L’intelligence artificielle : enjeux et débats

Introduction :

Les ordinateurs modernes sont tellement puissants qu’ils peuvent désormais faire tourner des programmes très complexes visant à simuler certains aspects de l’intelligence humaine. On appelle intelligence artificielle (IA) l’ensemble de ces théories et techniques mises en œuvre en vue de réaliser des machines capables de simuler l’intelligence. Cette discipline s’est développée de manière fulgurante ces dernières années. Ses applications se multiplient et s’étendent à tous les pans de la société.

Nous présenterons dans une première partie l’essor du *big data*, qui a servi de socle au développement de l’intelligence artificielle. Puis nous étudierons l’apprentissage automatique réalisé par les outils d’intelligence artificielle, avant d’aborder dans une troisième partie les enjeux éthiques et sociétaux posés par ces nouvelles technologies.

1 | *Big data*

L’expression ***big data*** désigne la récente explosion du nombre de données générées et exploitées informatiquement depuis plusieurs années.

Définition

Big data **:**

Données de tous types, générées en grandes quantités, souvent en temps réel, à partir des comportements des utilisateurs d’outils et de services informatiques.

a. Données massives

Désignées en français par **« données massives »** ou **« mégadonnées »**, les informations disponibles sous forme numérique ne cessent de se développer.
De nombreuses entreprises spécialisées se sont développées pour traiter, stocker, analyser et valoriser ces données. Le traitement de ces données massives est un processus largement automatisé.

À retenir

Trois propriétés caractérisent historiquement les données massives :

- le **volume**, une quantité massive de données ;
- la **variété**, ces données peuvent être des signaux, des textes, des images, des sons, des vidéos, dans différents formats ;
- la **vélocité**, la vitesse de génération, de collecte, de stockage, de misa à jour et de diffusion des données, de plus en plus en temps réel.

Les coûts de stockage informatique des données ont grandement diminué avec l'essor de l'économie numérique, permettant de conserver des volumes toujours plus importants de données.
Les données audio et surtout vidéo, omniprésentes aujourd'hui, requièrent des volumes de stockage bien plus importants que les données essentiellement textuelles de l'informatique traditionnelle.

Exemple

Un espace de stockage de $1\,\mathrm{Go}$ permet de stocker :

- plus de 20 fois les œuvres complètes de Victor Hugo (plus de 20 000 pages de texte) ;
- environ 200 photos de 12 mégapixels au format JPEG ou 15 à 30 environ au format RAW ;
- quelques minutes seulement de vidéo HD.

→ Les internautes étant à la fois consommateurs et producteurs de contenus et de données, ils contribuent par leurs publications et leurs navigations sur Internet à l'essor des données massives.

 Exemple

> Ce sont environ 300 heures de vidéo qui sont mises en ligne chaque minute sur le site YouTube.

L'essor du *big data* a été accompagné par le développement du ***cloud***, l'**informatique en nuage**. Des ressources informatiques colossales sont proposées en ligne pour le stockage ou le traitement de données.
La puissance de calcul et le volume de stockage sont aisément modulables en quelques clics pour accompagner une montée en charge ou en volume au gré des besoins, et permettre si nécessaire des traitements en temps réel par les entreprises réalisant l'analyse de ces données.

→ Avec le *cloud*, les entreprises n'ont donc plus besoin de posséder ou d'exploiter des serveurs physiques. Des serveurs virtuels sont à leur disposition pour effectuer les nombreux calculs d'analyse. Cela permet d'acheter ou plus exactement de louer de la puissance à la demande, parfois très ponctuellement, pour absorber un pic de calculs. Cette puissance est disponible et facturée de manière élastique et en fonction de la puissance consommée.

Les comportements et les cheminements des internautes sur le web sont analysés dans les moindres détails par une myriade d'entreprises et de services spécialisés.

Cette masse de données est organisée en ensembles cohérents dotés de caractéristiques communes, appelés **jeux de données** (*datasets* en anglais). Ils peuvent être analysés, mais aussi servir de matière première pour nourrir des algorithmes d'apprentissage automatique, comme nous le verrons dans la partie suivante du cours.

Des jeux de données sont constitués dans tous les domaines d'activité.

 Exemple

- Les chaînes industrielles collectent les données issues de capteurs présents sur les chaînes de production.
- Les laboratoires d'analyse ADN analysent et recoupent les données génétiques de toutes sortes.
- Les centres commerciaux collectent des informations sur leurs visiteurs et analysent leurs comportements.

