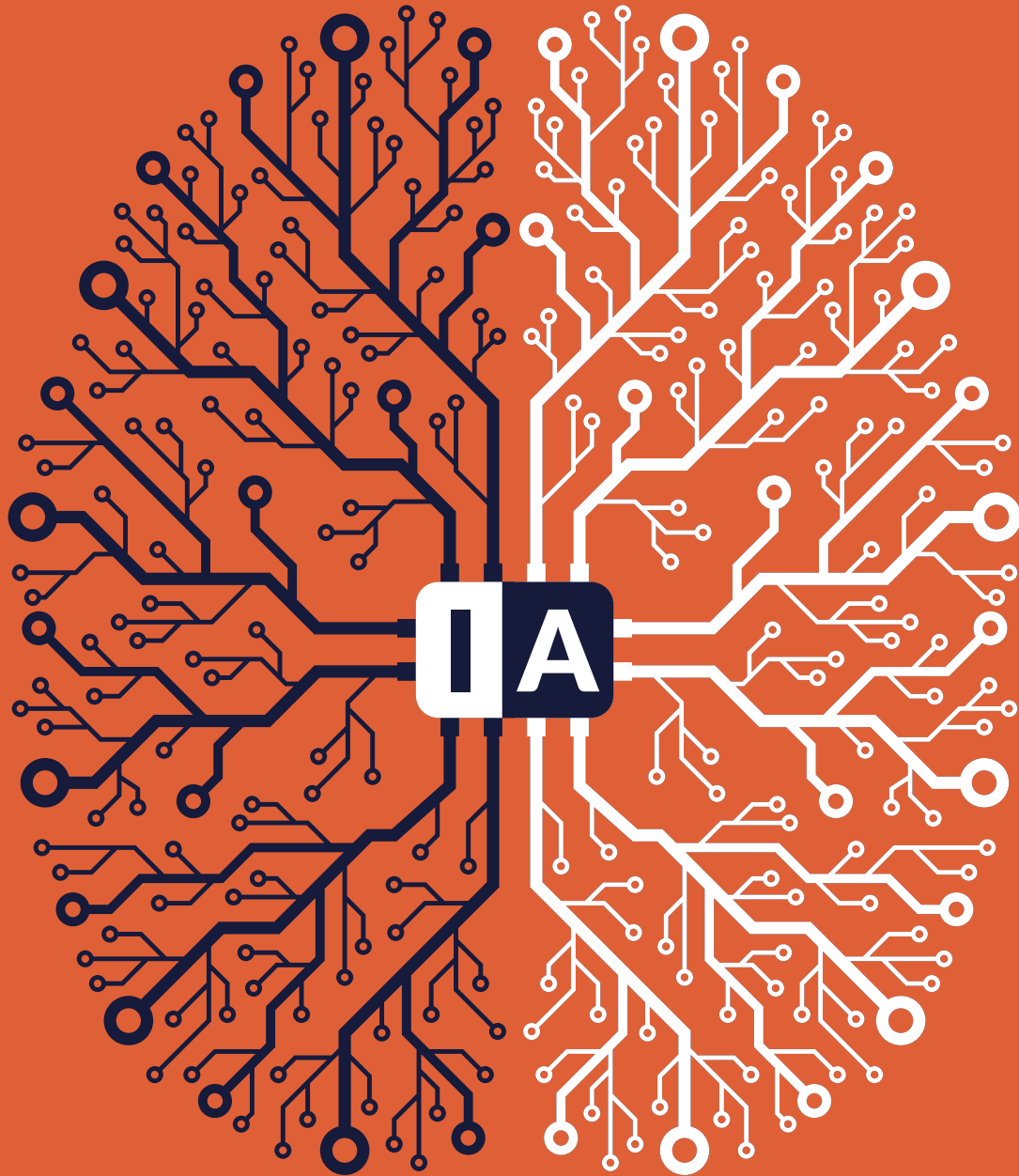




INTELLIGENCE ARTIFICIELLE

Restez maître de votre futur !

Avant-propos

S'il est un sujet qui convoque les passions et nourrit toutes les réflexions dans le monde des nouvelles technologies, c'est bien l'intelligence artificielle. Le sigle IA s'est définitivement installé dans le langage. Au-delà de l'état de l'art technologique, les enjeux sont immenses et principalement sociétaux.

Les fervents soutiendront que c'est l'avancée technologique majeure depuis l'invention de la machine à vapeur, les détracteurs évoqueront la fin des libertés individuelles et collectives.

Ces points de vue extrêmes et sans nuance créent, insidieusement, une sorte de mystère autour de son origine, de son utilisation et de ses apports.

En tant qu'expert du domaine, Business & Decision apporte légitimement son éclairage en matière d'intelligence artificielle. En témoignant des expériences ancrées dans le réel, Business & Decision ambitionne d'associer ses valeurs d'entreprise à la pratique concrète de ces nouvelles technologies.

C'est tout l'objectif de ce livre blanc.

Qui dit IA, dit data, carburant de l'IA. C'est en faisant preuve de pédagogie et de transparence que les craintes et les doutes peuvent être levés quant à l'exploitation des données et à l'usage des algorithmes.

L'IA comporte des risques, indiscutablement. Notre rôle est de les éclairer, de les étayer et d'alerter si nécessaire. Mais nous ne devons jamais perdre de vue que l'IA est un outil qui doit rester au service de l'humain. Nous portons la conviction que les risques peuvent être évalués et les écueils évités, en particulier pour les organisations, lorsque les principes de responsabilité, d'éthique et d'inclusion font partie intégrante des projets d'IA.

Les bénéfices de l'intelligence artificielle sont nombreux et encore trop peu mis en lumière. En matière de santé, d'écologie et d'empreinte carbone, de politiques de la ville, d'industries et de services, ce vecteur d'innovation sans précédent ouvre des perspectives d'amélioration pour tout un chacun. À condition que les usages se réalisent en toute transparence et suscitent la confiance.

Sur le long chemin d'apprentissage de tout ce que l'intelligence artificielle peut potentiellement offrir, nous partageons, avec nos clients, partenaires et experts, nos expériences et nos connaissances. Nous nous enrichissons de nos doutes et parfois de nos erreurs. Mais nous ne perdrons jamais de vue les principes et les valeurs qui guident notre action.

Valérie Lafdal,
Directrice Générale Business & Decision France
Directrice Générale déléguée Groupe Business & Decision

1

Une définition de l'intelligence artificielle

Naissance de l'intelligence artificielle _____ P. 5

Définition de l'IA _____ P. 8

2

AI : Apprentissage Intensif

L'environnement d'une IA _____ P. 11

Les modes d'apprentissage _____ P. 14

L'IA demain : où en est la recherche ? _____ P. 25

3

Les enjeux de l'IA

Pourquoi les entreprises doivent y aller _____ P. 28

Développer une IA digne de confiance _____ P. 30

Évaluer votre maturité face à l'IA _____ P. 32

Intégrer et travailler avec l'IA dans l'entreprise _____ P. 34

Conclusion _____ P. 38

1 —

Une définition de l'intelligence artificielle

Une définition de l'intelligence artificielle

En 2020, le métier de spécialiste en intelligence artificielle (IA*) occupait la première place du classement LinkedIn des jobs émergents aux États-Unis. En France, la fonction d'ingénieur en intelligence artificielle se classe quant à elle en deuxième position des métiers les plus recherchés¹. Les recrutements de spécialistes IA ont augmenté de 74 % au cours des quatre dernières années. Une tendance qui n'est pas près de s'inverser : les dépenses mondiales devraient en effet doubler au cours des quatre prochaines années pour dépasser les 110 milliards de dollars en 2024². En Europe, c'est la France qui occupe la première place des pays investisseurs depuis 2019³. Mais pourquoi l'intelligence artificielle, termes apparus dès les années 50, suscite-t-elle aujourd'hui un tel engouement ? Retour sur quelques principes fondateurs.

LES RECRUTEMENTS
DE SPÉCIALISTES IA
ONT AUGMENTÉ
DE 74%
AU COURS DES
4 DERNIÈRES ANNÉES.

Naissance de l'intelligence artificielle

Quand on évoque l'intelligence artificielle, de quoi parle-t-on réellement ? Selon la définition de l'Encyclopédie Larousse⁴, il s'agit d'un « *ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine* ». Entre mythes et réalités, il est nécessaire de retracer l'histoire de l'IA pour mieux en comprendre les enjeux au XXI^e siècle.

Le contexte des années 1940-1950

Dès le milieu des années 1940, les neurologues Warren McCulloch⁵ et Walter Pitts⁶ publient leurs premiers essais sur ce qu'ils considèrent, à l'époque, comme une modélisation mathématique du cerveau.

Leurs travaux inspirent alors nombre de chercheurs comme Franck Rosenblatt⁷, John McCarthy⁸ et Marvin Minsky⁹ qui ambitionnaient de modéliser et de reproduire électroniquement le fonctionnement du cerveau. À l'époque, à la vue des premiers résultats d'apprentissage obtenus avec les neurones artificiels (appelés plus tard « perceptrons » par Rosenblatt), les chercheurs se persuadent alors dès la fin des années 1950 que la mise « en réseau » de ces neurones électroniques permettrait, à l'instar du fonctionnement du cerveau, de résoudre des problèmes d'une complexité infinie.

Pourquoi le terme « intelligence artificielle » ?

À l'époque, afin de financer toutes leurs recherches, McCarthy et Minsky eurent l'idée de créer le terme « intelligence artificielle » qui fut présenté pour la première fois à la conférence de Dartmouth (Dartmouth Summer Research Project on Artificial Intelligence) dans le courant de l'été 1956. L'IA fut alors non seulement créée officiellement en tant que discipline académique, mais elle suscita également un fort intérêt auprès des investisseurs américains.

Un faux départ ?

McCarthy et Minsky obtinrent, à la suite de cette conférence de Dartmouth, des financements très conséquents. Mais l'euphorie ne durera que jusqu'à la fin des années 1960, années à partir desquelles la recherche dans le domaine rentrera progressivement dans ce qu'on appelle « le premier hiver de l'IA ». La raison : les résultats étaient en effet à l'époque extrêmement décevants par rapport aux promesses initialement faites à Dartmouth.

À cela, plusieurs explications : outre la puissance de calcul qui n'était pas au rendez-vous, l'IA a montré des restrictions directement liées aux limites intrinsèques qu'une machine peut avoir face à l'intelligence humaine, comme l'explique Luc Julia (voir l'encadré p7). Les explications à cet excès d'ambition sont maintenant un peu mieux cernées et il est important, pour ne pas retomber dans un nouvel hiver de l'IA, de bien les comprendre.

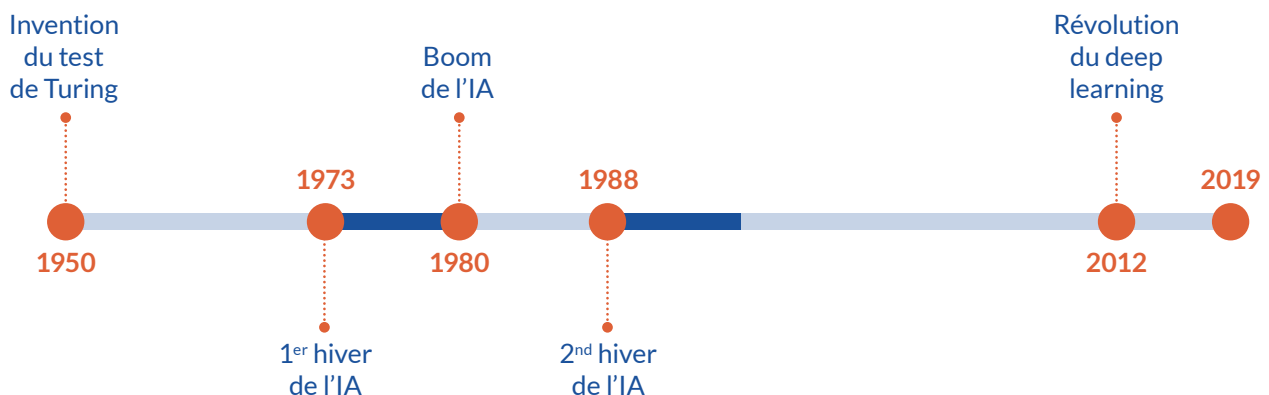
Tout d'abord, la modélisation mathématique de départ était certainement bien trop simpliste. Plus la recherche sur le cerveau et les neurosciences avance, plus elle fait apparaître son lot de complexités inattendues et de questions sans réponse. Et quand bien même on parviendrait à comprendre le fonctionnement complet de la structure, ce qui

est très loin d'être le cas, il nous manque encore le plus important : la compréhension de la fonction, c'est-à-dire de manière extrêmement vulgarisée, « le software » qui le fait fonctionner.

