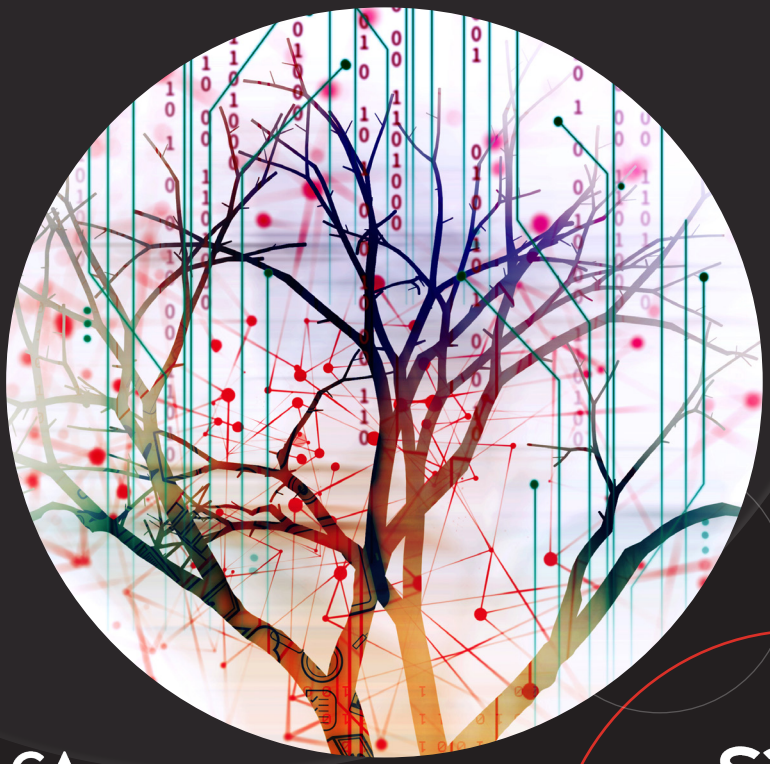
Cahier

n°38

l'Academie
SCIENCES TECHNIQUES COMPTABLES FINANCIERES

INTELLIGENCE ARTIFICIELLE ET CONFIANCE

Règlementation, enjeux, risques, audit et certification



SEPTEMBRE 2021

ISACA
PARIS-FRANCE CHAPTER

sage

EDITOS

Intelligence artificielle (IA) : ensemble des théories et des techniques développant des programmes informatiques complexes capables de simuler certains traits de l'intelligence humaine (raisonnement, apprentissage...).

Source Dictionnaire Le Robert

Vous trouvez inutile de rappeler ce qu'est l'IA tant les outils développés grâce à elle sont nombreux ? Et pourtant... Il paraît plus que jamais nécessaire de prendre le temps de l'explication et de s'appuyer davantage sur les faits et les données, pour maitriser cette innovation majeure. C'est à cette condition que l'IA constituera un gisement de valeur incontestable, libéré des principaux risques qu'elle pourrait engendrer. Il nous incombe - citoyens, entreprises et pouvoirs publics - de construire le cadre d'une IA digne de confiance. Et même si cet exercice demande de s'arrêter quelques temps, cela est nécessaire pour avancer côte à côte avec une technologie qui porte en elle autant de risques irréversibles que d'avancées sociétales majeures.

C'est dans cette optique que le groupe de travail de l'Académie des Sciences et Techniques Comptables et Financières, a lancé ses travaux en 2017. Professionnels du chiffre, avocats, professeurs, consultants, organisations, associations et institutions parties prenantes au débat, se sont réunis pour réfléchir à la règlementation, aux enjeux, aux risques, et à la façon d'auditer un algorithme pour être à même de le certifier « conforme ». Les membres du groupe de travail ont étudié les projets de règlements européens ainsi que les travaux menés et référentiels disponibles, au niveau international. Ils ont analysé les textes et les situations identifiées pour détecter les limites, les biais, les risques, les responsabilités et ainsi dégager les bonnes pratiques et le cadre nécessaire à l'auditabilité et à la certification.

Alors que l'IA questionne plus que jamais sur les sujets de transparence, de traçabilité, de fiabilité, de sécurité, ce nouveau Cahier de l'Académie vous propose de faire le point sur les étapes nécessaires pour disposer d'une évaluation objective de la conception, de la réalisation et du fonctionnement, des applications basées sur l'IA. Les auteurs s'interrogent également sur la gouvernance et l'audit de ces algorithmes et les bonnes pratiques à généraliser par les développeurs de ces systèmes mais aussi par leurs promoteurs et par les entreprises qui les commercialisent et les exploitent.

Nous sommes tous concernés par ces réflexions. Les data n'ont pas de frontières et les algorithmes leur donnent une puissance de feu et un potentiel de performance qui font d'elles un enjeu central pour nos sociétés. Si les algorithmes sont entrés par effraction dans nos pratiques courantes, il est indispensable de prendre la main sur l'encadrement de leurs développements et de leurs usages. La transformation numérique va vite et il nous appartient de nous adapter à ce nouveau rythme pour éviter tout dérapage. Car avant de tirer profit de l'IA, de la valoriser et de la protéger, encore faut-il en maîtriser la progression et les enjeux.

William NAHUM
Président fondateur de l'Académie des Sciences et Techniques Comptables et Financières

Trop souvent les systèmes à base d'Intelligence Artificielle sont des « boites noires ». Or, ces systèmes prennent des décisions qui peuvent avoir de graves conséquences et il est légitime de se demander de quelle manière ils ont procédé pour parvenir à ces résultats. Sur quelles bases et avec quelles données ? Par quel cheminement arrivent-ils à une solution ?

Pour donner confiance aux utilisateurs et aux décideurs mettant en œuvre ces nouveaux systèmes, il sera nécessaire de les contrôler efficacement et de les auditer. Mais comment s'y prendre ? Quelles démarches suivre ? Quels contrôles effectuer ? Autant de questions ouvertes, car les référentiels d'audit informatique actuels ne répondent que partiellement à cette préoccupation.

Pour répondre à cette attente l'ISACA-AFAI, qui représente le chapitre français de l'association internationale ISACA et qui regroupe 140 000 auditeurs informatiques dans le monde, et l'Académie des Sciences et Techniques Comptables et Financières se sont rapprochés pour créer un groupe de travail chargé de faire le point sur l'état de l'art dans le domaine de l'Intelligence Artificielle et de proposer des démarches d'audit opérationnelles que les futurs contrôleurs d'algorithmes pourront mettre en œuvre. Une première étape a été franchie par la publication de ce document de référence. D'autres doivent suivre dans les mois et les années à venir.

Vincent Manière,
Président de l'ISACA-AFAI

Serge Yablonsky,
Président d'honneur de l'ISACA-AFAI

Aux côtés de l'Académie depuis sa création, Sage s'est toujours engagé à soutenir et accompagner la profession comptable et financière aussi bien dans la réflexion sur les grands enjeux qui nourrissent notre quotidien, que dans les réponses à y apporter.

Dans un monde où le changement permanent est devenu la règle, l'innovation que nous qualifions chez Sage d'utile joue un rôle majeur pour nous permettre à tous de mieux épouser certains changements touchant à nos métiers.

La thématique de ce 38ème cahier, autour de l'Intelligence Artificielle et de la Confiance, résonne chez Sage comme une évidence dans le cadre d'une démarche que nous menons depuis toujours.

L'intelligence Artificielle est un formidable levier de simplification et de productivité pour les professions comptables, auquel nous avons toujours lié la responsabilité qui est la nôtre de gérer la confiance par des actes précis. Nous agissons en permanence dans le respect des réglementations et dans l'anticipation des risques et des biais, mais aussi par l'évaluation et les échanges permanents avec nos clients, y compris en phase de conception, de nos solutions dotées d'IA. Autant d'actions parmi d'autres qui sont pour nous l'assurance de l'intégration réussie de l'IA dans nos solutions, auxquelles nous ajoutons la réflexion autour de l'éthique afin d'être toujours en accord avec nos valeurs de partage, d'ouverture et de bienveillance.

Nous positionnons ainsi l'Intelligence Artificielle comme un réel apport permettant notamment aux experts-comptables, allégés ainsi des tâches à faible valeur ajoutée, de mieux se concentrer sur de nouvelles missions fortes de sens dans le cadre de la mutation de leur métier.

Ce sens, nous le retrouvons dans les actions de l'Académie au service de la réflexion, des échanges et de la sensibilisation sur les grands enjeux qui sont les nôtres, et c'est donc avec enthousiasme que nous nous associons pleinement à cette démarche.

Sabine DUCROT-CISS
Senior Product Marketing Director
SAGE

Sommaire

AVANT-PROPOS

Les enjeux de l'Intelligence Artificielle sont considérables et celle-ci va profondément changer la société dans les années à venir. Elle va considérablement améliorer le fonctionnement des applications informatiques actuelles et surtout elle permettra le développement d'un grand nombre de nouvelles applications. C'est un facteur de croissance important et en ces temps d'incertitude, c'est un avantage important.

L'OCDE a publié en 2019 un document de référence sur l'Intelligence Artificielle : « Recommandation du Conseil sur l'intelligence artificielle »[1]. « C'est la première norme intergouvernementale pour les politiques relative à l'Intelligence Artificielle et elle constitue un socle sur lequel s'appuyer pour la réalisation d'analyses complémentaires ». L'objectif est de bâtir des systèmes centrés sur l'humain et l'équité tout en garantissant leur transparence et l'explicabilité des décisions prises.

En Février 2020 la Commission Européenne a publié un Livre Blanc reprenant les grandes orientations de la Recommandation de l'OCDE : *Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance*[2]. « L'IA est une technologie stratégique qui offre de nombreux avantages aux citoyens, aux entreprises et à la société dans son ensemble, à condition qu'elle soit éthique, durable, axée sur le facteur humain et respectueuse des valeurs et droits fondamentaux ». Ce document met en avant les sept exigences suivantes :

- " facteur humain et contrôle humain,
- robustesse technique et sécurité,
- respect de la vie privée et gouvernance des données,
- transparence,
- diversité, non-discrimination et équité,
- bien-être sociétal et environnemental, et
- responsabilisation."

[1] OCDE, Mai 2019, Recommandation du Conseil sur l'intelligence artificielle, OECD/LEGAL/0449 Voir : https://www.oecd.org/fr/presse/quarante-deux-pays-adoptent-les-nouveaux-principes-de-l-ocde-sur-l-intelligence-artificielle.htm

[2] Livre blanc, Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance, Bruxelles, 19 février 2020, COM(2020) 65 final. Voir : https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf

En octobre 2020, pour instaurer la confiance dans les systèmes à base d'Intelligence Artificielle, le Parlement Européen a voté trois résolutions contenant des recommandations destinées à la Commission Européenne concernant :

- Les aspects éthiques de l'Intelligence Artificielle[3] ;
- Le régime de responsabilité civile[4] ;
- Les droits de propriété intellectuelle ;

ainsi que deux propositions de règlement concernant ces sujets.

Ces trois textes mettent l'accent sur la nécessité de :

- Couvrir un périmètre élargi (l'Intelligence Artificielle, les robots et les technologies connexes).
- Respecter les points suivants :
  - des exigences a minima de transparence, d'équité, de fiabilité, de robustesse, de sécurité, de traçabilité, de vérifiabilité, d'interopérabilité… ;
  - des principes comme la nécessité et les limitations, la proportionnalité, les précautions, la distinction… ;
  - des droits à respecter : les droits de l'homme, le droit à l'information, le droit de recours, les droits de propriété intellectuelle, le droit à la vie privée et au secret des affaires… ;
- Mettre en œuvre une approche basée sur les risques différenciés et des critères clairs et concrets d'évaluation de ces risques. Ce sont des risques liés au secteur, à la technologie, la finalité du système, la gravité de son impact… ;
- Disposer d'un modèle de gouvernance clair et cohérent fixant les responsabilités et les obligations de rendre compte ;
- Obtenir une certification par des tiers de confiance impartiaux pour les systèmes à base d'Intelligence Artificielle à hauts risques selon des critères d'évaluation spécifiques ;
- Identifier les bonnes pratiques à respecter pour les systèmes d'Intelligence Artificielle à hauts risques permettant leur certification.