La collecte n'est que la première étape du processus de traitement des données.

b. Besoins massifs d'analyses

Les entreprises spécialisées dans l'analyse des données ont pleinement exploité les opportunités offertes par la disponibilité croissante de jeux de données toujours plus nombreux et toujours plus complets.
Ce travail d'analyse est facilité par la souplesse et la possibilité de changer d'échelle offertes par l'informatique en nuage.

À retenir

La surabondance de données disponibles est telle depuis plusieurs années que l'aide de la machine s'est vite avérée indispensable pour leur analyse.
Dans le même temps, l'augmentation de la puissance de calcul informatique et la baisse de son coût ont rendu possible la mise au point et l'exploitation en production de modèles d'intelligence artificielle à base de **réseaux de neurones artificiels**.

Conçus dès les années 1980, ces réseaux de neurones étaient, à l'époque, impossibles à mettre en œuvre avec la puissance informatique disponible.

Le *big data* du XXI^{e} siècle change la donne et ouvre la voie à un développement spectaculaire de l'intelligence artificielle, rapidement mise à contribution pour analyser différemment la vaste quantité de données devenues disponibles.

La discipline connait depuis plusieurs années une croissance très rapide, tant sur le plan de la recherche que sur les applications industrielles.

Avec l'intelligence artificielle, le traitement et l'usage des données n'est cependant pas identique à celui effectué en programmation informatique traditionnelle.

2 Apprentissage automatique

Intéressons-nous maintenant à l'**apprentissage automatique par la machine**.

Définition

> **Apprentissage machine ou « apprentissage automatique » :**
>
> L'apprentissage machine consiste à conférer aux ordinateurs la capacité d'apprendre à partir des données.

→ L'intelligence artificielle combine des bases mathématiques, notamment probabilistes et statistiques, et des bases informatiques, parmi lesquelles les réseaux de neurones artificiels.

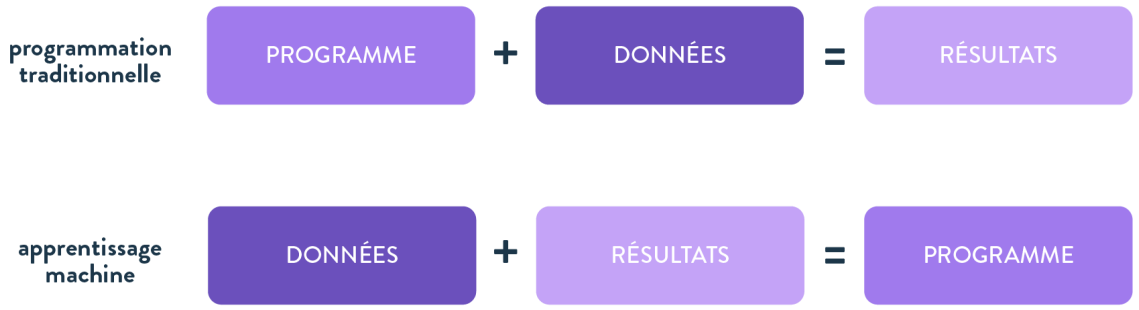
À l'instar des neurones biologiques dont ils sont inspirés, les neurones artificiels sont capables d'adapter leur configuration pour effectuer un apprentissage et fournir une meilleure réponse les fois suivantes.

a. Paradigme d'apprentissage

La programmation informatique traditionnelle consiste généralement à définir très précisément le comportement attendu du programme en fonction de règles définies dans les moindres détails par le développeur sous forme de code.

En intelligence artificielle, le processus est différent et peut sembler déroutant : la machine apprend à partir des données.

Comparaison des paradigmes d'apprentissage



Cette approche est applicable à tout type de données : textes, images, sons, vidéos, etc.
L'analyse des données permet ensuite de proposer des outils de classification, de traitement du langage naturel, de reconnaissance vocale, ou encore de génération automatique de sous-titres, pour ne citer que quelques-unes des nombreuses applications concrètes de l'intelligence artificielle.

