

Glossaire de l'Intelligence Artificielle

document de travail

<https://huit.re/glossaire-IA>

source : <https://www.techtarget.com> par [Margaret Rouse](#)

Ce document regroupe un ensemble de définition proposée par Margaret Rouse sur le blog techtarget.com. Il n'a pas vocation à être exhaustif ni parfaitement rigoureux d'un point de vue scientifique mais présente un ensemble de notions accessible à tous.

Algorithme	2
Algorithme évolutionniste ou évolutionnaire	2
Analyse syntaxique (voir aussi TLN/NLP)	2
Apprentissage non supervisé (voir aussi Machine Learning)	3
Apprentissage supervisé (voir aussi Machine Learning)	4
Chatbot	4
Comment fonctionnent les chatbots	4
Exemples d'utilisations de chatbots	5
Pourquoi les chatbots sont importants	5
Cloud computing	6
Data Scientist	6
Deep learning (réseau neuronal profond)	7
Intelligence artificielle (IA, AI)	8
Types d'intelligence artificielle	9
IA explicable (XAI)	9
Machine Learning	10
Réseau de neurones artificiels (RNA)	10
Fonctionnement des réseaux de neurones artificiels	11
Apprentissage des réseaux neuronaux	11
Types de réseaux neuronaux	12
Applications des réseaux de neurones artificiels	13
Historique des réseaux neuronaux	13
Traitement du langage naturel (TLN)	14
Utilisations du traitement du langage naturel	14
Fonctionnement du traitement du langage naturel	14
Importance du TLN	15

Algorithme

Un algorithme est une procédure ou une formule qui permet de résoudre un problème. Ce mot est dérivé du nom du mathématicien Muhammad ibn Musa al-Kharezmi qui était à la cour royale à Bagdad et a vécu de 780 à 850 environ. Les travaux de ce mathématicien seraient également à l'origine du mot algèbre. Les programmes informatiques peuvent être considérés comme des algorithmes sophistiqués. Dans les domaines des mathématiques et de l'informatique, un algorithme signifie généralement une petite procédure qui résout un problème récurrent.

Algorithme évolutionniste ou évolutionnaire

Un [algorithme](#) évolutionniste ou évolutionnaire (AE) met en œuvre des mécanismes inspirés de la nature pour résoudre des problèmes comme le feraient des organismes vivants. Il appartient aussi bien à l'informatique bio-inspirée qu'à l'informatique évolutionniste. Les AE s'appuient sur les concepts de la théorie de l'évolution. Dans ces algorithmes, les solutions jouent le rôle des organismes individuels d'une population. Le panel des solutions possibles à un problème est d'abord une proposition aléatoire. La population est ensuite testée pour son aptitude à résoudre un problème correctement et rapidement. Puis les individus les plus adaptés sont sélectionnés pour la reproduction. Et le cycle recommence : on évalue le degré d'adaptation de la population et on élimine les individus les moins adaptés.

Comme les mécanismes des AE simulent les organismes vivants et leur évolution, ils font appel à des fonctions telles que la sélection, la reproduction, la mutation et la recombinaison. Le processus d'évolution qui consiste à choisir les meilleures solutions possibles à un problème selon un critère d'adaptation simule la sélection naturelle décrite par Darwin. Les solutions algorithmiques qui donnent les meilleurs résultats parmi les options proposées se reproduisent ; les moins adaptées sont éliminées et ne sont pas reproduites. Grâce à des tests mesurant le degré d'adaptation d'après les performances, l'optimisation s'opère sur plusieurs générations avec des fonctions comme la mutation. Les AE sont un excellent outil d'optimisation de solutions. Toutefois, ces algorithmes ne trouvent pas nécessairement la solution optimale. En effet, ils ne cessent de trouver des solutions dont ils comparent les performances, ce qui peut ou non faire émerger LA meilleure solution. Il est également important de noter que les AE sont plutôt gourmands en ressources de calcul, notamment à cause de la complexité inhérente à la détermination du degré d'adaptation. Il est possible de réduire cette complexité par l'approximation de l'adaptation.

Analyse syntaxique (voir aussi TLN/NLP)

L'analyse syntaxique consiste à analyser des éléments d'un texte de façon ordonnée. En linguistique, l'analyse syntaxique scinde les mots et les expressions en différentes parties afin d'étudier les relations et les significations. Par exemple, on demande parfois aux élèves d'effectuer l'analyse syntaxique d'une phrase en séparant le sujet et le prédicat, puis les expressions dépendantes, les modificateurs et ainsi de suite.

En général, effectuer l'analyse syntaxique d'un écrit ou d'un discours signifie simplement l'interpréter.

En informatique, l'analyse syntaxique consiste à diviser un énoncé en langage informatique en plusieurs parties exploitables par l'ordinateur. Dans un compilateur, un analyseur syntaxique est un programme qui prend chaque déclaration écrite par un développeur et la découpe en morceaux (par exemple, la commande principale, les options, les objets cibles, leurs attributs, etc.). Ces morceaux peuvent ensuite être utilisés pour développer d'autres actions ou créer des instructions formant un programme exécutable.

Apprentissage non supervisé (voir aussi Machine Learning)

L'apprentissage non supervisé consiste à apprendre à un [algorithme](#) d'intelligence artificielle (IA) des informations qui ne sont ni classées, ni étiquetées, et à permettre à cet algorithme de réagir à ces informations sans supervision.