[3] Parlement Européen 8 Octobre 2020, Rapport contenant des recommandations à la Commission Européenne concernant un cadre d'aspects éthiques en matière d'intelligence artificielle, de robotique et de technologies connexes. Voir : https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_FR.pdf

[4] Rapport sur un régime de responsabilité civile pour l'intelligence artificielle : A9-0178/2020 et texte adopté : 2020/2014(INL). Voir : https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_FR.html .

En avril 2021, la Commission Européenne a proposé au Parlement Européen et au Conseil un texte pour une nouvelle règlementation concernant les systèmes à base d'Intelligence Artificielle. C'est le : *Artificial Intelligence Act*[5]. Les éléments essentiels du texte sont les suivants :

- La réduction du périmètre de la nouvelle réglementation par rapport au texte proposé par le Parlement Européen. Les robots et les technologies connexes sont exclus. De même certains éléments sont sortis du cadre général et pris en compte directement dans les législations sectorielles ;
- La définition de l'Intelligence Artificielle est plus précise et plus évolutive, intégrant certaines approches et techniques spécifiques aux Intelligence Artificielle précisées dans une annexe comme les techniques d'apprentissage, les approches basées sur la logique et la connaissance, et les approches statistiques ;
- Une approche fondée sur 4 niveaux de risques :
  - Les domaines où les risques sont inacceptables et ou le recours à l'Intelligence Artificielle est interdit. Ceci concerne quatre types d'Intelligence Artificielle pour certains cas d'usage spécifiés : la reconnaissance faciale à distance, l'utilisation de techniques « subliminales » pour influencer les personnes, l'exploitation des vulnérabilités de certains groupes et la notation sociale ;
  - Les domaines à risque élevé. Ce sont les Intelligence Artificielle à hauts risques avec des obligations strictes à respecter :
    - Ce sont certains usages dans 8 domaines à hauts risques au-delà des critères qui pourraient être inclus dans les législations sectorielles :
      - la biométrie ;
      - les infrastructures sensibles ;
      - l'éducation ;
      - l'emploi ;
      - les services publiques essentielles ;
      - les forces de l'ordre ;
      - l'immigration ;
      - la justice et les processus démocratiques.

[5] Proposal for a Regulation laying down harmonised rules on artificial intelligence : Artificial Intelligence Act, 21 Avril 2021. Voir : https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence pour le texte du projet de régulation et son annexe.

- Les exigences attendues et les obligations strictes à respecter concernant 5 volets :
  - les données et la gouvernance des données ;
  - la documentation et la rétention ;
  - la transparence et le droit à l'information ;
  - la surveillance humaine ;
  - la robustesse, l'exactitude et la sécurité.

- Les domaines avec des risques limités. Ces systèmes d'Intelligence Artificielle ont des obligations en matière de transparence. Ce sont des systèmes prévoyant des interactions humaines, des applications biométriques ou avec des contenus truqués.
- Les domaines à risques minimes. Ces systèmes à base d'Intelligence Artificielle sont à utilisation libre. A priori c'est la grande majorité des systèmes d'Intelligence Artificielle.

- Les systèmes d'Intelligence Artificielle à hauts risques doivent nécessairement être certifiés sous le contrôle de l'opérateur. Cette certification est basée sur l'évaluation de trois aspects du contrôle interne :
  - Existence d'un système d'assurance qualité ;
  - Vérification du contenu de la documentation ;
  - Analyse du processus de conception et de développement.

Anticipant les problèmes que soulèvent les systèmes à base d'Intelligence Artificielle, dès 2017-2018 un groupe de travail s'est constitué autour de l'Académie des Sciences et des Techniques Comptables et Financières et de l'ISACA-AFAI, chapitre français de l'ISACA International. Au départ, le groupe s'est fixé pour objectif de réfléchir sur l'audit des algorithmes. Constatant l'évolution des préoccupations concernant les applications recourant à l'Intelligence Artificielle, il a évolué vers une réflexion sur les conditions de la confiance qu'il est possible d'accorder dans les systèmes à base d'Intelligence Artificielle. C'est un problème complexe et pour y faire face le groupe de travail a fait appel à des experts de quatre métiers différents : l'audit, l'informatique, le droit et la connaissance des approches de l'Intelligence Artificielle.

Notre réflexion se situe dans la logique de ces différents documents en cherchant à favoriser le développement d'applications recourant à l'Intelligence Artificielle qui soient dignes de confiance. Le présent document cherche à identifier les enjeux d'un système à base d'Intelligence Artificielle de confiance, les spécificités associées à une telle application d'Intelligence Artificielle et les conditions nécessaires pour y arriver. Ainsi, nous y évoquons différentes typologies de systèmes d'Intelligence Artificielle, des risques spécifiques, des règlementations en vigueur, des aspects

éthiques, de la nature des bonnes pratiques d'Intelligence Artificielle à mettre en œuvre et le rôle essentiel d'un audit, voire d'une certification de conformité à certaines exigences faite par un tiers impartial pour donner un niveau de confiance raisonnable. D'autres documents sont en préparation.

INTRODUCTION

Depuis quelques années, nous constatons que certains types de développements informatiques ont soulevé des inquiétudes et notamment celles du contenu des algorithmes. Des choses importantes seraient cachées aux utilisateurs ou des biais amèneraient des décisions non-conformes.

Ce type de préoccupation est très clairement apparu lorsque les dysfonctionnements d'Admission Post Bac (APB) et de sa nouvelle version Parcoursup ont été constatés. Le développement de l'Intelligence Artificielle et notamment le Deep Learning ne peut que renforcer la suspicion à l'encontre des algorithmes car on ne sait plus très bien ce qu'ils font[6] ? Peut-on améliorer la traçabilité des opérations ? Que se passera-t-il si les ordinateurs devenaient plus intelligents que les hommes ? Est-ce qu'ils prendraient le pouvoir et asserviraient l'humanité ? Sommes-nous voués à l'esclavage ? Et que se passerait-il si les robots se reprogrammaient d'eux-mêmes ?... Ainsi, on assiste à une libération de tous les fantasmes possibles et imaginables.

L'Intelligence Artificielle suscite beaucoup d'espoirs et de nombreuses craintes. Mais elle représente encore un marché limité. Il est évalué en 2019 à 11,3 milliards de dollars par Tractica mais on estime qu'il va croître rapidement dans les prochaines années pour atteindre 89,8 milliards de dollars en 2025.

De nombreuses applications d'Intelligence Artificielle ont été développées dans un grand nombre de domaines comme les voitures autonomes, les aides au diagnostic dans la médecine, l'analyse des risques et le scoring dans la banque et l'assurance, la prédiction des décisions de justice, la régulation du trafic dans les transports en commun, les drones et les systèmes d'armes,... Sans compter les jeux, comme le jeu de Go avec DeepMind.

Depuis quelques années, on constate que l'Intelligence Artificielle mobilise une grande partie des efforts de recherche notamment le domaine du « machine learning ». L'opinion publique est intéressée par ses progrès : elle suscite de nombreux espoirs mais cette technologie provoque de nombreuses réticences. Certains craignent qu'elle permette « la domination de l'humanité par les robots ». Au-delà de ces fantasmes, il est vrai qu'il existe des risques qu'il serait dangereux de

[6] Contrairement à ce qui est souvent indiqué, plusieurs experts considèrent en effet qu'APB comme Parcoursup, sont des systèmes qui ne font pas appel à l'Intelligence Artificielle. Ces logiciels reposent sur l'algorithme de Gale-Shapley dit des mariages stables. Lloyd Shapley a reçu le prix Nobel d'économie en 2012 pour la découverte de cet algorithme.

sous-estimer. C'est le rôle de la gouvernance des algorithmes et en particulier de l'Intelligence Artificielle.

Depuis des années les Etats se sont intéressés à l'Intelligence Artificielle pour répondre à leurs besoins, notamment dans le domaine militaire, mais aussi pour développer la recherche dans ce domaine et pour en réguler les usages. Les organisations internationales comme l'OCDE, l'Union Européenne... ont développé une réflexion pour avoir une Intelligence Artificielle éthique, responsable et digne de confiance.

Ainsi l'OCDE propose de définir l'Intelligence Artificielle de la manière suivante : « C'est un système automatisé qui, pour un ensemble donné d'objectifs définis par l'homme, est en mesure d'établir des prévisions, de formuler des recommandations, ou de prendre des décisions influant sur des environnements réels ou virtuels. Les systèmes d'Intelligence Artificielle sont conçus pour fonctionner à des degrés d'autonomie divers »[7]. Cette définition est largement acceptée notamment par les 36 pays membres de l'OCDE et 8 autres pays, principalement d'Amérique Latine.

Cette définition de l'Intelligence Artificielle est fondamentale et elle a été reprise par la Commission Européenne dans un Livre Blanc qui va fixer les orientations des Etats dans les années à venir concernant l'Intelligence Artificielle[8]. On la retrouve dans les rapports et le projet de règlement européen établis par le Parlement Européen et la Commission Européenne[9] : « un système d'intelligence artificielle est un logiciel qui est développé avec une ou plusieurs des techniques et approches... et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prévisions, des recommandations, ou des décisions influençant les environnements avec lesquels ils interagissent. »

[7] OCDE, Mai 2019, Recommandation du Conseil sur l'intelligence artificielle, OECD/LEGAL/0449. https://www.oecd.org/fr/presse/quarante-deux-pays-adoptent-les-nouveaux-principes-de-l-ocde-sur-l-intelligence-artificielle.htm

[8] Livre blanc, Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance, Bruxelles, 19 février 2020. Voir : https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf

[9] Proposal for a Regulation laying down harmonised rules on artificial intelligence : Artificial Intelligence Act, Bruxelles,21 Avril 2021. Voir : https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence pour le texte du projet de régulation et son annexe.

Ces documents s'inspirent en grande partie du travail fait par l'Université de Montréal qui a publié en 2017 la Déclaration de Montréal pour « un développement responsable de l'Intelligence Artificielle »[10]. Effectivement, ces documents font apparaître différents problèmes d'éthiques et de gouvernance qui peuvent entraîner des résistances, voire des rejets. Ils se traduisent par une prise de conscience collective des impacts potentiels de l'Intelligence Artificielle[11]. Pour éviter ces dérives et créer un niveau de confiance suffisant dans cette technologie, il est nécessaire de définir un certain nombre d'actions à mener notamment pour mieux maîtriser les algorithmes.

Dans ce but, il est souhaitable de recourir à un tiers de confiance indépendant pour évaluer les systèmes recourant à l'Intelligence Artificielle. Le rapport de Cédric Villani « Donner un sens à l'Intelligence Artificielle »[12] recommande d'effectuer des audits des algorithmes et pour cela suggère : « *La constitution d'un corps d'experts dotés des compétences requises semble nécessaire pour procéder à des audits d'algorithmes et de bases de données sur pièce, et procéder à des tests par tout moyen requis* ».

Cela revient à créer des commissaires aux algorithmes sur le modèle des commissaires aux comptes. En effet, il existe des auditeurs informatiques et des professionnels du « chiffre » qui ont les compétences nécessaires pour mener à bien ce type de missions d'audit :

- Connaissance des entreprises et de leurs enjeux ;
- Démarche d'audit basée sur l'évaluation des risques ;
- Diligences professionnelles ;
- Proximité avec le management des projets ;
- Connaissance fonctionnelle des applications et la maîtrise des données ;
- Connaissance du domaine du numérique ;
- Maîtrise des problèmes de sécurité informatique ;
- ...