À retenir

> Le développeur d'une solution d'intelligence artificielle définit globalement les modalités d'apprentissage et le résultat attendu, et la machine recherche par elle-même le moyen de restituer au mieux ce qu'elle apprend progressivement à partir des données au terme du processus d'apprentissage.

b. Processus d'apprentissage

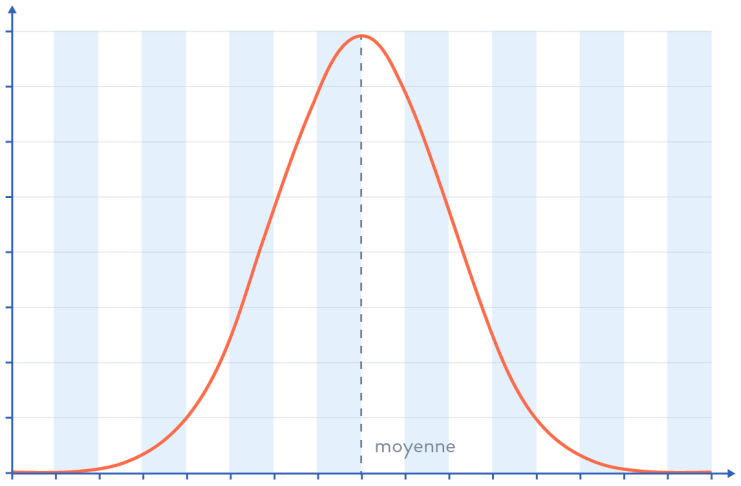
Il existe différentes modalités d'apprentissage par la machine.
Nous nous intéressons plus particulièrement à l'**apprentissage supervisé** : il repose sur des données préalablement annotées ou étiquetées par des humains. Cela signifie que les données à analyser (texte, image, son, etc.) sont fournies à la machine avec les résultats attendus.

Cet apprentissage est progressif, et fonctionne par essais et corrections d'erreurs. Au départ, le **modèle** ne connaît aucune donnée et ne peut compter que sur le hasard pour tenter de deviner les bonnes réponses, mais il est conçu pour s'ajuster progressivement afin correspondre de mieux en mieux aux données au fur à mesure de leur exploration.

Définition

Modèle :

Un modèle est une représentation mathématique d'une donnée ou d'un comportement.
Par exemple, on cite souvent le modèle gaussien pour représenter une donnée.



À retenir

La base de cet apprentissage est l'**inférence bayésienne**, une méthode de calcul consistant à ajuster un modèle basé sur des probabilités extraites à partir de données connues.

Définition

> **Inférence bayésienne :**
>
> L'inférence bayésienne est un procédé appliquant à des données inconnues un calcul probabiliste basé sur des données connues.

Le processus d'apprentissage s'effectue par itérations (c'est-à-dire par répétitions successives) sur des données découpées en lots, avec ajustements progressifs des prédictions effectuées par le modèle pour en améliorer la fiabilité.

Les données annotées utilisées pour l'entraînement du modèle sont réparties en deux catégories :

- les **données d'apprentissage** ;
- les **données de validation**.

Les proportions peuvent varier, mais sont généralement de l'ordre de 80 % des données utilisées pour l'apprentissage et 20 % pour la validation de la fiabilité du modèle.

En s'entraînant avec suffisamment de données et des données suffisamment variées pour être représentatives de celles qu'il aura à traiter une fois mis en production, le modèle pourra fournir des performances de bon niveau en termes de fiabilité.

Les données de validation permettent de mesurer la fiabilité du modèle et de déterminer quand il est suffisamment bien entraîné pour pouvoir être utilisé en production.

C. Matrice de confusion

L'évaluation du modèle consiste à vérifier s'il produit bien le comportement attendu et si ses prédictions peuvent être considérées comme suffisamment fiables.

Il se peut en effet que le modèle commette des erreurs, et celles-ci peuvent être de différents types.
Nous l'illustrons de manière concrète avec un exemple de modèle entraîné

pour détecter une tumeur maligne, c'est-à-dire un cancer, à partir de données d'imagerie médicale.

On attend du modèle qu'il analyse les données et qu'il propose une classification binaire de l'image indiquant si la tumeur proposée est maligne ou seulement bénigne.

Le modèle peut prédire correctement la classification de la tumeur dans deux cas de figure :

- tumeur maligne correctement classifiée comme telle ;
- tumeur bénigne correctement classifiée comme telle.

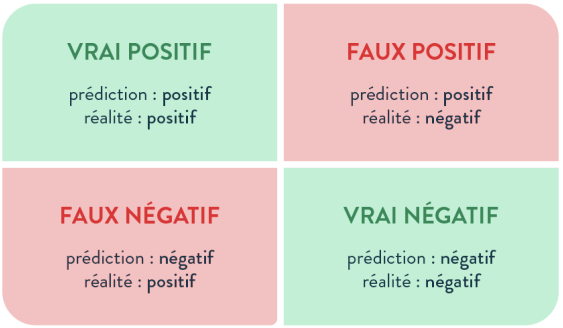
Le modèle peut aussi effectuer une classification incorrecte de la tumeur dans deux cas de figure :

- tumeur maligne incorrectement classifiée comme bénigne ;
- tumeur bénigne correctement classifiée comme maligne.

