

# Approche d'une vision éthique de l'intelligence artificielle par et pour les entreprises

# | Sommaire

Edito	Page 3
Introduction : l'IA technologie en plein essor mais en quête de sens et de confiance	Page 4
Les recommandations	Page 7
Un peu d'histoire : les saisons de l'IA	Page 8
Intelligence artificielle, de quoi parle-t-on ?	Page 10
Pourquoi les questions d'éthique sont-elles au cœur des débats ?	Page 13
Une IA de confiance	Page 20
Les réglementations qui s'appliquent déjà	Page 24
Un exemple : la justice	Page 29

# | Annexes

Les 23 principes d'Asilomar	Page 31
Liste des personnes auditionnées	Page 33
Bibliographie	Page 34

# | Edito

En voulant travailler sur l'éthique de l'intelligence artificielle, la Commission Mutations technologiques et impacts sociétaux du MEDEF est au cœur de sa mission : explorer avec et pour ses adhérents les nouvelles frontières du possible et de réfléchir sur le souhaitable.

La raison d'être du MEDEF « *agir pour une croissance responsable* » nous place au-delà de l'« *affectio societatis* » juridique pour nous inviter à faire « société », c'est-à-dire à vivre ensemble.

Est-ce que ce vivre ensemble s'étend aux machines ?  
Est-ce la place du vivant parmi les machines qui est interrogé ?

Car l'intelligence dite artificielle n'est pas un outil informatique de plus : c'est une mutation complète de nos organisations et donc des entreprises dont il est question.

A l'heure où de plus en plus de systèmes décisionnels sont modifiés, nous devons prendre un temps d'arrêt : celui de la réflexion pour intégrer ce questionnement : l'IA est-elle éthique ?

Dans sa conception ? Dans son utilisation ?

Ces questions nous ne sommes pas les seuls à nous les poser en Europe et dans le monde, mais il est crucial que les entreprises s'en emparent afin que nous puissions en comprendre les mécanismes de base et aider à leur utilisation et leur adoption pour le bien de tous.

Faire définitivement de l'IA une « Tech for good » !

# | Introduction

## L'IA une technologie en plein essor mais en quête de sens et de confiance

Conçue dans les années 1950, à partir des publications d'Alan Turing, mathématicien et cryptologue britannique, auteur de travaux qui fondent scientifiquement l'informatique, et considérée très tôt comme une technologie prometteuse, l'Intelligence artificielle (IA) a déjà vécu bien des saisons et des vicissitudes. On lui prête trois printemps et deux hivers technologiques. Sa terminologie trompeuse nourrit les fantasmes de la science-fiction ou frustre les prophètes du transhumanisme qui s'impatientent de la voir supplanter les capacités du cerveau humain. Certes, l'IA a remporté sur l'homme bien des épreuves dans des domaines d'excellence, comme peuvent en témoigner ses victoires sur le champion mondial du jeu de go ou ses remarquables performances au concours national de l'internat de médecine en Chine, mais sans jamais dominer sa matière ni comprendre ses résultats.

L'IA se rapproche davantage d'une intelligence assistée que d'une intelligence spontanée et artificielle indépendante de la volonté humaine. Avec l'essor des sciences cognitives, la croissance exponentielle des données, l'explosion des capacités de calcul, le développement de l'internet des objets et du cloud, l'IA franchi un nouveau cap et devient désormais le vecteur de la 4<sup>ème</sup> révolution industrielle. **L'IA permet d'entrer pleinement dans l'économie de la connaissance en valorisant toutes les données.**

Le foisonnement des débats qu'elle provoque conduit autant à croire dans les formidables potentiels de croissance de cette technologie qu'à se méfier des risques humains qu'elle peut engendrer. Alors qu'en 2015, le marché de l'intelligence artificielle pesait 200 millions de dollars, on estime qu'en 2025, il s'élèvera à près de 90 milliards de dollars<sup>1</sup>. D'importants gains de productivité sont attendus avec l'IA. Selon MC Kinsey, l'automatisation boostée à l'IA va améliorer les taux de productivité globale de 0,8 à 1,4% chaque année sur les 50 prochaines années, soit un taux deux à trois fois plus élevé que ceux enregistrés dans les décennies précédentes avec l'expansion de la robotique et de l'IT. Technologie disruptive pour l'ensemble des métiers, l'IA se propage en effet dans divers domaines : éducation, santé, transport, justice, travail, culture, sécurité, et s'installe particulièrement dans les secteurs de la banque, de la logistique et des transports.

Une telle révolution technologique a également un impact sur l'emploi. L'IA remplacera ou du moins devrait transformer 40% des emplois d'ici 10 à 15 ans d'après plusieurs enquêtes convergentes dont une étude récente de l'OCDE. L'automatisation et le développement des robots ne signifient pas pour autant la fin du travail. Mais il va falloir travailler autrement avec les machines et autres robots. Les dirigeants d'entreprises et leurs managers y sont deux fois plus prêts que leurs salariés<sup>2</sup>.

**Le MEDEF a pris part à ces débats à travers une série de manifestations organisées depuis 2015 par la Commission Innovation**, avec la participation régulière de Cédric Villani auteur du rapport à l'origine de la stratégie nationale de l'IA à laquelle le MEDEF

---

<sup>1</sup> <https://www.tractica.com/newsroom/press-releases/artificial-intelligence-software-market-to-reach-89-8-billion-in-annual-worldwide-revenue-by-2025/>

<sup>2</sup> Source : enquête Malakoff Médéric – Humanis – 2018

a également contribué en décembre 2017, en prenant position sur les conditions de développement de l'IA pour l'entreprise<sup>3</sup>.

Le MEDEF a récemment complété cette position par une enquête menée en collaboration avec l'Association des Centraliens sur les usages de l'IA et les attentes des entreprises<sup>4</sup>.

Cette enquête montre que la pénétration de l'IA dans les TPE, les PME et ETI est encore assez limitée. Les entreprises sont davantage utilisatrices que créatrices d'IA. L'impact de l'IA sur l'amélioration de la productivité est bien identifié par les entreprises, y compris par celles qui n'utilisent pas l'IA dans leurs activités. C'est surtout une « **IA faible** » c'est-à-dire l'Intelligence artificielle concentrée sur une tâche précise qui commence à se déployer dans l'entreprise. Cet usage modifie déjà profondément les pratiques et les business models. La reconnaissance faciale et vocale, l'interprétation d'image, la prédiction de pannes permettent d'introduire rapidement des produits et services pour défier la concurrence et établir des liens plus étroits avec le client.

Encore minoritaires dans la pratique, les entreprises sont de plus en plus nombreuses à inscrire dans leur priorités le recours à l'IA dans les deux à trois années à venir. Elles s'y préparent en portant une attention particulière au degré de confiance que l'on peut accorder à l'IA en l'état de son développement.

L'IA apparaît encore, même aux experts, comme une « boîte noire » qui contient encore bien des mystères à percer. Or, les clients ont besoin d'être rassurés sur l'exploitation de leurs données dans le respect de leur vie privée. Les personnels d'entreprise s'interrogent sur la place, le sens et la qualité du travail face aux robots et sur l'évolution de leurs métiers.

**Tous revendiquent une IA de confiance respectueuse de la personne humaine et de son intégrité. L'IA de confiance nécessite le respect d'une éthique dans sa gestion et son usage.**

Mais l'éthique n'est pas le droit. Le MEDEF, dans sa contribution à l'élaboration de la stratégie nationale en IA, l'a souligné : il faut réguler sans brider. L'éthique est donc aussi une affaire d'autorégulation des comportements et de gestion préventive de l'utilisation de l'IA dont les principes sont progressivement posés par les institutions publiques tant nationales qu'européennes. La mise en œuvre pilotée par le gouvernement de la stratégie en IA « *AI for Humanity* » se fonde également sur l'éthique et l'équilibre entre la performance et l'humain.

Récemment, la Commission européenne publiait deux rapports issus des travaux d'un groupe de 50 experts à haut niveau établissant d'une part des recommandations pour une intelligence artificielle de confiance et d'autre part des recommandations d'investissement. La présidente de la Commission européenne ayant intégré à son programme l'élaboration dans les premiers jours de son mandat d'une régulation ad hoc sur l'intelligence artificielle, **le MEDEF souhaite éclairer les conditions de mise en œuvre de tels principes.** Quel niveau de transparence peut être garantie sur la conception, la réalisation, l'usage de l'IA ? Comment s'assurer de la transparence des algorithmes ? Est-il toujours possible en cas d'erreur de décision dictée par l'IA d'analyser l'origine de l'erreur ? Peut-on certifier la fiabilité de l'IA ? Jusqu'où peut-on déléguer des tâches à l'IA ? Comment la contrôler ? Comment éviter les biais et maîtriser les processus de décision ? Un audit a posteriori est-il préférable et suivant quel protocole ? En réponse à ces questions de bonnes pratiques sont-elles déjà à

<sup>3</sup> Contribution du MEDEF sur les conditions de développement de l'intelligence artificielle pour l'entreprise (décembre 2017) : <https://www.innover-en-france.com/file/197799/>

<sup>4</sup> Les [résultats](#) ont été présentés et débattus lors des 4<sup>èmes</sup> rencontres de l'innovation MEDEF – Centraliens en janvier 2019

l'œuvre dans des entreprises, voire au sein de secteurs professionnels ? Sont-elles reproductibles ?