Ces missions impliquent un grand nombre d'acteurs comme : les décideurs, les sponsors, les concepteurs, les développeurs, les data scientists (Big Data), les clients, les consommateurs, les usagers, les régulateurs, les tiers de confiances, etc.

---

[10] L'Université de Montréal a lancé en novembre 2017 une démarche de coconstruction de la Déclaration de Montréal pour un développement responsable de l'Intelligence Artificielle Voir : https://www.declarationmontreal-iaresponsable.com/contexte.

[11] Voir à ce sujet le serment d'Holberton et Turing souscrit par un grand nombre de chercheurs en Intelligence Artificielle : https://www.holbertonturingoath.org/

[12] « Donner un sens à l'Intelligence Artificielle » Voir : https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de.

Le présent rapport a pour but d'explorer ce domaine en analysant :

- Notre approche de l'Intelligence Artificielle ;
- Les bénéfices de l'Intelligence Artificielle ;
- Les risques et les incertitudes de l'Intelligence Artificielle ;
- Le droit de l'Intelligence Artificielle ;
- L'éthique de l'Intelligence Artificielle ;
- Vers une Intelligence Artificielle digne de confiance.

Cet ouvrage est le premier d'une série. Nous approfondirons les différents points abordés dans ce document et au-delà. Un livre blanc sur les problématiques des applications à base d'Intelligence Artificielle de confiance est en cours de rédaction. D'autres suivront notamment sur les démarches d'audit et de certification.

CHAPITRE 1

NOTRE APPROCHE DE L'INTELLIGENCE ARTIFICIELLE

Il est très difficile de définir l'Intelligence Artificielle. De très nombreuses définitions ont été données mais aucune n'est vraiment satisfaisante. De cet ensemble, il se dégage une idée simple : l'Intelligence Artificielle consiste à faire accomplir par un ordinateur ce que le cerveau humain fait habituellement : prendre en compte et traiter des données, les comprendre, prendre des décisions, s'exprimer de manière compréhensible,...

Dans une première approche de l'Intelligence Artificielle on constate qu'elle concerne de nombreux domaines de l'informatique comme les différentes technologies, les bases de données, les règles de gouvernance, les algorithmes, les données, les machines apprenantes (Machine Learning), etc. Ces domaines concernent aussi des aspects différents de la société allant de la conception de nouveaux services à l'évolution de la démocratie dans les années à venir.

Lorsqu'on analyse les différentes fonctions mises en œuvre par un système recourant à de l'Intelligence Artificielle, on est très vite amené à détailler le ou les algorithmes à la base de ce type de traitement. A ce propos, il faut prendre garde au fait que très souvent, on confond l'algorithme et le programme informatique qui permet son exécution. « *Un algorithme est une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre une classe de problèmes* »[13]. Il se traduit par un « *ensemble de codes qui permet l'exécution automatique des tâches* » et qui constitue un ou plusieurs programmes. L'algorithme est une suite d'opérations permettant de traiter un problème alors que le programme est constitué par une liste d'instructions indiquant à l'ordinateur comment exécuter l'algorithme.

Il existe un débat pour savoir si tel ou tel système relève de l'Intelligence Artificielle ou n'en relève pas et la commission qui a établi ce rapport n'y a pas échappé. Ainsi le système d'inscription des étudiants en première année universitaire, Parcoursup, a été l'objet de longues discussions dans les médias et dans l'opinion publique qui considèrent ce système comme un exemple parfait d'Intelligence Artificielle et donc accusent cette technologie des éventuels problèmes rencontrés. Le rapport parlementaire établi par le Comité Ethique et Scientifique de Parcoursup (CESP) pour l'évaluer ne considère pas que c'est un algorithme d'Intelligence Artificielle[14]. En effet, il a recours

[13] Définition de Wikipédia : https://fr.wikipedia.org/wiki/Algorithme .

[14] Ce rapport a été publié en Janvier 2019 https://www.enseignementsup-recherche.gouv.fr/cid138009/rapport-du-comite-ethique-et-scientifique-de-parcoursup.html

à l'algorithme Gale et Shapley dit des mariages heureux[15] qui est basé sur une logique classique et sur des règles utilisant des instructions conditionnelles[16].

Un système d'Intelligence Artificielle repose sur des algorithmes, c'est-à-dire selon Donald Ervin Knuth, pionnier de l'algorithmique moderne, « *d'une suite finie et non ambiguë d'opérations ou d'instructions permettant, à l'aide d'entrées, de résoudre un problème ou d'obtenir un résultat, ces sorties étant réalisées selon certaines règles* »[17]. Ce terme fait l'objet d'une définition normalisée par l'ISO comme étant un « *ensemble fini de règles bien définies pour la résolution d'un problème* »[18]. L'idée fondamentale est de traiter de l'information, par une série d'opérations qu'on désigne sous le terme d'algorithmes, de façon à y associer une action pertinente, à l'image de ce qu'un individu réalise à l'aide de son cerveau[19].

Un système d'Intelligence Artificielle a pour but de mimer l'intelligence humaine. Il comprend des algorithmes mais aussi des données, des paramètres et des procédures, comme tout système d'information, qui lui sont associés. Il est pour cela important de définir le périmètre du système algorithmique et les paramètres indépendamment de l'algorithme. Les données sont très importantes. C'est le point de départ de nombreux systèmes d'Intelligence Artificielle notamment ceux recourant aux technologies de Machine Learning qui ont besoin d'analyser de grandes masses de données afin de fixer les paramètres du système avant de pouvoir effectuer tout traitement à proprement parler.

Les données constituent en fait le cœur de tout système basé sur l'Intelligence Artificielle. Elles peuvent provenir de bases de données préexistantes, de réseaux sociaux, de capteurs IoT, de saisies effectuées à l'aide de pages Web s'affichant sur des PC, des tablettes ou des smartphones... La collecte, le contrôle et le stockage des données sont des points cruciaux surtout s'ils fonctionnent en temps réel avec des bases de données réparties, des entrepôts de données ou des « data lake ». Enfin, il ne faut pas sous-évaluer la puissance de traitement nécessaire à l'exploitation de ces données. Il est donc dans ce cas nécessaire de disposer de serveurs puissants, dotés de processeurs spécialisés.

---

[15] Voir https://fr.wikipedia.org/wiki/Algorithme_de_Gale_et_Shapley . Cet algorithme a été présenté en 1962. Lloyd Shapley a reçu le prix Nobel pour la découverte de cet algorithme.
[16] Nous ferons référence dans ce rapport à cette application qui a bénéficié d'une large publicité bien qu'elle ne soit pas de l'Intelligence Artificielle mais qui met en œuvre des algorithmes très puissants et intéressants.
[17] S. Abiteboul, G. Dowek, *Le temps des algorithmes*, Ed. Le Pommier 2017.
[18] Norme ISO 2382:2015, Technologies de l'information -- Vocabulaire -- Partie 28: Intelligence artificielle.
[19] S. Dehaene, Apprendre ! *Les talents du cerveau, le défi des machines*, Paris, Odile Jacob, 2018.

On sait le rôle important joué par les GAFAM[20] dans le développement de l'Intelligence Artificielle. Ce rôle est, pour partie, lié à leurs investissements massifs dans les développements consacrés à cette technologie et dans leurs parcs de serveurs. Ceux-ci ont ainsi été capables de stocker des masses de données considérables. L'accès à ces données constitue un avantage certain. Pour cette raison, ces entreprises sont devenues incontournables. De manière plus générale, toute entreprise qui maîtrise l'accès aux données, aux réseaux de communication, à la puissance de ses traitements, à son stockage, et aux services et technologies recourant à l'Intelligence Artificielle bénéficie d'un avantage concurrentiel déterminant par rapport à ses concurrents.

[20] GAFAM : Google Apple Facebook Amazon Microsoft

CHAPITRE 2

LES GRANDES TYPOLOGIES DES SYSTEMES D'INTELLIGENCE ARTIFICIELLE

Il existe de très nombreux types de systèmes d'Intelligence Artificielle. Certains sont relativement simples et d'autres sont très complexes. Ils se différencient aussi par les enjeux et les risques associés. Ils ne sont pas de la même nature ni du même niveau. Pour les classer, plusieurs caractéristiques peuvent être prises en compte, notamment :

- Le périmètre du système ;
- Les technologies mises en œuvre ;
- Les fonctions de base assurées.

Il y a lieu d'examiner ces trois points :

- **Le périmètre du système.** On distingue habituellement :
  - **Les systèmes d'Intelligence Artificielle faible.**

    Ils effectuent des tâches précises comme par exemple jouer aux échecs, traduire un texte, comprendre la parole d'un locuteur, effectuer un diagnostic médical,...
  - **Les systèmes d'Intelligence Artificielle forte.**

    Ils sont basés sur des transferts de connaissances entre différents domaines et pour certains experts, ils peuvent même disposer d'une « conscience ». Cela relèverait alors apparemment de la science-fiction mais c'est peut-être ce que nous réserve le futur, par exemple avec la voiture autonome fonctionnant sans conducteur.
- **Les technologies mises en œuvre.** On distingue différents types d'algorithmes allant du plus simple au plus sophistiqué :
  - **Les algorithmes ne recourant pas à l'Intelligence Artificielle.**

    Ce sont les algorithmes les plus nombreux et représentent la quasi-totalité des traitements comme : la sélection de données, les tris, les calculs algébrique ou trigonométrique, les échanges de données entre systèmes, le cryptage, les sauvegardes, etc. Ces programmes reposent sur des séquences d'instructions précises prédéfinies à l'avance. Ils n'impliquent pas de processus d'apprentissage. C'est, par exemple, le cas de l'algorithme de classement Gale-Shapley des étudiants effectué par Parcoursup.
  - **Les systèmes experts.**

    Ce sont des programmes répondant à des questions à l'aide d'une base de règles, d'une

base de connaissances et d'un solveur qui est un moteur de gestion des règles. Ces bases représentent le savoir d'un groupe d'experts reconnu dans un domaine particulier. Les connaissances sont représentées sous forme de graphes. La principale difficulté des systèmes experts est de codifier le savoir des experts.

- **Les systèmes à base de logique floue ou *fuzzy logic*.**

  Ce sont des systèmes permettant de traiter des informations floues qui ne sont ni totalement exactes ni totalement fausses. Il existe de très nombreuses applications de logique floue : les freins ABS, le pilotage de fours industriels, la gestion des boîtes de vitesse automatiques, la stabilisation des appareils photo pour éviter les « bougés », la stabilisation des cameras pour stabiliser l'image, la gestion de processus (cimenterie, traitement des eaux, méthaniseur,...), le pilotage des machines à laver, la gestion de la circulation automobile (pilotage des feux rouges), le réglage des injecteurs des moteurs, la reconnaissance de l'écriture manuscrite,... Comme on le constate, les applications de logique floue se retrouvent dans de nombreux produits et de nombreux systèmes.

- **L'apprentissage automatique ou machine learning *« supervisé »*.**

  Dans ce cas avant d'être opérationnel, le système doit effectuer un apprentissage en traitant un grand nombre de données préalablement labellisées manuellement. C'est par exemple une collection de photos avec leur description. L'apprentissage permet au système de comprendre les liens entre les données et le résultat prévu. A l'issu de cet apprentissage basé sur des classements ou des régressions, il est possible de produire un résultat qui soit en ligne avec le modèle appris. C'est la phase de production à partir du modèle appris. Les décisions qu'il est amené à produire sont basées sur des probabilités déduites du processus d'apprentissage. En fait, ce type d'algorithme utilise les techniques statistiques classiques et des calculs d'optimisation de fonction de coût de revient. Des banques utilisent des modèles de ce type pour évaluer la solvabilité des demandeurs de prêts et prévoir, le cas échéant, le risque d'insolvabilité.