Pour illustrer ce point, il est possible de prendre à titre d'exemple l'hypothèse, de plus en plus admise par certains chercheurs en biologie, que des millions d'années d'évolution de la vie seraient inscrites dans chacune de nos cellules, y compris et en premier lieu dans nos neurones. Vous conviendrez aisément que si tel est effectivement le cas, il semble alors impossible de la modéliser avec une quantité finie de mémoire. Voilà pourquoi de nombreux chercheurs en IA se sont récemment éloignés du concept de départ de McCarthy et de Minsky, bien que le nom « intelligence artificielle » soit quant à lui resté.

Après une phase durant laquelle les systèmes experts* ont dominé les années 80, l'IA a connu une autre crise – appelée « second hiver de l'IA » – entre 1987 et 1993, liée à la perception que les investisseurs et les agences gouvernementales en avaient. Il faudra attendre les années 2000, et plus particulièrement 2012, pour que le Deep Learning* (apprentissage profond, voir glossaire) redonne une véritable impulsion à la discipline.

LES ANNÉES
1970'S
MARQUENT
LE 1^{er} HIVER
DE L'IA.



L'avis de l'expert

Luc Julia,
CTO de Samsung Electronics,
cocréateur de Siri et auteur de « L'intelligence artificielle n'existe pas »
First Éditions - 2019

Prenons le cas de la voiture autonome : l'autonomie de niveau 5 (conduite complètement autonome sans l'aide de conducteur dans toutes les circonstances, ndlr) n'existera jamais. Pourquoi ? Parce qu'il y a toujours une situation imprévue à laquelle l'humain pourra s'adapter alors que la machine, elle, en sera incapable. Une voiture autonome sur la Place de l'Étoile à Paris à 18h n'avancera pas parce que le code de la route n'y est pas appliqué. Or la machine n'invente jamais : elle ne fait que mettre en œuvre ce qu'on lui a appris ou ce qu'elle a observé dans les modèles* qui lui ont été présentés.

Quelles différences entre l'Homme et l'IA ?

Contrairement aux idées reçues, l'intelligence artificielle n'a donc pas vocation (ni même la capacité) à remplacer l'humain. Il existe ainsi à ce jour trois différences majeures qui dissocient l'homme de l'IA :

L'inédit • L'édito • L'éthique

Voilà quelles sont les trois limites majeures de l'IA pour **Michael Deheneffe**,
directeur de la Stratégie & Innovation, Business & Decision :

« L'IA est en effet incapable de réagir face à l'inédit :

*si elle n'a pas appris, elle ne peut pas réagir. Un grain de sable qui se glisse
dans les rouages a alors la capacité de perturber tout son écosystème.*

La deuxième limite repose sur l'édito, à savoir que l'IA n'a pas de point de vue.

*L'homme comprend et crée ; l'IA apprend et exécute. Il n'y a en réalité aucune
intelligence réelle dans l'IA. Enfin, l'intelligence artificielle ignore toute notion
d'éthique ou de justice. Elle est incapable de faire la différence entre le bien et le mal,
et, en ce sens, de choisir par elle-même la meilleure approche. »*

Définition de l'IA

Au regard des différences majeures entre l'intelligence artificielle et l'humain, il apparaît donc que, à court ou à moyen terme, l'IA ne peut exister qu'en support de l'homme. Un positionnement vers lequel tend la définition moderne de l'IA.

Définition de Stuart Russel & Peter Norvig

Si on se base sur la définition parue en 2010 dans l'ouvrage Stuart Russel & Peter Norvig, "Artificial Intelligence, a modern Approach", l'intelligence artificielle est une discipline consacrée à la construction d'agents* autonomes qui sont des entités (matérielles et/ou logicielles), dotés d'un certain degré d'autonomie, capables de percevoir leur environnement (grâce à des capteurs) et d'agir sur celui-ci (via des actionneurs). Ils peuvent également analyser et prendre des décisions en vue d'atteindre des objectifs.

L'ouvrage de Stuart Russel & Peter Norvig précise que l'IA ainsi définie évolue dans un environnement déterminé, ce qui est aussi parfois défini comme une « IA contextuelle » ou une « IA faible » par rapport au concept « d'IA généraliste » ou « d'IA forte ». En réalité, la notion d'IA forte n'a pas réellement de sens, au moins dans l'état actuel de la recherche, puisqu'une IA, comme nous l'avons montré précédemment, ne sait pas faire abstraction de son environnement. Changer une intelligence artificielle d'environnement implique de devoir refaire à minima une très grande partie de

son apprentissage au sein de ce nouvel environnement. Nous partirons donc dorénavant du postulat que **toute IA est contextuelle**.

L'intelligence artificielle selon Business & Decision

À la lumière de ce qui précède, Business & Decision propose la définition suivante de l'intelligence artificielle : « *Discipline consacrée à la construction d'agents **qui interagissent** avec l'environnement, peuvent **apprendre** à partir de données initiales fournies et/ou de données collectées lors des interactions pour réaliser une fonction **dans un environnement déterminé**, avec un certain degré d'autonomie.* »

En résumé, l'IA consiste à capter les données, à les analyser et à interagir avec son environnement (selon la définition de Peter Norvig). Mais nous ajoutons à cela la notion **d'interaction** et celle de capacité **d'apprentissage** dans un **environnement déterminé**. En ce sens, l'intelligence artificielle repose donc sur un triptyque : **interactions, analyse et apprentissage** favorisant la prise de décision.

« L'IA, c'est l'homme qui donne quatre compétences à la machine.

La perception tout d'abord, afin de lui permettre de reconnaître des formes, des images, du texte...

Puis l'interaction, qui peut s'exprimer sur de multiples modes, depuis le simple affichage de messages sur un écran jusqu'aux robots humanoïdes les plus évolués.

Ces deux premières compétences permettent à l'IA de "vivre" dans le monde réel.

Les deux autres compétences sont relatives au domaine cognitif. L'analyse tout d'abord, qui se décline sur trois niveaux : descriptif, prédictif et prescriptif. Et enfin, l'apprentissage.

Cette dernière compétence est fondamentale car elle est à la base de toutes les autres.

L'IA ne sait en réalité faire qu'une seule chose : apprendre. »

Mick Levy,

directeur de l'Innovation Business, Business & Decision

L'avis de l'expert

Luc Julia,

CTO de Samsung Electronics, cocréateur de Siri

et auteur de « L'intelligence artificielle n'existe pas » - First Éditions - 2019

« Le terme IA a été utilisé la première fois en 1956. Mais malheureusement, les scientifiques se sont attaqués au problème le plus difficile, problème que nous n'avons toujours pas résolu : comprendre le langage naturel* (NLP ou NLU). Résultat, les recherches ont été abandonnées dès les années 60 et nous sommes entrés dans le premier hiver de l'IA. Nous avons donc fait des systèmes experts pendant 30 ans, et ce jusqu'en 1997, année durant laquelle un système expert a été capable de battre pour la première fois Gary Kasparov aux échecs. Puis internet a permis de faire émerger d'importants volumes de données nous permettant de reprendre les recherches sur les réseaux de neurones. En 2016, le champion du monde de Go a lui aussi été battu par DeepMind de Google. Mais ce n'est pas pour autant qu'il faut comparer l'IA et l'intelligence : l'intelligence humaine est un continuum. Le cerveau est infini. Donc non, l'intelligence artificielle ne va pas remplacer l'humain. Descartes a dit : "Le langage est le propre de l'homme". Ce n'est donc pas celui de la machine ! En ce sens, l'intelligence artificielle qui fait peur, qui est censée prendre le pouvoir sur l'homme, n'existe pas ! En revanche, il y a bien une IA qui existe et sur laquelle nous travaillons depuis 30 ans. Si l'IA généraliste ou "forte" n'existe pas, les intelligences artificielles quant à elles, dans le sens des IA faibles ou contextuelles, existent. Même en utilisant les techniques mathématiques et statistiques* d'aujourd'hui, nous ne parviendrons jamais à nous approcher d'une intelligence humaine. Au mieux, nous pouvons parler d'intelligence augmentée – ou d'apprentissage intensif – à travers les outils que nous utilisons pour nous améliorer, physiquement ou intelligemment, à l'image du marteau : à l'être humain de décider s'il souhaite s'en servir à bon ou à mauvais escient... comme pour l'intelligence artificielle. »

2

—

AI : *Apprentissage Intensif*

AI : Apprentissage Intensif

Comme souligné précédemment, l'état de nos connaissances nous permet aujourd'hui de concevoir une intelligence artificielle qui ne fonctionne que dans un type d'environnement déterminé. Les différentes études menées nous conduisent à prendre en compte six caractéristiques déterminantes de l'environnement sur une IA¹⁰ ainsi que trois modes d'apprentissage.

L'ENVIRONNEMENT
D'UNE INTELLIGENCE
ARTIFICIELLE EST ESSENTIEL
CAR IL FAIT PARTIE INTÉGRANTE
DU PROCESSUS DE CONCEPTION
ET D'APPRENTISSAGE.

L'environnement d'une IA

L'environnement d'une intelligence artificielle est essentiel car il fait partie intégrante du processus de conception et d'apprentissage. En effet, l'IA évolue selon l'environnement dans lequel elle a été créée. C'est pourquoi il est important de considérer l'environnement de conception de l'intelligence artificielle pour la rendre efficace au regard des objectifs fixés. Découplée de son environnement, elle perd alors tout son sens et surtout son efficacité.

Pour construire « correctement » une IA de façon à répondre aux enjeux de l'organisation, il est alors nécessaire de prendre en compte six caractéristiques principales capables d'affecter son environnement (et par conséquent l'IA elle-même).

1 | Environnement accessible observable versus non-accessible non-observable

On dit qu'un environnement est accessible (ou observable) si **toutes les données** de l'environnement sont accessibles à l'IA, comme par exemple aux échecs ou au jeu de Go. En situation réelle, ce n'est majoritairement pas le cas, une partie de l'environnement restant souvent cachée à l'IA.

2 | Environnement déterministe versus incertain

Si l'état courant et l'action menée par l'IA permettent d'obtenir une certitude prévisible sur l'état de l'environnement qui suit l'action – comme par exemple le fait de couper un moteur arrête une machine –, on dit que l'environnement est déterministe.

Inversement, si l'action effectuée par l'IA peut conduire à plusieurs résultats possibles – tel que le freinage d'une voiture dont l'efficacité dépend de l'état de la route –, on parle alors d'un environnement incertain.

Les environnements fortement multifactoriels sont par nature incertains car la multiplication des variables au-delà de plusieurs milliers rend notamment le contrôle complet de la situation par l'IA difficile voire parfois impossible. L'algorithme devra alors faire comme si ce qu'il ne maîtrise pas était dû au hasard, ce qui impacte lourdement le mode d'apprentissage.

3 | Environnement épisodique versus séquentiel

Derrière ce concept se cache une caractéristique importante concernant la façon dont l'IA va pouvoir traiter les données et les actions. Dans un environnement épisodique, chaque action entreprise par l'IA est découpée en « mini-épisodes » qui ne pourront dépendre que de l'état actuel de l'environnement.

Autrement dit, chaque action de l'IA est un nouvel épisode résultant d'une nouvelle perception totalement indépendante des précédentes et des suivantes. L'exemple typique est celui d'une machine qui trie les bonnes pièces des mauvaises à chaque fois sans aucun lien avec les pièces précédentes. Le temps n'est alors qu'une variable comme les autres : la notion d'anticipation n'a donc pas réellement d'importance dans un environnement épisodique.

Dès qu'il peut y avoir des conséquences au-delà de l'état qui suit immédiatement la prochaine action, on dit que l'IA opère dans un environnement « séquentiel » ou « temporel ». Un grand nombre d'applications de la vie courante se situent dans un environnement séquentiel et la variable temps y joue un rôle central.

