L'IA, un défi pour nos législations vie privée

Poullet, Yves

Published in:
Kunstiche Intelligenz und Datenschutz

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

L'IA un défi pour nos législations vie privée,

Actes des 3èmes journées suisses du droit de la protection des données, FRIBOURG, 2.10.2020

Yves POULLET

Professeur émérite à la Faculté de droit de Namur
Professeur associé à l'UCLille
Membre de l'Académie royale de Belgique

1. « *Considérant que l'intelligence artificielle, la robotique et les technologies connexes, qui sont susceptibles d'ouvrir de nouvelles perspectives aux entreprises et de profiter aux citoyens, mais aussi d'avoir une incidence directe sur tous les aspects de nos sociétés, y compris les droits fondamentaux et les valeurs et principes sociaux et économiques de base, et d'exercer une influence durable sur l'ensemble des secteurs d'activité, connaissent un développement et une diffusion rapides*; ». Ainsi s'exprime la récente Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes.[1] Sans doute, l'opinion oscille entre la crainte d'une technologie qui, plus qu'hier encore, nous épiera et, fait nouveau, nous prédira, celle mise en œuvre par la société Cambridge Analytica[2] et l'espoir, mis dans la même technologie, celle qui avec un certain succès aide à combattre la maladie, notamment la propagation de la Covid.[3].

   Notre propos est d'analyser le rôle que peuvent jouer les législations de protection des données, en particulier le Règlement général de protection des données européen[4], pour la protection de nos libertés mais également les limites de leur apport tant intrinsèques, c'est-à-

---

1 P9_TA(2020)0275, 2020/2012/ INL, Introduction, p. 2.

2 Le scandale *Facebook/Cambridge Analytica* créé par la fuite de données des communications échangées sur le réseau social de Facebook et transmises à *Cambridge Analytica* depuis 2014 concerne les données à caractère personnel de 87 millions d'utilisateurs *Facebook*. Ces données ont permis à la société anglaise de profiler ces clients de Facebook, notamment sur leurs préférences politiques. Certains hommes politiques ont utilisé ces profils pour influencer les intentions de votes de leurs citoyens. Sur ce scandale, parmi de nombreux articles de journaux, lire « *Cambridge Analytica : 87 millions de comptes Facebook concernés* », *Le Monde*, 4 avril 2018

3 « L'intelligence artificielle (IA) est utilisée pour venir porter appui à la lutte contre la pandémie virale touchant le monde entier depuis le début de l'année 2020. La presse et la communauté scientifique se font écho des grands espoirs reposant sur la science des données et l'IA pour affronter le coronavirus (D. Yakobovitch, How to fight the Coronavirus with AI and Data Science, Medium, 15 février 2020). » CAHAI, *IA et lutte contre le coronavirus*, document disponible sur le site : IA et lutte contre le coronavirus Covid-19 (coe.int), (dernière consultation, le 12 janvier, 2021)

4 Nous n'avons malheureusement pas eu le temps d'analyser la toute récente loi suisse fédérale de protection des données du 20 septembre 2020, texte disponible sur le site : 17.059 | Loi sur la protection des données. Révision totale et modification d'autres lois fédérales | Bulletin officiel | Le Parlement suisse (parlament.ch), consulté le 12 janvier 2021). La loi suisse sur la protection des données a été modernisée et contient des changements importants pour les PME. La mise en œuvre de la nouvelle réglementation n'est prévue que pour la fin de 2021.

dire liées au contenu du texte face à la réalité de l'intelligence artificielle[5], qu'extrinsèques, dans la mesure où un telle législation n'envisage pas l'ensemble des risques encourus par la société et offre donc une protection limitée. Notre propos reprendra certaines réflexions déjà publiées[6]. Une brève introduction sur les aspects techniques, la typologie des systèmes et les acteurs de l'IA (Chapitre I) précédera l'analyse des risques liés aux applications de cette technologie (Chapitre II). Ces deux éléments nous permettront de comprendre à la fois le rôle du RGPD et ses limites (Chapitre III).

Chapitre I : Quelques réflexions sur le donné technologique, la typologie et les acteurs de l'IA

2. Notre propos n'est pas de reprendre ici un long exposé technique, ce qu'un autre orateur n'a pas manqué de faire de manière précise et excellente lors de cette troisième journée mais uniquement de pointer quelques spécificités des systèmes IA dont l'incidence détermine les nouveaux enjeux rencontrés par notre société et nos législations en particulier de protection de données. Commençons par cette déclaration provocante de l'INRIA[7], *« .« L'IA n'existe pas mais la puissance combinée des données disponibles, d'algorithmes et de ressources de calcul ouvre de formidables perspectives dans de nombreux domaines*. » Cette affirmation est forte venant d'un organisme reconnu pour avoir développé tant d'applications et d'algorithmes d'IA. Il est donc préférable de parler de '*machine learning'* ou d'apprentissage machine'[8] . L'appellation prônée met l'accent sur le fait qu'il faut s'interdire de voir dans la technologie du '*machine learning*' une quelconque