- **L'apprentissage automatique dit machine learning *« non supervisé »*.**

  Dans ce cas, le système a la capacité de trouver de lui-même des modèles à partir de données qui n'ont pas été, au préalable, labellisées. C'est un apprentissage basé sur l'identification des éléments qui pourraient avoir une forte incidence sur d'une part les résultats et d'autre part l'identification de groupes de données qui seraient homogènes. Le système cherche à déterminer la structure sous-jacente des données et le ou les modèles associés. C'est, par exemple, le cas de systèmes recherchant des fraudes comptables ou de systèmes qui s'attachent à identifier les causes d'une maladie à la

suite de ce que les médecins appellent l'anamnèse. C'est la phase de consultation où le patient va décrire les symptômes qu'il perçoit et le soignant va recueillir un certain nombre d'informations générales le concernant.

- **L'apprentissage par renforcement ou « *reinforced learning* ».**

  C'est un mode d'apprentissage orienté objectif reposant sur une démarche de type essai et erreur, basée sur des récompenses pour les bonnes réponses et des sanctions pour les mauvaises réponses. Il est ainsi possible d'établir une stratégie permettant de résoudre le problème qui doit être traité. Le système est alors capable d'effectuer des prédictions, de les valider en fonction des résultats obtenus et de s'ajuster de manière continue afin de s'améliorer. Cette démarche permet d'augmenter l'efficacité du processus d'apprentissage automatique. C'est la stratégie adoptée par certains jeux comme la V2 du jeu de Go développée par AlphaGo.

- **Le « deep learning »**

  C'est un sous-ensemble de l'apprentissage automatique. Ce mode d'apprentissage repose sur des réseaux de neurones[21] permettant d'identifier une chaîne complexe de liens de corrélations qui sont des indicateurs probables de causalités. C'est, par exemple, le cas des systèmes capables de faire de la reconnaissance de formes, de visages[22] ou de la parole, la navigation des véhicules autonomes, le pilotage des drones, ou des outils de veille et d'analyse du web et des réseaux sociaux comme Talkwalker.

- **Les fonctions de base.** Les systèmes recourant à l'Intelligence Artificielle reposent sur différentes fonctions :
  - **L'analyse descriptive** permet de comprendre ce qui s'est passé et donc restituer le déroulement des événements et des décisions qui ont été prises.
  - **L'analyse diagnostique** explique pour quelles raisons un résultat donné a été obtenu. Généralement, il est obtenu grâce à l'identification de différentes corrélations.
  - **L'analyse prédictive.** C'est le résultat des modèles d'apprentissage. Compte tenu du contexte et des derniers événements qui viennent de survenir, elle permet de décrire ce qui va se passer.

[21] Les réseaux de neurones sont des programmes informatiques qui simulent le fonctionnement des milliards de neurones qui constituent le cerveau humain.

[22] Il existe plusieurs niveaux dans la reconnaissance de personnes : 1er niveau : reconnaitre les lignes droites et les courbes du visage, 2ème niveau : identifier les yeux, le nez et les oreilles, 3ème niveau : identifier les visages et reconnaitre les personnes.

- **L'analyse prescriptive.** Cette fonction sert à décrire ce qu'il faut faire pour qu'un objectif particulier se réalise. Ce sont des solutions opérationnelles qui peuvent être mises en œuvre par un humain ou directement par le système.

Les évolutions dans le domaine de l'Intelligence Artificielle sont actuellement très rapides. De très nombreux traitements sont dès aujourd'hui opérationnels et donnent des résultats très positifs. Il est dans ces conditions très important de s'assurer que ces systèmes soient correctement maîtrisés et que les futurs usages soient rapidement contrôlés. La situation actuelle des développements en matière de système recourant à l'Intelligence Artificielle est la suivante :

- Les applications actuelles sont plutôt de l'Intelligence Augmentée. Elles sont capables d'effectuer des tâches relativement simples nécessitant un faible niveau d'intelligence.
- On note une absence très générale d'outils d'analyse et d'audit des systèmes d'Intelligence Artificielle existants. Ceci crée une suspicion à l'encontre de toutes ces applications.
- De manière générale on constate un manque de traçabilité des décisions prises par ces systèmes et donc le sentiment de se retrouver dans des situations non-satisfaisantes. De plus comme il y a des insuffisances en matière de transparence et de traçabilité cela se traduit par un manque de confiance dans ces systèmes avec un risque de duperie, voire d'actes malveillants.

On note quatre tendances émergentes actuellement :

- La puissance de traitement des processeurs augmente rapidement. On a recours de plus en plus à des processeurs spécialisés : GPU (Graphic Processor Unit), NPU (Neural Processor Unit)... À terme l'informatique quantique va permettre d'analyser de très grandes quantités de données et ainsi de détecter des tendances fines ;
- Les environnements de développement des logiciels d'Intelligence Artificielle apportent aux développeurs un grand nombre de fonctions de base qui sont nécessaires et qui vont leur permettre de se concentrer sur la réalisation du système ;
- Les risques de dérives possibles concernant les libertés et la vie privée (*privacy*) impactent les habitudes de consommation des personnes, les comportements, la détection des transgressions, le classement social, etc. Ces risques doivent être mis sous contrôle ;
- Au-delà, un certain nombre de personnes rêvent à l'Homme Augmenté et, plus généralement, au transhumanisme. C'est un courant de pensée important en Californie qui a un certain nombre de prosélytes en France.

CHAPITRE 3

LES BENEFICES DE L'INTELLIGENCE ARTIFICIELLE

Le développement des systèmes basés sur l'Intelligence Artificielle est source de nombreuses innovations. Ils permettent la création de nouveaux produits et de nouveaux services et ils offrent aussi de nouvelles possibilités pour traiter des fonctions traditionnelles comme la gestion de la relation avec les clients, le traitement de différents contentieux, la comptabilité, le contrôle de gestion, etc.

Pour apprécier l'importance des bénéfices liés aux applications reposant sur l'Intelligence Artificielle, il est utile d'analyser différents cas de mise en œuvre et de faire ressortir les apports qu'ils permettent de réaliser :

- **Les assistants personnels virtuels.**
  Ce sont des applications comme Alexa d'Amazon ou Google Assistant mais aussi des logiciels comme Siri d'Apple, ou Cortana de Microsoft. Ils sont capables d'interaction vocale avec les personnes afin d'effectuer une recherche sur Internet, de se connecter à un site Web, de lancer l'écoute d'une chanson ou d'un morceau de musique, d'exécuter une liste de tâches, de régler des alarmes, d'écouter des podcasts et des livres-audio, de permettre de prendre connaissance de la météo, du trafic routier et d'autres informations en temps réel. Ils peuvent également contrôler plusieurs appareils domestiques intelligents, régler le chauffage ou la climatisation, allumer ou éteindre les lumières etc. Les assistants personnels virtuels permettent de dégager les avantages suivants :
  - Éviter de devoir dactylographier les commandes informatiques au clavier et à terme éviter de mettre des claviers et des écrans sur des appareils domestiques et même des matériels professionnels ;
  - Automatiser les tâches (comme la gestion de la température du chauffage) et augmenter la productivité (comme d'effectuer des tâches répétitives) ;
  - Faciliter la vie des personnes. Grâce à la reconnaissance de la parole il est possible de lancer des commandes à la voix tout en faisant autre chose ;
  - Permettre à des handicapés moteurs ou à des mal-voyants d'améliorer leur vie en commandant des systèmes à la voix.

- **La voiture autonome.**
  Il est possible de faire rouler un véhicule en toute sécurité sans intervention humaine. Il existe à côté des voitures autonomes des situations intermédiaires avec une autonomie partielle comme des systèmes d'aide à la conduite automatisée, en ville ou sur autoroute, ou encore des systèmes de parking automatisé des véhicules. Ces dispositifs concernent les voitures particulières, mais aussi les autobus, les camions, les métros et les trains. Les avantages de ce type de dispositif sont nombreux :
  - Améliorer la sécurité des véhicules en diminuant le nombre d'accidents et donc réduire le nombre de morts et de blessés sur la route et diminuer le montant des dommages matériels ;
  - Réduire les embouteillages et améliorer le flux des véhicules dans le cadre d'une « smart city » :
    - Réduire les temps perdus par les passagers ;
    - Diminuer la consommation d'essence et la pollution ;
    - Embaucher moins d'agents chargés de contrôler la circulation ;
    - Réduire la signalisation...
  - Diminuer les tarifs d'assurance ;
  - Réduire le nombre de contraventions et de délits routiers ;
  - Interdire la conduite sans permis ou dans un état incompatible (fatigue, ivresse, maladie...) ;
  - Eviter la recherche de place de parking et faciliter la réduction de l'espace nécessaire.
- **L'amélioration de la qualité de service.**
  L'Intelligence Artificielle permet d'améliorer le fonctionnement de nombreuses applications classiques :
  - Google utilise l'Intelligence Artificielle de nombreuses manières différentes notamment pour faciliter et améliorer la qualité des recherches effectuées à l'aide du moteur de recherche. Autre application utile : améliorer la qualité des traductions dans Google Traduction ;
  - Amazon conseil des livres à ses clients en fonction de leurs centres d'intérêt en analysant leurs achats précédents et de leurs recherches à l'aide d'un algorithme d'Intelligence Artificielle (C'est une des première application grand public de l'Intelligence Artificielle) ;

- Netflix et YouTube utilisent l'Intelligence Artificielle pour conseiller à leurs utilisateurs des vidéos correspondant à leurs préférences recensées à l'aide des vidéos précédemment visionnées ;
- Les ressources humaines de certaines entreprises utilisent l'Intelligence Artificielle pour améliorer la qualité des recrutements. Ces systèmes permettent de sélectionner dans la masse des CV reçus par l'entreprise les personnes les plus susceptibles de correspondre aux postes recherchés.

L'Intelligence Artificielle ne change pas la nature des fonctions mise en œuvre par l'application mais elle permet d'améliorer de manière significative leur fonctionnement. Le principal gain consiste à améliorer le confort de ses utilisateurs et l'efficacité des traitements opérés.

- **L'amélioration de la productivité.**
  Les systèmes recourant à l'Intelligence Artificielle permettent d'exécuter plus rapidement un certain nombre d'opérations administratives ou industrielles :
  - Automatisation des processus. Il est possible d'effectuer des tâches complexes plus rapidement avec du personnel moins qualifié. C'est notamment le cas de la comptabilité, la prise de commandes, la gestion des opérations bancaires ou des tâches effectuées par les compagnies d'assurance,...
  - Robots plus efficaces. Dans l'industrie et dans la logistique, il est possible d'automatiser un grand nombre de tâches notamment grâce à la reconnaissance de forme et à la commande vocale.

  Les gains liés à l'amélioration de la productivité sont principalement de trois ordres :
  - Baisse des coûts de revient. Encore faut-il s'assurer que les gains sur les coûts de personnel sont supérieurs aux coûts supplémentaires liés à la mise en œuvre d'applications recourant à l'Intelligence Artificielle ;
  - Suppression des goulets d'étranglement. Dans tous les processus, on constate des ralentissements des opérations dus au manque de personnels qualifiés ou à l'afflux soudain d'un volume important d'opérations. Ces goulets tendent à disparaître grâce à l'Intelligence Artificielle mais il est nécessaire de disposer d'une puissance de traitement suffisante ;
  - Réduction des délais. La mise en place de systèmes puissants permet de réduire significativement les délais de traitements et ainsi de répondre plus rapidement aux attentes des clients.