→ On parle de **vrais positifs** et de **vrais négatifs** quand le modèle effectue des prédictions correctes, et de **faux positifs** et **faux négatifs** quand le modèle effectue des prédictions erronées.

L'ensemble de ces types de prédictions peut être regroupé sous forme d'une matrice appelée **matrice de confusion**.

Matrice de confusion



Les prévisions correctes du modèle sont en vert :

- les vrais positifs sont les cas où la tumeur était maligne et a correctement été classifiée comme telle ;
- les vrais négatifs sont les cas où la tumeur était bénigne et a correctement été classifiée comme telle.

Les prévisions incorrectes du modèle sont en rouge :

- les faux positifs sont les cas où la tumeur était bénigne mais a été incorrectement classifiée comme maligne ;
- les faux négatifs sont les cas où la tumeur était maligne mais a été incorrectement classifiée comme bénigne.

→ Chaque type d'erreur peut avoir des conséquences sérieuses, et il est important d'évaluer la performance globale du modèle et de poursuivre son entraînement jusqu'à obtenir un résultat satisfaisant.

⚠ Attention

> On rappellera aussi que la machine apporte une aide au diagnostic, mais n'a pas vocation, du moins à court terme dans le domaine médical, à

remplacer le diagnostic du médecin.

Toutefois les intelligences artificielles sont de plus en plus utilisées pour toutes sortes d'applications, ce qui n'est pas sans poser des questions d'éthique.

3 Aspects éthiques

L'usage croissant de l'intelligence artificielle pose un certain nombre de questions et constitue un enjeu non négligeable pour l'avenir de la société.

Les progrès technologiques fulgurants de cette discipline sont souvent plus rapides que la réflexion et la législation.

a. Caractérisation des enjeux

Les enjeux liés à l'intelligence artificielle sont multiples.

À retenir

Quel que soit le type d'activité, le fait de déléguer des contrôles, des analyses et même des décisions à un programme informatique n'est pas anodin. Cela constitue une **forme d'externalisation de l'intelligence humaine**, bien que limitée à une tâche précise.

Les enjeux sont variables en fonction de la tâche dévolue à une intelligence artificielle : l'enjeu est limité si elle a trait à un jeu vidéo (exemple : modalités de déplacement d'un personnage). Il peut entraîner des conséquences beaucoup plus graves si une intelligence artificielle décide de recommander l'arrestation d'une personne par des forces de police.

Exemple

En janvier 2020, un Américain a été arrêté et détenu pendant trente heures par la police de Détroit sur la base d'une ressemblance déterminée par une intelligence artificielle. Le programme avait établi une correspondance erronée entre la photo du permis de conduire de

> Robert Williams et celle d'un voleur captée par caméras de vidéo-surveillance sur le lieu d'un cambriolage.

Les entreprises spécialisées rivalisent de solutions pour la surveillance permanente de l'espace public par différents dispositifs de captation et d'analyse des visages, des silhouettes, des bruits, des plaques d'immatriculation, du trafic automobile ou encore des mouvements de foules.

D'ambitieux programmes à base d'intelligence artificielle sont également développés dans les domaines de la finance ou encore de la santé, exploitant le volume sans cesse croissant de données disponibles et de possibilités d'analyses de celles-ci.

Ces entreprises assurent généralement que leurs modèles sont performants et fiables, mais ces affirmations ont été régulièrement démenties par des tests indépendants et des cas d'erreurs médiatisés. Ainsi, la **fiabilité des modèles** peut parfois être remise en question.
Par ailleurs, d'autres problématiques peuvent aussi émerger lorsque la collecte et l'utilisation des données ne sont pas faites dans le respect de la vie privée. Cela soulève la question de la **manipulation des données** : si elle n'est pas correctement encadrée, cette manipulation pourrait peser sur les libertés individuelles, notamment si sa généralisation se fait dans une logique de contrôle des citoyens (exemple : suivi par reconnaissance faciale des individus).

→ Des voix se font entendre pour appeler à un encadrement de ces usages. Des appels à **moratoire** ou à **bannissement** ont été lancés dans plusieurs pays, en particulier pour la **reconnaissance faciale**.