4 | Environnement statique versus dynamique

Si l'environnement dans lequel opère l'IA ne change pas au cours du temps, mais seulement du fait des actions liées aux agents qui y sont présents, on parle d'un environnement statique. Inversement, dès lors que le temps joue un rôle dans l'évolution de l'environnement, comme par exemple des objets en mouvement non contrôlés, on parle d'un environnement dynamique.

5 | Environnement discret versus continu

Il s'agit ici d'une notion proche de celle utilisée en mathématiques. Si l'environnement prend une succession dénombrable (qu'on peut numérotter) d'états, on parle d'un environnement discret.

Dans le cas inverse, on a un environnement continu. L'assiette d'un avion est par exemple typiquement un environnement continu car il y a une infinité non dénombrable d'inclinaisons possibles.

6 | Environnement multi versus mono-agent

Il s'agit simplement du nombre d'agents (au sens large de ce qui peut agir) qui peuvent influencer sur l'environnement pris en compte par l'IA. Le cas mono-agent est trivial puisque seule l'IA peut agir sur l'environnement (cas finalement assez rare).

Les cas multi-agents possèdent plusieurs variantes, selon que les agents sont en concurrence ou, au contraire, coopèrent.

L'avis de l'expert

Didier Gaultier,
Directeur AI et DataScience, Business & Decision

Les caractéristiques des environnements dans lesquels opère une IA sont finalement nombreuses, parfois imbriquées et souvent complexes. Ce qui n'est pas observé peut agir à tout moment sur l'environnement de l'IA et modifier ainsi l'effet direct de ses actions, l'environnement de l'IA tend alors à devenir incertain.

De même, dès lors qu'un environnement est séquentiel, chaque action de l'IA aura des répercussions possibles sur les suivantes. Il faudra alors doter l'IA de capacités prédictives afin de lui permettre d'anticiper des situations pouvant se produire bien plus tard. Un exemple simple est celui d'un véhicule qui prend de la vitesse, action qui changera nécessairement le temps et la distance de freinage, notamment en cas de rencontre avec un obstacle potentiel plus tard.

La complexité du développement d'une IA dépend directement de la complexité de l'environnement : si on cumule un environnement non observable, incertain, séquentiel et dynamique – ce qui correspond à de nombreuses situations réelles –, la tâche d'une IA peut s'avérer d'une grande complexité. La technique actuelle ne permet pas de dissocier une IA de son environnement. Il est donc essentiel de savoir analyser le type d'environnement auquel on est confronté avant de décider le mode d'apprentissage et le type d'algorithme* qui devra être utilisé.

Les modes d'apprentissage

À l'instar des environnements de l'intelligence artificielle, il n'existe pas un seul et unique mode d'apprentissage. On en dénombre trois principaux possédant chacun leurs caractéristiques et leurs cas d'application.

Apprentissage supervisé

FONCTIONNEMENT

Ce mode d'apprentissage, dont la genèse remonte vraisemblablement à Carl Friedrich Gauss en 1795, représente aujourd'hui très largement le mode d'apprentissage le plus utilisé en intelligence artificielle.

On distingue deux types d'apprentissages supervisés selon le type de variables* que l'on souhaite prédire. S'il s'agit d'une ou de plusieurs variables numériques*, on parle de « régression ». Au contraire, s'il s'agit d'une variable catégorielle* (qui prend ses valeurs dans une liste prédéfinie), on parle alors de classification ou de classifieur.

La notion d'exemples annotés (on dit aussi parfois « labellisés » ou « étiquetés ») est fondamentale dans ce type d'algorithme, puisqu'il consiste à effectuer l'apprentissage du modèle à partir d'exemples dont on connaît au préalable la réponse attendue. On dit qu'ils sont annotés car on fournit la réponse à l'algorithme sous forme d'annotation sur un échantillon. En réalité, cette annotation est contenue dans la variable cible à prédire, mais **uniquement pour les éléments connus** qui servent de point de départ (dont l'ensemble forme ce qu'on appelle l'échantillon d'apprentissage). Ensuite, l'algorithme se débrouille tout seul pour prédire la valeur des éléments inconnus.

Pour un classifieur, le modèle doit prédire une ou plusieurs réponses dans des listes déterminées à l'avance. On parle alors de classification supervisée (chaque valeur possible d'une réponse correspondant alors à une classe prédéterminée). Ses applications très larges vont de la prédiction d'une simple valeur numérique à des opérations de robotique d'une extrême complexité.

CAS D'USAGE

Le cas de l'optimisation portuaire est un des exemples les plus parlants, voire même spectaculaires, de l'apprentissage supervisé. Nous l'avons notamment mis en œuvre pour l'un des plus grands ports de marchandise européens.

En effet, le chargement et le déchargement des porte-conteneurs nécessitent une organisation et une logistique sans faille. Si l'opérateur portuaire ne termine pas dans les temps ces opérations (par exemple, si le porte-conteneur a dû passer par des files d'attente non prévues), le port sera soumis à de multiples pénalités, sans oublier le fait que les retards ont tendance à se propager d'un chargement (ou déchargement) à un autre. Le total de ces pénalités peut s'avérer astronomique si la logistique du port est sans cesse retardée.

Or il est toujours difficile de prévoir l'heure exacte d'arrivée (*Estimated Time of Arrival*, ETA) d'un navire à quai. De multiples facteurs comme la météo, l'encombrement des routes, des files d'attente mal gérées, ou tout simplement une absence de prévision peuvent complètement désorganiser la logistique portuaire. C'est là que l'intelligence artificielle intervient.

Le projet réalisé par Business & Decision consiste à utiliser un algorithme supervisé qui effectue dans une première étape une prédiction précise de l'ETA. Les données utilisées sont évidemment très nombreuses telles que la météo, l'encombrement des routes, les différentes files d'attente, etc.

« L'APPRENTISSAGE SUPERVISÉ (SUPERVISED LEARNING EN ANGLAIS) EST UNE TÂCHE D'APPRENTISSAGE AUTOMATIQUE CONSISTANT À APPRENDRE UNE FONCTION DE PRÉDICTION À PARTIR D'EXEMPLES ANNOTÉS. »*

*SOURCE : WIKIPÉDIA

Les bénéfices obtenus sont eux aussi extrêmement nombreux. Parmi lesquels : une réduction drastique des retards, des pénalités multiples qui en découlent et la relation gagnant-gagnant pour l'ensemble des acteurs portuaires (le port, l'opérateur, les compagnies maritimes, les douanes...).

D'un point de vue financier, les gains apportés par l'IA dans un tel cas peuvent amplement dépasser le million d'euros par an. En conséquence, un tel projet est rentabilisé en quelques trimestres.

APPLICATIONS

Les applications de l'apprentissage supervisé sont multiples. On peut citer pêle-mêle :

MODES D'UTILISATION

En résumé, l'apprentissage supervisé doit être utilisé :

- Quand on doit réaliser une prédiction (variable à prédire) dont on connaît parfaitement les caractéristiques, qu'il s'agisse d'une valeur dans une liste ou d'une valeur numérique (la variable à prédire doit alors être structurée).
- Quand on dispose d'un historique connu et suffisant de la variable à prédire. On parle d'historique labellisé (dont on connaît la valeur à prédire sur un échantillon d'apprentissage).

L'octroi de crédit bancaire	Le calcul de risque	L'estimation de l'efficacité des campagnes marketing
La prédiction des pannes	Le calcul de prime en assurances	L'aide au diagnostic
La détection d'anomalies	La prédiction de l'attrition (perte des clients)	La détermination d'un cours de marché, d'un prix de vente pour une plateforme de trading, etc.
La détection de fraudes		

Le cas particulier des séries temporelles

L'apprentissage supervisé prend une forme particulière lorsqu'il s'agit d'étudier l'évolution des données dans le temps. On parle alors de « séries temporelles ». Les algorithmes utilisés dans les séries temporelles supervisées sont différents des algorithmes supervisés « classiques ». Les contraintes sur les données d'apprentissage sont également plus importantes. Les données doivent en effet vérifier la règle « *Interval Data* » qui stipule que les mesures doivent être faites à une cadence régulière (toutes les minutes ou toutes les heures par exemple). Ne pas respecter cette règle de base conduit très souvent à rendre les données difficilement ou non exploitables par les algorithmes spécialisés en séries temporelles. Comme vu précédemment, l'axe temps se voit conférer une importance toute particulière en IA. Son altération ou son manque d'exactitude peut très vite empêcher une IA de fonctionner correctement¹¹.

Apprentissage non supervisé

FONCTIONNEMENT

À l'inverse de l'apprentissage supervisé, les données d'apprentissage dans ce mode-ci ne sont pas annotées à l'avance (ou pas labellisées). L'algorithme doit donc se débrouiller tout seul pour déterminer le travail à faire. Ce mode d'apprentissage doit sa parenté à Karl Pearson en 1901. Il se base principalement sur la détection de similarités et de différences entre les individus. Il en existe de nombreuses variantes.

CAS D'USAGE

Le cas d'usage des **typologies de points de vente** présenté ici a été réalisé par Business & Decision auprès de plusieurs enseignes de distribution. Il s'agit de réaliser dans un premier temps une base de données de tous les points de vente d'une enseigne. On va alors inclure dans ce Data Hub les caractéristiques physiques du point de vente et celles de sa zone de chalandise : la proximité d'une station de transport en commun, les caractéristiques socio-économiques de la population, l'historique des ventes... En utilisant par exemple les données anonymisées dans le respect du RGPD, on peut même y intégrer les caractéristiques des personnes de passage dans la zone, la fréquentation, ainsi que les faits marquants qui pourraient influencer le fonctionnement du point de vente (la liste donnée ici n'est pas exhaustive).

Une fois ces données très soigneusement mises en forme dans le Data Hub, un algorithme non supervisé va segmenter la base des points de vente. La finesse de ces segments peut être ajustée en fonction de l'objectif à atteindre. Dans le cas présent, une enseigne de B2C a réussi à doubler la taille d'un de ses plus importants points de vente grâce à cette procédure. Comment ? En appliquant successivement une analyse non supervisée puis supervisée du potentiel de vente. L'algorithme supervisé s'est notamment appuyé sur les segments initialement donnés par l'algorithme non supervisé et sur les données du Data Hub pour effectuer ses prédictions. En optimisant les capacités du point de vente (caractéristiques physiques) et son attractivité, il s'est avéré qu'il était possible de doubler le nombre de ventes journalières.

MODES D'UTILISATION

En résumé, l'apprentissage non-supervisé doit être utilisé :

- Quand on doit réaliser une aide à la décision ou une prédiction dont on ne connaît pas parfaitement toutes les caractéristiques, et que l'on cherche justement à créer de nouveaux indicateurs structurés pertinents.
- Quand les caractéristiques présentes sont plus importantes qu'un historique de la variable à prédire.
- Quand on dispose de beaucoup d'informations, notamment dans des environnements multifactoriels.
- Quand le problème se trouverait simplifié ou résolu en trouvant de nouvelles caractéristiques simples et communes à différents éléments, et en les regroupant selon leurs caractéristiques.



APPLICATIONS

Le nombre d'applications de l'analyse non supervisée est infini. Il est par exemple possible de mieux comprendre pourquoi tel client est monté en gamme plus facilement qu'un autre. Les causes de prolifération d'un virus peuvent être analysées à l'aide de ce type d'algorithme (car elles sont souvent multifactorielles), ouvrant ensuite la porte à une analyse supervisée pour faire de la prédiction épidémiologique.

Une des principales applications en B2C repose sur la segmentation intelligente des clients pour lancer des campagnes marketing. La puissance et la précision des algorithmes non supervisés va parfois au-delà des attentes. En 2015, une enseigne américaine de grande distribution (Target aux environs de Minneapolis) avait ainsi réussi à prédire qu'une cliente de 16 ans était enceinte (alors qu'elle n'en avait pas l'information directe). Elle lui avait alors envoyé des promotions adaptées (vêtements de grossesse, etc.) provoquant l'ire de ses parents envers le magasin... avant de découvrir finalement que leur fille était réellement enceinte.

Au-delà de ces exemples, l'apprentissage non supervisé doit souvent être vu comme une première

analyse qui va permettre de mieux faire fonctionner des analyses supervisées par la suite. En effet, il n'exige pas, à l'inverse de l'analyse supervisée, de disposer d'un historique de données labellisées. Dans le cas d'une détection de fraudes à l'assurance, Business & Decision a pu aider un client qui ne disposait pas d'un historique de fraudes permettant de faire de la prédiction supervisée. L'apprentissage non supervisé associé à différents scores d'anomalies a ainsi permis d'isoler dans des segments les dossiers qui devaient être étudiés à la main afin d'y trouver potentiellement une fraude. Une fois analysés, ces clusters ont révélé plus de 100 000 € de fraude avérée. Mais surtout, ils ont pu par la suite servir de base d'apprentissage pour une analyse supervisée.