Dans ce mode d'apprentissage, le système d'IA peut regrouper des informations non triées en fonction de leurs similitudes et de leurs différences, même si aucune catégorie n'est indiquée. Les systèmes d'IA capables d'utiliser l'apprentissage non supervisé sont souvent associés à des modèles d'apprentissage génératifs, mais ils peuvent aussi fonctionner avec une approche basée sur la récupération (souvent associée à l'[apprentissage supervisé](#)). Les approches d'apprentissage supervisé ou non supervisé sont utilisées notamment dans les [chatbots](#), les véhicules autonomes, les programmes de [reconnaissance faciale](#), les systèmes experts et les robots.

Dans l'apprentissage non supervisé, les données fournies au système d'IA ne sont ni étiquetées, ni classées, et les algorithmes du système traitent les données sans aucun entraînement préalable. La sortie dépend des algorithmes codés. L'introduction dans un système d'une approche d'apprentissage non supervisé est un moyen d'expérimenter l'intelligence artificielle.

Les algorithmes d'apprentissage non supervisé peuvent exécuter des tâches de traitement plus complexes que les systèmes d'apprentissage supervisé, mais ils peuvent aussi être plus imprévisibles. Même si un système d'IA d'apprentissage non supervisé parvient tout seul, par exemple, à faire le tri entre des chats et des chiens, il peut aussi ajouter des catégories inattendues et non désirées pour y classer des races inhabituelles, créant la confusion au lieu de mettre de l'ordre.

Apprentissage supervisé (voir aussi Machine Learning)

L'apprentissage supervisé, dans le contexte de l'intelligence artificielle ([IA](#)) et de l'[apprentissage automatique](#), est un système qui fournit à la fois les données en entrée et les données attendues en sortie. Les données en entrée et en sortie sont étiquetées en vue de leur classification, afin d'établir une base d'apprentissage pour le traitement ultérieur des données.

Les systèmes d'apprentissage automatique supervisé alimentent les [algorithmes](#) d'apprentissage avec des quantités connues qui étayeront les futures décisions. Supervisé ou [non supervisé](#), l'apprentissage automatique est utilisé notamment pour les [chatbots](#), les véhicules autonomes, les programmes de [reconnaissance faciale](#), les systèmes experts et les robots. Les systèmes d'apprentissage supervisé sont associés pour la plupart à une intelligence artificielle basée sur la récupération, mais ils peuvent aussi reposer sur un modèle d'apprentissage génératif.

Les données utilisées pour l'apprentissage supervisé sont une série d'exemples comprenant des paires composées de sujets en entrée et de sorties attendues (appelées également signaux de supervision). Prenons l'exemple d'un système d'apprentissage supervisé pour le traitement d'images dans lequel on introduit des photos de véhicules appartenant aux catégories voitures et camions. Après un temps d'observation suffisant, le système doit être capable de faire la distinction entre plusieurs images non étiquetées et de les catégoriser ; une fois cet objectif atteint, l'apprentissage peut être considéré comme terminé.

Les modèles d'apprentissage supervisé présentent certains avantages sur les modèles non supervisés, mais ils ont aussi des limites. Par exemple, ils sont plus susceptibles de prendre des décisions auxquelles les humains peuvent s'identifier parce qu'elles reposent sur des données fournies par les humains. Mais dans le cas d'une méthode basée sur la récupération, les systèmes d'apprentissage supervisé ont des difficultés à traiter les nouvelles informations. Si un système qui connaît les catégories voitures et camions, par exemple, reçoit une image de vélo, il devra la placer dans l'une ou l'autre de ces deux catégories, ce qui sera incorrect. Alors que si le système était génératif, il ne saurait pas forcément reconnaître un vélo, mais il serait capable de l'identifier comme appartenant à une autre catégorie.

Chatbot

Un chatbot (à prononcer « tchate-botte »), appelé parfois assistant virtuel, est un programme informatique qui simule une conversation (« chat » en anglais) avec une personne, à l'écrit ou à l'oral. Quand un utilisateur lui pose une question ou formule une commande, le chatbot lui répondra ou exécutera l'action demandée.

Ce sont en quelque sorte les porte-parole de [l'intelligence artificielle \(IA\)](#). Cette forme accessible d'IA est souvent mise en oeuvre par les entreprises dans leurs services SAV et d'assistance. On les trouve également dans les secteurs des appareils et des applications grand public.

Comment fonctionnent les chatbots

Les chatbots comme [ELIZA](#) et PARRY figurent parmi les premières tentatives de créer des programmes capables de tromper une personne, ne serait-ce que temporairement, et lui faire croire qu'elle tient une conversation avec quelqu'un. L'efficacité de PARRY a été mesurée au début des années 1970 à l'aide d'une version du test de Turing : l'attribution correcte par les testeurs des interactions à quelqu'un ou à un chatbot correspond aux résultats que donneraient des réponses aléatoires.

Les chatbots se sont beaucoup améliorés depuis. Ils exploitent les technologies d'IA notamment le Deep Learning, le traitement automatique du langage et les [algorithmes](#) d'apprentissage automatique ([Machine Learning](#)). Les plus généralistes ont besoin d'énormes volumes de données pour apprendre une langue. Plus un utilisateur interagit avec la machine, plus la reconnaissance de la parole ou du texte améliore sa réponse. Les chatbots peuvent être « stateless » ou « statefull ». Dans le premier cas, le chatbot aborde chaque interaction comme s'il s'agissait d'un nouvel utilisateur. Dans le deuxième, le chatbot est plus sophistiqué : les interactions passées lui servent à contextualiser ses prochaines réponses.