[5] On note que le Livre blanc de la Commission sur l'IA « Excellence and Trust » l'envisage ouvertement : « *Les développeurs et les déployeurs d'IA sont déjà soumis à la législation européenne sur les droits fondamentaux (c'est-à-dire la protection des données, le respect de la vie privée, la non-discrimination), sur la protection des consommateurs, et sur la sécurité des produits et la responsabilité du fait des produits. Les consommateurs entendent jouir du même niveau de sécurité et bénéficier des mêmes droits, qu'un produit ou un système soit fondé sur l'IA ou non. Or certaines spécificités de l'IA (telles que l'opacité) peuvent compliquer l'application et le contrôle de l'application de cette législation. Il faut, dès lors, examiner si la législation actuelle est en mesure de faire face aux risques liés à l'IA et si son respect peut être assuré efficacement, si elle doit être adaptée, ou si une nouvelle législation s'impose. » (Livre blanc,,* Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions du 19 février 2020 intitulée «Intelligence artificielle - Une approche européenne axée sur l'excellence et la confiance» (COM(2020)0065), p. 10.

[6] Ainsi, nous renvoyons le lecteur à la lecture de l'ouvrage publié par l'auteur, Le RGPD face à l'intelligence artificielle », Cahier du CRIDS, n°49, Larcier, Bruxelles, 2020 et au rapport B.FRENAY – Y. POULLET, « Profiling and Convention 108+ : Report on developments after the adoption of Recommendation (2010)13 on profiling », rapport établi pour le Conseil consultatif de la Convention n°108, novembre 2019 dans le cadre de la révision de la recommandation de 2010 sur le profilage, rapport à paraître.

[7] C'est par ses mots que l'INRIA débute son analyse de l'intelligence artificielle : INRIA, *Intelligence artificielle*, texte disponible à l'adresse : https://www.inria.fr/fr/intelligence-artificielle

[8] L'expression « traitement utilisant des procédés d'apprentissage automatique » (*machine learning*) signifie un traitement utilisant des méthodes particulières d'intelligence artificielle basé sur des approches statistiques pour donner aux ordinateurs la capacité d' « apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune.
Sur le 'machine learning' et les différents types des systèmes de *machine learning* (notamment, systèmes supervisés, systèmes de '*deep learning'),* lire B. FRENAY – Y. POULLET, *Profilage et la Convention 108+ : Rapport sur l'évolution de la situation après l'adoption de la Recommandation (2010)13 sur le profilage*, Rapport au Comité consultatif de la Convention n°108, Conseil de l'Europe, , Strasbourg 7 novembre 2019, T-PD(2019)07rev., p. 8 : « *Parmi les techniques d'intelligence artificielle, le machine learning a la spécificité de pouvoir exploiter les données disponibles pour permettre la création d'intelligences artificielles apprenantes. Plus particulièrement, lorsque les données sont des images, du son ou du texte, il est courant d'utiliser le deep learning, un sous-domaine du machine learning, qui permet de créer des réseaux de neurones adaptés à la modélisation de ces types de données.* »

intelligence au sens humain et français[9] du terme, c'est-à-dire une capacité de compréhension semblable à celle humaine. Les algorithmes exécutent, ils ne pensent pas. Il n'en reste pas moins vrai que ces systèmes dits IA ou *machine learning* simulent l'intelligence humaine puisqu'ils sont capables d'interpréter l'environnement et de prendre ou suggérer des décisions. Par ailleurs, ce choix terminologique écarte des théories comme celle de la 'singularité technologique », chère à VINGE[10]. Ce dernier auteur considère que l'évolution exponentielle de la technologie informatique sera-t-elle qu'elle dépassera l'entendement humain et à une date qu'il situe aux alentours de 2035 au plus tard, l'homme aura créé une intelligence supérieure à la sienne.

3. Pour la facilité, nous utiliserons cependant le terme 'intelligence artificielle ' (en abrégé IA) dans la mesure où ce terme plus évocateur de la puissance de nos ordinateurs est universellement consacré[11]. La proposition de règlement issue de la Résolution du Parlement européen en donne la définition suivante[12] : « « *un système qui est soit fondé sur des* ***logiciels (***(depuis le *credit rating*, le profilage, la reconnaissance faciale, la traduction automatisée, les *chatbot*s ..)***, soit intégré dans des dispositifs matériels (***les robots: voiture autonome, enceintes connectées, drônes, soldat augmenté, … ***)***, *et qui affiche un* ***comportement simulant l'intelligence****, notamment en collectant et traitant des données, en analysant et en* ***interprétant son environnement et en agissant****, avec un certain degré d'****autonomie****, pour atteindre des objectifs spécifiques. ».* Cette définition souligne l'importance des robots comme applications de l'IA, qu'ils soient humanoïdes ou non et caractérise l'IA par son autonomie[13], caractéristique sur laquelle nous revenons *infra,* n°.