**Autant de questionnements que la Commission Mutations technologique et impacts sociétaux a voulu éclairer en menant en coopération avec la Commission Innovation une série d'auditions réalisées entre juin et novembre 2019** auprès de fédérations et d'entreprises, dont les enseignements ont conduit à émettre quelques recommandations pour le déploiement éthique de l'IA dans les entreprises.

# | Recommandations

## Gouvernance nationale :

- | Décliner au sein du Conseil national consultatif d'éthique une enceinte pour les sciences numériques (p. 16).

## Gouvernance dans les entreprises :

- | Encourager les entreprises à inclure les questions d'éthique des systèmes informationnels dans leurs outils de gouvernance (p. 15)
- | L'éthique des systèmes d'IA ne doit pas reposer seulement sur les concepteurs des IA mais engager des équipes pluri disciplinaires (p. 16)
- | Les « éducateurs » d'IA doivent être sensibilisés à l'éthique (p. 18)
- | Ne pas faire du contrôle humain un critère systématique de garantie (p. 19 et 22)

## Création d'outils de pilotage de la gouvernance de l'IA :

- | Structurer les questions proposées par le groupe d'experts européens par thématiques (p. 21)
- | Intégrer une évaluation des risques ex ante et / ou ex post et décliner les obligations en fonction du niveau d'impact de l'IA (p. 19)
- | Privilégier l'auditabilité de la robustesse et de la répétabilité plutôt que la certification des IA (p. 23)
- | Eviter une régulation du type « RGPD » pour l'IA, mais privilégier une mise en perspective des législations en place et leur adaptation si nécessaire (p. 23)

# | Un peu d'histoire

Est-il pertinent de doter les machines d'une intelligence ?

## **Aristote :**

*« Si chaque instrument pouvait, par ordre ou par pressentiment, accomplir son œuvre propre, et si les navettes tissaient d'elles-mêmes et les plectres jouaient de la cithare, alors les maîtres d'œuvre n'auraient plus besoin de manœuvres, ni les maîtres d'esclaves. Les vrais hommes abandonneraient alors les tâches viles, si indignes d'eux, pour ne plus se consacrer qu'aux activités de citoyens et à la recherche du savoir et de la sagesse qu'il procure. » (Aristote, La Politique, chapitre 4).*

La capacité pour une machine à avoir une capacité cognitive proche de celle de l'humain est questionnée par la loi d'Asimov exposée pour la première fois dans les années 40 :

- I Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ;
- I Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ;
- I Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

## **Alan Turing :**

En 1956, plusieurs des plus grands chercheurs de l'époque dans ce domaine de recherche, se réunissent dans le New Hampshire lors d'une conférence d'été donnée sur le campus du Dartmouth Collège. Le terme "Intelligence Artificielle" est adopté pour désigner ce qui devient une discipline académique à part entière : *« La science qui consiste à faire faire aux machines ce que l'homme ferait moyennant une certaine intelligence ».*

Cette science suscitait alors autant d'attentes que de craintes, la plupart des spécialistes prédisant le développement d'une intelligence semblable à celle de l'homme dans les vingt années à venir.

Dans la même période (1950), Alan Turing, mathématicien et cryptologue britannique, auteur de travaux qui fondent scientifiquement l'informatique, soulève la question d'apporter aux machines une forme d'intelligence dans son livre *Computing Machinery and Intelligence*.

Il décrit alors un test aujourd'hui connu sous le nom de « Test de Turing » dans lequel un sujet interagit à l'aveugle avec un autre humain, puis avec une machine programmée pour formuler des réponses sensées. Si le sujet n'est pas capable de faire la différence, alors la machine a réussi le test et selon l'auteur, peut véritablement être considérée comme « intelligente ».



# Les saisons de l'IA :

## Un âge d'or :

Aux Etats-Unis, l'Agence de Recherche du Pentagone (ARPA puis DARPA) finance activement les recherches dans ce domaine et les universités créent des laboratoires spécialisés, notamment le AI Lab au MIT (Massachusetts Institute of Technology) fondé par John McCarthy et Marvin Minsky, pionniers de l'intelligence artificielle. En Europe l'intelligence artificielle intéresse aussi les universités et les gouvernements. En France, Jacques Pitrat en est l'un des pionniers.

## Des « Hivers » :

- I La première modélisation des réseaux de neurones aboutit à un échec. Les concepteurs espéraient créer une machine intelligente capable de comprendre le langage naturel. Le programme est abandonné dans les années 60, c'est le premier hiver de l'IA. Mais elle renaît dans les années 80 avec les systèmes experts, des programmes qui reproduisent un raisonnement logique et qui peuvent résoudre des problèmes. En 1997, c'est l'apogée : Deep Blue arrive à battre Garry Kasparov, champion du monde des échecs ;
- I Dans les années 1980, l'apprentissage automatique (« Machine Learning ») se développe. L'ordinateur commence à déduire des « règles à suivre » rien qu'en analysant des données ;
- I Parallèlement, des algorithmes « apprenants » sont créés qui préfigurent les futurs réseaux de neurones ;
- I Dans les années 2000, le big data et de nouvelles puissances et infrastructures de calcul, permettent à certains ordinateurs d'explorer des masses de données sans précédent ; c'est l'apprentissage profond (« *deep learning* »).

## Un nouveau printemps :

En 2016, AlphGo parvient à battre le champion du monde du jeu de Go, Lee Sedol. L'accélération computationnelle en application de la loi de Moore cette fois, permet à ce jour d'avoir des ordinateurs qui traitent des données exaflopiques alors même que nous sommes en train de passer dans l'ère quantique.

La capacité quantique permet(tra) de réaliser des calculs encore plus rapides puisque là où un ordinateur traite des bits classiques qui ne peuvent opérer qu'avec une unité d'information, soit l'état 0, soit l'état 1, l'ordinateur quantique utilise les propriétés quantiques de la matière pour représenter **à la fois l'état de 1 et de 0**. Les qubits (bits quantiques) sont en mesure de prendre deux configurations simultanément au lieu d'une. En pratique, ils pourront par exemple permettre de trouver la fameuse "factorisation" en nombres premiers d'un entier donné, un calcul capital en cryptographie. Mais ils pourraient aussi servir dans des domaines tels que l'aérospatial, la chimie ou l'intelligence artificielle : l'optimisation instantanée d'un ensemble de satellites, la gestion des données pour faire circuler des voitures autonomes.

C'est précisément cette étrange propriété qui leur confère leur incroyable rapidité.

La résolution du **problème du voyageur de commerce**, est un problème d'optimisation qui, étant donné une liste de villes, et des distances entre toutes les paires de villes, détermine un plus court chemin qui visite chaque ville une et une seule fois et qui termine dans la ville de départ. La résolution de ce type de problème prend un temps qui croît de manière exponentielle avec le nombre de villes.

# | Intelligence artificielle, de quoi parle-t-on ?

## Algorithme vs Intelligence artificielle

L'utilisation du mot « intelligence » peut entraîner une erreur sémantique : l'intelligence (artificielle) n'en n'est pas une : elle n'a pas de libre arbitre, pas de créativité et, à ce jour, cela n'apparaît pas même au fur et à mesure de l'augmentation des capacités. Cette appellation nous entraîne dans des représentations anthropomorphiques qui peuvent nous induire en erreur.

Pourtant, il est patent que l'on a cherché à utiliser un mot différent afin de distinguer ce qu'un programme informatique permet de faire de ce que l'IA peut ou pourrait faire.

Un algorithme est une succession d'actions ou d'étapes programmées permettant d'aboutir à un résultat. La précision des étapes est décrite dans des suites de lignes ou de codes qui ne laissent à la machine que l'exécution des instructions. Il faut que le programmeur découpe chacune des étapes, et s'agissant d'un programme informatique qu'il soit traduit en commandes écrites.

L'intelligence est l'ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle permettant de connaître, comprendre et de s'adapter à une situation, à choisir des moyens d'action en fonction des circonstances.

L'intelligence dite « artificielle » se placerait dans l'intermédiaire, ou proche de cette dernière définition, en pouvant reproduire tout ou partie des capacités cognitives des êtres vivants.