Aujourd'hui, une entreprise qui dote ses services d'un chatbot n'a que peu de code à écrire. En effet, plusieurs prestataires permettent aux développeurs de créer des interfaces de conversation quasiment clefs en mains pour tout type d'application.

Les applications de chatbot mobiles comme Siri d'Apple ou Cortana de Microsoft sont plutôt qualifiés d'assistants virtuels.

Exemples d'utilisations de chatbots

Le recours aux chatbots augmente aussi bien sur le marché B2B que grand public. La qualité s'améliorant, les consommateurs ont moins de réticence à interagir avec eux. Entre technologie de pointe et transition sociétale vers des communications textuelles plus passives, les chatbots occupent un ancien créneau du téléphone.

Les chatbots servent depuis des années dans des applications de [messagerie instantanée \(IM\)](#) et des jeux interactifs en ligne, mais n'ont que récemment pénétré les services grand public (B2C) et aux professionnels (B2B). Dans le gaming, si le robot est assez sophistiqué, un joueur peut ne pas se rendre compte qu'il interagit avec un programme informatique. Dans le commerce, les chatbots assistent les consommateurs qui font leurs achats en ligne. Ils peuvent répondre à des questions simples ou fournir des informations utiles (comme les frais de livraison et la disponibilité). Dans les services client, ils assistent également les agents pour répondre aux demandes récurrentes, soit directement au client, soit en conseillant l'agent humain (on parle alors « d'agent augmenté »). Si le bot est en « front » (contact direct avec le client), si la conversation devient trop complexe, elle peut être transférée à un conseiller.

Les chatbots servent aussi d'assistants virtuels à interface vocale et aident ainsi le consommateur dans sa vie quotidienne. Apple, Amazon, Google et Microsoft proposent tous ce type d'assistant sous une forme ou une autre ; les applications, comme Siri d'Apple et Cortana de Microsoft, ou les produits comme Alexa d'Amazon ou Google Home, sont tous des chatbots personnels généralistes.

Aujourd'hui, les tâches effectuées par les assistants virtuels de ce type restent simples.

Pourquoi les chatbots sont importants

Les économies et l'efficacité qui découlent des chatbots, notamment de leurs réponses aux questions récurrentes, sont des atouts pour des entreprises qui cherchent à augmenter la productivité de leurs services de ventes et d'assistance.

L'abandon progressif des formes traditionnelles de communication par les consommateurs devrait se traduire par un recours accru aux méthodes de communication à base de Chat. Les assistants virtuels / chatbot sont de plus en plus utilisés pour s'occuper de tâches simples.

Dans les centres d'appels, les conseillers - soulagés des questions basiques - peuvent se consacrer à des dossiers à plus fort potentiel ou plus épineux. L'entreprise fait baisser ses coûts (les employés coûtent plus cher) et peut en plus fournir un niveau de service standard à ses clients même quand aucun conseiller humain n'est disponible - y compris la nuit ou le week-end, un bot étant par définition disponible 24h sur 24, 7 jours sur 7.

Cloud computing

Le Cloud computing est un concept général qui désigne la mise à disposition de services hébergés sur Internet. Le Cloud computing permet aux entreprises de consommer les ressources informatiques à la demande (comme elle le ferait d'un service public tel que l'électricité), en leur évitant de créer et de gérer des infrastructures en interne.

Plusieurs avantages pour les entreprises et les utilisateurs découlent du Cloud. Trois de ces avantages sont les suivants :

- Mise à disposition en libre-service : Les utilisateurs peuvent se procurer des ressources informatiques à la demande pour presque tous les types de workloads.
- Elasticité : Les entreprises peuvent monter en puissance lorsque leurs besoins augmentent, puis de nouveau diminuer leur capacité de traitement lorsque la demande baisse.
- Paiement à l'usage : Les ressources informatiques sont évaluées de façon très fine, les utilisateurs ne payant que les ressources qu'ils utilisent.

Les services de Cloud computing peuvent être [privés](#), [publics](#) ou [hybrides](#).

Les services de Cloud privé sont fournis par le datacenter d'une entreprise à ses utilisateurs internes. Ce modèle offre souplesse et confort d'emploi, tout en garantissant la gestion, le contrôle et la sécurité. Les services peuvent ou non être facturés aux clients internes par le biais d'une refacturation.

Dans le modèle de Cloud public, c'est un prestataire tiers qui fournit le service de Cloud sur Internet. Les services de Cloud public sont facturés à la demande, généralement à la minute ou à l'heure. Les clients ne payent que les cycles de CPU, le stockage ou la bande passante qu'ils consomment. Les principaux fournisseurs de Cloud public sont Amazon Web Services (AWS), Microsoft Azure, IBM et Google.

Data Scientist

Le terme Data Scientist est un titre professionnel qui désigne un employé ou un consultant en informatique décisionnelle ([BI, Business Intelligence](#)) qui excelle dans l'analyse des

données, particulièrement des grandes quantités de données, et ce afin d'aider une entreprise à en tirer un avantage concurrentiel.

Le titre de Data Scientist est parfois déprécié comme n'étant pas assez spécifique. Il peut être perçu comme un synonyme d'analyste des données, un cran plus haut.

Quoi qu'il en soit, le poste est de plus en plus reconnu dans les grandes entreprises qui s'intéressent à l'interprétation du [Big Data](#), défini comme la vaste quantité de données structurées, semi-structurées et non structurées qu'elles produisent.