   La proposition de règlement insiste sut le fait que l'IA s'intègre dans un **système** qui s'appuie en aval sur des bases de données ou mégadonnées dans lesquelles les algorithmes procéderont à des corrélations d'autant plus pertinentes et fines que ces mégadonnées seront riches en données[14] et, souvent nourries elles-mêmes par ces données collectées grâce à l'internet des objets. La constitution de ces mégadonnées et leur traitement par les systèmes d'IA résultent de l'augmentation continue tant de la capacité de stockage de nos serveurs, que de leur puissance, qui permet de collecter et stocker des données à l'infini et de disposer de systèmes algorithmiques de plus en plus performants. A la capacité quasi infinie de nos ordinateurs tant

---

[9] La langue anglaise propose une autre compréhension du mot 'intelligence', à savoir ce qui permet à l'humain de mieux comprendre les choses, ainsi les vastes bases de données qui '*collect intelligence'* (Collins Dictionary) et dans lesquelles une information peut être trouvée comme celles tenues par les '*intelligence agency'*. Ainsi, le Secret Intelligence Service (SIS), également connu sous la dénomination de MI6 (pour Military Intelligence, section 6), est le service de renseignements extérieurs du Royaume-Uni. Son rôle est de produire des renseignements sur les sujets concernant les intérêts vitaux du Royaume-Uni en matière de sécurité, défense, politique étrangère et politique économique.

[10] V. VINGE, *Un feu sur l'abîme (en anglais A Fire under the Deep),* Robert Laffont, coll. « Ailleurs et Demain », 1994. L'auteur ressuscite le mythe prométhéen mais lui donne une autre portée, ce n'est plus l'homme qui défie Dieu mais la machine qui défie son créateur, l'homme.

[11] Nous savons que la quantité de données générées par le réseau internet est **gigantesque**. Le moteur de recherche Google traite plus de 40 000 recherches par seconde, chaque minute environ 4 150 000 vidéos sont regardées sur YouTube, les utilisateurs d'Instagram postent 46 740 photos par minute.

[12] Proposition de règlement annexée à la Résolution déjà citée, article 4 a)

[13] Que le même article point c) définit comme suit: « *autonomie», un système d'intelligence artificielle (IA) qui fonctionne en interprétant certaines données entrées et en utilisant un ensemble d'instructions prédéterminées, sans se limiter à de telles instructions, bien que le comportement du système vise à atteindre l'objectif qui lui a été assigné et qu'il soit soumis aux contraintes de cet objectif et d'autres choix de conception pertinents posés par son développeur*; ».

[14] On cite volontiers l'exemple d'ImageNet*,* cet outil, qui sert à de nombreux outils IA en matière de reconnaissance faciale, contient 1.200.000 images appartenant à 1000 classes différentes

dans la collecte que dans le traitement des données, s'ajoute l'existence de terminaux, à la fois, de plus en plus petits et de plus en plus nombreux[15]. Leur présence, dans notre environnement (*ubiquitous ou ambient computing*) et, pour certains, dans nos corps (*body implants ou 'under the skin'*) génèrent, collectent et traitent des données, soit en local (*edge computing*[16]), tantôt de manière centralisée. Ces capteurs placés dans nos voitures connectées, nos poches, nos salons, nos murs, nos rues, notre cerveau, ces connexions multiples aux plateformes de communication (réseaux sociaux) ou d'informations (moteur de recherche) permettent de connaître nos goûts, de prédire nos préférences, par exemples de destinations touristiques, notre état de santé, nos capacités financières, nos sentiments et, par exemple, d'évaluer les risques que nous, individuellement, présentons pour une compagnie d'assurances.

4. Venons-en aux caractéristiques de l'IA. La première est certes le caractère probabiliste de son mode de fonctionnement. Nos systèmes experts traditionnels[17] reposent sur la mise en forme algorithmique de la logique suivie par les experts pour traiter d'une situation et prendre une décision : ainsi, le *credit scoring* dans un système expert traditionnel repose sur la représentation du poids accordé *a priori* par les experts à une série d'éléments qui permettent de juger de la solvabilité de la personne. L'algorithme fonctionne donc de manière causale puisqu'il ne fait qu'appliquer la « loi » décrétée par les experts à un cas concret après vérification de la présence ou non des critères retenus et de leur pondération. A l'inverse, l'IA fonctionne différemment. Reprenons l'exemple de l'évaluation des crédits, une *big data* notera toutes les décisions prises avec l'ensemble des circonstances de l'octroi ou non du refus de crédit. C'est la machine elle-même qui, passant en revue ces décisions, repérera pour un crédit demandé les similitudes ou non avec les cas déjà décidés. Bref, c'est sur une base statistique tirée d'une comparaison avec des résultats déjà engrangées que la décision ou le projet de décision sera pris par la machine. On note la différence entre un système fondé sur la probabilité et la référence à un vaste nombre de décisions et celui traditionnel **causaliste** et **déterministe** car fondé sur la logique fût-elle erronée d'un raisonnement humain. On ajoute que, dans la mesure où le premier fonctionne sur des règles exprimées à l'avance et donc parfaitement transparents, le second est probabiliste et bien souvent opaque dans la mesure où celui qui utilise le système IA ignore bien souvent les corrélations effectuées entre toutes les données conservées par la mégadonnées et par des réseaux de neurones qui se superposent et s'entrecroisent.

[15] Les tags des objets connectés munis de RFID (Radio Frequency IDentifiers) comme les cartons d'emballage des produits dans les grands magasins, les capteurs dans les murs, les badges portés par les travailleurs, etc. Selon un rapport produit pour la Commission européenne, chaque citoyen, en 2025, rencontrerait chaque jour 4800 objets connectés.