On constate d'ailleurs qu'une analyse non supervisée doit la plupart du temps faire l'objet d'une interprétation très poussée par un data scientist. Dans le cas précédent, c'est une intervention humaine qui a permis d'identifier les fraudes réelles dans les clusters, et de parvenir in fine à une base d'apprentissage supervisée.

LA PUISSANCE
ET LA PRÉCISION
DES ALGORITHMES
NON SUPERVISÉS
VONT PARFOIS AU-DELÀ
DES ATTENTES.

Quelques exemples d'applications pêle-mêle :

L'exploration automatisée des données	Le ciblage des campagnes marketing	La réalisation de sondages ou d'enquêtes
La simplification des bases de données (réduction du nombre de variables)	L'identification de types d'anomalies, de pannes, de fraudes	Le classement de produits ou de points de vente en classe
La segmentation d'une base clients	La spécification de tableaux de bord de pilotage d'activité	La détermination des types de fraudes possibles
		Le profiling des clients qui quittent (ou vont quitter) la marque

Apprentissage renforcé

FONCTIONNEMENT

Le mode d'apprentissage renforcé a été introduit pour la première fois par Richard Ernest Bellman en 1957.

C'est un mode d'apprentissage fascinant à plus d'un titre puisque c'est celui qui a permis au programme Alpha-Go de battre en octobre 2015, Fan Hui, le champion français de Go, puis en mai 2017, le champion du monde, Ke Jie.

Ce mode fascine bon nombre de marketeurs dans la mesure où il sert de base à de nombreux moteurs de recommandation du type Next Best Offer ou Next Best Action. Il est parfois (de manière parfois abusive) également appelé « auto-apprentissage » ou encore « active learning ». L'apprentissage renforcé doit impérativement être « coaché » pour pouvoir progresser. Dans le cas d'une partie d'échecs, le coaching provient de l'information « partie gagnée » ou « partie perdue ». L'IA doit donc perdre des milliers de fois (généralement plus) avant de gagner la moindre partie.

Ce mode d'apprentissage, contrairement aux deux modes précédents (supervisé et non supervisé) apprend au fil de l'eau (au lieu d'apprendre préalablement à son exploitation). Ses applications sont nombreuses et en perpétuelle évolution. Elles sont en pratique limitées par le fait que les erreurs font partie intégrante de son ADN. En effet, dès lors qu'une situation nouvelle

se produit, l'algorithme va devoir explorer tout l'univers des erreurs possibles et tâtonner avant de trouver la réponse la plus appropriée.



Si vous deviez mettre ce type d'algorithme au volant d'une voiture autonome, il faudrait qu'il puisse faire tomber la voiture plusieurs dizaines de fois dans un fossé (et y survivre) pour assimiler le fait qu'un fossé est un obstacle à éviter. En revanche, dans un environnement relativement maîtrisé comme un jeu de société ou une recommandation produit à un client, il peut apporter des réponses inédites auxquelles on ne s'attend pas forcément (comme cela a été le cas d'Alpha-Go devenu imbattable).

Dans le cas d'un moteur de recommandations marketing, le coach n'est autre... que le client lui-même. L'information qui sert de coach au moteur est en effet l'acte d'achat. Si le client achète, cela envoie une information de référence à l'algorithme. Dans le cas contraire, cette information est également prise en compte. Certains moteurs

de recommandations peuvent être pré-entraînés avant d'être réellement mis en service. Le fait de recommander un produit erroné à un client de temps en temps est un prix à payer finalement « acceptable » en vue d'obtenir un algorithme capable de s'adapter en permanence au marché.

CAS D'USAGE

Prenons le cas d'un moteur de recommandations d'offres en ligne de type NBO sur le site marchand d'une grande enseigne de distribution, réalisé par Business & Decision. Le moteur de recommandations présente au visiteur une première offre (sur laquelle il peut cliquer). Puis, on lui propose de procéder à une recherche. Dès que le client a entré les premières caractéristiques de sa recherche, le moteur de recommandations lui propose plusieurs offres personnalisées correspondant mieux à ses affinités et à ses critères.

Ce type de moteur NBO a été mis en marche en parallèle et en compétition avec un moteur de règles plus traditionnel dans ce genre d'application. **Au bout de quelques mois, et en comparaison avec le moteur de règles, l'enseigne a obtenu des taux de conversion multipliés par deux et un revenu internet généré multiplié par trois :** non seulement le site web vendait deux fois plus, mais également avec un panier moyen nettement plus élevé. Le moteur de règles a alors été « débranché » pour laisser le moteur de recommandations faire son travail.

Un des avantages de cette solution est qu'elle est capable de s'adapter à des changements de marché puisqu'elle n'arrête jamais de progresser. Et elle réalise

le rêve secret de tout commercial puisqu'elle progresse quand même à chaque fois que le client n'achète pas !

MODES D'UTILISATION

Pour utiliser le mode d'apprentissage renforcé, les conditions suivantes doivent être réunies :

- L'échec d'une action ne doit avoir que peu de conséquences graves (dans l'état actuel de la technologie, il faut se résoudre à exclure ce mode d'apprentissage pour le pilotage d'engins à grande vitesse).

EN QUELQUES MOIS,
LES TAUX DE CONVERSION
ONT ÉTÉ
MULTIPLIÉS PAR 2
ET LES CHIFFRE D'AFFAIRES
GÉNÉRÉS PAR INTERNET
MULTIPLIÉS PAR 3.

- Ce mode d'apprentissage doit avoir en permanence à sa disposition un dispositif de coaching qui lui indique, au bout d'un certain nombre d'actions, s'il est dans l'erreur ou non (achat de la part d'un client, partie gagnée, coach humain, etc.).

- Il faut disposer d'un temps relativement long pour l'apprentissage, pendant lequel le moteur subira un nombre excessivement élevé d'échecs.

- L'environnement peut être certain ou incertain, le mode d'apprentissage renforcé sait en effet bien gérer et tirer parti d'une incertitude sur le résultat.

APPLICATIONS

Nous sommes loin d'avoir exploré toutes les applications possibles, et ce d'autant plus que la technologie dans ce domaine progresse régulièrement. Il faut reconnaître que ce mode d'apprentissage, à l'instar du moteur de recommandations cité en exemple, peut parfois être mis en parallèle avec un autre mode pendant la phase d'apprentissage initiale. Il est aussi à noter qu'on peut à un moment donné « figer » la phase d'apprentissage si on estime que l'environnement n'évoluera pratiquement plus.

Dans les applications courantes, on peut donc citer :

La conduite autonome des véhicules de type « lents » ne présentant aucun risque majeur de dégâts en cas de collision pendant l'apprentissage. L'exemple typique est le robot aspirateur ou nettoyeur domestique qui, grâce à ce mode, va se montrer capable d'apprendre la géographie du terrain. Il est sans danger car, avec sa vitesse lente, ses capteurs lui signalent la typologie des obstacles avant tout risque de collision. Dans le même type d'appareils, on peut citer la cireuse pour parquet et la tondeuse à gazon autonome.

Dans cette catégorie rentrent également tous les programmes en capacité de jouer à des jeux de société divers et variés.

L'application majeure de ce type d'algorithme reste bien entendu les moteurs de recommandation en marketing. On trouve les moteurs de type Next Best Offer, les plus courants, qui proposent en général un ou plusieurs produits à l'achat, et les moteurs de type Next Best Action. Ces derniers vont plus loin puisqu'ils peuvent proposer autre chose qu'un achat, comme par exemple de parler à un commercial ou à un agent afin de résoudre ou d'anticiper un problème.

Un système de groupe d'ascenseurs ou de gestion de file d'attente automatisée peut être géré suivant ce principe (si toutefois l'échec est limité en conséquence à une attente plus longue, mais exclut des dommages matériels ou des pannes).

Il existe de nombreuses applications en robotique qui permettent à un robot d'apprendre à marcher, à courir, à sauter, etc.

La priorisation des tâches est aussi un sujet pour lequel ce mode d'apprentissage peut être utilisé.

Enfin, on se doit de citer dans le domaine de la recherche des applications au contrôle de véhicules semi-autonomes. En « figeant » le mode d'apprentissage au bout d'un certain temps, le système ne risque pas de se mettre à explorer des solutions nouvelles potentiellement risquées, en mode d'exploitation. Ce mode a notamment été expérimenté par Tesla, mais fait encore largement partie du domaine actuel de la recherche, notamment lorsqu'il est couplé avec des réseaux de neurones.

NOUS SOMMES
LOIN D'AVOIR EXPLORÉ
TOUTES LES APPLICATIONS
POSSIBLES.



Cas particulier de l'apprentissage profond

FONCTIONNEMENT

Les premières recherches sur l'apprentissage profond, fondées sur des réseaux de neurones dits « convolutifs* », remontent aux années 1970. Le concept ne commence cependant à être pris réellement au sérieux par la communauté qu'en 2012, à la suite des travaux menés par Yann Le Cun à partir des années 2000. Leur fonctionnement repose sur une analogie avec le cortex visuel.

Les applications en analyse de données non structurées sont quasiment infinies. Elles vont de la reconnaissance d'images ou de vidéos à l'interprétation des données sémantiques et textuelles, en passant par la reconnaissance vocale. L'apprentissage profond, souvent nommé par sa traduction anglaise d'origine « *deep learning* », est la technologie qui a véritablement fait « renaître » l'intelligence artificielle. Il faut noter que l'apprentissage profond peut être lui-même de type supervisé, non supervisé ou renforcé.

CAS D'USAGE

Le cas d'usage suivant, réalisé par Business & Decision, sort des sentiers battus et prend place dans l'univers de l'habillement et du prêt-à-porter. Il s'agit d'un moteur de recommandations un peu particulier, puisqu'il est capable, à travers la photo d'un client, d'identifier des articles qui pourraient convenir à son style vestimentaire.

Il se démarque en cela des moteurs de recommandations classiques (appelés « look alike ») qui ne prennent en compte que les achats passés des autres clients et internautes. Certes, ceux-ci peuvent donner des idées valables mais, ne pouvant se baser que sur les achats déjà réalisés, ils n'arrivent pas à couvrir tout le champ des possibles, notamment dans des domaines tels que l'habillement dans lequel les combinaisons sont quasi infinies. Ici au contraire, le *deep learning* est capable de proposer des suggestions innovantes et néanmoins pertinentes en se basant sur le style, la couleur, la taille, la texture, etc.



MODES D'UTILISATION

Une des caractéristiques du *deep learning* est qu'il implique, en mode supervisé, de disposer d'une base d'apprentissage très importante pour pouvoir donner des résultats pertinents. Cette base de données d'apprentissage doit être impérativement « labellisée » (ou étiquetée). En somme, si on souhaite lui faire reconnaître des images d'objets ou d'animaux, chaque image qu'on lui fournira pendant la phase d'apprentissage devra être associée à un « label » qui indiquera de quel type d'objet ou d'animal il s'agit.

Cette application du *deep learning* est actuellement la plus répandue et la plus opérationnelle. D'autres modes d'apprentissages du *deep learning*, comme l'apprentissage non supervisé voire auto-supervisé, existent. Mais ils font encore partie du domaine de la recherche fondamentale et n'ont donc pas leur place ici. Une fois la phase d'apprentissage initiale terminée, il est possible d'alimenter le moteur de *deep learning* avec des images non labellisées. Il sera alors en mesure d'identifier les images en question, à condition bien sûr de lui avoir fourni suffisamment d'images de ce type lors de la phase d'apprentissage.

Il faut garder à l'esprit que le *deep learning* n'improvise pas et qu'il ne saura pas réagir face à des situations nouvelles. C'est d'ailleurs pour cela que l'on parle d'intelligence artificielle faible (par opposition à des IA dites fortes qui sauraient théoriquement improviser en fonction du contexte, mais qui n'existent pas à ce stade de notre développement technologique actuel).