## Les systèmes cognitifs artificiels

**L'intelligence artificielle souhaite se rapprocher des capacités cognitives du vivant et principalement celui du cerveau humain en ce qu'il ne lui est pas forcément nécessaire d'avoir toutes les étapes pour comprendre et raisonner.** Comme l'a dit Marvin Minsky lors de la conférence de Darmouth précitée : l'intelligence artificielle est *« la science qui consiste à faire faire aux machines ce que l'homme ferait moyennant une certaine intelligence »*.

Plutôt que de parler « d'intelligence artificielle » certains préfèrent donc parler de « systèmes cognitifs artificiels » pour décrire l'ensemble des technologies qui font référence à des moyens et des mécanismes d'acquisition des connaissances mettant en jeu la mémoire, le langage, le raisonnement, l'apprentissage, l'intelligence, la résolution de problème, la prise de décision, la perception ou l'attention.

L'IA a pour but de détailler l'intelligence en actions cognitives, de les simuler et de les utiliser pour des usages spécifiques.

1. Comprendre son environnement ;
2. Reasonner ;
3. Interagir.

A chaque étape de la capacité cognitive, la machine est capable de répliquer plus ou moins bien :

- | La donnée : la reconnaître, être capable de la figurer, de s'en souvenir...
- | L'information : demande à savoir modéliser et donner une signification à la donnée ;
- | La connaissance : capacité d'apprendre, de mémoriser de faire des corrélations afin de trouver une bonne réponse et prévoir la réponse à des situations répétitives et/ou similaires ;
- | L'intelligence : faculté à s'interroger et à formuler des règles abstraites et de les relier à d'autres.

**Toutefois, sur des capacités prises séparément, les « machines » sont arrivées à une équivalence des capacités humaines suivantes :**

- | En 2016 : reconnaissance d'objet taux d'erreur < à 4% ;
- | En 2017 : reconnaissance vocale → capacité de dictée ;
- | En mars 2017 : traduction simultanée (MIT).

A ce jour, les capacités sont souvent déployées en silo et ne se complètent pas nécessairement les unes les autres : la complexité des capacités cognitives humaines reste, à ce stade, inégalée.

## L'apprentissage :

Pour automatiser l'apprentissage on alimente la machine avec des exemples de la tâche à accomplir, l'homme « entraîne » en quelque sorte la machine à : reconnaître un chat (d'un tigre) un mélanome, un autre véhicule etc... ces « apprentissages » permettent « de faire agir les ordinateurs sans qu'ils aient à être explicitement programmés ».

Il existe différents types d'algorithmes d'apprentissage <sup>5</sup>: on les classe en trois grandes catégories (certaines méthodes peuvent aussi être intermédiaires, ou combinées) :

- | **Apprentissage supervisé** : les données utilisées doivent être annotées préalablement par des « experts ». Lors d'une première phase, dite d'apprentissage, la machine construit un « modèle » de données annotées (un ensemble de règles, un arbre de décision, un ensemble de matrices comme dans les réseaux de neurones, etc.). Ce modèle est ensuite utilisé dans une seconde phase, dite de reconnaissance, dans laquelle par exemple l'algorithme reconnaît l'élément dans une nouvelle image.
- | **Apprentissage non supervisé** : aucun expert n'est requis pour annoter les données. L'algorithme découvre par lui-même la structure des données, en classant ces données en groupes homogènes.

---

<sup>5</sup> CERNA Éthique de la recherche en apprentissage machine : [http://cerna-ethics-allistene.org/digitalAssets/53/53991\\_cerna\\_ethique\\_apprentissage.pdf](http://cerna-ethics-allistene.org/digitalAssets/53/53991_cerna_ethique_apprentissage.pdf)

- | **Apprentissage par renforcement** : le but est d'apprendre, à partir d'expériences, ce qu'il convient de faire en différentes situations, de façon à optimiser une récompense quantitative au cours du temps. L'algorithme procède par essais et erreurs, chaque erreur l'amenant à améliorer sa performance dans la résolution du problème. Le rôle des experts se limite ici à définir les critères de succès de l'algorithme.

Cette « éducation » met donc en exergue l'importance du formateur.

# | Pourquoi les questions d'éthique sont-elles au cœur des débats ?

Stephen Hawking, physicien théoricien britannique, «réveille» la communauté scientifique en 2014 en disant que : *« Parvenir à créer des intelligences artificielles serait le plus grand accomplissement de l'histoire humaine. Malheureusement, il pourrait aussi s'agir du dernier, à moins que nous apprenions à éviter les risques. »*.

S'en suit une rencontre Beneficial AI organisé par le Future of Life Institute (FLI, institut pour le futur de la vie) à Asilomar, en Californie où sont arrêtés les « 23 principes d'Asilomar »<sup>6</sup> qui reçoit l'adhésion de 2000 signataires, dont l'astrophysicien Stephen Hawking lui-même, qui se veut être un guide de référence pour un développement éthique de l'intelligence artificielle.

## L'éthique en science

Depuis l'injonction de Rabelais : *« science sans conscience n'est que ruine de l'âme »*, la question de la morale et de l'éthique se pose avec acuité à chaque saut technologique.

Jean-Gabriel Ganascia, professeur à l'université Pierre et Marie Curie, informaticien, spécialiste d'intelligence artificielle et Président du comité d'éthique du CNRS<sup>7</sup>, a souligné lors son audition qu'en sciences on se fonde sur une approche par les principes (accord sur les principes de la protection de la vie privée, la non-discrimination etc...). Or, l'innovation entre souvent en contradiction avec ces principes ; l'éthique doit gérer les tensions entre des principes légitimes et les possibilités d'innover. Dans le cas de l'intelligence artificielle un obstacle à cette articulation vient de l'extrême rapidité avec laquelle cette mutation se fait ! En effet, l'étymologie du mot « éthique » vient du grec ancien « ethos » (l'étable) ou autrement dit le caractère habituel, la manière d'être, les habitudes d'une personne. L'éthique est liée aux coutumes. Mais il n'y a pas de coutumes en la matière car il n'y pas d'identité entre la société des parents et celles des enfants : le chantier est donc immense.

En tout état de cause, il lui semble primordial de ne pas règlementer sur les visions prospectives qui pourraient être erronées (notamment des « visions anthropomorphiques » en ce qui concerne l'IA).

## L'éthique en France

En France, les évolutions technologiques dans le domaine médical et particulièrement dans celui de la génétique entraînent la création en 1983 du Conseil Consultatif national d'Éthique<sup>8</sup> pour les sciences de la vie et de la santé.

---

<sup>6</sup> Les 23 principes d'Asilomar <https://futureoflife.org/ai-principles/?cn-reloaded=1>

<sup>7</sup> <http://www.cnrs.fr/comets/spip.php?article169>

<sup>8</sup> <https://www.ccne-ethique.fr/>

Le législateur a créé une institution de réflexion ayant pour objet de l'éclairer sur les impacts du franchissement de nouvelles limites du possible par les progrès de la science et de leur impact sur le « Contrat social » et l'acceptabilité de ces progrès.

Aujourd'hui même, dans le cadre du projet de loi relatif à la bio éthique (adopté en première lecture à l'Assemblée Nationale), en cours de débat devant le parlement, l'article 11 est rédigé ainsi :

#### Article 11

« Art. L. 4001-3. – I. – Lorsque, pour des actes à visée préventive, diagnostique ou thérapeutique, est **utilisé un traitement algorithmique de données massives, le professionnel de santé** qui communique les résultats de ces actes **informe la personne de cette utilisation et des modalités d'action de ce traitement.**

« II. – L'adaptation des paramètres d'un traitement mentionné au I pour des actions à visée préventive, diagnostique ou thérapeutique concernant une personne **ne peut être réalisée sans l'intervention d'un professionnel de santé.**

La rapidité de résolution de problèmes par l'intelligence artificielle que l'intelligence humaine, y compris soutenue par la puissance de calcul informatique, peut créer une angoisse allant jusqu'au rejet potentiel. Car les intelligences artificielles questionnent notre manière de fonctionner en société : qui prend les décisions ? Comment ? Pour quelle responsabilité et quelle humanité ?

Comment accepter d'être conduit par un véhicule autonome, d'avoir une prescription médicale délivrée par une IA..., voilà des choix majeurs ! Structurellement l'adoption des technologies d'IA va poser des questions éthiques :

- I Par leur création (l'éthique by design) dès la conception ;
- I Par leurs usages : pourquoi s'en sert-on ?
- I Sociétale : pour quelle transformation de la société et pour quel coût social ?

➔ **Les entreprises ne sont pas exemptes de ce questionnement dans la mesure où elles utilisent, et utiliseront de plus en plus d'outils d'IA, et qu'elles en créeront également.**

## En entreprise

Dans une étude menée sur plus de 1500 dirigeants, 500 organisations et 4000 consommateurs dans le monde, Capgemini Research Institute<sup>9</sup> nous révèle que 86% des dirigeants sont conscients de situations où l'IA a pu poser des problèmes éthiques dans ces 2 à 3 dernières années ; presque la moitié des consommateurs se sont sentis exposés à des situations où l'IA aurait osé un problème éthique et que trois-quarts d'entre eux souhaitent de nouvelles réglementations à cet égard.