- **La qualité des décisions d'investissements.** Les systèmes d'évaluation des investissements à base d'Intelligence Artificielle permettent d'améliorer la qualité des projets d'investissements tel que le montage des opérations industrielles. Il est aussi possible d'améliorer la gestion de portefeuille grâce à l'Intelligence Artificielle. Les gains générés par ces systèmes sont de deux ordres :
  - Améliorer la qualité des choix effectués et enrichir les options offertes ;
  - Accroître la rentabilité globale des investissements.
- **La gestion des intérêts des différentes parties prenantes.**
  Parcoursup est un exemple parfait de système gérant les intérêts de différentes parties prenantes bien que ce ne soit pas un système recourant à l'Intelligence Artificielle à proprement parler. Il y a de nombreuses parties prenantes. Il offre aux élèves la possibilité de ne pas hiérarchiser leurs vœux d'orientation à l'inscription des nouveaux bacheliers à l'université, dans les instituts ou dans les classes préparatoires. Les gains liés à cette application sont multiples :
  - Simplifier la procédure d'inscription en évitant les longues files d'attente devant les guichets administratifs des Universités et préparer le travail d'inscription de l'administration ;
  - Améliorer l'orientation des étudiants en leur évitant de s'engager dans des filières pour lesquelles ils n'ont pas le niveau requis et ni les connaissances nécessaires ou de s'aventurer dans des secteurs dont les débouchés sont limités ;
  - Augmenter la rapidité d'affectation des étudiants et donc diminuer leurs inquiétudes et leurs angoisses liées à l'allongement du temps d'attente pour la plus grande majorité des futurs étudiants ;
  - Définir une meilleure organisation des établissements pour faire face à l'afflux des étudiants en embauchant des enseignants, en trouvant de nouveaux locaux... Ce dispositif permet aussi de s'organiser pour faire face à une baisse éventuelle des inscriptions ;
  - Limiter les biais sociaux ou territoriaux et donc réduire les iniquités entre les candidats.
- **La reconnaissance d'images.**
  C'est notamment le cas du domaine médical pour diagnostiquer les cancers de la peau ou les diagnostics radiologiques,...
  - Augmenter le nombre de bons diagnostics et améliorer la possibilité d'effectuer un diagnostic précoce ;

- Améliorer la qualité et l'efficacité des traitements préconisés en évitant les incompatibilités entre les médicaments ;
- Diminuer le coût global de la santé ;
- Accroître le nombre d'années pendant lesquelles la population peut vivre en bonne santé ;
- Améliorer et sécuriser le travail des professionnels de santé.

Le domaine de la sécurité recourt aussi à l'Intelligence Artificielle notamment pour pratiquer la reconnaissance faciale mais aussi pour détecter des comportements à risques.

Cette liste d'applications recourant à l'Intelligence Artificielle est partielle et ne recouvre qu'une partie des types d'avantages possibles. Néanmoins, les bénéfices de l'Intelligence Artificielle sont importants et expliquent, en grande partie, l'engouement actuel des entreprises pour l'utilisation de l'Intelligence Artificielle.

Pour cette raison, il est important de communiquer de manière objective sur les avantages des systèmes recourant à l'Intelligence Artificielle sans céder aux mythes et aux rêves. Les bénéfices que l'on peut en tirer, le niveau de risques correspondant (réduction des erreurs, moins de morts, baisse des coûts,...), son acceptabilité par les utilisateurs, et le moindre niveau d'effort nécessaire (qui sont à apprécier dans chaque contexte particulier) sont à mettre en avant. Cela permet d'avoir un développement responsable et d'arriver à une Intelligence Artificielle digne de confiance.

Il est en particulier important de tenir compte des bénéfices dont peuvent profiter les différentes parties prenantes concernées du fait des projets basés sur l'Intelligence Artificielle. Les sponsors, les concepteurs et les data-scientists ont des objectifs et des enjeux différents de ceux des développeurs et des programmeurs. Cela joue sur la finalité du projet, sur les choix technologiques effectués (le machine learning, le deep learning,...), sur la manière de collecter les données ainsi que sur la mise en œuvre de l'application.

Pour apprécier correctement les enjeux, il est nécessaire de tenir compte du niveau des coûts de développement et de l'importance des coûts d'utilisation. En effet, le montant des investissements qui doivent être effectués pour développer des applications d'Intelligence Artificielle, notamment en matériel, peut être très significatif.

En conséquence, les bénéfices attendus de l'Intelligence Artificielle sont très nombreux et variés :

- Simplification des procédures ;
- Automatisation des processus ;
- Facilitation des opérations pour les usagers ;

- Accroissement de la productivité ;
- Baisse des coûts de revient ;
- Augmentation de la sécurité des systèmes ;
- Amélioration de la qualité et de l'efficacité des opérations ;
- Analyse des comportements des clients et des prospects ;
- Augmentation des chiffres d'affaires et des marges ;
- Amélioration de la rentabilité des investissements ;
- Anticipation et suppression des goulets d'étranglement ;
- Réduction des délais et amélioration de la réactivité ;
- Limitation des biais,
- ...

Cependant les entreprises ne pourront engranger ces bénéfices que si l'Intelligence Artificielle est correctement maîtrisée.

CHAPITRE 4

LES RISQUES ET LES INCERTITUDES DE L'INTELLIGENCE ARTIFICIELLE

Comme évoqué, les usages de l'Intelligence Artificielle se développent rapidement et ils présentent des avantages certains. Mais il existe aussi des risques potentiels qu'il est nécessaire de prendre en compte.

On doit d'abord tenir compte des risques inhérents à tous les projets et notamment ceux recourant à l'informatique : périmètre fonctionnel incertain, fonctions mal définies, sous-évaluation des opérations à effectuer, réalisation imparfaite, tests insuffisants,...

Il y a ensuite les risques liés à l'exploitation et à la maintenance des systèmes : complexité des opérations, puissance de traitement insuffisante, sécurité incertaine, dérive fonctionnelle, etc.

Enfin il y a aussi les risques liés aux biais et aux erreurs. Ces derniers risques sont variés et sont souvent dus à des confusions et des ambiguïtés. Ils sont différents selon le type d'Intelligence Artificielle mis en œuvre :

- Les données prises en compte ne sont pas représentatives, sont imparfaites ou elles n'ont pas été validées et finalement ne sont pas fiables. Elles peuvent aussi être biaisées[23]. Souvent ces données sont mal perçues ou mal interprétées. Elles peuvent aussi ne pas être statistiquement significatives, recourir à des variables ayant un faible coefficient de corrélation ou un faible niveau de confiance. Par exemple, Ryanair effectuait régulièrement des enquêtes auprès de ses clients proposant des questions fermées avec cinq réponses possibles mais celles-ci ne couvraient pas l'ensemble du spectre des réponses envisageables (de satisfaisant à pas satisfaisant du tout). En conséquence, n'ont répondu que les clients qui se sont retrouvés dans les réponses disponibles. L'analyse a montré que cette pratique conduisait à un biais de représentativité et à des erreurs d'interprétation.
- De même, il existe à côté de biais involontaires des biais volontaires. Par exemple, Parcoursup, (même si ce logiciel n'est pas un système d'Intelligence Artificielle) présente un cas de correction de biais qui pourrait poser problème au regard d'un éventuel commissaire aux algorithmes. En effet, un redressement des notes des élèves est appliqué selon les moyennes des lycées concernés. Or, il existe des lycées laxistes et d'autres qui notent les

[23] Un biais est un défaut de collecte de données qui donne une image faussée de la réalité. Le biais peut aussi être une erreur de méthode ou le non-respect des règles d'un protocole ce qui engendre des résultats erronés.

élèves de manière plus sévère. Ceci fait que les élèves des lycées laxistes sont avantagés au détriment des élèves des lycées qui ont noté sévèrement. Pour éviter cela, des ajustements sont effectués afin de corriger ces situations.

- Il existe des risques de biais et de discriminations dus à des erreurs. En effet, l'Intelligence Artificielle repose sur l'identification de corrélations qui constitue, comme dans les comportements humains, la base des processus d'apprentissage. Mais il peut arriver qu'on relie des variables qui, en définitive, ne sont pas liées entre elles. Ceci peut mener à des schémas de pensée trompeurs et faussement logiques.
- On confond souvent les risques absolus et les risques relatifs. Les conséquences ne sont pourtant pas les mêmes. Ainsi en matière de médicaments, il y a ceux qui sont déconseillés avec des effets secondaires et d'autres qui sont catégoriquement prohibés. Il en est de même en matière de nutrition, de produits phytosanitaires, de violences... Il faut se rappeler qu'il existe toujours des risques tolérables et des risques extrêmes. On doit admettre qu'il subsiste toujours le risque qu'une petite erreur puisse se traduire par un résultat considérable et erroné comme l'on évoque en météorologie l'effet papillon : un battement d'aile ici provoque un ouragan aux antipodes.

La corrélation est un outil statistique remarquable mais en aucun cas, une bonne corrélation entre deux variables ne présume un lien de causalité. C'est seulement un indice de l'existence probable d'une causalité. L'expérience montre qu'il existe de nombreuses corrélations trompeuses. La détermination d'un lien de causalité entre deux variables doit toujours reposer sur une approche méthodique. Cette dernière se fonde sur une réflexion théorique et des vérifications expérimentales. Pour cela on identifie différentes populations et elles sont testées pour voir si les résultats sont stables.

Pour s'assurer qu'il n'y a pas de biais inacceptables, il est nécessaire de vérifier méthodiquement qu'il s'agit bien d'une causalité et non pas d'une corrélation simplement due au hasard. C'est fondamental. Pourtant l'expérience montre qu'en pratique, ces vérifications sont très rares et, lorsqu'elles sont effectuées, elles sont très souvent réalisées d'une manière statistiquement peu rigoureuse.

Ainsi, il existe très souvent, des confusions entre ces différents concepts notamment lorsque des exemples sont donnés pour illustrer ces problématiques qui sont, en réalité, eux-mêmes potentiellement erronés ou biaisés.

La lutte pour le contrôle des biais est un élément déterminant de l'éthique de l'Intelligence Artificielle. Elle repose sur des principes simples : l'équité, la non-discrimination, le libre

consentement, l'autonomie, la responsabilité,.... Or, on constate que la manière dont les médias et les réseaux sociaux évoquent ces problèmes ne peut que renforcer la confusion concernant les choix éthiques à pratiquer. Les débats actuels autour de la voiture autonome ou des systèmes d'aide au recrutement recourant à l'Intelligence Artificielle font apparaître un haut degré de confusion. Le contrôle des biais pouvant être induits par des systèmes recourant à l'Intelligence Artificielle est absolument nécessaire pour que les personnes qui les utilisent aient confiance.

Ainsi, en matière de reconnaissance faciale, un biais a été mis en évidence concernant la reconnaissance des personnes de couleurs. En effet, les systèmes reconnaissent mieux les « blancs » que les « noirs ». Les systèmes recourant à l'Intelligence Artificielle seraient-ils racistes ? L'analyse montre qu'il ne s'agit pas réellement d'un biais mais d'une erreur liée au processus d'apprentissage. En l'occurrence, on a utilisé une base de données contenant les visages de 800 millions de personnes dont 700 millions de « blancs » et seulement 100 millions de « noirs ». Le système d'Intelligence Artificielle présente donc un taux d'erreurs plus élevé dans le cas de reconnaissance des « noirs » que des « blancs ».

Ceci est dû au fait que le moment où l'apprentissage pouvait être considéré comme suffisamment fiable pour être utilisé de manière opérationnelle n'a pas été évalué. Ceci pose une importante question éthique : si le taux de fiabilité concernant deux populations différentes, comme par exemple des hommes et des femmes, ou des « noirs » et des « blancs », n'est pas le même, peut-on alors recourir de manière opérationnelle à un système d'Intelligence Artificielle ? Quel taux de fiabilité est acceptable ? Doit-on arriver à un taux d'erreur identique dans les deux populations avant de pouvoir utiliser l'application ? Sinon, à quel moment l'écart de fiabilité entre ces deux populations sera considéré comme suffisamment faible pour pouvoir l'utiliser à bon escient ?

Dans le cadre des systèmes d'Intelligence Artificielle mis en œuvre dans les voitures autonomes, on peut imaginer que ce type de problème s'applique à la reconnaissance de nombreux types de populations (hommes, femmes, enfants, cyclistes, chats, chiens, oiseaux, drapeaux flottants, feuilles d'arbre,...) pour lesquels le nombre de données traitées lors de l'apprentissage conduit à des taux de non-reconnaissance variés dus à des fiabilités différentes du processus de reconnaissance. Combien faudra-t-il écraser de chiens et de chats avant de les détecter et de les éviter ? Dans ces conditions peut-on mettre en exploitation sans risque ce type de système recourant à l'Intelligence Artificielle ?