[16] Ainsi, une nouvelle génération de casques policiers serait capable de reconnaître sur les lieux mêmes d'intervention de repérer les personnes fichées par la police, grâce à des systèmes de reconnaissance faciale implantés dans les casques et reliés à la banque de données de la police.

[17] Certains auteurs parlent à leur propos d'intelligence symbolique. « L'intelligence artificielle symbolique correspond aux techniques d'IA basées sur des représentations « symboliques » (lisibles par l'homme) de haut niveau. L'IA symbolique a été l'approche dominante de la recherche sur l'intelligence artificielle du milieu des années 1950 (c'est à dire de la création de l'intelligence artificielle) jusqu'à la fin des années 1980. En intelligence artificielle symbolique, on modélise les raisonnements humains et on les codes sous forme d'algorithme. » (R. RICHARD, Votre premier cours d'intelligence artificielle, Définitions, disponible à l'adresse : Intelligence artificielle symbolique : Définition, Exemples, Fonctionnement (24pm.com) (site consulté le 12 janvier 2021)

5. Sans doute, l'opacité sera moins grande dans certains types d'IA que d'autres[18]. A cet égard, on distinguera les systèmes suivant qu'ils sont ou non supervisés[19], c'est-à-dire que la qualité des réponses est systématiquement vérifiée du moins pendant une période de test . Dans les systèmes d'IA, nous dit Lambert, « *l'apprentissage se réalise en soumettant les entrées du réseau de neurones à une base de données d'exemples sur laquelle il va s'entraîner (à reconnaître par exemple telle forme géométrique ou adopter une bonne stratégie dans un jeu). Selon les réponses que le réseau va fournir à des exemples qui sont présentés à ses entrées, on va essayer de minimiser l'erreur de ses réponses en modifiant certains paramètres du réseau (les poids des connexions entre les neurones). Il s'agit ici d'un apprentissage supervisé, puisqu'un sujet humain dirige l'apprentissage en indiquant quelles sont les 'bonnes' réponses. Mais, aujourd'hui, on pense de plus en plus à des réseaux capables d'apprentissages par renforcement ou encore d'apprentissages non supervisés, c'est-à-dire aptes à extraire, par eux-mêmes, des formes, patterns, ou régularités non prescrites d'avance. Il s'agit ici de l'émergence progressive, au cours de fonctionnement du réseau, de catégories ou classes nouvelles, auxquelles l'humain n'aurait pas pensé, ou pu découvrir, en raison des limites de ses capacités d'analyse ou de perception. »*[20]. On connaît également la catégorie des systèmes dits *'deep learning'* précisément caractérisés par cette opacité : « *Il s'agit d'un traitement effectué par un grand nombre de neurones artificiels (imitant de façon très simplifiée les neurones biologiques) qui, par leurs interactions, permettent au système d'apprendre progressivement à partir d'images, de textes ou d'autres données. L'apprentissage repose sur des principes mathématiques généraux. Le résultat de l'apprentissage est une représentation (par exemple, 'Cette image contient des éléments différents'), une décision (par exemple, 'Cette image représente Jeanne Dupont') ou une transformation (par exemple, la traduction d'un texte dans une autre langue).* »[21] Ainsi, on établit un lien entre l'opacité et l'autonomie des systèmes. Est dit autonome, *« un système d'intelligence artificielle (IA) qui fonctionne en interprétant certaines données entrées et en utilisant un ensemble d'instructions prédéterminées, sans se limiter à de telles instructions, bien que le comportement du système vise à atteindre l'objectif qui lui a été assigné et qu'il soit soumis aux contraintes de cet objectif et d'autres choix de conception pertinents posés par son développeur. ».* Moins un système est supervisé, plus il mêle des couches de réseaux de neurones, plus il sera opaque et autonome, c'est-à-dire capable de décisions *a priori* hors du contrôle de celui qui l'utilise voire de celui qui l'a conçu.

6. A ces caractéristiques essentielles, on en ajoutera deux : la fragilité et la « prédictivité ». La première concerne différents aspects : la fragilité tient tout d'abord à la possibilité par une attaque externe ou interne de fausser les résultats du fonctionnement du système. Ainsi, l'ajout de quelques pixels aux images peut brouiller partiellement ou totalement un système de

---

[18] Sur le 'machine learning' et les différents types des systèmes de *machine learning* (notamment, systèmes supervisés, systèmes de '*deep learning'),* lire B. FRENAY – Y. POULLET, *Profilage et la Convention 108+ : Rapport sur l'évolution de la situation après l'adoption de la Recommandation (2010)13 sur le profilage*, Rapport au Comité consultatif de la Convention n°108, Conseil de l'Europe, , Strasbourg 7 novembre 2019, T-PD(2019)07rev., p. 8 : « *Parmi les techniques d'intelligence artificielle, le machine learning a la spécificité de pouvoir exploiter les données disponibles pour permettre la création d'intelligences artificielles apprenantes. Plus particulièrement, lorsque les données sont des images, du son ou du texte, il est courant d'utiliser le deep learning, un sous-domaine du machine learning, qui permet de créer des réseaux de neurones adaptés à la modélisation de ces types de données.* »

[19] « *There are also* ***intermediary situations*** *in profiling. For example,* ***semi-supervised*** *algorithms can be used to carry out supervised profiling, even if the expected response is known only for a limited number of individuals. This means that a large population of individuals can be used to build a classification model, even though the correct category is known only for a low percentage of them". B.* FRENAY- Y. POULLET, *Report on Profiling and data Protection,* C.of Europe, 2019)

[20] D. LAMBERT, *La robotique et l'intelligence artificielle,* Paris, Ed. Lessius, 2019.