L'entreprise ne pourra pas faire d'impasse sur les outils d'intelligence artificielle et beaucoup s'en sont déjà saisis. En permettant d'automatiser des décisions, les capacités d'augmenter la rapidité de celles-ci et les possibilités de proposer de nouvelles solutions s'accroissent considérablement. Des entreprises sont même en

<sup>9</sup> [https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2019/07/CRI-AI-in-Ethics\\_Web.pdf](https://www.capgemini.com/fr-fr/wp-content/uploads/sites/2/2019/07/CRI-AI-in-Ethics_Web.pdf)

train de créer des réponses à base d'IA à des situations propres à leur domaine et créent ainsi de nouveaux outils qu'elles sont en capacité de revendre à leur secteur. Différents exemples nous ont été présentés qui ne semblent pas a priori questionner les notions d'éthiques :

- | Cybersécurité : des outils qui détectent les intrusions sur les systèmes avec de l'intelligence artificielle,
- | Aviation : création de solutions de gestion des retards d'avions.
- | Santé : l'analyse des devis dentaire ou optiques...

Mais d'autres en posent plus directement :

- | Les intelligences qui scannent automatiquement les CV de candidatures à un emploi.
- | Des systèmes intégrés de surveillance (existent déjà pour des postes critiques où l'attention de l'opérateur est vitale, l'outil peut détecter une baisse d'attention)
- | Création de prix très personnalisés<sup>10</sup>

Enfin, des questions et inquiétudes sur les transformations de l'emploi induites par cette révolution technologique sont placées au cœur des réflexions. Le Syntec numérique a communiqué l'étude sur l'intelligence artificielle réalisée par leur Observatoire de branche (OPIIEC) « Formations et compétences sur l'Intelligence Artificielle en France » et le MEDEF est partie prenante de la création d'un EDEC de l'IA par le biais de la coalition numérique.

**Recommandation : encourager les entreprises à inclure les questions d'éthique des systèmes informationnels dans leurs outils de gouvernance.**

## Des solutions transposables ou déjà mises en place

Parmi les entreprises auditionnées la plupart ont mis en place ou démarré un processus interne de revue de leurs outils d'IA :

**IBM** a créé un comité international de déontologie

**Thales** a fait une charte interne :

- | Pour répondre aux préoccupations des salariés
- | Pour attirer les talents

Pour David Gruson, directeur de programme, membre du Comité exécutif du **Groupe Jouve**, les 5 clés de régulation positive de la robotisation en santé sont :

- | L'information du patient : existe déjà complété par le fait que la recommandation vient d'une IA → garantie humaine de l'IA ;

<sup>10</sup> Fondapol : Vers des prix personnalisés à l'heure du numérique Octobre 2019



- | La garantie humaine ;
- | Le collège de garantie
- | Le pilotage avec révision par la FDA : maintien de la validation si processus de garantie de révision humaine.

Pour **Thales** : dans le B to B nous sommes dans la course alors que nous ne le sommes pas dans le B to C.

Il faut que la France comprenne bien les enjeux B to B.

## Ethique des usages de l'IA

L'éthique des usages vise à s'interroger sur les usages du numérique que font les clients, les collaborateurs, mais aussi les managers et les partenaires d'une entreprise. Il s'agit donc de mener une réflexion éthique sur l'usage que fait l'homme des ressources technologiques qu'il a à sa disposition.

L'IA et les algorithmes en général consistent à automatiser des tâches accomplies par les humains et même agir et décider en lieu et place des humains. L'intelligence artificielle permet donc de faire des tâches usuellement opérées par un humain et qui existent donc déjà en tant que telles, leur adoption sous format IA est sans doute déjà acquise.

En revanche, l'intelligence artificielle permet aussi de proposer des services ou des biens qu'il n'était pas possible de proposer ou à des coûts prohibitifs : la question se pose de savoir s'il est nécessaire de les regarder spécifiquement ? Faut-il lister/légiférer sur des finalités interdites (par exemple en assurance pas de tarifications individuelles) ?

- | **La reconnaissance faciale** ou l'usage de données biométrique (voix, ADN...) font partie des sujets sensibles. Si les données personnelles sensibles sont protégées spécifiquement dans le cadre de la régulation des données personnelles (niveau d'information, d'acceptation formelle), leur usage n'est pas pour autant prohibé, limité ou contrôlé dans leur finalité.

A ce jour, l'ouverture d'un smartphone se fait déjà par la reconnaissance faciale : n'est-ce pas déjà un pas de franchi ? Pourtant son usage par des services publics à des fins d'authentification pose questions : sur la reconnaissance faciale, le Conseil national du numérique a été saisi et opère une série de consultations publiques.

- | **Les chatbots : les agents conversationnels** (bots) ou *chatbots*, qui sont des agents de traitement automatique de conversation en langage naturel sont de plus en plus présents en tant qu'assistants personnels ou interlocuteurs dans des transactions commerciales. Il arrive aussi qu'ils soient majoritaires dans des chats avec les humains.

Mais l'échange entre humains ne se limite pas à la capacité de compréhension du langage parlé, l'intonation de la voix, son volume décèle des signes de stress, peur, frustrations, ou enthousiasme, joie, ... qui disent quelque chose sur le ressenti : comment le *chatbot* sait que le client est content, impatient, perdu ?



Mais leur utilisation ne se limite pas à des interactions basiques (renseignements, échanges sur une proposition commerciale, suivi de dossier...) mais certains vont plus loin comme la représentation de personnes décédées ?

- ▶ Faut-il poser un cadre général à ces usages ? Une revue préalable ?
- ▶ Fixer des dispositifs généraux comme respecter l'intégrité mentale et physique des humains !

**Recommandation : décliner au sein du Conseil consultatif une enceinte pour les sciences numériques<sup>11</sup>.**

## Ethique by design

**L'éthique by design s'intéresse à la phase de conception des outils numériques.**

Elle concerne la technologie dans sa complexité technique et le savoir-faire des ingénieurs, programmeurs, etc... Elle aborde en particulier la déontologie des concepteurs numériques quels qu'ils soient (développeurs, designers numériques, chef de projet...) qui ont une responsabilité éthique dès la conception, dans la mesure où les données ou les algorithmes, peuvent ou non reproduire des biais humains, faire apparaître de nouvelles discriminations ou en reproduire à plus grande échelle, aboutir à des injustices, etc<sup>12</sup>...

Si « code is law », cela veut-il dire que ces questions d'éthique doivent être incluses dans le système ?

Toutefois, il ne faudrait pas que l'éthique soit confisquée par un seul groupe : celui des concepteurs d'IA.

**Recommandation : l'éthique des systèmes d'IA ne doit pas reposer seulement sur les concepteurs des IA mais engager des équipes pluri disciplinaires.**

### Les biais

La question du biais des données revient souvent et pour cause : le « stock » des données utilisées pour alimenter les IA ou les éduquer vont être le reflet de la société et donc des biais de discriminations qui existent.

Faut-il réécrire l'histoire ? Faut-il éliminer dès le départ les biais ? Faut-il corriger ex post ?

Celui ou celle qui entraîne les algorithmes a donc une responsabilité importante et son éducation aux enjeux éthiques semble indispensable<sup>13</sup>.

<sup>11</sup> Alors que cette position est établie, le CCNE (Comité consultatif national d'éthique) a été chargé de constituer un comité pilote d'éthique du numérique : [https://www.ccne-ethique.fr/sites/default/files/communiquelancement\\_comite\\_numerique.pdf](https://www.ccne-ethique.fr/sites/default/files/communiquelancement_comite_numerique.pdf)

<sup>12</sup> CIGREF et Syntec numérique, Ethique et numérique, oct. 2018 <https://www.cigref.fr/ethique-numerique-un-referentiel-pratique-pour-les-acteurs-du-numerique>

<sup>13</sup> Le risque de renforcement des biais est traité par le livre de Cathy O'Neil « Weapons of Math Destruction » qui traite des impacts sociétaux et du risque de renforcement des inégalités [https://en.wikipedia.org/wiki/Weapons\\_of\\_Math\\_Destruction](https://en.wikipedia.org/wiki/Weapons_of_Math_Destruction)

## **Recommandation : les « éducateurs » d'IA doivent être sensibilisés à l'éthique.**

### **L'enfermement ou la prédiction auto réalisatrice**

Le risque existe que des personnes soient enfermées dans des « profils types » qui systématisent le traitement de leur situation. En vous « profilant », le système vous catégorise et peut conduire à faire se réaliser ce qu'il a prédit. L'exemple de l'orientation professionnelle est un des domaines où le risque existe notamment.