Il est possible de citer d'autres exemples de différence de taux de fiabilité dans différents sous-groupes. Chez Amazon, un système d'Intelligence Artificielle facilite le recrutement en analysant les CV reçus et en tenant compte des éléments disponibles sur le passé de la personne. Or, on

constate que le système rejette un pourcentage plus élevé de dossiers de femmes que de dossiers d'hommes. Est-ce que le système d'Intelligence Artificielle serait « machiste » ? Est-ce qu'il ne reflète pas plutôt les choix potentiellement biaisés qu'a fait l'entreprise dans le passé, du fait par exemple, des préjugés des recruteurs ? Ce phénomène préexistant serait amplifié et poussé à l'extrême par les traitements systématiques permis par les algorithmes d'Intelligence Artificielle.

Si le critère du genre est éliminé du processus de choix, et que les autres corrélations sont retenues, on constate que l'algorithme sélectionne alors entre 80 et 95 % de candidats masculins. Ceci s'expliquerait principalement par le fait que l'entreprise reçoit un plus grand nombre de CV masculins dû au simple fait qu'il y a plus d'étudiants ayant un diplôme d'informatique ou de système d'information que d'étudiantes. Ces biais ou ces discriminations sont-elles éthiquement acceptables ? Doit-on corriger ce biais en favorisant l'égalité des opportunités ou l'égalité des résultats avec 50 % d'hommes et 50 % de femmes embauchés ? On voit bien ici l'importance de prendre en compte les biais et les corrélations entre les facteurs ainsi que les conséquences que ces biais peuvent impliquer.

En matière de sélection des candidatures, on est tenté de penser que le processus d'apprentissage de l'Intelligence Artificielle est mature à partir du moment où on constate un taux d'échec de recrutement très faible pour les hommes. Ceci est dû au fait que le nombre de données disponibles les concernant est très élevé. Dans ces conditions, les résultats obtenus pourraient être utilisés de manière opérationnelle. Par opposition, les données concernant les femmes étant nettement moins nombreuses, le processus d'apprentissage est imparfait. Cela se traduit par un taux d'échec très élevé. Il serait peut-être nécessaire de définir un processus de recrutement spécifique pour les femmes, en attendant d'avoir une quantité de données suffisantes à traiter dans l'application.

Le choix des variables à prendre en considération est très important. Par exemple, des analyses mal conçues montrent que le taux de démence de la population augmente lorsque les personnes vivent près de routes très fréquentées alors que chacun sait que les facteurs clés expliquant la démence sont l'obésité, le fait d'être fumeur, le niveau d'éducation, etc.

Ces problèmes de biais et de discriminations liés à des niveaux de maturité insuffisants des systèmes d'Intelligence Artificielle sont fondamentaux. Il est pour cela important de mettre en œuvre une Intelligence Artificielle qui soit « éthique ». Cette préoccupation apparaît de plus en plus clairement dans la succession des documents de l'OCDE, de la Commission et du Parlement Européen sur le contrôle de l'Intelligence Artificielle. Sans un effort important de clarification, il y a des risques sérieux de prendre de mauvaises décisions. Pour éviter ces mauvais choix, différentes mesures sont envisageables :

- Il est souvent proposé de mettre en place des « contre-biais » (c'est la notion de « nudging »[24] due à Richard Thaler, ce qui lui a valu un prix Nobel) pour compenser les biais d'origine. Est-il éthique de considérer que l'introduction de biais ne serait pas éthique mais que l'introduction de contre-biais (qui restent néanmoins des biais) le serait ? La mise en œuvre de contre-biais ne serait-il pas contradictoire avec l'application de plusieurs principes comme l'autonomie, la responsabilité et le consentement ?
- Ces confusions sont dues au fait qu'il existe de fortes appréhensions quant au remplacement de l'homme par la machine ;
- A cela s'ajoute le fait qu'il existe de nombreux autres types de risques comme les deep fakes, les cyberattaques, la non-application du RGPD, etc.

Autre question posée : qui va subir les conséquences des risques liés aux applications d'Intelligence Artificielle ? Il est pour cela nécessaire d'identifier les différentes parties prenantes concernées par ces applications. Or, on constate souvent dans ces projets d'application une confusion entre le rôle et la responsabilité des différentes parties prenantes tel que les sponsors des systèmes, leurs concepteurs, les développeurs et les programmeurs, les data-scientists,... notamment en ce qui concerne les différents aspects à risque comme les biais, les erreurs, la gestion des opérations,... Pour cela, il est nécessaire d'effectuer une analyse des risques et plus particulièrement :

- Les risques doivent être clairement identifié et analysés ;
- Leur probabilité de survenance doit être estimée ;
- Les conséquences de ces risques doivent être évaluées ;
- Des mesures doivent être proposées pour diminuer le niveau de risque accepté.

Cette démarche doit permettre de dégager des recommandations de bonnes pratiques qui serviront à mieux maîtriser les applications recourant à l'Intelligence Artificielle.

[24] Le nudge est souvent considéré comme une technique de manipulation. Est-elle acceptable lorsqu'elle est utilisée à des fins positives ?

CHAPITRE 5

LE DROIT DE L'INTELLIGENCE ARTIFICIELLE

La montée en puissance des traitements algorithmiques dans nos vies quotidiennes rend de plus en plus nécessaire la prise en compte d'un cadre juridique protecteur des droits fondamentaux (non-discrimination, sécurité, transparence, neutralité et intégrité intellectuelles, maîtrise par l'utilisateur, etc.). Dans son livre blanc sur l'Intelligence Artificielle publié en février 2020[25], la Commission européenne a souligné l'importance de trouver des définitions suffisamment souples pour tenir compte des progrès techniques tout en étant suffisamment précises pour garantir la sécurité juridique nécessaire.

La notion d'algorithme conditionne le droit actuellement applicable et va permettre de construire sur cette base de nouveaux droits.

Pour l'heure, la notion de traitement de l'information doit être rapprochée de la notion de « *traitement automatisé de données à caractère personnel* » qui lui, fait l'objet d'un encadrement juridique et fournit un socle utile à la régulation des algorithmes :

- **La régulation relative aux données personnelles.**
  En France, la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés a donné très tôt un cadre juridique à l'informatique et au traitement des données à caractère personnel.

  Son article 1er resté intact lors de la refonte issue du RGPD[26], prévoit que « *l'informatique doit être au service de chaque citoyen* » et qu'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* ». Ce corpus normatif entretient des liens très étroits avec les techniques d'Intelligence Artificielle probabilistes qu'il réglemente à travers l'usage des données à caractère personnel. Ainsi :
  - Les bases de données d'apprentissage doivent être exploitées de manière licite ;
  - Les décisions prises sur le fondement exclusif d'un traitement automatisé, c'est-à-dire déléguées à la machine, font l'objet d'un encadrement dédié ;

[25] Livre blanc, *Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance*, Bruxelles, 19 février 2020, COM(2020) 65 final. Voir : https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf

[26] La loi du 6 janvier 1978 a fait l'objet d'une refonte à la suite de l'entrée en application du RGPD par la loi n° 2018-493 du 20 juin 2018 et l'ordonnance n° 2018-1125 du 12 décembre 2018. Voir : https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/.

- Les obligations de transparence, d'information et de compréhension permettent à la personne concernée de se positionner par rapport aux traitements algorithmiques dont elle peut faire l'objet.

La loi n'interdit pas les processus de prise de décision entièrement automatisés cependant elle prévoit une obligation d'information et de transparence accrue. Elle porte sur l'existence d'une prise de décision automatisée et sur les informations permettant de comprendre la logique sous-jacente de l'algorithme.

- **La régulation relative à la transparence.**
  La récente loi pour une République numérique[27] a fait entrer les codes sources dans la liste des documents communicables comme étant des documents administratifs. Ceci fait qu'elle consacre le droit d'accès aux codes sources des algorithmes publics et aux bases de données des administrations.

  Le Code des relations entre le public et l'administration (CRPA) pose un principe de transparence des algorithmes au terme duquel les administrations ont l'obligation d'avertir leurs usagers (art. L. 311-3-1) lorsqu'une décision individuelle est prise sur le fondement d'un traitement algorithmique, et doivent publier les règles définissant les principaux traitements algorithmiques utilisés pour de telles décisions (art. L. 312-1-3).

  En 2018, une profonde réforme de l'algorithme d'admission post-bac (APB) a été réalisée à la suite d'un rapport de la Cour des Comptes jugeant « opaques » les mécanismes de l'algorithme devenu un « outil de sélection et d'orientation »[28]. Cette réforme a donné lieu à la nouvelle plateforme Parcoursup. La loi du 8 mars 2018 relative à l'orientation et à la réussite des étudiants a rendu publics les algorithmes nationaux de la procédure d'accès à l'université. Cependant, notons que les critères locaux manqueraient encore de transparence.

  La CNIL[29] ainsi que le rapport Villani[30] proposent d'ouvrir les « boîtes noires » pour accroître

[27] Loi 2016-1321 du 7 octobre 2016 pour une République numérique : Version consolidée sur Légifrance. Voir : https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/.
[28] Rapport Cour des Comptes, Admission post-bac et accès à l'enseignement supérieur : Un dispositif contesté à réformer, octobre 2017. Voir : https://www.ccomptes.fr/fr/publications/apb-et-acces-lenseignement-superieur-un-dispositif-conteste-reformer .
[29] CNIL, « Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle », Synthèse du débat public, Décembre 2017. Voir : https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de .
[30] Rapport Villani, « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne » du 28 mars 2018. Voir : https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation .

la transparence et faciliter l'audit des systèmes autonomes.

- **La responsabilité des producteurs de systèmes d'Intelligence Artificielle.**
  Les risques d'algorithmes malfaisants sont déjà présents (espionnage, chatbot discriminant, véhicule autonome défaillant, etc.). La responsabilité des producteurs de systèmes d'Intelligence Artificielle est immense en ce domaine.

  Lawrence Lessing, professeur à la Harvard Law School et fondateur du *Center for Internet and Society*, constate que « *le code régule. Il implémente – ou non – un certain nombre de valeurs. Il garantit certaines libertés, ou les empêche. Il protège la vie privée, ou promeut la surveillance* »[31]. Les producteurs de systèmes d'Intelligence Artificielle ont un pouvoir « *discrétionnaire* » dans la rédaction du code de leurs programmes, comme le résume Lawrence Lessing : « *Code Is Law* »[32].

  Pour se protéger des dérives dans l'utilisation des algorithmes tant à l'égard des personnes physiques (biais et discriminations dans la prise de décisions) que des entreprises (notamment à cause de pratiques anticoncurrentielles[33]), des règles doivent être instaurées.
- **Vers un droit des algorithmes.**
  Il faudra nécessairement créer un droit particulier, le droit classique étant inopérant pour encadrer les systèmes algorithmiques. Des solutions existent, mais elles sont totalement inadaptées aux systèmes à base d'Intelligence Artificielle. Ceci fait qu'elles peuvent être aussi dangereuses que bénéfiques. Les principes de ce nouveau droit concerneront[34] nécessairement :
  - Le droit à la protection de la vie humaine : ce droit pose la question des critères qui peuvent être retenus par l'Intelligence Artificielle d'une voiture autonome placée dans une situation de dilemme impliquant la mort d'autrui ;
  - Le droit à l'intimité : jusqu'où peut-on s'immiscer dans la connaissance d'un individu ?

[31] Laurence Lessig, « Code Is Law. On Liberty in Cyberspace » : Harvard Magazine, janvier 2000, disponible en anglais sur http://harvardmagazine.com/, traduction française disponible sur Framablog.org/. Voir https://framablog.org/2010/05/22/code-is-law-lessig/

[32] Lauwrence Lessig, « Code Is Law. On Liberty in Cyberspace », précité.