[21] V. BARRA, « Apprentissage automatique et réseaux de neurones », in *Tangente*, hors-série n° 68, octobre 2018, p. 44-47

reconnaissance faciale. Comme le note le Groupe d'experts européens[22], « *les systèmes d'IA, à l'instar de tous les systèmes logiciels, devraient être protégés face aux vulnérabilités qui pourraient permettre à des adversaires de les exploiter (par exemple, piratage). Des attaques pourraient cibler les données (empoisonnement des données), le modèle (fuite de modèle) ou l'infrastructure sous-jacente, tant matérielle que logicielle. Lorsqu'un système d'IA fait l'objet d'une attaque, par exemple d'une attaque antagoniste, le comportement des données ainsi que du système peut être modifié, ce qui conduit le système à prendre des décisions différentes voire à s'arrêter. Les systèmes et les données peuvent également être corrompus en raison d'interventions malveillantes ou de l'exposition à des situations imprévues.* ». La robustesse du système implique la sécurité du système. Par ailleurs, la fragilité du système provient comme dans pour tout logiciel d'erreurs éventuels de programmation. Dans le cas des systèmes IA, s'ajoute le risque de biais[23]. Selon l'explication de Castets- Renard[24], « *dans le cadre des algorithmes informatiques, un biais renvoie à un défaut dans le raisonnement ou le résultat d'un algorithme. En ce sens, on peut y voir un parallèle avec le biais cognitif bien connu de l'Homme. Il s'agit d'un dysfonctionnement du schéma de pensée, du raisonnement, qui influence nos choix ou notre vision de la réalité* ». Toutefois, les conditions dans lesquelles les images sont collectées et associées à des catégories peuvent avoir un impact non-négligeable sur les modèles qui seront développés et utilisés par les autres acteurs. La répartition de l'origine géographique des images atteste d'un clair biais de représentativité qui explique que certaines images soient mal reconnues par les réseaux de neurones entraînés sur ImageNet. Le même problème a été observé avec plusieurs systèmes commerciaux de reconnaissance faciale entraînés sur des collections comportant principalement des individus mâles de type caucasien[25]. Ces biais sont conscients ou inconscients. Ainsi, on a pu reprocher à nombre de systèmes d'IA utilisés en matière d'emploi une discrimination de genre introduite sans le vouloir par le personnel informaticien mâle des firmes concepteurs de ces algorithmes[26] Des exemples sont souvent cités : la reconnaissance faciale.

7. En ce qui concerne la dernière caractéristique, la « prédictibilité » signifie la possibilité grâce à l'IA de prévoir le comportement d'autrui. B. SCHRODER écrit[27] : « *Nous sommes maintenant capables, de résoudre de toutes nouvelles classes de problèmes, telles que la reconnaissance d'image ou la transcription de la voix. Nous pouvons prédire des événements non modélisables. Par exemple, Cornell University détecte la survenance de l'état dépressif de patients bipolaires en analysant les changements dans la frappe de messages sur l'écran d'un smartphone. Un algorithme de Microsoft prévoit un diagnostic futur de cancer du pancréas ou de poumon par l'analyse de l'historique des mots clés entrés dans un moteur de recherche.* ». L'IA permet non seulement de mieux appréhender ce que nous avons fait, de nous reconnaître mais également de se tourner vers l'avenir et de prédire nos comportements futurs. Elle nous rend, jusqu'à un certain point du moins, transparents et prédictibles au

[22] HLGE (High Level Group of experts) on AI, *LIGNES DIRECTRICES EN MATIERE D'ETHIQUE pour UNE IA DIGNE DE CONFIANCE*, 8 avril 2019, n° 67, texte disponible sur le site : Ethics guidelines for trustworthy AI - Publications Office of the EU (europa.eu) (consulté pour la dernière fois le 13 janvier 2021)

[23] Par biais, on entend : toute perception personnelle ou sociale préjudiciable d'une personne ou d'un groupe de personnes sur la base de ses caractéristiques personnelles.

[24] C. CASTETS-RENARD, « Régulation des algorithmes et gouvernance du machine learning : vers une transparence et « explicabilité » des décisions algorithmiques », Dossier spécial Intelligence Artificielle, Revue Droit & Affaires, Paris II, nov. 2018.

[25] Sur cet exemple et d'autres, lire B. FRENAY et Y. POULLET, rapport cité, p. 11 et s.