Dans le domaine informationnel, cet enfermement crée une bulle qui peut vous donner l'impression d'être informé lorsqu'un large pan de l'actualité ne vous est pas adressé puisqu'elle ne sera pas considérée comme pouvant vous plaire ou vous intéresser. Ce n'est plus vous qui choisissez en conscience une ligne éditoriale mais celle-ci qui s'impose à vous à votre insu et joue du « biais de confirmation ».

Par exemple : une intelligence artificielle qui vous recommande des livres n'aurait pas, a priori, de risque décisionnel ou sociétal majeur, sauf si elle conduit à enfermer les personnes dans une pensée unique voire à les manipuler.

### **Individualisation et personnalisation vs mutualisation et vie collective<sup>14</sup>**

Les services numériques nouveaux permettent une instantanéité et une personnalisation extrême dont les effets à grande échelle n'ont pas été anticipés ou évalués. Divers risques ont ainsi été évoqués lors des auditions :

- I Laurence Devillers, professeur à Sorbonne Université, chercheuse en Intelligence artificielle nous a parlé lors de son audition du « nudging » ou « incitation à faire » qui consiste à donner une seule alternative et faire comme la machine le dit.
- I Le risque d'aller vers une « algocratie » : où l'on déléguerait de plus en plus de tâches bénignes et de décisions, mais qui mises bout à bout constituent notre être. Si la technocratie est le système tant critiqué dans lequel les décisions sont prises en privilégiant les données techniques par rapport aux facteurs humains et sociaux, l'« algocratie » serait celui où les décisions sont prises sur le fondement d'algorithmes ?
- I Les effets collatéraux indésirables : l'hyper personnalisation conduit à prendre des décisions individuelles qui mises bout à bout sont néfastes à la communauté : par exemple le système GPS qui indique de prendre un autre chemin qui se révèle inadapté pour des passages réguliers.

Il conviendra d'inclure dans la gouvernance des IA une évaluation des risques sociétaux quand les résultats poussés à leur extrêmes arrivent à des situations individuelles contraires aux droits humains ou collectifs ou contre productifs.

Face à ces nouveaux questionnements il ne semble pas possible de trouver de proposer des réponses immédiates. A priori, il n'y a pas de possibilité d'automatiser l'éthique. De plus, cette dernière consiste justement à faire des choix en fonction de

---

<sup>14</sup> Comment permettre à l'homme de garder la main ? Enjeux éthiques des algorithmes et de l'IA CNIL  
Décembre 2017

critères de valeurs dans des cas qui n'étaient pas prévus et où des dilemmes se font jour.

Une segmentation des différents types d'IA et des différents types de risques serait nécessaire afin d'anticiper autant que possible, et corriger dès que nécessaires, les effets indésirables pour les individus et la collectivité.

A cet égard, pour les systèmes décisionnels, beaucoup de nos interlocuteurs ont parlé de la nécessité d'une garantie humaine (en matière de santé ou quand un risque létal est possible). Pourtant il ne semble pas utile que cette disposition, tout à fait adéquat dans certains cas, soit imposée comme une règle générale.

#### **Recommandations :**

- ▶ **Intégrer une évaluation des risques ex ante et/ou ex post et décliner les obligations en fonction du niveau d'impact de l'IA ;**
- ▶ **Ne pas faire du contrôle humain un critère systématique de garantie.**

## **Une RSE de l'IA ?**

Enfin, au-delà des usages et de la constitution même de l'IA, une éthique sociétale interroge cette technologie et les impacts du numérique à un niveau macroscopique. Elle traite ainsi, notamment, de l'acceptabilité des innovations et solutions numériques, de l'économie et du design de l'attention, et de l'empreinte écologique et énergétique du numérique, de l'inclusion sociale, etc...

En addition aux questions liées aux biais des jeux de données, des questions d'usage et de responsabilité, les questions de l'impact environnemental et social de l'IA doivent être évoquées.

Ce sont des sujets que nous n'avons pas encore creusés suffisamment pour proposer des recommandations mais nous pouvons acter qu'ils sont majeurs et doivent être regardés.

# | Une IA de confiance ?

Le rapport européen s'intitule dans sa version française : « Lignes directrices en matière d'**éthique pour une IA digne de confiance** »<sup>15</sup>

Quels seraient les critères qui permettraient de donner confiance dans la technologie et qui ne se transformeraient pas en nouvelles contraintes réglementaires potentiellement contraires à l'innovation et néfastes à l'émergence de solutions européennes compatibles avec nos valeurs, nous conduisant irrémédiablement prendre sur « étagère » des solutions déjà toutes faites ?

Nozha Boujemaa, directrice Scientifique et de l'Innovation de Median Technologies, qui a été très impliquée dans le groupe de travail de la Commission européenne attire notre attention sur le fait que le sujet était de mettre en pratique de l'IA de confiance.

L'IA de confiance est plus large et plus opérationnelle que l'éthique de l'IA car elle recouvre aussi la **performance, la qualité de robustesse et de répétabilité.**

Ainsi, dans une approche pragmatique et aussi pédagogique, le groupe d'experts européens a proposé une liste de questions que la Commission européenne a soumise à consultation. Cette liste de plus de 100 questions est séquencée dans plusieurs domaines :

- | Robustesse technique : résilience, sûreté de fonctionnement, sécurité, reproductibilité ;
- | Respect de la vie privée et gouvernance des données : qualité, intégrité ;
- | Transparence : traçabilité ;
- | Diversité et non-discrimination : absence de biais, accessibilité, conception, participation des parties prenantes ;
- | Bien être sociétal et environnemental ;
- | Responsabilité : audibilité, recours...

## **Quels outils d'analyse et d'évaluation du niveau de confiance :**

Toutefois à l'épreuve de leur relecture il semble que toutes ces questions ne soient pas forcément pertinentes eu égard au domaine d'activité/ au type d'IA en présence / à l'usage auquel elle est destinée.

Dès lors la question se pose de savoir si globalement, ou pour chaque outil IA, l'entreprise devra le passer au tamis de tout ou parti ce questionnaire ?

Est-il possible de tirer de cette liste un schéma, y compris lui-même programmable, de construction / révision de l'IA ?

Une chose apparaît clairement la typologie des risques mérite une différence dans l'approche. Mais nous avons vu également que, même peu risquée, une IA peut se révéler néfaste si son échelle d'usage conduit à changer son environnement.

---

<sup>15</sup> [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60427](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60427)

Il faut sans doute trouver une approche plus synthétique est-ce :

- | Une approche par les risques ?
- | Une approche par les volumes ?
- | Autre ?

**Recommandation : structurer les questions proposées par le groupe d'experts européens par thématiques**

### **La confiance dépendante du rôle de l'humain ? de la loi ? :**

En santé, non seulement les textes sont en cours d'évolution pour le prévoir mais la « garantie » humaine semble être évidente pour les professionnels de ce secteur que nous avons reçus.

Pour David Gruson directeur du programme santé, membre du comité exécutif du groupe Jouve et fondateur de l'initiative académique et citoyenne Ethik-IA, les 5 clés de régulation positive de la robotisation en santé sont :

- | L'information du patient : existe déjà complété par le fait que la recommandation vient d'une IA
- | La « garantie humaine »
- | Collège de garantie
- | Pilotage avec révision (par ex la FDA : maintien de la validation si processus de garantie de révision humaine)

Pour Nozha Boujemaa, une IA dans un domaine médical par exemple doit être aussi efficace dans un hôpital donné que dans un autre : les critères de robustesse et de répétabilité apparaissent comme majeurs pour estimer que les outils d'intelligence artificielle mis à la disposition de professionnels sont pertinents et utiles. A cet égard elle attire notre attention sur l'approche prise par la FDA américaine<sup>16</sup> : autorité de régulation avec les autorisations de mise sur le marché de médicament ont fait un rapport « software as medical device » ils ont fait circuler une consultation internationale. Le travail en cours sur ce sujet mais à ce stade n'ont pas établi de synthèse.

L'avis n°129 du comité consultatif national d'éthique préconise d'inscrire le « principe fondamental d'une garantie humaine du numérique en santé » au niveau législatif.

Cette garantie pourrait notamment être assurée « par la reconnaissance de préserver la maîtrise finale du professionnel de santé, en interaction avec le patient, pour prendre les décisions appropriées en fonction de chaque situation spécifique ».

---

<sup>16</sup> <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#regulation>

Le comité consultatif national d'éthique estime également que toute personne doit pouvoir être informée préalablement du recours à un algorithme d'aide à la décision médicale dans son parcours de prise en charge en santé.

*Pour Thales*, dans les systèmes critiques l'IA reste un support de l'intervention humaine, il n'est pas question de perdre le contrôle dans le système autonome.