Un Data Scientist maîtrise un ensemble de compétences incluant [l'analytique](#), l'apprentissage automatique (Machine Learning), l'exploration de données (data mining) et l'analyse statistique. Il dispose parallèlement d'une expérience des algorithmes et du codage.

Mais sa compétence la plus importante n'en demeure pas moins sa capacité à expliquer la signification des données dans un langage facilement compréhensible par tous.

Deep learning (réseau neuronal profond)

L'apprentissage profond, ou « deep learning », est un aspect de l'intelligence artificielle ([IA](#)) qui imite la méthode d'apprentissage qu'utilisent les êtres humains pour acquérir certains types de connaissances. Sous sa forme la plus simple, le deep learning peut être considéré comme un moyen d'automatiser l'analytique prédictive.

Alors que les [algorithmes](#) traditionnels de l'apprentissage automatique sont linéaires, ceux du deep learning sont empilés dans une architecture d'une complexité et d'une abstraction croissantes.

Pour comprendre le deep learning, imaginez un jeune enfant dont le premier mot est « chien ». Il apprend ce qu'est un chien (et ce qui n'est pas un chien) en pointant des objets du doigt et en disant le mot « chien ». Ses parents lui répondent « Oui, c'est un chien » ou « Non, ce n'est pas un chien ».

A mesure que l'enfant continue de montrer des objets, il comprend de mieux en mieux quelles sont les caractéristiques que possèdent tous les chiens. Ce qu'il fait, sans le savoir, c'est clarifier une abstraction complexe (le concept de chien) en construisant une architecture dans laquelle chaque niveau d'abstraction est créé avec les connaissances acquises au niveau précédent.

Les programmes informatiques qui utilisent le deep learning suivent en grande partie le même processus. Chacun des algorithmes de l'architecture applique une transformation non linéaire aux données en entrée et utilise ce qu'il apprend pour créer un modèle statistique en sortie. Les itérations se poursuivent jusqu'à ce que la sortie ait atteint un niveau de précision acceptable. Le qualificatif « deep » (profond) s'explique par le nombre de couches de traitement par lesquelles les données doivent passer.

Dans [l'apprentissage automatique](#) classique, le processus d'apprentissage est supervisé et le programmeur doit être très, très explicite lorsqu'il indique à l'ordinateur le type d'éléments qu'il doit rechercher pour déterminer si une image représente un chien ou pas. C'est un processus laborieux appelé extraction de caractéristiques, et le taux de réussite de l'ordinateur dépend entièrement de la capacité du programmeur à définir de manière précise le jeu de caractéristiques d'un « chien ». L'avantage du deep learning est que le programme crée lui-même le jeu de caractéristiques sans supervision. C'est non seulement plus rapide, mais c'est généralement plus précis.

Au départ, le programme informatique peut être alimenté par des données d'apprentissage, par exemple une série d'images à chacune desquelles un humain aura associé la mention « chien » ou « pas chien » à l'aide de métabalises. Il utilise les informations fournies par ces données pour créer le jeu de caractéristiques du chien et élaborer un modèle prédictif. Dans ce cas, le modèle que l'ordinateur crée en premier peut estimer que dans une image, chaque élément doté de quatre pattes et d'une queue doit être étiqueté comme étant un « chien ». Bien sûr, le programme ne connaît pas les étiquettes « quatre pattes » ou « queue », il recherche simplement des motifs de pixels dans les données numériques. A chaque itération, le modèle prédictif que crée l'ordinateur devient plus complexe et précis. Parce que ce processus imite la pensée humaine, le deep learning est parfois appelé apprentissage neuronal profond ou réseau neuronal profond. A la différence du petit enfant, qui mettra des semaines, voire des mois, à comprendre le concept de « chien », un programme informatique qui utilise des algorithmes de deep learning peut, après avoir reçu un jeu de données d'apprentissage, trier des millions d'images, en identifiant avec précision et en quelques minutes celles qui représentent des chiens.

Pour que le niveau de précision soit acceptable, les programmes de deep learning doivent avoir accès à des quantités phénoménales de données d'apprentissage et de puissance de traitement, deux conditions qui étaient difficiles à remplir pour les programmeurs avant l'avènement du Big Data et du Cloud computing. Les programmes de deep learning étant capables de créer des modèles statistiques complexes directement depuis leurs propres sorties itératives, ils peuvent élaborer des modèles prédictifs précis à partir de grandes quantités de données non structurées et sans étiquette. Cette capacité est importante dans le contexte de la propagation de l'Internet des objets ([IoT](#), Internet of Things), car la plupart des données que les humains et les machines créent ne sont ni structurées, ni étiquetées. Les cas d'utilisation du deep learning sont tous les types d'applications d'analytique du [Big Data](#), en particulier celles qui sont axées sur le traitement du langage naturel, la traduction, le diagnostic médical, les signaux boursiers, la sécurité des réseaux et l'identification des images.

Voici une illustration très simple du fonctionnement d'un programme de deep learning. Cette vidéo réalisée par LuLu Art Group présente la sortie d'un programme de deep learning qui a été alimenté par des données brutes de capture des mouvements. Elle montre à quoi ressemble le concept abstrait de « danse » selon l'analyse prédictive du programme.

<https://youtu.be/QnaKyc1Mpmo>

À chaque itération, le modèle prédictif du programme est devenu plus complexe et plus précis. <https://www.youtube.com/watch?v=W1oRgDPxEkc&>

Intelligence artificielle (IA, AI)

L'intelligence artificielle ou IA (Artificial Intelligence -AI- en anglais) vise à permettre à des machines, et plus particulièrement à des systèmes informatiques, de simuler les [processus cognitifs](#) humains.

