.org/wiki/International_Conference_on_Machine_Learning?oldid=786613415
21. https://fr.wikipedia.org/wiki/Neural_Information_Processing_Systems
22. <http://www.protection-des-donnees.fr/expertises/big-data/>
23. <http://www.revuedlf.com/droit-ue/la-protection-des-donnees-personnelles-face-aux-algorithmes-predictifs/>
24. <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//FR>



L'ARTCI scrute le paysage des TIC afin de déterminer de nouveaux sujets d'informations. Ces sujets permettent d'analyser l'actualité du secteur, de mieux comprendre les enjeux de la régulation et l'impact des TIC dans la vie de tous les jours.

Avril 2019
Abidjan, Côte d'Ivoire