**Recommandation : ne pas faire du contrôle humain un critère systématique de garantie**

De même, a priori, les dispositifs doivent se conformer aux dispositions législatives en vigueur, mais peut-on le garantir en ce qui concerne l'IA ? Pas forcément, car ils peuvent être en partie imprévisibles. L'ajout de prescriptions pourrait éviter des effets inacceptables, ces prescriptions ne peuvent pas être apprises en tant que telles mais doivent être imposées.

### Quels critères spécifiques pour l'IA ?

Les questions posées peuvent s'appliquer non pas seulement aux IA mais aux traitements automatiques de données au sens large. Dès lors ne faut-il pas avoir une étape plus générale sur la gouvernance informatique ? Notamment la loyauté et l'équité des systèmes informatiques est un socle de confiance de base essentiel :

- I **La loyauté des systèmes** informatiques qui signifie que ces systèmes se comportent, au cours de leur exécution, comme leurs concepteurs le déclarent. Si par exemple, ceux-ci déclarent qu'un système n'archive pas les données personnelles de ses utilisateurs, il ne doit pas le faire.
- I **L'équité d'un système** informatique consiste en un traitement juste et équitable des usagers.
- I **La conformité** : Un système numérique doit demeurer conforme à son cahier des charges, et son cahier des charges doit être conforme à la législation. La conformité d'un système à son cahier des charges signifie que le système est conçu pour effectuer des tâches spécifiées en respectant les contraintes explicitées dans ce cahier. Les spécifications du cahier des charges formalisent souvent une interprétation restrictive de la loi, à défaut de pouvoir traduire les finesses de cette dernière. La conformité doit être vérifiée avant que le système ne soit utilisé, en analysant son code et ses données. Par exemple, la conformité inclut le respect des règles de protection des données personnelles pour un système d'analyse de données ou le respect du code de la route pour la voiture autonome.

### La certification vs l'auditabilité

La complexité de l'IA fait que les concepts de transparence et l'explicabilité ont été quasi unanimement rejetés car mal adaptés à l'IA dont les résultats ne sont pas tous explicables, voire pas du tout.

**La transparence d'un système** signifie que son fonctionnement n'est pas opaque, qu'il est possible, par exemple pour un utilisateur, de vérifier son comportement. Cette transparence s'appuie notamment sur la traçabilité, la mise à disposition d'informations sur ses actions suffisamment détaillées (mémorisées dans un journal) pour qu'il soit possible après coup de suivre ses actions. La traçabilité est essentielle d'une part pour les attributions de responsabilité, où elle permet le cas échéant de fonder un recours juridique, et d'autre part pour diagnostiquer et corriger les erreurs. Elle permet également d'expliquer le fonctionnement d'un système à partir des traces laissées par celui-ci ; on parle d'explicabilité.

En ce qui concerne **l'explicabilité** : la question nécessite de savoir à qui l'on s'adresse : parler d'explicabilité mais pour qui ? pour le développeur ? pour l'utilisateur ? Expliquer ce n'est pas possible aujourd'hui c'est même un sujet de recherche (cf les travaux de Transalgo : reprise par le SGPI<sup>17</sup>).

Quant à la **certification**, Jean-Gabriel Ganascia professeur à l'université Pierre et Marie Curie, informaticien, spécialiste d'intelligence artificielle, président du COMETS (Comité d'éthique du CNRS), nous dit qu'elle s'applique mal à l'IA : en effet, si on veut établir un lien de confiance il faudrait des organismes d'homologation, et de certification extérieure et qui seraient capables d'analyser un système. Mais comme indiqué précédemment l'IA inclut une part de non explicabilité et en devenant apprenante, il faudrait renouveler cette certification très fréquemment ce qui conduit à freiner l'innovation (trop long trop coûteux). Pour ce chercheur ce n'est pas la voie idoine.

En revanche beaucoup se sont prononcés pour **l'auditabilité** : que les IA puissent être contrôlables dans la robustesse et la répétabilité de leurs résultats.

**Recommandation : privilégier l'auditabilité de la robustesse et de la répétabilité plutôt que la certification des IA**

L'ensemble des personnes auditionnées ont convergé vers le souhait d'éviter que l'éthique soit réduite à un jeu de contraintes « liste de cases à cocher » pesant, inadapté, incompréhensible, inutilisable, trop complexe pesant sur les équipes opérationnelles ; en plus des nombreuses normes de conformité déjà existantes. Plus généralement les régulations existantes devraient être ajustées si nécessaires et leur lecture facilitée par une aide récapitulative.

**Recommandation : pas de régulation type « RGPD » pour l'IA mais une mise en perspective des législations en place et leur adaptation si nécessaire**

<sup>17</sup> <https://www.transalgo.org/>



# | Les réglementations qui s'appliquent déjà, et celles qui seraient à réexaminer

## La protection des données personnelles

La loi informatique et liberté et la loi régissent les modalités de collecte, traitement et conservation des données personnelles (voir tableaux ci-après). Or l'intelligence artificielle nécessite beaucoup de données, parfois personnelles, et sur de longue durée : cela semble a priori incompatible avec l'IA.

### | Article 1<sup>er</sup> du chapitre 1 de la loi informatique et liberté de 1978 modifiée en Juin 2018

**L'informatique** doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

### | Article 6

**Il est interdit de traiter des données à caractère personnel** qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique ou **de traiter des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique**, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

### | Article 21

**Aucune décision produisant des à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur de données à caractère personnel (...).**

On constate que la loi prévoit et contraint déjà fortement les entreprises. A cet égard, beaucoup des dispositifs législatifs existent et ne sont donc pas à réécrire. En revanche, le développement de l'IA interroge la vision initiale de la loi au regard des nouveaux développements et de sa compatibilité avec la nécessité d'entrer dans la course mondiale pour la capacité de développement d'IA.

## Responsabilité civile et Intelligence artificielle avons-nous les bons outils ?

La Cour d'Appel de Paris dans un rapport publié en Avril 2019 avec l'Université Versailles Saint-Quentin-en-Yvelines s'est penchée sur la question (page 107) :

*« Des différentes réflexions qui ont eu lieu ou sont encore en cours sur ce sujet en constante évolution et développement, deux courants se dessinent clairement :*



- | le recours à l'un des fondements existants de la responsabilité civile objective suffirait à régler la grande majorité des cas,
- | seule, la reconnaissance de la personnalité juridique aux robots permettrait de répondre aux enjeux de l'intelligence artificielle, sans qu'à ce stade aucun n'emporte totalement la conviction. »

### **Les fondements existants de la responsabilité civile objective suffiraient ?**

Jusqu'à ce jour, le cadrage classique de la responsabilité du fait des choses et le régime spécial de responsabilité du fait des produits défectueux ont jusqu'ici permis d'encapsuler globalement les effets juridiques de l'intelligence « artificielle » et ce tant que celle-ci reste « faible ». La Directive 85/372/CE sur les produits défectueux et sa transposition dans le droit français a rempli ses objectifs en matière de sécurité juridique de l'ensemble de acteurs y compris des consommateurs dans le domaine de nouvelles technologies. Les systèmes logiques (logiciels) ont été inclus par la jurisprudence européenne dans les potentiels produits défectueux (décembre 2017).

Ainsi, la complexité des produits ne change pas les principes fondamentaux et la logique de responsabilité définies par ce texte. Il faut, prioritairement, éviter de complexifier la tâche du consommateur et de l'utilisateur dans l'identification de sous-responsabilités.

Toutefois, la responsabilité du producteur du fait des produits défectueux peut être engagée pour défaut de conception ou d'exécution d'un produit. Que répondre aux « objections suivantes : *quid de la détermination de la défectuosité au moment de la commercialisation (si le système évolue entre sa commercialisation et la survenance du dommage), et de l'établissement du lien de causalité entre le dommage et le défaut et plus globalement à l'application de ce régime aux biens immatériels.* » ? Dans les cas de l'IA apprenante, quid du dommage et de son indemnisation ? Car « la question de la cause d'exonération fondée sur le risque de développement au champ d'application jusqu'ici très limité pourrait prendre une autre dimension dans ce nouveau contexte » nous dit la Cour d'Appel. Selon notre analyse, c'est au concepteur d'encadrer la faculté d'apprentissage, afin de garantir tout au long du cycle de vie du produit un niveau de sécurité adéquat.

### **Il faudrait reconnaître une personnalité juridique aux robots ?**

D'où « La tentation de certains (v. la résolution du Parlement européen du 16 février 2017) de pousser le raisonnement jusqu'à l'extrême en cas d'intelligence artificielle autonome dite « forte » de reconnaître la personnalité juridique aux robots autonomes les plus sophistiqués ».