[26] J. CHARPENET et C. LEQUESNE-ROTH, « Discrimination et biais genrés. Les lacunes juridiques de l'audit algorithmique », D. 2019, p. 1852

[27] B. SCHRODER, *Vie Privée, transparence et démocratie,* Actes du Colloque du REHNAM, Namur le 28 novembre 2019, Y. Poullet (ed.), Cahier du CRIDS, n° 50, 2020

moment même où la complexité du fonctionnement des algorithmes qui constituent le système, la richesse des données sur lesquelles ils travaillent (quelles corrélations ? entre quelles données ? avec quels poids ?) rendent le système non transparent.

8. Tout projet de développement et d'exploitation d'un système d'IA constitue une opération complexe et peut mettre en cause de nombreux acteurs. On le conçoit aisément même si on peut imaginer que toutes les fonctions énumérées ci-dessous soient réalisées en interne. La mise sur pied d'une *big data* requerra bien souvent la collecte de données auprès de fournisseurs agrégeant leurs données ou permettant l'accès à leurs propres bases de données, ainsi, tous les constructeurs de voitures peuvent décider de mettre en commun leurs bases de données anonymes ou non reprenant des données d'utilisation. Ces opérations de *data sharing* sont particulièrement prisées par l'Union européenne[28] qui, faute des GAFAM, ne disposent pas mégadonnées. A ces fournisseurs de données, on ajoute les fournisseurs d'algorithmes, que ceux-ci soient disponibles en *open source* ou non, gratuitement ou non. On ajoute que bien souvent, un consultant jouera le rôle d'interface entre le fournisseur et celui qui souhaite exploiter le système afin de 'customiser' les algorithmes aux spécificités du domaine d'application pour lequel il servira. Le fonctionnement du système implique une période de *testing* sur les données de l'entreprise ou de l'utilisateur[29] et, à des intervalles réguliers, la maintenance de manière à corriger les biais et erreurs. Il ne peut donc être question d'imputer à la seule responsabilité de celui qui exploite le système IA et que la récente résolution du Parlement européen en matière de responsabilité des systèmes d'IA[30] qualifie d' « opérateur frontal »[31], les conséquences des erreurs, des biais, de l'absence de sécurité voire du non-respect des dispositions des législations de protection des données.

9. Quelques réflexions complémentaires sont sans doute nécessaires en ce qui concerne le rôle des plateformes, en particulier des GAFA, qu'elles gèrent les réseaux sociaux ou constituent un outil indispensable pour assurer l'accès à l'information comme les moteurs de recherche ou à la communication comme les opérateurs de messagerie électronique. De par l'étendue de leurs activités exercées à travers de nombreuses filiales ou sociétés apparentées, de par le fait qu'elles constituent un passage obligé (*gatekeepers)*, elles disposent de données en nombre indéfini qu'elles peuvent croiser, notamment via les cookies ou autres identifiants. On ne sera dès lors pas surpris qu'elles investissent massivement dans l'intelligence artificielle, par exemple à des fins de marketing ou de recherches (ainsi, les recherches menées par une filiale

[28] Proposition de RÉGLEMENTATION DU PARLEMENT EUROPÉEN ET DUCONSEIL sur la gouvernance européenne des données (Loi sur la gouvernance des données), Bruxelles 25 nov. 2020, COM(2020)767 finale. Sur les questions de protection des données posées par le partage de données, lire le remarquable document publié par l'ICO (le DPA anglais) « *Data sharing Code of Practice* », disponible sur le site : co.org.uk/for-organisations/data-sharing-a-code-of-practice/data-sharing-agreements/.

[29] Ainsi, le chirurgien qui utilise un robot pour ses opérations chirurgicales, paramétra en fonction de sa pratique le robot.

[30] Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle (2020/2014(INL)). Cette résolution s'appuie sur le rapport du HLGE européen on Liability and AI, H.L.G.E. on artificial intelligence and liability, " LIABILITY FOR ARTIFICIAL INTELLIGENCE AND OTHER EMERGING DIGITAL TECHNOLOGIES", novembre, 2019

[31] Ainsi, la proposition de règlement (article 3 e) et f)) contenue dans la Résolution du parlement européen distingue deux types d'opérateur en matière d'IA, d'une part l' «opérateur frontal», c'est-à-dire toute personne physique ou morale qui exerce un certain contrôle sur un risque associé à l'exploitations et au fonctionnement du système d'IA et tire profit de son exploitation; et d'autre part, l' «opérateur d'amont», c'est-à-dire, « toute personne physique ou morale qui, de manière continue, définit les caractéristiques de la technologie et fournit des données ainsi qu'un service de soutien en amont essentiel et exerce donc également un certain contrôle sur le risque lié à l'exploitation et au fonctionnement du système d'IA; »

d'Alphabet : Calypo qui travaille sur le vieillissement). Outre leurs propres développements en IA à des fins internes, elles représentent sur le marché un fournisseur de données : ainsi une société de prêt à porter souhaite obtenir les adresses électroniques des personnes susceptibles d'être intéressées par leurs produits et à partir des critères que cette société désignera, elle ira, via des API et dans le cadre d'un accord bien balisé, fouiller dans les bases de données de la plateforme et acquerra d'elle les données *ad hoc*. Autre scénario, la plateforme, à partir du projet de la société, élaborera, comme un fournisseur de services, le système IA capables d'extraire de ses bases de données les données adéquates. Ces divers scénarios illustrent l'importance du rôle joué par les plateformes dans le cadre du développement des applications de l'IA.[32]

Chapitre II : Des vertus de l'IA aux risques liés à l'IA.