Il faut également remarquer que le processus d'apprentissage du *deep learning* requiert des ressources en calcul extrêmement importantes. Ces ressources vont effectuer des opérations aussi simples que des additions et des multiplications mais en quantités gigantesques. Il faut garder à l'esprit qu'une banale image 4K contient déjà plus de huit millions de pixels*. Or le *deep learning* va devoir en traiter parfois bien plus de 30 à la seconde, de surcroît en couleur - multipliant par trois le nombre de pixels - alors que chaque pixel va à son tour générer un nombre d'opérations à donner le vertige.

IL FAUT GARDER
À L'ESPRIT
QUE LE DEEP LEARNING
N'IMPROVISE PAS.
IL NE RÉAGIT PAS EFFICACEMENT
DANS DES SITUATIONS
NOUVELLES ET INCONNUES.

En conséquence, les CPU* (Central Processing Unit) même les plus puissants s'avèrent de plus en plus insuffisants pour réaliser cette masse de calculs. Ils sont donc souvent remplacés par des GPU* (Graphics Processing Unit) qui savent faire du calcul massif en parallèle. Google a même conçu des TPU* (Tensor Processing Unit) qui sont des GPU ultra spécialisés dans cette tâche. La phase d'apprentissage du *deep learning* doit donc s'effectuer le plus souvent sur des infrastructures Big Data spécialisées. Il est à noter que les services des différents fournisseurs cloud proposent souvent des services adaptés en puissance de calcul et en GPU dédiés à l'apprentissage. Le déploiement et l'usage du *deep learning* une fois cette phase initiale effectuée nécessitent heureusement une puissance de calcul nettement moindre, et pourra s'effectuer sur des architectures plus conventionnelles et surtout moins gourmandes en énergie.

APPLICATIONS

Les applications du deep learning sont tellement vastes qu'il ne s'agit pas ici d'en faire une liste exhaustive.

On peut néanmoins citer pêle-mêle :

L'analyse sémantique de texte appelée aussi *Natural Language Processing* (NLP) ou *Text Mining*.

La surveillance vidéo de sites publics (gares, aéroports) ou industriels.

La correction orthographique et la traduction de textes.

La surveillance de la croissance de cultures de bactéries ou de virus pour la fabrication de vaccins.

La reconnaissance d'images dans le domaine industriel pour l'automatisation de processus.

L'imagerie médicale et l'aide au diagnostic.

La conduite de véhicules autonomes ou semi-autonomes (voitures, bateaux, avions, hélicoptères, drones).

Les jeux de société.

L'aide à la composition musicale.

La robotique (la fabrication de robots autonomes).

L'aide à la conception graphique.

L'optimisation de processus de fabrication.

Les chatbots*.

Les voicebots* et les assistants personnels de type Alexa, Siri, Cortana, l'assistant Google, etc.

L'analyse du contenu de réunions et l'élaboration de documents de synthèse.

Tableau comparatif des modes d'apprentissage

	MODE D'APPRENTISSAGE					
	Supervisé	Non supervisé	Renforcé	Profond - Supervisé	Profond - Non Supervisé	Profond - Renforcé
Types de données en entrée	 Structurées	 Structurées	 Structurées	 Non structurées	 Non structurées	 Structurées et non structurées
Apprentissage en continu	 Non	 Non	 Oui	 Non	 Non	 Oui
Transfert d'apprentissage*	 Non	 Non	 Non	 Oui	 Oui	 Oui
Complexité	 Faible à moyenne	 Moyenne	 Élevée	 Très élevée	 Extrêmement élevée	 Extrêmement élevée
Durée de l'apprentissage	 Secondes à heures	 Minutes à jours	 En continu	 Minutes à jours	 Minutes à jours	 En continu
Transparence et explicabilité	 Très bonne	 Bonne	 Faible	 Très faible	 Nulle	 Nulle
Risque de biais	 Fort	 Faible à Moyen	 Faible à Moyen	 Fort	 Moyen	 Moyen
Type de ressources	CPU	CPU	CPU	CPU, TPU ou GPU	CPU, TPU ou GPU	CPU, TPU ou GPU

L'IA demain : où en est la recherche ?

Les startups françaises de l'IA sont les championnes d'Europe des investissements. En 2019, elles ont levé le double de fonds par rapport à 2018¹², soit 1 268 millions de dollars (ou plus de 1 046 millions d'euros). À noter que près des trois quarts des fonds proviennent de l'hexagone et seulement 16 % d'Europe.

Le marché mondial de l'IA est en plein essor : les montants levés par les startups spécialisées ont été multipliés par trois entre 2010 et 2016. En seulement 6 ans, les investissements mondiaux sont ainsi passés de 600 millions à

EN 2019,
LES STARTUPS FRANÇAISES
DE L'IA ONT LEVÉ
LE DOUBLE DE FONDS
PAR RAPPORT À 2018¹²,
SOIT L'ÉQUIVALENT
DE 1 268 MILLIONS
DE DOLLARS.

1,8 milliard de dollars¹³ (soit près d'1,5 milliard d'euros). Sans surprise, les entreprises américaines sont en tête – plus particulièrement les GAFAM –, totalisant 62 % des investissements du secteur... contre 3 % pour la France.

Toutefois, malgré ces inégalités, tous les secteurs, tous les pays et toutes les entreprises devraient bénéficier des atouts de l'intelligence artificielle. Mais attention, car les risques d'accroître ces inégalités sont bien plus grands pour celles et ceux qui n'investiraient pas suffisamment tôt dans l'IA... ou qui s'y prendraient mal.

L'avis de l'expert

Françoise Soulié-Fogelman,
Scientific Advisor, Hub FrancelA

Les small data représentent l'un des sujets importants du moment. Nous avons besoin pour les meilleurs algorithmes d'IA d'un très grand nombre de données qu'il faut collecter, connecter puis étiqueter. L'enjeu est donc de trouver des algorithmes beaucoup plus économes en données. C'est actuellement le principal point faible de l'IA. Concrètement, comment procéder si l'on veut faire en sorte que mon système d'IA ait un taux d'erreur le plus minime possible ? Un autre sujet important tient au besoin d'auditer mon système et donc, en particulier, d'accéder au suivi précis des data. Mais on ne garde aujourd'hui aucune "trace" documentée de ces données. L'auditabilité doit être très opérationnelle. C'est un point fondamental qui aujourd'hui pose problème. L'avenir de la recherche en IA reposera sans doute en partie dans l'apprentissage auto-supervisé. Il faudra parvenir aux small data. On ne peut plus se contenter d'algorithmes gourmands de big data quand on sait que, par exemple, seules Facebook et certaines entreprises chinoises peuvent véritablement faire de la reconnaissance faciale car elles ont à disposition plusieurs milliards d'images ! Enfin, une règle éthique consiste aussi à dire que tout le monde peut participer. Or la donnée est aujourd'hui une barrière d'entrée à l'IA pour les startups ou les petites entreprises. Une des clés, ce sont les small data. Voilà un sujet sur lequel axer des recherches.

3

—

Les enjeux de l'IA

Les enjeux de l'IA

L'IA appliquée au contexte de l'entreprise s'impose comme un accélérateur de développement mais surtout comme une extension nécessaire des compétences humaines. Contrairement à la définition de Dartmouth de 1956, l'IA n'a pas vocation à remplacer l'humain, mais à le compléter et à l'accompagner !

En 2020, la valeur du marché de l'IA était estimée à 17,3 milliards de dollars¹⁴ (soit plus de 14 milliards d'euros). En 2025, l'estimation s'élève à près de 90 milliards de dollars (soit plus de 74 milliards d'euros). En cinq ans, la valeur du marché va être multipliée par plus de quatre. Attention donc à ne pas rater le train de l'IA !

EN 5 ANS,
LA VALEUR DU MARCHÉ
DE L'IA
VA ÊTRE MULTIPLIÉE
PAR PLUS DE 4¹⁴.

« Il faut trouver la meilleure complémentarité entre intelligence humaine et artificielle. L'IA va permettre d'aller vers une intelligence augmentée. L'idée est de garder la décision finale à chaque fois que la responsabilité du côté humain est engagée. Il faut aussi accepter l'idée que l'humain n'est pas en capacité d'analyser des milliards d'informations, et que l'usage de l'IA va amener énormément de valeur pour l'intelligence humaine. Donc il faut considérer l'IA comme un formidable assistant de l'humain. »

Mick Levy,
directeur de l'Innovation Business, Business & Decision

Pourquoi les entreprises doivent y aller

Les quatre grandes familles de cas d'usage

Il existe quatre manières de tirer toute la valeur des données à disposition pour les transformer en bénéfices substantiels.

- **La première** repose sur le fait d'améliorer l'efficacité opérationnelle et de faire mieux. On trouvera dans cette rubrique toutes les formes d'optimisation logistiques et opérationnelles, l'optimisation des coûts, des rendements de production, la réalisation d'économies, l'accélération des traitements, etc.
- **La deuxième** consiste à renforcer la relation client, à personnaliser les messages et les offres jusqu'à établir des liens à la fois plus intimes et plus pertinents. Soit, en un mot, à faire plus.
- **La troisième** réside dans la maîtrise et l'anticipation des risques. Ce domaine s'étend aux sujets de la conformité aux réglementations et de la lutte contre la fraude.
- **La quatrième**, et dernière, relève de la création de nouveaux produits et services qui ne sont pas envisageables sans l'IA. Cette association entre data et IA représente un socle d'innovation considérable pour les entreprises et collectivités.

À l'heure où les données constituent le nouvel or noir du XXI^e siècle, leur exploitation par l'IA devient une priorité pour toutes les fonctions de l'entreprise et tous les secteurs d'activité. Désormais, les entreprises doivent devenir AI-ready et ne pas prendre de retard dans l'implémentation de l'intelligence artificielle.

Mais bien que l'IA commence à être intégrée ponctuellement dans les progiciels métier, dès lors qu'elle touche à ce que l'on peut appeler le « core business » de l'entreprise, la tendance générale est au « Build » et au « sur mesure ». D'une part, pour s'adapter au plus près de l'environnement et de l'écosystème data de l'entreprise. D'autre part, pour conserver la maîtrise de ses propres données, ce type d'application nécessitant une certaine centralisation des données et l'utilisation de Data Lakes* d'entreprise et de Data Hubs*. Nous sommes à l'aube de l'ère de l'IA d'entreprise !

ALORS QUE LA CRÉATION
DE DONNÉES
PAR LES ENTREPRISES
NE REPRÉSENTAIT
QUE 30 %
DU VOLUME EN 2016,
CETTE PROPORTION DEVRAIT
DOUBLER D'ICI 2025¹⁵.

Les bénéfices de l'IA

Les usages qui sont faits de l'intelligence artificielle peuvent produire un fort retour sur investissement. L'IA est en premier lieu un fabuleux outil d'aide à la décision qui, en ce sens, viendra compléter la BI (Business Intelligence) d'aujourd'hui. En suivant les recommandations d'une IA, une chaîne de distribution a ainsi pu doubler le chiffre d'affaires de son plus gros point de vente.

L'IA est aussi un fabuleux outil de pilotage et de contrôle d'activité. À l'heure où les fraudes à la carte bancaire explosent, l'IA permet d'en détecter et d'en prévenir une très grosse majorité. Dans une chaîne de fabrication, pour autant que les données soient disponibles, l'IA permet de détecter à l'avance une pièce qui va s'avérer défectueuse. Être capable de déceler quelques transactions frauduleuses dans des milliards de transactions, de relever des anomalies dans des centaines de millions de produits, voilà un domaine où l'IA excelle et s'avère un allié incontournable à l'humain.

Autre cas : la prédiction est un domaine dans lequel l'IA s'avère également très utile. Aujourd'hui, une IA peut à titre d'exemple prédire avec précision l'heure

d'arrivée d'un porte-container dans un port, rendant possible l'optimisation de toute la logistique portuaire et, par conséquent, l'économie de dizaines de millions d'euros.

Enfin, au fur et à mesure que la technologie devient plus mature, des outils de prescription à base d'IA commencent à faire leur apparition. Une IA sera notamment capable de suggérer le remplacement d'une pièce du bloc moteur d'une grue avant que celle-ci ne tombe en panne, évitant la paralysie complète d'un chantier pendant plusieurs jours.

En conséquence, alors que près de sept salariés français sur dix¹⁷ affirment ne pas s'engager dans leur travail (un chiffre qui s'élève à 85 % dans le monde), l'IA apparaît comme un soutien efficace, notamment pour prendre le relais sur les tâches les plus monotones, mécaniques et répétitives. C'est là qu'entre en jeu l'automatisation des processus ou RPA* rendue possible grâce à cette intelligence artificielle. Pour les entreprises, c'est un atout essentiel afin d'éviter le burn out et le bore out... ainsi que des pertes économiques estimées à près de 7 milliards de dollars de revenus.