Ces processus comprennent l'apprentissage (acquisition d'informations et de règles liées à leur utilisation), le raisonnement (application des règles pour parvenir à des conclusions approximatives ou précises) et l'autocorrection. Les applications spécifiques de l'IA sont notamment les systèmes experts, la reconnaissance vocale et la vision artificielle.

On doit la première occurrence du terme d'IA à John McCarthy, chercheur américain en informatique, à la Dartmouth Conference de 1956 qui vit naître la discipline. Aujourd'hui, le terme recouvre aussi bien l'automatisation robotisée des processus (Robotic Process Automation, RPA) que la robotique proprement dite. Il a récemment gagné en visibilité en partie à cause des [Big Data](#), c'est-à-dire de la vitesse, du volume et de la diversité des données collectées par les entreprises. Plus apte que l'homme à faire ressortir des tendances des données, l'IA permet aux entreprises d'exploiter un maximum d'informations.

Types d'intelligence artificielle

Il y a plusieurs façons de classer les systèmes d'IA. En voici deux exemples.

Le premier est que l'intelligence artificielle peut être considérée comme faible ou forte.

L'IA faible est un système d'intelligence artificielle conçu pour reproduire une tâche précise à laquelle il est formé. Les assistants personnels virtuels comme Siri d'Apple en sont une forme. L'IA forte, dite aussi intelligence artificielle générale, est un système doté de capacités cognitives humaines générales qui, lorsqu'on lui présente une tâche inhabituelle, est assez intelligent pour trouver une solution. Bien que controversé, le test de Turing, élaboré par le mathématicien du même nom en 1950, est une méthode qui sert à déterminer si un ordinateur pense comme un humain.

Le deuxième exemple nous vient d'Arend Hintze, professeur en biologie intégrative et ingénierie informatique à la Michigan State University. Il classe l'IA en quatre types allant de celui des systèmes actuels aux systèmes sensibles à venir. Ses catégories sont :

- Type 1 : machines réactives. Chacun se souvient de Deep Blue, le programme d'IBM qui a battu Garry Kasparov aux échecs dans les années 1990. Deep Blue identifie les pièces sur l'échiquier et émet des prédictions, mais il n'a aucune mémoire et n'utilise pas ses expériences passées pour formuler les prédictions futures. Il analyse les coups possibles, les siens et ceux de son adversaire, et choisit le plus stratégique. Deep Blue et AlphaGO de Google ont été conçus à des fins précises : ils ne sont pas facilement transposables à une autre situation.
- Type 2 : machines à mémoire restreinte. Ces systèmes d'IA s'appuient sur leurs expériences passées pour prendre les décisions suivantes. Certaines fonctions décisionnelles des véhicules autonomes suivent ce modèle. Les observations servent à contextualiser les actions du futur proche, par exemple une voiture qui change de file. Ces observations ne sont pas stockées de manière permanente.
- Type 3 : théorie de l'esprit. Il s'agit d'un concept de psychologie qui se rapporte à la compréhension des gens en tant qu'êtres ayant des pensées, des désirs et des raisons propres qui les poussent à prendre leurs décisions. Ce type d'IA n'existe pas encore.
- Type 4 : conscience de soi. Dans cette catégorie, les systèmes d'IA ont une identité, une conscience. Ces machines douées de conscience connaissent leur état actuel et utilisent ces informations pour inférer ce que les autres ressentent. Ce type d'IA n'existe pas encore.

IA explicable (XAI)

L'IA explicable ou XAI (eXplainable Artificial Intelligence) est une forme d'[intelligence artificielle](#) prévue pour décrire son objet, sa logique et sa prise de décision de manière

intelligible à une personne lambda. Souvent évoquée de pair avec l'apprentissage profond ou « [deep learning](#) », elle joue un rôle primordial dans le modèle d'apprentissage automatique non discriminatoire, redevable et transparent dit FAT ML (Fairness, Accountability and Transparency in Machine Learning).

La XAI renseigne globalement sur la prise de décision d'un programme d'intelligence artificielle en dévoilant :

- Les points forts et les faiblesses
- Les critères précis retenus pour arriver à une décision
- Les motifs qui l'ont conduit à telle décision plutôt qu'aux autres
- Le niveau de confiance approprié selon les différents types de décision
- Les types d'erreur qu'il est susceptible de commettre
- La façon de corriger les erreurs

Un des grands objectifs de la XAI est la redevabilité / responsabilité algorithmique. Jusqu'à maintenant, les systèmes d'IA ont été par essence des boîtes noires. Si on en connaît les données en entrée et en sortie, les algorithmes qui mènent à une décision sont généralement propriétaires ou peu intelligibles, même quand les mécanismes de logique internes sont accessibles gratuitement en [open source](#).

Étant donné que l'intelligence artificielle est de plus en plus répandue, il est plus que jamais important de savoir comment traiter les distorsions et la question de la confiance. Notons par exemple qu'une des clauses du règlement général sur la protection des données ([RGPD](#)) de l'UE instaure le droit à l'explication.

Machine Learning

L'apprentissage statistique (Machine Learning) est un type d'intelligence artificielle qui confère aux ordinateurs la capacité d'apprendre sans être explicitement programmés. Cette technologie s'appuie sur le développement de programmes informatiques capables d'acquérir de nouvelles connaissances afin de s'améliorer et d'évoluer d'eux-mêmes dès qu'ils sont exposés à de nouvelles données.