[33] La Commission européenne a infligé à Google une amende de 2,42 milliards d'euros pour abus de position dominante sur le marché des moteurs de recherche en favorisant son propre service de comparaison de prix («Google Shopping»).

[34] Il existe 15 droits différents concernant les systèmes à base d'Intelligence Artificielle : Le droit à l'auto-développèrent, Le droit à la dignité, Le droit à l'indépendance, Le droit à l'intimité, Le droit de l'accès aux algorithmes, Le droit à d'alerte, Le droit à la compréhension, Le droit à l'information, Le droit à la médiation, Le droit à la neutralité, Le droit de décision en dernier ressort, Le droit à l'explicabilité, Le droit à l'impartialité, Le droit de jouabilité, Le droit du principe de responsabilité autonome.

- Le droit à la décision : il s'agit du droit le plus difficile à appréhender, celui de la décision en dernier ressort[35].

Les systèmes à base d'Intelligence Artificielle s'intéressent à ces trois principes et apportent des solutions beaucoup plus adaptées que celles proposées par les humains[36].

Quoiqu'il en soit, nous entrons dans une ère de dépendance à la technologie. Un certain nombre d'auteurs qualifient le phénomène « d'algocratie » pour décrire le pouvoir qu'ont pris les plateformes, au travers de leurs algorithmes de filtrage et de tri exercés dans nos vies[37]. Il est donc plus que nécessaire de revenir aux principes fondamentaux de respect de la dignité humaine et de non-discrimination tels que posés par la Charte des droits fondamentaux de l'Union européenne[38].

[35] Alain Bensoussan, « Intelligence artificielle : quelle régulation pour le code ? » : Blog expert sur le site du Figaro.fr, 11 décembre 2017. Voir : https://blog.lefigaro.fr/bensoussan/2017/12/intelligence-artificielle-quelle-regulation-pour-le-code.html

[36] Alain Bensoussan, J. Bensoussan, IA, robots et droit, Ed. Bruylant 2019.

[37] Patrick Cingolani, « Ubérisation, turc mécanique, économie à la demande : où va le capitalisme de plateforme ? » : The Conversation.com, 26 août 2016. Voir : https://theconversation.com/uberisation-turc-mecanique-economie-a-la-demande-ou-va-le-capitalisme-de-plateforme-64150

[38] Charte des droits fondamentaux de l'Union européenne. Voir : https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:12016P/TXT&from=FR

CHAPITRE 6

L'ETHIQUE DE L'INTELLIGENCE ARTIFICIELLE

Le développement de la réflexion sur la gouvernance des systèmes d'Intelligence Artificielle se traduit par une préoccupation croissante de la dimension de l'Ethique. Les textes de référence mettent de plus en plus en avant cette dimension.

Le texte de référence de l'OCDE, « *Recommandation du Conseil sur l'Intelligence Artificielle* », n'évoque pas directement l'éthique mais mentionne des « *valeurs centrées sur l'humain et l'équité* ». Il est précisé : « *Les acteurs de l'Intelligence Artificielle devraient respecter l'état de droit, les droits de l'homme et les valeurs démocratiques tout au long du cycle de vie des systèmes d'Intelligence Artificielle. Ces droits et valeurs comprennent la liberté, la dignité et l'autonomie, la protection de la vie privée et des données, la non-discrimination et l'égalité, la diversité, l'équité, la justice sociale, ainsi que les droits des travailleurs reconnus à l'échelle internationale*. » (page 6). Ce n'est pas la définition classique de la notion d'éthique mais le périmètre ainsi décrit par cette liste recouvre une bonne partie de la notion de gouvernance des sociétés libérales : « l'état de droit », « les droits de l'homme », « les valeurs démocratiques », « la liberté », « la dignité », « l'autonomie », « la protection de la vie privée et des données », « la non-discrimination », « l'égalité », « la diversité », « l'équité », « la justice sociale » et « les droits des travailleurs ».

Le Livre Blanc de la Commission Européenne intitulé « *Une approche européenne axée sur l'excellence et la confiance* », propose une démarche voisine en proposant « de créer un « ***écosystème de confiance*** » unique en son genre. Pour cela, il devra garantir le respect des règles de l'UE, notamment celles qui protègent les droits fondamentaux et les droits des consommateurs, en particulier pour les systèmes d'Intelligence Artificielle à haut risque exploités dans l'UE. La création d'un écosystème de confiance est un objectif stratégique en soi, qui devrait susciter chez les citoyens la confiance nécessaire pour adopter les applications d'Intelligence Artificielle et donner aux entreprises et aux organismes du secteur public la sécurité juridique voulue pour innover au moyen de l'Intelligence Artificielle» (page 3). Cette approche est moins large que celle proposée par l'OCDE mais les objectifs sont voisins : *« le respect des règles de l'UE », « les droits fondamentaux », « les droits des consommateurs », « un écosystème de confiance », « la confiance nécessaire pour adopter les applications d'Intelligence Artificielle » et « la sécurité juridique nécessaire aux entreprises et aux organismes du secteur public ».*

Le Parlement Européen a adopté en Octobre 2020 un « *Cadre pour les aspects éthiques de l'Intelligence Artificielle, de la robotique et des technologies annexes* ». Il élargit la notion d'éthique

appliquée à l'Intelligence Artificielle : « *le développement, le déploiement et l'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes à haut risque, notamment, mais pas exclusivement, par des êtres humains, devraient toujours obéir à des principes éthiques et être conçus pour respecter et favoriser l'action humaine et le contrôle démocratique, et permettre aux êtres humains de reprendre le contrôle dès que nécessaire en mettant en œuvre des mesures de contrôle appropriées* » (page 9). Les points clés mis en avant sont le respect « *des principes éthiques* », le « *respecter et favoriser l'action humaine et le contrôle démocratique* », et fait que des « *êtres humains doivent contrôler* » les systèmes. L'objectif est de : « *promouvoir l'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes au sein de l'Union en veillant à ce qu'elles soient développées, déployées et utilisées de manière conforme aux principes éthiques* » (page 35).

Mais que sont ces principes d'éthiques ? Les textes de références ne le disent pas. En Octobre 2018 le Gigref et le Syntec ont publié un rapport : « Ethique et Numérique » qui précise pages 10 et 11 :

L'Intelligence Artificielle pose des problématiques éthiques spécifiques. Les algorithmes d'apprentissage machine sont en pleine expansion dans de nombreux secteurs. Ces algorithmes, qui apprennent à partir de nombreux exemples, manquent de transparence et d'outils de traçabilité permettant d'expliquer leurs résultats. D'où l'expression « boîte noire », souvent utilisée pour qualifier l'opacité de certains systèmes. L'explicabilité des algorithmes nécessite la prise en compte de plusieurs niveaux :

- **La sélection des jeux de données :**
  certains jeux de données qui servent de base d'apprentissage aux algorithmes peuvent véhiculer des biais cognitifs. Certaines bases de données peuvent par exemple, dans un domaine particulier, contenir un biais culturel et historique en termes de représentativité homme – femme. Il semble important de pouvoir expliquer le contenu des données sélectionnées et utilisées par des algorithmes apprenants afin de s'assurer de leur neutralité.
- **Le suivi de l'apprentissage :**
  il existe un risque de reproduction d'injustices ou de discriminations dans l'apprentissage machine. C'est pourquoi la supervision de l'apprentissage machine est particulièrement importante. Certains développeurs mettent en place des processus d'évaluation spécialement dédiés à la question de la neutralité dans l'apprentissage, en réservant des phases de test avant un déploiement opérationnel.

- **L'acceptabilité sociale :**
  certains systèmes algorithmiques ont un impact social non négligeable ; ils peuvent par exemple influencer massivement des comportements politiques par les « *bulles de filtres* ».

Cette approche est différente de l'approche classique des auditeurs centrés sur la conformité. « *En effet, la conformité consiste à agir en accord avec une norme, une loi, à quelque chose qui est extérieur à soi et qui a force d'autorité. Il est donc de la responsabilité de chacun de respecter la loi sous peine de sanctions.*

L'éthique, en revanche, est une réflexion personnelle ou collective (à l'échelle d'une entreprise par exemple) qui consiste à se donner à soi-même ses propres lignes de conduite. Cette réflexion incarne des valeurs ou des principes servant de guide aux actions. L'éthique est un acte de responsabilisation (et non uniquement de responsabilité), d'engagement et d'intégrité ». (page 5 du rapport Ethique et Numérique du Cigref-Syntec)

Comme le mentionne le rapport de Cédric Villani *« Donner un sens à l'Intelligence Artificielle » : « Dans ces cas où la norme est inexistante, muette ou insuffisante, la responsabilité morale du développeur est accrue ».*

Dans ces conditions l'éthique de l'Intelligence Artificielle est constituée par toutes les mesures permettant d'éviter les dérives liées à l'absence ou au non-respect des normes appliquées dans une société civilisée.

CHAPITRE 7

VERS UNE INTELLIGENCE ARTIFICIELLE DIGNE DE CONFIANCE

Pour répondre à la problématique d'une Intelligence Artificielle digne de confiance, il sera nécessaire de mettre en place des moyens adaptés, c'est-à-dire des structures organisationnelles, des directives et des référentiels, des comportements, des processus, etc. Il est surtout important de disposer de référentiels de bonnes pratiques en la matière. Elles devront être identifiées voire, le cas échéant, adaptées.

Par ailleurs, il est nécessaire d'obtenir une assurance indépendante faite, par exemple par des commissaires aux algorithmes, qui pourrait fournir le niveau de confiance souhaité. De même, il convient de disposer de moyens adaptés pour fournir le niveau d'assurance attendu comme, par exemple des référentiels d'audit, des outils d'audit, des compétences nécessaires, des règles d'éthique, etc.

Les moyens nécessaires pour obtenir des systèmes à base d'Intelligence Artificielle digne de confiance sont des référentiels, des outils et des bonnes pratiques.

Démarche et dispositif de gouvernance à mettre en place pour évaluer les options, arbitrer et fixer les orientations en matière de systèmes d'Intelligence Artificielle dignes de confiance

Plusieurs parties prenantes sont concernées par les systèmes à base d'Intelligence Artificielle. Chacun a des intérêts propres qui peuvent être parfois divergents, mouvants, conflictuels, contradictoires, à court ou long terme,... Chacun souhaite créer de la valeur qui lui soit propre c'est-à-dire qui lui confère des avantages ou des bénéfices en maintenant un niveau de risque acceptable et mobilisant un niveau de ressources acceptable, de manière responsable. Chacun aura donc ses propres attentes en matière d'exigences.

Pour répondre à cette situation, il s'agira dans un premier temps d'évaluer les besoins, les conditions et les options de création de valeur pour les différentes parties prenantes, en vue de déterminer des objectifs et des exigences à atteindre convenus entre elles. Ensuite on doit s'attacher à fixer une orientation après une étape d'arbitrage, de priorisation et de prise de décision. Enfin on va piloter la performance et la conformité au regard des orientations, des objectifs et des exigences convenus préalablement.

Un dispositif de gouvernance performant devra ainsi être mis en place pour réaliser cette première étape.

Des dispositifs de management et opérationnel qui soient performants

Il sera ensuite nécessaire de planifier, de développer ou d'acquérir, d'exploiter et de surveiller les activités et les moyens en cohérence avec l'orientation, les objectifs et les exigences convenus préalablement. Des dispositifs de management et opérationnel qui soient performants devront ainsi être mis en place pour réaliser cette deuxième étape en liaison et en cohérence avec le dispositif de gouvernance.

Des bonnes pratiques spécifiques pour répondre aux exigences et spécificités des systèmes à base d'Intelligence Artificielle

Des bonnes pratiques génériques et contextuelles pour chacun des types de dispositifs devront être identifiées pour chacune des exigences ou préoccupations à traiter. De nombreux référentiels fournissent des exemples de bonnes pratiques usuellement utilisées pour traiter les préoccupations et exigences relatives aux systèmes d'information et au numérique. Ces bonnes pratiques pour la plupart d'entre elles peuvent s'appliquer aux systèmes à base d'Intelligence Artificielle. On peut citer COBIT 2019, ITIL V4 et ISO.