L'IA DEVRAIT PERMETTRE
D'AMÉLIORER LA RENTABILITÉ
DES ENTREPRISES
DE PLUS D'UN TIERS
EN 2035¹⁶.

L'avis de l'expert

Ada Sekirin,
Directrice International, Business & Decision

Les entreprises qui iront plus loin dans la réflexion sur l'éthique seront plus compétitives. Le grand public notamment est particulièrement sensible au sujet. En ce sens, l'éthique et la protection des données utilisées à des fins commerciales peuvent devenir un facteur concurrentiel dans le futur. Aujourd'hui, les entreprises sont prêtes à utiliser l'IA d'un point de vue marketing, mais sans remplacer l'intelligence humaine par l'IA.

Développer une IA digne de confiance

Quels sont les risques de l'IA ?

Bien évidemment, tous ces bénéfices sont réels et rendent l'IA très attractive auprès des entreprises. Toutefois, elle n'est pas dénuée de risques. Selon Françoise Soulié-Fogelman, Scientific Advisor chez Hub FrancelA, l'IA présente trois risques majeurs :

- **Les biais algorithmiques**

Si les biais de genre, sexiste ou raciste, sont relativement bien connus¹⁸, il en reste un grand nombre d'inconnus car le raisonnement humain comporte naturellement des biais. Résultat, les discriminations de toutes sortes ont la vie dure. Mais est-ce à l'IA de corriger la vision du monde issue de données réelles ? La vraie question serait plutôt de détecter les discriminations créées par cette IA et de les mesurer : les data scientists ne peuvent

sans doute pas redresser les données, mais ils peuvent mesurer les résultats qui en résultent. C'est pourquoi, à terme, ils devront être responsables des modèles qu'ils produisent. Le profil actuel du data scientist tendra donc à s'accompagner de profils spécialistes dédiés à des tâches très pointues : analyse des biais, des discriminations, de l'éthique...

L'avis de l'expert

Didier Gaultier,

Directeur AI et DataScience, Business & Decision

Quand le RGPD représente un frein pour évaluer et tester les IA

Les IA complexes présentent, on l'a vu, un problème de transparence dans leurs décisions : plus elles sont complexes et plus elles font appel aux Big Data, moins elles sont transparentes et traçables. Une des solutions pour y remédier consiste à utiliser des tests statistiques. Par exemple, pour tester la non-discrimination par le genre dans un processus IA, on injecte en entrée du modèle* des échantillons représentatifs tirés aléatoirement d'hommes et de femmes, et on analyse statistiquement le comportement de l'IA en sortie par rapport à ces échantillons.

Jusque-là, tout va bien. Mais alors, comment tester la non-discrimination politique, religieuse, ethnique ou encore syndicale ? C'est impossible puisque le RGPD interdit de collecter et de conserver des fichiers comportant ce type de données dites sensibles. Aujourd'hui, en Europe, compte tenu de la réglementation en vigueur et des complexités de l'IA, nous sommes dans une impasse pour la résolution de ce type de problèmes. On peut raisonnablement penser qu'une adaptation du RGPD à l'IA sera indispensable dans le futur.

• La transparence

Plus de 70 % des Français sont en attente d'informations sur l'impact positif de l'IA dans la santé et l'éducation¹⁹. Mais en cas d'impact négatif,

il est important de mettre en place des procédures de contrôle et de communiquer sur les risques.

L'avis de l'expert

Michael Deheneffe,

directeur de la Stratégie & Innovation, Business & Decision

La crise actuelle est un bon révélateur des attentes et des impératifs. C'est une loupe sur les comportements. Il en ressort une très forte augmentation de l'exigence de transparence tant chez les individus qu'au sein des organisations. Ainsi, 77 % des Français accordaient une grande importance à la communication des marques pendant la crise sanitaire sur la manière dont leurs produits pouvaient les aider à mieux vivre la situation. 50 % en ont même fait un facteur de décision quant à leurs futures décisions d'achat²⁰. Demain, certaines pratiques de fraude par exemple seront inacceptables. L'IA va alors être mise à l'épreuve au regard de son utilité et de ses objectifs. En ce sens, l'IA devra donc être employée pour optimiser ce qui a de la valeur et minimiser ce qui n'en a pas. Voilà ce vers quoi les entreprises doivent tendre : une IA sans perdant !

• La responsabilité

L'organisation autour de l'IA en entreprise doit être pluridisciplinaire pour apporter un maximum de points de vue différents et ainsi mieux maîtriser le risque d'engager la responsabilité de l'entreprise. C'est pourquoi les projets liés à l'IA sont souvent cogérés en interne par différents départements, à commencer par une association entre

EN 2019,
33 % DES PROJETS D'IA
ET DE MACHINE LEARNING*
ÉTAIENT ALORS
GÉRÉS UNIQUEMENT
PAR L'INFORMATIQUE.

services de l'IT et directions métiers dans 23 % des cas. L'IT est amenée à gérer seule ces projets dans seulement 19 % des cas²¹, et plus rarement (11 %), c'est aux départements de data science que revient ce rôle. C'est là une évolution majeure puisqu'en 2019, 33 % des projets d'IA et de machine learning* étaient alors gérés uniquement par l'informatique.

L'avis de l'expert

Françoise Soulié-Fogelman,
Scientific Advisor, Hub FrancelA

La première question que les entreprises doivent se poser porte sur les risques : le produit qu'elle va commercialiser présente-t-il des risques pour les utilisateurs ? Pour y répondre, il est nécessaire de mettre en place une approche multidisciplinaire dans l'entreprise, avec des acteurs de niveaux différents : managers, responsable des risques, data scientists, informaticiens... Il faut casser les silos ! Cette pluralité d'opinions et de corpus de connaissances fait que le sujet est assez complexe. Nous avons besoin de mettre en œuvre des formations sur l'IA de confiance et d'établir les responsabilités face à un sujet aussi multiple.

Une fois l'équipe définie, le projet passe alors par une étape de conception, puis de production. Dès lors, il s'avère indispensable de suivre le produit tout au long de son cycle de vie avec une préoccupation constante : l'éthique ! En ce sens, l'IA implique de déployer un système de contrôle et un outil capable d'automatiser cette évaluation éthique continue sur l'ensemble des produits. Ce processus fait partie intégrante du pacte entre consommateurs et entreprises.

Mais aujourd'hui, force est de constater que si la plupart des startups et grands groupes français ont entamé leur réflexion sur le sujet d'une IA éthique et responsable, nombre de PME et d'ETI répondent encore aux abonnées absentes ! Or elles représentent 80 % des emplois. Avant d'avancer dans leurs projets d'intelligence artificielle, elles doivent encore gagner en maturité digitale !

Évaluer votre maturité face à l'IA

Un des éléments à prendre en compte dans le déploiement des projets IA est la maturité de l'entreprise. La maturité d'une entreprise ou d'une organisation face à l'IA est une résultante fortement multifactorielle prenant notamment en compte la culture, l'organisation, les moyens humains et technologiques, la gouvernance des données, les infrastructures, les projets, etc.

L'expérience de Business & Decision dans le domaine tend à prouver que la maturité sur la data en général et celle sur l'IA sont intimement corrélées. Dans la mesure où la data constitue le carburant de l'IA (voir notre dernier livre blanc « *Data Éthique / IA Éthique : les deux visages d'un futur responsable* ») et qu'une mauvaise qualité de données réduit à néant les chances d'aboutir de la plupart des projets d'IA, ce résultat n'est guère étonnant.

Avoir une donnée de qualité est certes une condition nécessaire mais pas suffisante pour mener à bien des projets d'IA. Pour réussir, certaines conditions sont requises *a minima* :

Avoir une très bonne compréhension des enjeux métier.

Avoir accès à des data scientists et des data engineers compétents.

Avoir accès à des données (internes et externes) de qualité et en relation avec ces enjeux.

Disposer d'une infrastructure suffisamment performante (cloud ou « on premise ») pour pouvoir supporter la consommation en stockage, mémoire et calcul, surtout dans la phase de conception de l'IA (la phase d'apprentissage est très consommatrice de ressources et de data).

Avoir su mettre en place une gouvernance des données et créer une culture d'entreprise favorables à l'IA.

Afin de vous permettre d'autoévaluer votre entreprise, Business & Decision vous propose un tableau d'évaluation. Il vous suffit d'essayer de repérer à quel niveau de maturité data et IA votre entreprise se trouve dans le tableau. Comme nous l'avons déjà souligné dans notre précédent livre blanc « *Data Éthique / IA Éthique : les deux visages d'un futur responsable* », il n'existe aucune mauvaise réponse. L'important pour faire le chemin est de savoir d'où l'on part, afin de mettre les ressources adéquates et de réaliser la bonne quantité d'efforts.

MODÈLE DE MATURITÉ DES DONNÉES

	Personnes	Processus	Technologies	Data & IA	
Intégré	• Culture de la qualité de données • Gouvernance de la donnée • Data Stewards clefs	• Chaîne de valorisation de la donnée optimisée • La qualité de donnée est partie prenante de tous processus métiers et DSI	• Plateforme collaborative pour la gestion couches sémantiques • MDM opérationnels	• Un grand nombre de projets IA viables et en production. La plupart des projets IA aboutissent dans les temps	• La donnée est gérée comme un actif clé • La donnée crée beaucoup de valeur via l'IA • Enrichissement maîtrisé du patrimoine de données
Proactif	• Comportement proactif sur la qualité de données • Positionnement clef des CDO • Data Stewards reconnus	• Qualité de donnée intégré aux processus métiers • Métadonnées centralisées • Stratégie qualité de données	• Outil dédié à la gouvernance des données. Plateforme de gestion des métadonnées centralisée	• Plusieurs projets IA en production un certain nombre d'entre eux prennent un retard sensible	• Information globalement cohérente et partagée • Vision partagée (métier et DSI) du patrimoine de données
Contrôlé	• Data Stewards en place • Émergence d'un CDO • Responsabilisation des Business Process Owner	• Règles métiers et dictionnaire de données documentés • Gestion partagée des métadonnées	• Plateforme qualité de données • MDM analytique & Partage des métadonnées • Glossaires métiers	• Quelques projets IA en production, la plupart prennent un retard sensible, certains projets sont abandonnés avant la fin	• Mesure de la qualité mais inégale suivant les domaines • Valeur de la qualité de données partagée • Pilotage de l'information par les métiers
Managé	• Perception de la nécessité de la qualité des données • Mise en place de Data Stewards	• Existence d'un glossaire métier partiel • Audit qualité de données • Méthodologies mise en place ponctuellement	• Outil de qualité de données • Nettoyage des données • Répertoire de métadonnées partiel	• Quelques projets IA mais très peu en production et une grande majorité d'entre eux sont arrêtés en chemin	• Capacité à évaluer le niveau et l'impact de la qualité des données • La qualité de données est un enjeu métier
Imprévisible	• Pas de prise de conscience sur la qualité des données • Pas de responsabilité clairement définie	• Pas de langage commun • Projet par projet • Pas de gestion des métadonnées	• Aucun outil.	• Pas ou très peu de projets IA, et la totalité des projets IA s'arrêtent en chemin	• Niveau de qualité inconnu

Intégrer et travailler avec l'IA dans l'entreprise

En faisant évoluer leur organisation pour mieux préparer leurs collaborateurs à la collaboration homme-machine, les entreprises pourraient augmenter leur chiffre d'affaires de 38 %²². Le sujet est donc essentiel pour les organisations.

Au sein de l'industrie du futur, l'humain et l'IA vont devoir associer leurs forces. À l'IA, l'automatisation et l'analyse de grands volumes de données, et à l'humain, la créativité, la sensibilité, l'empathie, l'émotion et l'éthique. Les entreprises pourront alors mettre en place une « grille » afin de déterminer quand employer l'IA et quand utiliser l'humain.

Leadership	Créativité	Empathie	Éthique	Apprentissage	Inférence	Maintenance	Interaction	Communication	Avatars	Transactions	Itérations	Applications	Calculs
Activités réservées à l'humain				Activités où l'humain et la machine collaborent						Activités dédiées ux machines			

Source : *Human + Machine: Reimagining Work in the Age of AI*, par Paul R. Daugherty et H. James Wilson

Instaurer une IA éthique et responsable

Qu'est-ce qu'une IA éthique ?