Le processus d'apprentissage automatique s'apparente à celui de l'exploration de données (data mining). En effet, il s'agit, dans les deux cas, d'analyser les données à la recherche de schémas récurrents. Cependant, au lieu d'extraire les données pour les soumettre à un traitement humain (comme c'est le cas dans les applications de data mining), l'apprentissage automatique utilise ces données pour améliorer la compréhension du programme lui-même.

Les programmes d'apprentissage automatique détectent des schémas dans les données et ajustent leur fonctionnement en conséquence. Par exemple, le fil d'actualité de Facebook change en fonction des interactions personnelles de l'utilisateur avec ses homologues. Ainsi, si un utilisateur identifie fréquemment un ami dans des photos, écrit sur son mur ou « aime » ses liens, le fil d'actualité montrera un plus grand nombre d'activités de cet ami, car il supposera qu'il s'agit d'un ami proche de l'utilisateur.

Réseau de neurones artificiels (RNA)

Dans le domaine des technologies de l'information, un réseau de neurones est un système logiciel et / ou matériel qui imite le fonctionnement des neurones biologiques. Les réseaux neuronaux, aussi appelés réseaux de neurones artificiels (RNA ou ANN en anglais), font partie des technologies d'[apprentissage profond](#) (ou « deep learning »), couvertes également par l'intelligence artificielle ([IA](#)).

Les applications commerciales sont souvent axées sur la résolution de problèmes complexes de traitement de signaux ou de reconnaissance de modèles. Parmi les exemples les plus connus depuis 2000, on trouve la reconnaissance de l'écriture manuscrite pour le traitement des chèques, la conversion de parole en texte, l'analyse des données d'exploration pétrolière, les prévisions météorologiques et la [reconnaissance faciale](#).

Fonctionnement des réseaux de neurones artificiels

Un réseau neuronal sous-entend normalement qu'un grand nombre de [processeurs](#) fonctionne en parallèle et en couches successives. La première couche reçoit en entrée les informations brutes, à la manière du nerf optique qui traite les données visuelles humaines. Chaque couche successive reçoit les données de la couche précédente plutôt que les données brutes, tout comme les neurones éloignés du nerf optique reçoivent les signaux des neurones voisins. La dernière couche produit le résultat.

Chaque [noeud](#) de traitement a sa petite bulle de connaissances, composée notamment de ce qu'il a vu et des règles programmées à l'origine ou définies par lui-même. Les couches sont étroitement interconnectées : chaque noeud d'une [couche n](#) est connecté à de nombreux noeuds de la couche n-1 (ses entrées) et de la couche n+1 qui seront à leur tour les entrées de ces noeuds-là. Il peut y avoir un ou plusieurs noeuds dans la couche de sortie dont provient la réponse lisible.

Les réseaux neuronaux sont remarquables par leur capacité d'adaptation : ils se modifient eux-mêmes en fonction de l'entraînement initial et les exécutions suivantes leur apportent encore plus d'informations sur le monde qui les entourent. Le modèle d'apprentissage le plus élémentaire est axé sur la pondération des flux d'entrée, autrement dit sur l'évaluation par chaque noeud de l'importance des entrées provenant de chacun de ses prédécesseurs. Les entrées qui contribuent à mener aux bonnes réponses ont un coefficient plus fort.

Apprentissage des réseaux neuronaux

Normalement, un réseau neuronal passe d'abord par une phase d'apprentissage, ou bien est alimenté par des données en nombre. L'entraînement consiste à lui fournir des données en entrée et à lui donner le résultat attendu. Par exemple, pour bâtir un réseau qui reconnaisse les visages d'acteurs, l'entraînement initial peut consister en une série de photographies de comédiens, de non-comédiens, de masques, de statues, de gueules d'animaux, etc. Chaque entrée est accompagnée de l'identification correspondante : le nom de l'acteur, la mention « non-acteur » ou « non-humain ». Grâce aux réponses fournies, le modèle ajuste ses pondérations internes pour apprendre à mieux faire son travail.

Prenons un exemple. Les noeuds David, Dianne et Dakota indiquent au noeud Ernie que l'image actuelle est une photo de Brad Pitt, mais le noeud Durango affirme qu'il s'agit de Betty White. Le programme d'entraînement confirme qu'il s'agit de l'acteur. Ernie va alors réduire le coefficient affecté aux entrées de Durango et augmenter ceux de David, Dianne et Dakota.

Pour définir les règles et prendre des décisions, c'est-à-dire décider de ce qu'il faut envoyer à la couche suivante d'après les entrées reçues de la précédente, les réseaux neuronaux suivent plusieurs principes, notamment l'apprentissage à base d'algorithme du gradient, la [logique floue](#) (ou « fuzzy logic »), les algorithmes génétiques et les méthodes bayésiennes. On peut leur donner quelques règles élémentaires sur les relations des objets dans l'espace en cours de modélisation.

Par exemple, un système de reconnaissance faciale peut recevoir les instructions : « les sourcils se trouvent au-dessus des yeux » ou « les moustaches se trouvent au-dessous du nez. Les moustaches se trouvent au-dessus et / ou sur les côtés de la bouche ». Le préchargement des règles accélère l'entraînement et produit plus rapidement un modèle plus puissant. Mais ce faisant, on intègre des hypothèses sur la nature des limites du problème, qui peuvent être non pertinentes, inutiles voire contre-productives : la décision quant aux règles à inclure, voire à leur inclusion elle-même, en est d'autant plus importante. Les réseaux de neurones et autres systèmes d'apprentissage automatique peuvent amplifier les préjugés / la partialité.