I. Les vertus de l'IA.

10. Comment expliquer le succès de cet outil tant auprès des acteurs privés que publics? Trois plus-values liées à son utilisation le justifient. Le numérique est, pour l'Etat comme pour les entreprises, une réponse, appuyée bien souvent par le droit[33], aux risques que cette technologie dans le même temps contribue à créer mais son apport ne se limite pas à cela. Il est 'gage' pour ces derniers, en particulier, de trois apports essentiels à la promotion de leurs intérêts : la première, la **sécurisation** des opérations **;** la deuxième, l'**optimisation** des processus et des décisions et la troisième, avec toutes les nuances que nous y mettons, l'**objectivation** des décisions que ces systèmes d'IA peuvent ou doivent prendre. Si pour l'entreprise comme pour l'Etat et ses administrations, l'humain est le premier facteur de risques à contrôler et à apprécier, le numérique permet, nous l'avons dit, de mieux le connaître et de le prédire. Bref, le numérique offre aux décideurs privés et publics un outil extraordinaire de prévision et de stratégie contre les risques associés au comportement humain. Au sein des technologies du numérique, il importe de souligner qu'en particulier, l'intelligence artificielle est perçue comme l'instrument le plus adéquat de réalisation par le numérique de trois objectifs recherchés par les Etats et les entreprises et en tout cas par le Droit

11. Le premier apport concerne la sécurisation. Ce terme vise tant la sécurité publique, que celle des investissements. Qu'il s'agisse pour un organisme de détecter et mettre à néant une cyberattaque contre des infrastructures ou bases de données privées ou publiques, qu'il s'agisse d'une cyberguerre menée par des drones ou autres *robotkillers* militaires, qu'il s'agisse de la détection de ces attaques, invasions ou armes de destruction (leurs 'mises à mort'), autant de justifications qui expliquent que l'entreprise, l'administration voire l'individu recourent aux algorithmes puissants des systèmes de *machine learning*. Par ailleurs, l'IA accroît la sécurité des infrastructures et systèmes informatiques. En ce sens, la robustesse de nos systèmes informatiques utilisés dans notre société fait de même l'objet d'une attention particulière. La plupart des contrôles d'accès qu'il s'agisse de biens ou services privés ou publics requièrent, dorénavant, la reconnaissance du demandeur d'accès et la vérification de son 'droit d'accéder' par des données biométriques ou des systèmes cryptographiques compliqués. De même, le contrôle par des logiciels de *testing* de la qualité des productions avant et après leur mise sur le

[32] Sur ces différents scenarios, lire EDPB, Guidelines 8/2020 on the targeting of social media users, Adopted on 2 September 2020, disponible sur le site de l'EDPB: edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/guidelines-082020-targeting-social-media-users_en. (dernière consultation 13 janvier 2021)

[33]

marché constitue une exigence également réclamée par chacun de nous surtout lorsque l'utilisation du bien met en péril des vies humaines, par exemple : la voiture intelligente, l'avion ou, de manière plus prosaïque, le bon fonctionnement d'un ascenseur.

12. Le deuxième apport est tout aussi important. Il pointe l'efficience et l'effectivité des actions de l'entreprise ou de l'administration. L'optimisation, notion jusqu'ici réservée à quelques académiques férus de mathématiques, est devenue la préoccupation majeure de nos gouvernants d'entreprises, d'administrations ou d'Etats[34]. Les capacités de l'intelligence artificielle nous conduisent à croire en sa possibilité, en matière d'emploi, de sélectionner le bon candidat ou de tracer pour le travailleur le bon chemin à suivre toute la journée et d'en contrôler le respect. En matière d'enseignement, l'IA permettrait, dès la maternelle, en vertu des « bons » critères, de déceler les futurs prix Nobel. En matière de *marketing*, les systèmes de profilage guidés par l'IA constituent une opportunité pour une entreprise d'approcher de façon idéale son client voire de la manipuler ou de décider du lieu idéal de la future implantation d'une succursale. On ajoutera que l'IA donne au médecin la capacité tantôt de programmer de manière optimale les soins de santé à prodiguer à ses patients tantôt d'intervenir chirurgicalement de la meilleure manière, tantôt enfin, d'analyser en quelques secondes les résultats de milliers de dossiers patients atteints d'une maladie et de noter les corrélations qui orienteront sa recherche. Demain, grâce à l'IA, le juge-robot remplacera les juges. Mieux que ces derniers, il connaîtra le droit et pourra donner efficacement et sans délai, la décision par définition incontestable sortie de l'ordinateur et de ses *big data*. En d'autres termes, la puissance de calcul de la machine, sa capacité à corréler de manière 'surhumaine' les infinies données recueillies, son incontestabilité vu l'opacité de son fonctionnement confèrent à la machine la vertu d'un *Deus ex Machina »*. Comment mieux parvenir à ces résultats, qu'en s'appuyant sur des algorithmes croisant le maximum de données historiques? On ajoute que cette optimalisation qu'autorisent les systèmes d'intelligence artificielle opéèe en continu dans la mesure où nombre de ces systèmes s'adaptent de manière permanente en fonction des données collectées.