Maxence Dhellemmes, directeur R&D, Business & Decision

*« L'IA éthique doit être bienveillante, juste pour l'utilisateur
tout en respectant les humains et la planète.*

*L'IA éthique est forcément un idéal. C'est un cap que l'on doit se fixer et au fil de l'eau
placer des garde-fous pour s'assurer que l'on suit la bonne voie.*

*L'IA éthique est une innovation : elle apporte un plus à l'utilisateur final. En résumé,
l'IA éthique repose donc sur une IA bienveillante en phase avec les besoins des utilisateurs.*

*Pour les entreprises, ce positionnement implique donc de se développer
tout en respectant les attentes "humaines". »*

Une IA éthique pour l'entreprise repose sur six actions majeures :

Nommer un **Chief Ethics Officer** pour accompagner le Data Protection Officer (DPO).

Un binôme qui garantirait ainsi l'intégrité et la qualité des données.

À l'image du concept de Privacy by Design, l'entreprise se doit de tendre vers un modèle **d'Ethics by design** pour intégrer l'éthique dès la conception de l'IA.

La réalisation **d'audits réguliers des algorithmes**.

Le contrôle doit s'accompagner d'un code de conduite ou **d'une charte** interne que tous les spécialistes data auront obligation de connaître et de suivre.

Il est nécessaire que **l'humain reste prioritaire**. Il faut donc prévoir un processus pour pouvoir reprendre la main sur l'IA à tout moment.

En matière d'IA, l'entreprise doit être « *human first* » autant que « *human centric* ».

L'avis de l'expert

Ada Sekirin,

directrice International, Business & Decision

Le principal enjeu repose sur la responsabilisation des créateurs de modèles algorithmiques. Si l'IA est aujourd'hui principalement employée en marketing, en e-commerce, en relation client, dans les médias, etc., demain, la prochaine vague concernera les processus de production, les usines 4.0... des domaines particulièrement sensibles. D'où la nécessité d'adopter une approche fondamentalement scientifique pour donner un cadre. L'absence actuelle d'une éthique globale ou générique acceptée par tous rend la programmation des algorithmes beaucoup plus complexe car elle impose de prendre position. Un "point de blocage" qui risque d'être un frein pour le développement de l'IA en entreprise dont l'utilisation reste encore très restreinte aujourd'hui. Le débat autour de l'IA doit donc être motivé par des cas concrets d'usage pour "débloquer" la situation et avancer.

Arborus et Orange dévoilent la première Charte internationale pour une intelligence artificielle inclusive

Le 21 avril 2020, le Fonds Arborus et Orange ont dévoilé la première Charte internationale pour une intelligence artificielle inclusive, sous le haut patronage du secrétariat d'État chargé du Numérique. Cette charte internationale se veut une référence pour les entreprises engagées en faveur de l'égalité des chances. Basée sur sept engagements, elle a notamment pour vocation de garantir une intelligence artificielle conçue et opérée de manière responsable et inclusive. Cette initiative, qui a donné lieu à la création d'un label GEEIS-AI, bénéficie de l'appui de Delphine O, Ambassadrice et Secrétaire Générale de la Conférence mondiale de l'ONU sur les femmes, et de Nicole Ameline, Vice-Présidente du CEDAW-ONU. Ce projet s'inscrit dans le cadre des recommandations internationales sur les droits des femmes et l'égalité. Le label GEEIS-Diversity, qui s'étend à tous les thèmes de la diversité ainsi que l'égalité professionnelle, a également été lancé en 2016. Orange a été la première entreprise à recevoir cette certification.

Peut-on (doit-on) réglementer l'IA ?

Le 19 février 2020, la Commission européenne dévoilait ses principes fédérateurs et ses mesures pour une transformation numérique profitable à tous, reflétant le meilleur de ce que l'Europe offre : l'ouverture, l'équité, la diversité, la démocratie et la confiance. L'objectif est triple : placer les citoyens au premier plan, ouvrir de nouvelles perspectives aux entreprises et encourager le développement de technologies fiables. L'Europe espère ainsi promouvoir une société ouverte et démocratique à travers une économie dynamique et durable.

Afin de façonner l'avenir numérique de l'Europe, la stratégie européenne pour les données repose sur

UNE
TRANSFORMATION
NUMÉRIQUE
PROFITABLE
À TOUS.

le développement d'une intelligence artificielle (IA) axée sur le facteur humain. Pour Thierry Breton, commissaire chargé du marché intérieur : « Notre société génère une masse considérable de données industrielles et publiques qui transformeront notre manière de produire, de consommer et de vivre. Je veux que les entreprises européennes et nos nombreuses PME aient accès à ces données et créent de la valeur pour les Européens, notamment en développant des applications d'intelligence artificielle. L'Europe possède tous les atouts nécessaires pour mener cette course aux mégadonnées et préserver sa souveraineté technologique, son leadership industriel et sa compétitivité économique au bénéfice des consommateurs européens. »

Pour en savoir plus sur la stratégie européenne en faveur d'une IA éthique et responsable :

https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_273



L'avis de l'expert

Françoise Soulié-Fogelman,
Scientific Advisor, Hub FranceIA

Le scandale Cambridge Analytica a démontré ce qu'il était possible de faire avec nos données. Par conséquent, il va y avoir de nouvelles règles du jeu autour d'une IA éthique impliquant, entre autres, un tiers de confiance qui regrouperait les données et en serait le garant. L'Europe a pour rôle vis-à-vis des utilisateurs de poser les grands principes, et de définir ce qui est autorisé ou non. Par exemple, une entreprise qui ne disposera pas de données suffisantes pourra alors se regrouper avec d'autres pour travailler sur certains sujets. Une mutualisation qui n'est pas sans poser des problèmes de protection des données. Et attention au retour de bâton face à des utilisateurs de plus en plus sensibles à la sécurité de leurs données. Une prise de conscience qui pousse vers une IA digne de confiance plus qu'une IA simplement éthique.



Conclusion

D'ici 2024, le marché mondial de l'intelligence artificielle devrait connaître une croissance annuelle moyenne de plus de 20 %²³. Parmi les principaux cas d'usage plébiscités actuellement par les entreprises et les organisations : l'aide à la décision pour le service client, les moteurs de recommandation, l'optimisation de la logistique, l'aide au diagnostic dans la santé, la ville intelligente, la cybersécurité, l'automatisation IT et la gestion des énergies durables.

Pourquoi une telle croissance exponentielle ? Parce que, insiste Ritu Jyoti, vice-président du programme d'intelligence artificielle chez IDC, l'IA permet aux entreprises de « *rester compétitives dans l'économie numérique. Elles adopteront l'IA – pas seulement parce qu'elles le peuvent, mais parce qu'elles le doivent* ». L'objectif est double : améliorer l'expérience client et optimiser l'efficacité des employés au travail pour tendre vers un « collaborateur augmenté ».

Et la France a bien compris ces enjeux : le gouvernement fait de l'intelligence artificielle une priorité sur le long terme à travers son plan de relance proposé en septembre 2020. On l'aura compris, l'intelligence artificielle n'est pas un simple « gadget » marketing mais bien une technologie de plus en plus performante, aboutie et indispensable à la stratégie des entreprises. **Toutefois si 34 % des moyennes et grandes entreprises françaises utilisent l'IA²⁴, combien d'entre elles sont-elles véritablement matures ?** D'un côté se trouveront les entreprises qui auront mis en place une IA de façon éthique et responsable, s'appuyant sur une gouvernance de données solide. De l'autre, se situeront celles qui n'auront pas assez investi, auront développé une IA à la va-vite avec une gouvernance de données mal définie, s'exposant à terme à des retours extrêmement douloureux de la part des clients, des employés et des actionnaires.

L'adoption de l'IA est avant tout une démarche créatrice de valeurs et bienveillante à moyen et à long terme pour les entreprises. Elle implique une évolution des cultures et des méthodes de travail. Cette impulsion nécessite une implication forte des directions générales et a également besoin d'être accompagnée. Quoi qu'il en soit, elle n'est surtout pas une simple affaire de technologie.

Le mot de la fin

Luc Julia, CTO de Samsung Electronics et cocréateur de Siri

*« Il est difficile d'encadrer l'IA dans une réglementation tant son champ d'investigation est vaste.
C'est pourquoi il est nécessaire d'éduquer pour en faciliter la compréhension.
Mais les millions de calculs effectués par l'intelligence artificielle rendent la tâche difficile.
L'Europe a été précurseur en matière de réglementation et de respect des données personnelles.
En ce sens, le California Consumer Privacy Act (CCPA) est une copie du RGPD.
Les individus doivent comprendre les impacts et, une fois informés, prendre la bonne décision
pour eux. L'histoire nous a montré que les révolutions industrielles posent toujours
à un moment donné des problèmes d'emploi. En 1790, à Lyon, les Canuts sont remplacés
par des machines à tisser. En 1820, le bassin d'emploi lyonnais était le plus dynamique
de France car les machines ont créé une économie florissante qui a généré des emplois.
Le meilleur accompagnement, c'est l'éducation. L'IA va changer nos métiers
mais elle ne va pas nous remplacer ! »*

Glossaire de data science et IA

Agent : en intelligence artificielle, un agent intelligent (AI) est une entité (matérielle et logicielle) avec un certain degré d'autonomie capable de percevoir son environnement (grâce à des capteurs) et aussi d'agir sur celui-ci (via des actionneurs) en vue de réaliser des objectifs. (Wikipédia)

Apprentissage profond (deep learning en anglais) : ensemble de méthodes d'apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de transformations non linéaires (comme les réseaux de neurones artificiels ndlr). Ces techniques ont permis des progrès importants et rapides dans les domaines de l'analyse du signal sonore ou visuel, et notamment de la vision par ordinateur, de la reconnaissance vocale, et du traitement automatisé du langage. Dans les années 2000, ces progrès ont suscité des investissements privés, universitaires et publics importants, notamment de la part des GAFAM (Google, Apple, Facebook, Amazon, Microsoft). (Wikipédia)

Apprentissage supervisé (supervised learning en anglais) : tâche d'apprentissage automatique consistant à apprendre une fonction de prédiction à partir d'exemples annotés ». (Wikipédia)

Chatbot (ou agent conversationnel) : agent logiciel qui dialogue avec un utilisateur. L'utilisateur est invité à formuler sa demande en langage naturel. Elle est alors affinée par un échange basé principalement sur des questions et des réponses préprogrammées dans une base de connaissances. Si l'échange est oral et non écrit, on parle alors aussi de voicebot.

CPU (en anglais, Central Processing Unit) : composant présent notamment dans les ordinateurs et dans de nombreux dispositifs électroniques qui exécutent les instructions machine des programmes informatiques. Avec la mémoire, c'est notamment l'une des fonctions qui existent depuis les premiers ordinateurs. Un processeur construit en un seul circuit intégré est un microprocesseur. (Extrait de Wikipédia)

Data engineering : discipline consacrée à la construction et la maintenance des systèmes de pipeline de données de l'organisation (O'Reilly). Cette discipline englobe les technologies propres aux big data, DataOps et MLOps.

Data lake ou lac de données en français : méthode de stockage de données massives utilisée par les big data (mégadonnées en français). Ces données sont gardées dans leurs formats originaux, peu transformées. Le lac de données donne la priorité au stockage rapide et volumineux de données hétérogènes en adoptant une architecture en cluster. Il n'est pas optimisé pour les requêtes SQL comme les SGBD relationnels classiques, contrairement au Data Hub dans lequel les données sont transformées, mises en qualité, vérifiées, triées et optimisées pour une exploitation par les data scientists et les métiers.

Data science ou science des données en français : Data science ou science des données en français : discipline au sens large qui permet l'extraction de connaissances à partir d'un ensemble de données (Wikipédia). Elle est composée de plusieurs disciplines, qui sont : le data engineering, la connaissance métier, la statistique, le machine learning et l'intelligence artificielle (Business & Decision 2020).

DataOps : méthodologie orientée processus, utilisée par les équipes de data engineering, pour améliorer la qualité et réduire le temps de cycle de l'analyse des données. Alors que DataOps a commencé comme un ensemble de bonnes pratiques, c'est maintenant devenu une approche nouvelle de data engineering.