La qualité et la représentativité des données utilisées pour entraîner les intelligences artificielles conditionnent la capacité du modèle entraîné à traiter les cas qui lui sont soumis avec de bonnes performances sur l'ensemble du domaine considéré.

b. Qualité des données d'apprentissage

Les jeux de données servant à développer les modèles d'intelligence artificielle doivent être annotés, et cette annotation est effectuée par des

humains. L'étiquetage des données est souvent sous-traité, *via* des plateformes spécialisées de micro-tâches en télétravail, à des petites mains invisibles de l'économie numérique, rémunérées quelques centimes pour chaque micro-tâche.

 Exemple

> Pour être capable d'analyser correctement une addition de restaurant, une intelligence artificielle aura besoin de données annotées et étiquetées selon certaines catégories définies par les concepteurs du programme.
> Il est nécessaire de disposer d'annotations en nombre suffisant pour que le modèle soit capable de traiter correctement toutes sortes d'additions, lesquelles peuvent être assez différentes d'un restaurant à l'autre.
> Quelques clics et un budget relativement contenu permettent de sous-traiter ce travail à de nombreux micro-travailleurs du numérique qui vont devoir multiplier les annotations à rythme souvent soutenu.

Au-delà des implications potentielles en termes de qualité, on peut s'interroger en termes d'éthique par rapport au niveau de rémunération de ces tâches, souvent peu payées alors qu'elles constituent pourtant le socle de « l'intelligence » du système.

c. Représentativité des données d'apprentissage

Des biais ont été constatés dans plusieurs outils d'intelligence artificielle.

 Définition

> **Biais :**
>
> Un biais est une distorsion des données en faveur ou à l'encontre de la valeur d'un paramètre.

 À retenir

> De nombreuses intelligences artificielles ont été entraînées avec des données d'apprentissage comportant des biais.

Ces intelligences artificielles reproduisent et amplifient parfois les biais avec lesquels elles ont été entraînées.

 Exemple

En 2016, Microsoft lançait un agent conversationnel (*chatbot* en anglais) nommé Tay qui publiait des contenus de manière autonome sur le réseau social Twitter.
Microsoft a cependant été contraint de désactiver très rapidement cet agent qui imitait le caractère polémique et même raciste de certains internautes.
Le biais ici est donc créé par le fait que l'algorithme s'est entraîné sur des données constituées de messages présentant majoritairement un caractère polémique.

La présence de biais est susceptible de générer des traitements différenciés.

 Exemple

Un algorithme entraîné pour analyser les CV de candidats à partir de recrutements antérieurs avait tendance à reproduire, à compétences égales, une discrimination envers les femmes par rapport aux hommes.

Les intelligences artificielles reproduisent les biais de leurs concepteurs ou des annotateurs de données.

 Exemple

- En 2017, l'intelligence artificielle de traduction automatique mise au point par Google exprimait des stéréotypes de genres. Quand on lui demandait de traduire depuis une langue non genrée comme le turc, où le pronom « *o* » signifie aussi bien « un » que « une », cette intelligence artificielle traduisait spontanément « un docteur » et « une infirmière ».
- En 2018, une association américaine de défense des libertés démontrait qu'une intelligence artificielle de reconnaissance faciale développée par

> Amazon et proposée aux forces de l'ordre identifiait par erreur une trentaine de membres du Congrès américain comme des délinquants recherchés, avec un taux d'erreur particulièrement élevé pour les personnes de couleur parmi eux.

Si on n'y prend pas garde, les intelligences artificielles répètent et amplifient parfois nos biais de société.

Conclusion :

Le développement de l'intelligence artificielle s'est appuyé sur les ressources du *big data*, offrant des possibilités inédites d'analyses de données, dont la quantité et la diversité est sans cesse croissante. Les machines ont appris à apprendre à partir de jeux de données et sont capables de corriger leurs erreurs pour proposer de meilleures performances.

Les enjeux liés à l'essor de l'intelligence artificielle sont majeurs pour l'avenir de notre société et suscitent de nombreux débats, notamment au sujet des usages policiers ou militaires, mais aussi à l'encontre des risques de surveillance généralisée des faits et gestes des citoyens. Au-delà des questions éthiques de certains usages, nous avons souligné l'existence récurrente de biais dans les solutions d'intelligence artificielle, pouvant entretenir ou renforcer des inégalités et discriminations. Il est donc essentiel de veiller à faire un usage raisonnable des technologies d'intelligence artificielle afin d'y remédier.