Cette approche ne nous semble pas être la bonne dans la mesure où l'AI n'a pas de volonté propre, le scénario de l'AI forte relevant de la spéculation. Par ailleurs, le problème ne serait que déplacé puisque qui dit personnalité juridique dit patrimoine et qu'il conviendrait alors de trouver les moyens d'abonder ce patrimoine par des personnes physiques ou morales. A ce sujet, le même rapport dit qu'on « peut douter de la pertinence de cet artifice » !

### **Retour à la responsabilité pour faute ?**

Le retour à l'application pure et simple d'une responsabilité pour faute est également examiné par la Cour d'Appel mais là encore son application à des situations où une intelligence artificielle dite forte aurait entraîné des dommages pourrait toutefois placer les victimes de ces dommages corporels dans une situation très difficile, car la

preuve de l'imputabilité du dommage risque de devenir particulièrement complexe en raison de la sophistication de l'intelligence artificielle. Le travail d'expertise va se déplacer du programmeur vers le spécialiste de l'organisation et du traitement des données et cela supposera de disposer des données d'apprentissage et d'en connaître l'origine pour déterminer l'imputabilité du dommage...

➔ **A ce stade de développement de l'intelligence artificielle, il semble que le droit de la responsabilité civile actuelle reste opérant, mais un travail en continu dans ce domaine est nécessaire afin de ne pas laisser d'éventuelles victimes dans des situations inacceptables.**

# Tableau

Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données	LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles
Article 22 Décision individuelle automatisée, y compris le profilage	Article 21 I.-L'article 10 de la loi n° 78-17 du 6 janvier 1978 précitée est ainsi rédigé :
<p>1. La personne concernée a <b>le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire.</b></p> <p>2. Le paragraphe 1 ne s'applique pas lorsque la décision :</p> <p>a) est nécessaire à la conclusion ou à <b>l'exécution d'un contrat</b> entre la personne concernée et un responsable du traitement ;</p> <p>b) est autorisée par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée ; ou</p> <p>c) est fondée sur <b>le consentement explicite</b> de la personne concernée.</p> <p>3. Dans les cas visés au paragraphe 2, points a) et c), le responsable du traitement met en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision.</p> <p>4. Les décisions visées au paragraphe 2 ne peuvent être fondées sur les catégories particulières de données à caractère personnel visées à</p>	<p>« Art. 10.-<b>Aucune décision de justice</b> impliquant une appréciation sur le comportement d'une personne <b>ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne.</b></p> <p>« <b>Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage, à l'exception :</b></p> <p>« 1° Des cas mentionnés aux <b>a et c du 2 de l'article 22 du règlement (UE) 2016/679</b> du Parlement européen et du Conseil du 27 avril 2016 précité, sous les réserves mentionnées au 3 du même article 22 et <b>à condition que les règles définissant le traitement ainsi que les principales caractéristiques de sa mise en œuvre soient communiquées</b>, à l'exception des secrets protégés par la loi, par le responsable de traitement à l'intéressé s'il en fait la demande ;</p> <p>« 2° Des décisions administratives individuelles prises dans le respect de l'article L. 311-3-1 et du chapitre Ier du titre Ier du livre IV du code des relations entre le public et l'administration, à condition que le traitement ne porte pas sur des données mentionnées au I de l'article 8 de la présente loi. Ces décisions comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du code des relations entre le public et l'administration. <b>Pour ces décisions, le</b></p>

<p>l'article 9, paragraphe 1, à moins que l'article 9, paragraphe 2, point a) ou g), ne s'applique et que des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée ne soient en place.</p>	<p><b>responsable de traitement s'assure de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard.</b></p> <p>« Par dérogation au 2° du présent article, aucune décision par laquelle l'administration se prononce sur un recours administratif mentionné au titre Ier du livre IV du code des relations entre le public et l'administration ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel. »</p>
--	---

# | Exemple : La justice<sup>18</sup>

## MAAD : modes algorithmiques d'analyse des décisions

L'article 10 de la loi du 6 janvier 1978 modifiée par la loi du 20 Juin 2018 (voir tableau supra), prévoit **qu'aucune décision produisant des effets juridiques à l'égard d'une personne ne peut être prise sur le seul fondement d'un traitement automatisé des données.**

Parmi les services offerts par les outils numériques dans le domaine du droit et de la justice, on peut distinguer notamment :

- | La résolution de litiges en ligne ;
- | L'analyse de tendances dans le but d'extraire des connaissances pour anticiper et orienter un processus décisionnel ;
- | La catégorisation des contrats selon différents critères et la détection de clauses contractuelles divergentes ou incompatibles ;
- | Les « chatbots » de renseignement du justiciable ou de support dans le déroulement du règlement des litiges ;
- | Les moteurs de recherche de jurisprudence ;
- | L'aide à la rédaction d'actes.

Les outils algorithmiques d'analyse des décisions répondent aux deux premières fonctions

Cette possibilité s'accroît non seulement avec les capacités de calcul et le développement des solutions d'IA mais aussi parce que la loi prévoit, depuis la loi sur une république numérique (article 21), l'ouverture des données de justice au public à titre gratuit et dans le respect de la vie privée de personnes concernées.

**Ces outils portent en germe la promesse d'une justice de qualité** (rapide, prévisible), avec une meilleure acceptabilité, le tout dans un contexte de confiance renforcée et dans l'intérêt conjugué des justiciables, de leurs conseils et des juges.

### Les cas d'usages :

Les MAAD permettent d'appréhender la potentialité de la réponse judiciaire compte tenu d'un certain nombre de variables ou de précédents. Ils sont des aides à la prise de décision qui peuvent être très efficaces dans des cas récurrents.

- | Ainsi par exemple au Canada, la province de Colombie britannique utilise les outils algorithmiques pour les contentieux de masse ou du quotidien (consommation, voisinage, licenciements...) et inférieurs à une certaine valeur.

---

<sup>18</sup> Synthèse d'éléments issus de la mission de recherche Université Nice Sophia Antipolis, ENS, CNRS, INRIA : « [Comment le numérique transforme le droit et la justice par de nouveaux usages et un bouleversement de la prise de décision. Anticiper les évolutions pour les accompagner et les maîtriser](#) ».

Toutefois cette technologie ne recouvre pas tous les contentieux : une règle de procédure civile les encadre. Ainsi ces technologies ne s'appliquent pas aux décisions qui nécessitent un traitement individuel et sont contingentées aux règlements de litiges juridiquement « analogues ».

### **Une nouvelle philosophie du droit ?**

Les Legal / reghteck doivent trouver leur passage entre « prophétie auto réalisatrice » qui prétendrait que la solution ayant été actée elle l'est une fois pour toute (Il ne s'agit plus de dire ce qui est probable mais ce qui doit être et on entre alors dans une dictature du droit) et outil d'aide à la décision.

Cette dictature pourrait d'autant plus effrayante que non seulement elle dirait la règle mais elle serait capable de dire, avant même que vous le sachiez, que vous ne la respectez pas et voir même prévenir les autorités de contrôle et vous punir. Une sorte de « contrôle radar permanent » à grande échelle.

Il faut donc réfléchir à une possibilité de remonter à l'autorité de régulation de telles solutions, les mises à jour législatives, les erreurs, les demandes d'actualisation.

### **La régulation à prévoir**

La confiance dans un tel système pourrait être fonction de sa transparence, or les systèmes sont par essence même peu ou pas transparents y compris à leur concepteur. La transparence supposerait alors une lisibilité des méthodes et une explicabilité des résultats.

Et même dans ce cas la question de leur autorisation (dans le cas de systèmes utilisés par les juges et donc par l'Etat) se pose. En effet, il faudra s'interroger des modes d'intégration des décisions au fur et à mesure qu'elles sont prises et fonction du niveau de l'autorité judiciaire qui la prend.

# | ANNEXES

## | 23 principes d'Asilomar<sup>19</sup>

2.000 signataires, dont l'astrophysicien Stephen Hawking et Elon Musk, patron de SpaceX, ont adopté un guide de référence pour un développement éthique de l'intelligence artificielle.