DevOps : pratique technique visant à l'unification du développement logiciel (Dev) et de l'administration des infrastructures informatiques (Ops). Elle se caractérise principalement par la promotion de l'automatisation et du suivi de toutes les étapes de la création d'un logiciel, depuis le développement, l'intégration, les tests, la livraison jusqu'au déploiement, l'exploitation et la maintenance des infrastructures. Les principes DevOps soutiennent des cycles de développement courts, une augmentation de la fréquence des déploiements et des livraisons continues, pour une meilleure atteinte des objectifs économiques de l'entreprise. (Wikipédia)

GPU (Graphics Processing Unit ou processeur graphique en français) : unité de calcul pouvant être présente sous forme de circuit intégré (ou puce) sur une carte graphique ou sur une carte mère, ou directement reliée au même circuit intégré que le microprocesseur qui assure les fonctions de calcul d'image et de vidéo. Un processeur graphique a généralement une structure hautement parallèle qui le rend efficace pour une large palette de tâches graphiques comme le rendu 3D, la gestion de la mémoire vidéo, le traitement du signal vidéo, mais également le calcul matriciel et notamment l'entraînement des réseaux de neurones artificiels. (Extrait de Wikipédia)

IA ou intelligence artificielle : discipline consacrée à la construction d'agents autonomes qui peuvent analyser et prendre des décisions (Peter Norvig 2010). Également appelée intelligence augmentée, c'est une branche de la data science consacrée à la construction d'agents qui interagissent avec l'environnement, qui peuvent apprendre à partir de données initiales fournies et/ou de données collectées lors des interactions pour réaliser une fonction dans un environnement déterminé, avec un certain degré d'autonomie (Business & Decision 2020).

IAOps – aussi appelée MLOps ou encore ProcessIA : méthodologie utilisée par les équipes de data science pour gérer, déployer et maintenir efficacement les projets d'IA en entreprise. Cette méthodologie, constituée principalement d'un ensemble de bonnes pratiques, est inspirée du croisement de la méthode CRISP étendue, utilisée depuis longtemps en data science, et des pratiques DevOps.

Machine learning : également appelé apprentissage automatique, apprentissage artificiel ou apprentissage statistique, il s'agit d'un champ d'étude de la data science qui se fonde sur des approches mathématiques et statistiques afin de donner aux ordinateurs la capacité d'apprendre à partir de données en utilisant des algorithmes. (Wikipédia)

Modèle : en data science ou en statistique, il s'agit d'une description mathématique approximative du mécanisme qui a généré les observations et qui répond à certaines hypothèses. Il s'exprime généralement comme une fonction (ou une famille de fonctions) de variables d'entrée $X_1 \dots X_n$. Chaque membre de la famille est une approximation possible de la réalité. L'inférence consiste donc à déterminer si un des membres s'accorde de manière acceptable avec toutes les données.

NLP (ou NLU, Natural Language Processing ou Understanding, également appelé Text Mining) : domaine multidisciplinaire impliquant la linguistique, l'informatique et l'intelligence artificielle, qui vise à créer des outils de traitement de la langue naturelle pour diverses applications, notamment les chatbots. (Wikipédia)

Pixel : plus petit élément de base d'une image numérique matricielle ou d'une vidéo. Son nom provient du terme anglais "picture element". Le nombre de pixels horizontaux et verticaux d'une image définit ce qu'on appelle la résolution. Par exemple : une image TV 4K en Ultra Haute Définition (UHD) représente 3 840 pixels horizontalement et 2 160 verticalement, soit quatre fois une définition full HD de 1 920 x 1 080.

Réseaux convolutifs : en apprentissage profond, un réseau de neurones convolutifs ou réseau de neurones à convolution (en anglais CNN pour Convolutional Neural Networks) est un type de réseau de neurones artificiels dans lequel le motif de connexion entre les neurones est inspiré par le cortex visuel des animaux.

Règlement Général sur la Protection des Données (RGPD ou GDPR, de l'anglais General Data Protection Regulation) : règlement de l'Union européenne qui constitue le texte de référence en matière de protection des données à caractère personnel. Il renforce et unifie la protection des données pour les individus au sein de l'Union européenne. (Wikipédia)

RPA (Robotic Process Automation) ou automatisation robotisée des processus : technologie de création d'agents par apprentissage du comportement d'un usager, notamment sur une interface graphique. Dans une démarche classique d'automatisation des processus, un développeur informatique écrit un programme informatique qui effectue un certain

nombre de tâches et interagit avec l'interface de programmation (API) de l'application. Dans une démarche d'automatisation des processus robotisés, le système apprend la liste des tâches à automatiser en observant le comportement d'utilisateurs humains. L'automatisation de processus robotisés est utilisée dans des entreprises pour automatiser des tâches répétitives.

Statistique : discipline qui étudie des phénomènes à travers la collecte de données, leur traitement, leur analyse, l'interprétation des résultats et leur présentation afin de rendre ces données compréhensibles par tous. (Wikipédia)

Système expert : outil capable de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier. Il s'agit de l'une des voies tentant d'aboutir à l'intelligence artificielle. Plus précisément, un système expert est un logiciel capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connues. Il peut servir notamment comme outil d'aide à la décision (Wikipédia).

TPU (Tensor Processing Unit) ou unité de traitement de tenseur : circuit intégré spécialisé dans le calcul matriciel, appartenant à la famille des GPU, pour une application développée spécifiquement par Google afin d'accélérer les systèmes d'intelligence artificielle utilisant des réseaux de neurones.

Transfert d'apprentissage (en anglais, *Transfer Learning*) : technique qui permet d'utiliser une IA qui a été pré-entraînée dans un environnement voisin mais légèrement différent de celui de son utilisation finale. L'apprentissage doit

alors malgré tout être complété et terminé dans l'environnement précis où l'IA va devoir opérer. Mais cette phase est néanmoins beaucoup moins longue et consommatrice d'énergie que la phase d'entraînement initial.

Variable booléenne : variable qui ne peut prendre que deux valeurs, vrai ou faux, un ou zéro. Son nom vient du créateur de la logique du même nom : George Boole, qui est à la base même du principe de fonctionnement de tous les ordinateurs modernes et de l'électronique numérique. Une variable booléenne est une variable dite structurée.

Variable catégorielle : variable qui prend ses valeurs dans une liste prédéfinie, comme par exemple une marque de voiture. On utilise aussi le terme « variable qualitative ». Une variable catégorielle est une variable dite structurée.

Variable non structurée : variable qui n'est ni numérique ni catégorielle, ni booléenne, ni une date. Cela peut être une variable de type texte (chaîne de caractère), qui contient un signal de type audio, vidéo, ou bien un log de machine.

Variable numérique : variable qui prend la forme d'un nombre entier (on parle alors de variable numérique discontinue) ou non (on parle alors généralement de variable numérique continue). On utilise aussi le terme « variable quantitative ». Une variable numérique est une variable dite structurée.

Variable structurée : variable numérique, catégorielle, booléenne ou de type temps / date.

Sources

- 1 Classement annuel des métiers émergents, LinkedIn, décembre 2019 – <https://business.linkedin.com/fr-fr/talent-solutions/emerging-jobs-report/emerging-jobs-report-france>
- 2 Selon IDC, août 2020 – <https://www.idc.com/getdoc.jsp?containerId=prUS46794720>
- 3 Selon une étude France Digitale - Roland Berger, Maddyness, novembre 2019 – <https://www.maddyness.com/2019/11/04/france-leader-investissements-ia/>
- 4 Encyclopédie Larousse – https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257
- 5 Warren McCulloch : https://fr.wikipedia.org/wiki/Warren_McCulloch
- 6 Walter Pitts : https://fr.wikipedia.org/wiki/Walter_Pitts
- 7 Franck Rosenblatt : https://fr.wikipedia.org/wiki/Frank_Rosenblatt
- 8 John McCarthy : https://fr.wikipedia.org/wiki/John_McCarthy
- 9 Marvin Minsky : https://fr.wikipedia.org/wiki/Marvin_Minsky
- 10 Voir « Intelligence Artificielle 3e édition » de Stuart Russel & Peter Norvig, éditions Pearson
- 11 Voir à ce sujet notre premier livre blanc, Data Éthique, IA Éthique, les deux visages d'un futur responsable – <https://fr.blog.businessdecision.com/livre-blanc-data-ethique-ia-ethique-futur-responsable/>
- 12 Selon une étude France Digitale – Roland Berger, novembre 2019 – <https://www.maddyness.com/2019/11/04/france-leader-investissements-ia/>
- 13 Selon le rapport « Future of Artificial Intelligence » réalisé par le Hub Institute en partenariat avec IBM, 2018 – <https://hubinstitute.com/Report/hubreport-future-ai>
- 14 Selon une prévision du chiffre d'affaires du marché de l'Intelligence Artificielle dans le monde de 2016 à 2025, en millions de dollars – <https://fr.statista.com/statistiques/638963/revenus-marche-intelligence-artificielle-monde/>
- 15 Selon une enquête IDC, 2018 – <https://www.decideo.fr/Big-Data-les-donnees-sont-partout-mais-insuffisamment-exploitees-a10425.html>
- 16 D'après une étude Accenture, 2018 – <https://www.ladn.eu/nouveaux-usages/etude-marketing/intelligence-artificielle-tout-le-monde-va-en-profiter-surtout-les-americains/>
- 17 <https://www.journaldunet.com/solutions/reseau-social-d-entreprise/1486875-en-2020-quelle-sera-la-place-de-l-intelligence-artificielle-dans-le-secteur-du-e-commerce/>
- 18 Voir à ce sujet notre premier livre blanc, Data Éthique, IA Éthique, les deux visages d'un futur responsable – <https://fr.blog.businessdecision.com/livre-blanc-data-ethique-ia-ethique-futur-responsable/>
- 19 Observatoire de la Notoriété et de l'Image de l'Intelligence Artificielle en France, Impact AI - Ifop, 2019 – <https://www.ifop.com/wp-content/uploads/2019/12/CP-Impact-AI-Barom%C3%A8tre-2019-VDEF.pdf>
- 20 Selon le Trust Barometer Edelman, avril 2020 – <https://www.elanedelman.com/research/trust-and-the-brand>
- 21 ZDnet, avril 2020 – <https://www.zdnet.fr/actualites/les-projets-autour-de-l-intelligence-artificielle-sont-plus-matures-en-2020-39901613.htm>
- 22 Selon une étude d'Accenture, 2018 – <https://www.accenture.com/fr-fr/company-news-release-artificial-intelligence-workforce>
- 23 Selon IDC, août 2020 – <https://www.idc.com/getdoc.jsp?containerId=prUS46794720>
- 24 Selon une enquête du JDN, Club Décision DSI et IT Research, septembre 2020 – <https://www.itresearch.fr/un-tiers-des-moyennes-et-grandes-entreprises-francaises-utilisent-lia/>

Remerciements

**Le livre blanc de Business & Decision est publié
sous la direction de Didier Gaultier, Directeur DataScience & AI, avec :**

Valérie Lafdal

Directrice générale France et Directrice générale déléguée Groupe,

Ada Sekirin

Directrice générale Europe

Michael Deheneffe

Directeur de la stratégie et de l'innovation,

Mick Levy

Directeur de l'Innovation Business

Nos remerciements particuliers pour leurs précieux témoignages et participations :

Faycal Boujemaa

Technology Strategist - Orange Labs Research

Philippe Ensarguet

Directeur Technique (CTO) - Orange Business Services

Jérôme Goulard

Directeur RSE Diversité & Éthique – Orange Business Services

Pascal Hauret

Performance Manager - Orange Business Services

Luc Julia

CTO & Senior Vice President of Innovation – Samsung

Françoise Soulié-Fogelman

Scientific Advisor - Hub FranceIA

Avec la participation de :

Romain Bernard

Manager DataScience Nord & Ile-de-France

Jérôme Dewever

Manager Conseil Expertise MDM

Maxence Dhellemmes

Directeur R&D

Christophe Hatterer

Manager Data Gouvernance

Cédric Missoffe

Directeur de l'agence Conseil & Expertise

Mondher Sendi

Expert AI - Computer Scientist

Stéphane Walter

Manager Conseil & Expertise Big Data

Réalisation :

L'agence Bespoke et Violaine Cherrier

Business & Decision

Cœur Défense A , 110 Esplanade Général de Gaulle, 92931 Paris La Défense Cedex

www.businessdecision.com

blog.businessdecision.com

 Business & Decision