De plus, les hypothèses avancées par ceux qui entraînent les [algorithmes](#) risquent de créer des réseaux neuronaux qui amplifient les préjugés culturels. Les ensembles de données subjectifs représentent un défi permanent lors de l'entraînement des systèmes qui trouvent eux-mêmes les réponses d'après les modèles extraits des données. Si les données en entrée de l'algorithme ne sont pas neutres, et presque aucune donnée ne l'est, la machine va propager des préjugés.

Types de réseaux neuronaux

On parle parfois de la profondeur des réseaux neuronaux, jusqu'à donner le nombre de couches séparant l'entrée de la sortie ou de couches dites cachées du modèle. C'est pourquoi le terme réseau neuronal ou de neurones s'emploie comme quasi-synonyme d'apprentissage profond. On s'y réfère aussi par le nombre de noeuds cachés du modèle ou d'entrées et de sorties de chaque noeud. Des variantes du réseau neuronal classique permettent diverses formes de propagation des informations entre les couches, vers l'amont et vers l'aval.

De plus, les hypothèses avancées par ceux qui entraînent les algorithmes risquent de créer des réseaux neuronaux qui amplifient les préjugés culturels. Les ensembles de données subjectifs représentent un défi permanent lors de l'entraînement des systèmes qui trouvent eux-mêmes les réponses d'après les modèles extraits des données. Si les données en entrée de l'algorithme ne sont pas neutres, et presque aucune donnée ne l'est, la machine va propager des préjugés.

Types de réseaux neuronaux On parle parfois de la profondeur des réseaux neuronaux, jusqu'à donner le nombre de couches séparant l'entrée de la sortie ou de couches dites cachées du modèle. C'est pourquoi le terme réseau neuronal ou de neurones s'emploie comme quasi-synonyme d'apprentissage profond. On s'y réfère aussi par le nombre de noeuds cachés du modèle ou d'entrées et de sorties de chaque noeud. Des variantes du réseau neuronal classique permettent diverses formes de propagation des informations entre les couches, vers l'amont et vers l'aval.

Cette vidéo montre la façon dont les réseaux neuronaux traitent les données et les prévisions qui en découlent. <https://www.youtube.com/watch?v=yEqxv9fKwRk>
La variante la plus simple est le réseau neuronal à propagation avant (ou acyclique). Ce type d'algorithme de réseau de neurones artificiels fait passer directement les informations

en entrée des noeuds de traitement vers les sorties. Il peut y avoir ou non des couches de noeuds cachées qui facilitent encore l'interprétation de leur fonctionnement.

Les réseaux neuronaux récurrents sont plus complexes. Ces algorithmes d'apprentissage profond enregistrent la sortie des noeuds de traitement et la réinjectent dans le modèle. On dit que le modèle apprend.

Les réseaux neuronaux convolutifs ou à convolution sont courants aujourd'hui, surtout dans le domaine de la reconnaissance d'images. Ce type particulier d'algorithme est utilisé dans la plupart des applications d'IA les plus sophistiquées, entre autres la reconnaissance faciale, la numérisation de textes et le traitement automatique des langues.

Applications des réseaux de neurones artificiels

La reconnaissance d'images est l'un des premiers domaines d'application réussie des réseaux neuronaux, mais la technologie s'est étendue à bien d'autres, à savoir :

- les [chatbots](#),
- le traitement automatique des langues, la traduction et la génération automatique de textes en langue naturelle,
- les prévisions boursières,
- la planification et l'optimisation des tournées de livraison,
- la découverte et l'élaboration de médicaments.

Il ne s'agit que de quelques domaines particuliers auxquels s'appliquent aujourd'hui les réseaux neuronaux. Les principaux usages concernent tout traitement qui suit des règles ou des modèles stricts, et dispose de volumes considérables de données. Si le volume de données est trop grand pour qu'une personne l'appréhende dans un laps de temps raisonnable, l'opération se prête parfaitement à l'automatisation par les réseaux de neurones artificiels.

Historique des réseaux neuronaux

Les réseaux de neurones artificiels remontent au tout début de l'informatique. En 1943, les mathématiciens Warren McCulloch et Walter Pitts construisent un système de circuits qui exécute des [algorithmes](#) simples pour imiter le fonctionnement du cerveau.

En 1957, le chercheur Frank Rosenblatt de l'université Cornell développe le perceptron, un algorithme de reconnaissance avancée de modèles qui ouvre la voie à la reconnaissance d'objets dans les images par les machines. Le perceptron n'ayant pas répondu aux attentes, les chercheurs ont délaissé les réseaux de neurones artificiels au cours des années 1960.

Marvin Minsky parle avec Seymour Papert de ses travaux sur les réseaux neuronaux.

En 1969, les chercheurs Marvin Minsky et Seymour Papert du MIT publient le livre *Perceptrons*, qui aborde différentes problématiques des réseaux neuronaux, notamment le manque de puissance des ordinateurs de l'époque empêchant de traiter les données nécessaires au fonctionnement normal des réseaux neuronaux. Pour beaucoup, ce livre est responsable de l'hiver prolongé de l'IA, une ère qui se caractérise par l'arrêt des recherches dans le domaine des réseaux neuronaux.