Néanmoins, il convient de les compléter de bonnes pratiques répondant aux exigences et préoccupations spécifiques aux systèmes à base d'Intelligence Artificielle telles que la transparence notamment l'explicabilité et la traçabilité ou l'équité et la non-discrimination. A cet effet, il existe des techniques et des outils spécifiques qui sont en cours de développement. Ils peuvent répondre à certaines exigences mais ils ont aujourd'hui leurs limites. Il est donc nécessaire d'identifier ces limites mais aussi les bonnes pratiques qui en l'état actuel permettraient néanmoins de fournir un bon niveau de confiance quant à certaines assertions.

La nécessité d'être rassuré par un tiers indépendant

Pour effectuer ces missions d'audit, il est indispensable de recourir à un professionnel compétent et indépendant. Il va commencer par cadrer l'audit puis ensuite définir les éléments nécessaires pour réaliser cet audit et enfin convenir de la nature des travaux d'audit à effectuer.

CHAPITRE 7

Définition de la proposition de valeur adaptée pour l'audit et la certification des systèmes à base d'Intelligence Artificielle

Les différents acteurs du système financier et économique qu'ils soient actionnaires, employés, banquiers, fournisseurs, clients, citoyens ou l'Etat lui-même ont besoin d'avoir confiance dans les éléments financiers qui sous-tendent leurs décisions d'investisseurs, de prêteurs, de partenaires, d'achats, etc. Sans confiance, il ne peut y avoir de croissance.

Pour favoriser cette confiance, les commissaires aux comptes, tiers indépendants, sont mandatés pour certifier la sincérité et la régularité des états financiers des entreprises privées et publiques, des associations et de l'Etat. C'est une mission d'intérêt général pour le compte de l'ensemble de l'écosystème. Ainsi, les grandes entreprises du CAC40 sont prêtes à payer plusieurs dizaines de millions d'euros en honoraires chaque année pour rassurer l'écosystème financier sur leurs comptes.

Compte tenu des enjeux de l'écosystème des systèmes à base d'Intelligence Artificielle et de l'impératif d'avoir des applications dignes de confiance, il apparaît nécessaire, tel que c'est le cas pour l'écosystème financier, de mandater un tiers de confiance indépendant, tel qu'un commissaire aux algorithmes, qui aurait une mission d'intérêt général de certification des assertions relatives à certaines exigences de qualité attendues des systèmes d'Intelligence Artificielle notamment celles à hauts risques. Ce tiers de confiance doit avoir une qualification professionnelle voire une certification type CISA[39].

Cette certification doit pouvoir créer de la valeur. Il s'agit donc de trouver le bon équilibre entre différentes approches :

- La nature des assertions attendues comme la transparence, l'équité et la non-discrimination, la fiabilité, la sécurité, la robustesse, la protection de la vie privée, la régularité,... ;
- La nature de l'opinion recherchée. Ce peut être une certification, un label, une attestation, un rapport,... ;
- Le niveau de confiance souhaité pour ces assertions ;
- Le niveau de risque acceptable de fournir une certification erronée par exemple de certifier qu'un système à base d'Intelligence Artificielle est équitable alors qu'elle ne le serait pas. Ce serait un faux positif. Mais ce peut aussi être un faux négatif en certifiant qu'elle ne serait pas équitable alors qu'elle le serait ;

[39] CISA : Certified Information Système Auditor. Certification des auditeurs informatiques assurée par l'ISACA.

- Le niveau d'efforts acceptable pour fournir ce type et ce niveau de confiance notamment un niveau d'honoraires soutenable.

A cet effet, il est nécessaire de définir précisément la démarche et les éléments qui permettront d'identifier, d'évaluer et de fixer les options de niveau de confiance qu'il convient de fournir dans un contexte donné.

Convenir des éléments qui permettront de définir l'objectif d'un audit donné

L'objectif d'un audit est de créer de la valeur en fournissant, à un certain nombre de parties prenantes, un certain niveau de confiance, quant à une opinion relative à certaines assertions de l'objet audité à une date ou pour une période donnée. Cette opinion leur serait utile compte tenu des avantages et des bénéfices qu'elles en retireraient et dans la mesure où le niveau de risque que cette opinion soit erronée et le niveau d'effort nécessaire pour y arriver soient acceptables.

Il s'agira donc de qualifier de manière précise l'objectif visé.

Cette qualification pourra résulter d'une obligation légale telle celle qui est prévue dans la proposition de règlementation de la Commission Européenne pour les systèmes recourant à l'Intelligence Artificielle à hauts risques qui fixe les contours de l'audit et de la certification attendue ou d'une demande contractuelle fixée par une lettre de mission.

Convenir de la nature et du niveau des travaux d'audit

Une fois le ou les objectifs et le périmètre d'un audit déterminés, c'est-à-dire une fois les différents éléments énoncés ci-dessus établis, il est alors possible d'identifier la nature et le niveau des travaux d'audit nécessaires pour répondre aux attentes de qualité des parties prenantes.

En fonction de l'objectif fixé, les travaux d'audit peuvent consister à valider la « qualité » d'une ou plusieurs assertions ou d'un résultat précis selon un standard de référence convenu d'avance mais ils peuvent aussi évaluer la « qualité » des moyens mis en œuvre selon un référentiel de bonnes pratiques convenu d'avance pour permettre d'aboutir au niveau de qualité de résultat attendu. Très souvent on assiste à une combinaison entre la qualité d'un résultat et la qualité des moyens.

L'opinion fournie par l'auditeur peut prendre plusieurs formes différentes : certification, attestation, opinion circonstanciée, etc.

La nature et le niveau des travaux d'audit pourront alors être identifiés. Ils dépendront de deux facteurs. Il y a d'abord le niveau de risque inhérent aux éléments audités. Ce sont, par

exemple, le risque que les moyens et les bonnes pratiques mis en œuvre par l'organisation ne permettent pas d'obtenir le niveau de qualité attendu de l'objet audité. Il y a ensuite le niveau de risque inhérent aux travaux d'audit. C'est le risque que les travaux d'audit conduisent à une opinion contraire à la réalité. L'auditeur émet une opinion positive alors qu'elle devrait être négative ou une opinion négative alors qu'elle devrait être positive.

Il faut donc déterminer quel niveau de risque est acceptable par les parties prenantes sachant que pour réduire le niveau de risque, il faudrait allouer des moyens supplémentaires qui soient soutenables. En général ils augmentent de manière exponentielle par rapport au gain du niveau de confiance attendu.

Les travaux d'audit à effectuer consistent, par exemple à identifier les règlementations en vigueur, les obligations contractuelles, les contraintes,... qui s'appliquent aux différentes exigences convenues d'avance devant faire l'objet d'une opinion d'audit. Ils consistent aussi à identifier et à évaluer les types de risques qui pourraient conduire à des systèmes à base d'Intelligence Artificielle non conformes aux exigences convenues d'avance. Ce peut être un périmètre fonctionnel mal défini, des défauts de conception, une mauvaise implantation, des tests insuffisants, l'inexactitude des données, des biais, des erreurs de programmation, des erreurs de saisie, des erreurs d'interprétation, etc.

Dans le cadre de la mission d'audit, il est aussi nécessaire d'identifier et d'évaluer les réponses mises en œuvre par les différentes parties prenantes pour y répondre sur l'ensemble du cycle de vie des systèmes à base d'Intelligence Artificielle. Ceci concerne en particulier l'application de bonnes pratiques de gouvernance, de management, de gestion opérationnelle ou de contrôle mises en œuvre pour l'ensemble des leviers. Ces derniers sont constitués par les structures organisationnelles, les directives données, le choix des référentiels, les outils mis en œuvre, les comportements, les compétences, les processus,...

De même, il faut s'assurer de l'effectivité des processus en procédant à des vérifications quant à leur mise en œuvre appropriée. On pourra ainsi examiner la qualité du cahier des charges et de la documentation technique, le paramétrage, les incidents d'exploitation, les résultats des jeux d'essais, etc. Il est aussi intéressant d'évaluer l'impact des anomalies identifiées sur l'opinion de l'auditeur notamment la nécessité d'émettre des réserves ou de ne pas en émettre,... Enfin il est nécessaire d'évaluer de manière globale si les travaux effectués sont suffisants pour émettre une opinion.

Pour disposer du système à base d'Intelligence Artificielle digne de confiance, il faut que la mise en œuvre de pratiques d'Intelligence Artificielle soit digne de confiance et que les auditeurs s'approprient des pratiques d'audit dignes de confiance.

COMPOSITION DU GROUPE DE TRAVAIL

Groupe de travail animé par :

- Alain BENSOUSSAN, avocat, Lexxing Alain Bensoussan Avocats
- Serge YABLONSKY, expert-comptable, commissaire aux comptes, CGEIT, CISA, CRISC, SYC Consultants, Président d'honneur de l'ISACA-AFAI

Groupe de travail actif :

- Marie-Agnès GRAS, Juriste propriété intellectuelle et droit du numérique
- Jean-Laurent LIENHARDT, expert-Comptable mémorialiste
- Camille ROSENTHAL-SABROUX, Professeur émérite Université Paris Dauphine-PSL, Lamsade, Administrateur de l'ISACA-AFAI
- Claude SALZMAN, consultant en Système d'Information, Président International du Club Européen de la Gouvernance des Systèmes d'Information, Administrateur de l'ISACA-AFAI
- Patrick STACHTCHENKO, chargé de cours en Gouvernance, Cyber Sécurité, Audit IA et Référentiels, CGEIT, CRISC et CISA, Ancien président de l'ISACA International.

L'Académie est une initiative de l'Ordre des experts-comptables

L'Intelligence Artificielle suscite de grands espoirs mais soulève aussi de nombreuses craintes. Les possibilités offertes par cette technologie ne se développeront et son emploi ne se généralisera que si toutes les personnes recourant à ces systèmes possèdent une confiance suffisante dans l'Intelligence Artificielle et dans son fonctionnement. Or, aujourd'hui l'Intelligence Artificielle pose de nombreux problèmes de conception, de mise au point, de fiabilité, de biais, de sécurité et de confidentialité.

Dès lors, comment instaurer la confiance nécessaire dans les systèmes à base d'Intelligence Artificielle ? Très certainement en effectuant une évaluation objective de la conception, de la réalisation et, périodiquement, du fonctionnement et de la maintenance de ces systèmes. C'est le but de la gouvernance et de l'audit des algorithmes de ces applications. Cette démarche repose sur la connaissance des bonnes pratiques qui doivent normalement être appliquées par les développeurs de ces systèmes mais aussi par leurs promoteurs, par les entreprises qui les commercialisent et les exploitent. Ce sont des enjeux considérables car, sans risquer de se tromper, il est possible d'affirmer que ces nouvelles applications auront un impact considérable sur l'économie de nos pays et plus généralement sur le fonctionnement de nos sociétés.

Ce cahier n°38 « Intelligence Artificielle et confiance » est le premier ouvrage d'une série de publications sur l'audit des systèmes reposant sur l'Intelligence Artificielle. Il a été établi par un groupe de travail commun à l'ISACA-AFAI, chapitre français de l'association internationale ISACA et l'Académie des Sciences et Techniques Comptables et Financières. Celui-ci regroupe des professionnels des quatre métiers concernés : audit, informatique, droit et Intelligence Artificielle.

www.lacademie.info

CONTACTS

Académie des Sciences et Techniques
Comptables et Financières

200-216 rue Raymond Losserand 75014 Paris
Tél. +33 (0)1 44 15 64 24

www.lacademie.info

William NAHUM
Président fondateur

Pierre VALENCIEN
Responsable de l'Académie

Marie-Amélie CALMAO
Chargée administrative