Ces 23 principes ont tous obtenu au moins 90% d'approbation de la part des spécialistes présents :

1. **Objectif de ces recherches** : Le développement de l'IA ne doit pas servir à créer une intelligence sans contrôle mais une intelligence bénéfique.
2. **Investissements** : Les investissements dans l'IA doivent être soutenus par le financement de recherches visant à s'assurer de son usage bénéfique, qui prend en compte des questions épineuses en matière d'informatique, d'économie, de loi, d'éthique et de sciences sociales.  
[Quelques exemples : « Comment rendre les futures IA suffisamment solides pour qu'elles fassent ce qu'on leur demande sans dysfonctionnement ou risque d'être piratées ? » ou encore « Comment améliorer notre prospérité grâce à cette automatisation tout en maintenant les effectifs humains ? »].
3. **Relations entre les scientifiques et les législateurs** : Un échange constructif entre les développeurs d'IA et les législateurs est souhaitable.
4. **Esprit de la recherche** : Un esprit de coopération, de confiance et de transparence devrait être entretenu entre les chercheurs et les scientifiques en charge de l'IA.
5. **Éviter une course** : Les équipes qui travaillent sur les IA sont encouragées à coopérer pour éviter des raccourcis en matière de standards de sécurité.
6. **Sécurité** : Les IA devraient être sécurisées tout au long de leur existence, une caractéristique vérifiable et applicable.
7. **Transparence en cas de problème** : Dans le cas d'une blessure provoquée par une IA, il est nécessaire d'en trouver la cause.
8. **Transparence judiciaire** : Toute implication d'un système autonome dans une décision judiciaire devrait être accompagnée d'une explication satisfaisante contrôlable par un humain.
9. **Responsabilité** : Les concepteurs et les constructeurs d'IA avancées sont les premiers concernés par les conséquences morales de leurs utilisations, détournements et agissements. Ils doivent donc assumer la charge de les influencer.
10. **Concordance de valeurs** : Les IA autonomes devraient être conçues de façon à ce que leurs objectifs et leur comportement s'avèrent conformes aux valeurs humaines.
11. **Valeurs humaines** : Les IA doivent être conçues et fonctionner en accord avec les idéaux de la dignité, des droits et des libertés de l'homme, ainsi que de la diversité culturelle.
12. **Données personnelles** : Chacun devrait avoir le droit d'accéder et de gérer les données le concernant au vu de la capacité des IA à analyser et utiliser ces données.

---

<sup>19</sup> Les 23 principes d'Asilomar <https://futureoflife.org/ai-principles/?cn-reloaded=1>



13. **Liberté et vie privée** : L'utilisation d'IA en matière de données personnelles ne doit pas rogner sur les libertés réelles ou perçue des citoyens.
14. **Bénéfice collectif** : Les IA devraient bénéficier au plus de gens possible et les valoriser.
15. **Prospérité partagée** : La prospérité économique permise par les IA devrait être partagée au plus grand nombre, pour le bien de l'humanité.
16. **Contrôle humain** : Les humains devraient pouvoir choisir comment et s'ils veulent reléguer des décisions de leur choix aux AI.
17. **Anti-renversement** : Le pouvoir obtenu en contrôlant des IA très avancées devrait être soumis au respect et à l'amélioration des processus civiques dont dépend le bien-être de la société plutôt qu'à leur détournement.
18. **Course aux IA d'armement** : Une course aux IA d'armement mortelles est à éviter.
19. **Avertissement sur les capacités** : En l'absence de consensus sur le sujet, il est recommandé d'éviter les hypothèses au sujet des capacités maximum des futures IA.
20. **Importance** : Les IA avancées pourraient entraîner un changement drastique dans l'histoire de la vie sur Terre, et doit donc être gérée avec un soin et des moyens considérables.
21. **Risques** : Les risques causés par les IA, particulièrement les catastrophiques ou existentiels, sont sujets à des efforts de préparation et d'atténuation adaptés à leur impact supposé.
22. **Auto-développement infini** : Les IA conçues pour s'auto-développer à l'infini ou s'auto-reproduire, au risque de devenir très nombreuses ou très avancées rapidement, doivent faire l'objet d'un contrôle de sécurité rigoureux.
23. **Bien commun** : Les intelligences surdéveloppées devraient seulement être développées pour contribuer à des idéaux éthiques partagés par le plus grand nombre et pour le bien de l'humanité plutôt que pour un État ou une entreprise.

# | Personnes auditionnées

## Fédérations/Organisations professionnelles :

**Fédération Française du Bâtiment / FFB** : **Alain PIQUET**, vice-président de la FFB et président de Destais SAS

**Fédération des Industries Electriques Electroniques et de Communication / FIEEC** : **Thierry TINGAUD**, vice-président Branche électronique et président du groupe Intelligence artificielle de la FIEEC – PDG de STMicroelectronics (ACSIEL Alliance Électronique)

**Fédération des Industries mécaniques / FIM** : **Benjamin FRUGIER**, directeur Développement des entreprises et des Projets

**Groupement des Industries Françaises Aéronautiques et Spatiales / GIFAS** : **Marko ERMAN**, Senior Vice President et Chief Technical Officer du groupe Thales et **Emmanuel BLOCH**, directeur Responsabilité d'entreprise de Thales.

**Fédération Française des Télécoms / FFT** : **Mari-Noëlle JÉGO-LAVEISSIÈRE**, directrice générale adjointe, directrice Technologie et Innovation et **Nicolas DEMASSIEUX**, directeur de la recherche d'Orange

**Syntec Numérique** : **Véronique TORNER**, administratrice, **Raphaëlle FRIJA**, déléguée aux usages numériques et **Philippine LEFÈVRE-ROTMANN**, déléguée aux relations institutionnelles

**Tech In France** : **Loïc RIVIÈRE**, délégué général de Tech In France, **Antoine DUBOIS**, Analyste – Atawao, Me **Jean-Sébastien MARIEZ**, Avocat associé – Société d'avocats De Gaulle Fleurance & Associés / DGFLA

## Personnalités qualifiées :

**Boston Consulting Group / BCG** : **Sylvain DURANTON**, directeur monde BCG GAMMA

**Centre national de la recherche scientifique / CNRS** : **Jean-Gabriel GANASCIA**, professeur à l'université Pierre et Marie Curie, informaticien, spécialiste d'intelligence artificielle, président du COMETS (Comité d'éthique du CNRS)

**Groupe Jouve** : **David GRUSON**, directeur du programme santé, membre du comité exécutif du groupe et fondateur de l'initiative académique et citoyenne Ethik-IA

**IBM France** : **Diane DUFOIX-GARNIER**, Government and regulatory affairs executive et **Emmanuel VIGNON**, directeur de la practice cognitive IBM Watson

**Malakoff Médéric Humanis** : **David GIBLAS**, directeur Innovation, Digital et Data - Membre du Comité exécutif et **Stéphane BARDE**, Directeur Data

**Median Technologies** : Dr. **Nozha BOUJEMAA**, directrice Scientifique et de l'Innovation

**Université de la Sorbonne** : Laurence **DEVILLIERS**, professeur, chercheuse en Intelligence artificielle

# | Bibliographie

- | Éthique de la recherche en apprentissage machine (ALLISTENE/Alliance des sciences et technologies du numérique – juin 2017) :  
[http://cerna-ethics-allistene.org/digitalAssets/53/53991\\_cerna\\_ethique\\_apprentissage.pdf](http://cerna-ethics-allistene.org/digitalAssets/53/53991_cerna_ethique_apprentissage.pdf)
- | Assemble nationale : étude d'impact du projet de loi bio éthique (24 juillet 2019)  
<http://www.assemblee-nationale.fr/15/projets/pl2187-ei.asp>
- | Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle - Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une république numérique (décembre 2017)  
[https://www.cnil.fr/sites/default/files/atoms/files/cnil\\_rapport\\_garder\\_la\\_main\\_web.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf)
- | Mission de recherche Université Nice Sophia Antipolis, ENS, CNRS, INRIA :  
[Comment le numérique transforme le droit et la justice par de nouveaux usages et un bouleversement de la prise de décision. Anticiper les évolutions pour les accompagner et les maîtriser.](http://www.gip-recherche-justice.fr/publication/comment-le-numerique-transforme-le-droit-et-la-justice-par-de-nouveaux-usages-et-un-bouleversement-de-la-prise-de-decision-anticiper-les-evolutions-pour-les-accompagner-et-les-maitriser/)  
<http://www.gip-recherche-justice.fr/publication/comment-le-numerique-transforme-le-droit-et-la-justice-par-de-nouveaux-usages-et-un-bouleversement-de-la-prise-de-decision-anticiper-les-evolutions-pour-les-accompagner-et-les-maitriser/>
- | Rapport sur la réforme du droit française de la responsabilité civile et les relations économiques  
Cour d'Appel de Paris (Avril 2019) Université de Versailles Saint-Quentin-en-Yvelines sur la responsabilité et l'intelligence artificielle : pages 107 et suivantes.  
[http://www.justice.gouv.fr/art\\_pix/Rapport\\_CA\\_PARIS\\_reforme\\_responsabilite\\_civile.pdf](http://www.justice.gouv.fr/art_pix/Rapport_CA_PARIS_reforme_responsabilite_civile.pdf)
- | Constructif : Intelligence artificielle de quoi parle-t-on (Octobre 2019 – FFB)  
[http://www.constructif.fr/bibliotheque/2019-10/intelligence-artificielle-de-quoi-parle-t-on.html?item\\_id=5704](http://www.constructif.fr/bibliotheque/2019-10/intelligence-artificielle-de-quoi-parle-t-on.html?item_id=5704)