Ce n'est qu'en 2010 que la recherche a repris. L'émergence des [Big Data](#), ces volumes colossaux de données qu'amassent les entreprises, et le traitement informatique parallèle ont donné aux data-scientists les données d'entraînement et les ressources informatiques indispensables aux réseaux de neurones artificiels complexes. En 2012, un réseau de

neurones a surpassé les performances humaines lors d'une tâche de reconnaissance d'images dans le cadre du concours ImageNet. Depuis lors, les réseaux de neurones artificiels suscitent l'engouement et la technologie ne cesse de s'améliorer.

Traitement du langage naturel (TLN)

Le traitement du langage naturel (TLN, ou NLP en anglais) est la capacité pour un programme informatique de comprendre le langage humain tel qu'il est parlé. Il fait partie des technologies d'[intelligence artificielle](#).

Le développement d'applications TLN est difficile parce que traditionnellement les ordinateurs sont conçus pour que les humains leur « parlent » dans un langage de programmation précis, sans ambiguïté et extrêmement structuré, ou à l'aide d'un nombre limité de commandes vocales clairement énoncées. Or le discours humain n'est pas toujours précis, il est souvent ambigu et sa structure linguistique peut dépendre d'un grand nombre de variables complexes, notamment l'argot, les dialectes régionaux et le contexte social.

Utilisations du traitement du langage naturel

La plupart des études menées sur le traitement du langage naturel tournent autour de la fonctionnalité de recherche, et plus particulièrement celle utilisée en entreprise. Il s'agit de permettre aux utilisateurs d'interroger des ensembles de données sous la forme d'une question qu'ils pourraient poser à une autre personne. La machine interprète les éléments importants de la phrase en langage humain, par exemple ceux qui peuvent correspondre à des fonctions spécifiques dans un ensemble de données, et renvoie une réponse.

Le traitement du langage naturel peut être utilisé pour interpréter du texte libre et le rendre analysable. Une énorme quantité d'informations est stockée dans des fichiers de texte libre, par exemple les dossiers médicaux des patients. Avant les modèles TLN reposant sur l'[apprentissage profond](#), l'analyse assistée par ordinateur ne pouvait pas accéder à ces informations qu'il n'était pas possible d'analyser de manière systématique, quelle que soit la méthode. Mais le TLN permet aux analystes de passer au crible des quantités massives de texte libre afin de chercher des informations pertinentes dans les fichiers.

Cette vidéo explique comment utiliser l'apprentissage profond pour élaborer des modèles TLN. <https://www.youtube.com/watch?v=lo0VfObzntA>

L'analyse des sentiments fait également partie des principaux cas d'utilisation du TLN. Elle permet aux [data-scientists](#) d'évaluer les commentaires publiés sur les réseaux sociaux pour voir comment se comporte la marque de leur entreprise, par exemple, ou d'examiner les notes des équipes du service clients pour identifier les domaines dans lesquels les clients souhaitent que l'entreprise apporte des améliorations.

La technologie de traduction automatique de Google et d'autres moteurs de recherche repose sur des modèles d'apprentissage profond de TLN. Elle permet aux [algorithmes](#) de lire du texte sur une page Web, d'interpréter sa signification et de le traduire dans une autre langue.

Fonctionnement du traitement du langage naturel

Les approches actuelles du TLN sont basées sur l'apprentissage profond, un type d'intelligence artificielle qui examine des structures de données et les utilise pour améliorer la compréhension d'un programme. Les modèles d'apprentissage profond exigent d'énormes volumes de données étiquetées pour apprendre et identifier les corrélations pertinentes, et l'assemblage de ce type de [Big Data](#) est actuellement l'un des principaux obstacles que rencontre le TLN.

Les approches précédentes du TLN étaient plus basées sur des règles, et consistaient à enseigner à des [algorithmes d'apprentissage automatique \(statistique\)](#) plus simples les mots et expressions à rechercher dans le texte, des réponses spécifiques étant générées lorsque ces expressions étaient trouvées. Mais l'apprentissage profond constitue une approche plus flexible, plus intuitive, dans laquelle les algorithmes apprennent à identifier l'intention du locuteur grâce à de nombreux exemples, un peu comme un enfant qui apprend à parler.

Importance du TLN

Les avantages du traitement du langage naturel peuvent être démontrés avec les deux phrases suivantes : « L'assurance cloud computing doit faire partie de chaque contrat de niveau de service » et « Avec un bon SLA, vous dormirez mieux... même dans le cloud ». Si vous utilisez le traitement du langage naturel pour lancer une recherche, le programme reconnaîtra cloud computing comme étant une entité, cloud comme une forme abrégée de [cloud computing](#) et [SLA](#) comme un acronyme désignant un contrat de niveau de service (« Service Level Agreement » en anglais).

Les éléments indistincts de ce type, qui apparaissent fréquemment dans le langage humain, ont longtemps posé des problèmes d'interprétation aux algorithmes d'apprentissage automatique. Aujourd'hui, grâce aux progrès de l'apprentissage profond et de l'intelligence artificielle, les algorithmes peuvent les interpréter efficacement.

Cette situation a des répercussions sur les types de données analysables. De plus en plus d'informations sont créées en ligne chaque jour, dont une bonne partie en langage humain naturel. Jusqu'à récemment, les entreprises ne parvenaient pas à analyser ces données. Mais les avancées du traitement du langage naturel permettent désormais d'analyser et d'assimiler les données provenant d'une plus vaste palette de sources.