    Cette optimisation n'est-elle pas, elle-même, souhaitée par la société et le droit ? L'intelligence artificielle est en effet un outil puissant d'application de la loi et donc d'effectivité de cette dernière. Ainsi, sans être complet, l'utilisation de l'IA est autorisée par le droit pour aider les plateformes à lutter contre les copies illicites d'œuvres ou la lutte contre la diffusion de messages illicites; l'administration identifiera plus facilement les bénéficiaires de telle avantage légal ou les personnes suspectes de fraudes fiscales ou sociales voire d'autres infractions. L'utilisation de l'IA n'est-elle, par ailleurs et dans certains cas, la traduction d'une exigence de la loi ? Ainsi, n'enjoint-t-on pas à la fois de connaître ses clients (*'Know your Customer')* tant pour éviter l'endettement excessif d'un candidat emprunteur que pour détecter des blanchiments d'argent ? [35]

13. Enfin, troisième apport, à la contestation que pourrait entraîner l'argument utilitaire de l'optimisation, les tenants de l'IA opposent le caractère 'objectif' des résultats obtenus suite aux opérations permises par l'intelligence artificielle. Ainsi, dans une entreprise, on vantera les mérites du choix 'neutre', opéré par la machine 'intelligente', en matière de recrutement

[34] . Comme l'écrit LAMBERT, « *dans bon nombre de sphères de la société contemporaine, nous sommes entrés dans une logique de l'optimisation, de la rationalisation ou de la standardisation des activités professionnelles, basée sur une évaluation constante des pratiques. Un travailleur, par exemple, doit être surveillé en permanence (d'où la nécessité de récolter en temps réel une foule de données à son égard : niveau de productivité, ranking en tous genres, etc.) et son activité doit être optimisée et alignée sur les standards « d'excellence » de l'entreprise* ». (D. LAMBERT, *La robotique et l'intelligence artificielle,* Paris, Ed. Lessius, 2019.

[35] Sur cette délégation du droit à l'IA et ses dangers, lire nos réflexions in

d'un nouvel employé ou, dans le système scolaire, pour évaluer la capacité de réussite scolaire d'un enfant. Ainsi, en matière de recrutement, la machine prendra sa décision ou émettra son conseil sur base de l'examen des mouvements faciaux des candidats lors de leur interview, de leur *curricula vitae*, d'une analyse graphologique automatique, etc., soit autant de faits et d'éléments dits objectifs et avérés (*'data do not lie')*.

Son 'raisonnement' sera, apparemment du moins, mené en toute indépendance loin de la subjectivité humaine qui, souvent, préside aux opérations de recrutement traditionnelles. Dans le même sens, qu'opposer à l'administration fiscale qui, dans le cadre de sa lutte contre la fraude, utilise un système nourri de données en provenance de nombreuses administrations voire d'entreprises ou associations privées et repère ainsi en toute neutralité les personnes suspectes? L'argument est d'autant plus pertinent que l'objectivation permise par les systèmes d'intelligence artificielle peut procurer des avantages à celui qui s'y soumet et rend dès lors l'utilisation de ces systèmes d'autant plus acceptable[36]. Sans doute, peut-on craindre que cette prétendue objectivité des décisions prises par la machine ne conduise à une déshumanisation[37] et en tout cas à une automatisation galopante des processus de décision et surtout à une acceptation quasi-automatique de la validité et de la pertinence de ces décisions par ceux qui y ont recours : en d'autres termes, qu'elle n'amène ces derniers à désinvestir leur fonction de décideurs ou, du moins, à se 'déresponsabiliser', prenant comme alibi les vertus de la machine.

14. **Du côté des individus**, de nombreuses explications sont avancées pour comprendre ce recours de plus en plus massif aux technologies les plus avancées du numérique. Sans prétendre à l'exhaustivité, on épingle chez l'utilisateur des services et des produits du numérique. La perte des repères traditionnels qui sécurisaient notre existence et nous faisaient exister. L'*« extimacy »,* en d'autres termes, la nécessité d'apparaître, technologie aidante, et de se raconter à longueur de blogs et de réseaux sociaux s'explique certes par cette volonté de retrouver une « ex-istence » en même temps qu'elle est une réponse à un besoin compulsif de communication de et avec l'autre dans un marché où chacun gagne et que l'on peut résumer ainsi : 'Je te parle et me raconte si tu me parles et te racontes.', besoin auquel l'individu résiste d'autant moins que ce service lui est offert gratuitement.

La lutte contre les sentiments d'insécurité et d'angoisse constitue une deuxième explication. Etre connecté, c'est pouvoir à tout moment se situer et espérer maîtriser ce qui fait notre angoisse : le succès des tests prédictifs *'on line'*, l'utilisation des produits *self quantified* comme le bracelet permettant à chaque instant de connaître son pouls, sa tension, son état de stress, etc. ou du 'GPS', ange gardien de notre conduite, s'explique de la sorte. A cette insécurité, s'ajoute l'argument du confort engendré par toute une série de technologies qui nous facilitent la vie (voir les outils de domotique qui permettent la gestion à distance de l'immeuble), nos choix (la présélection opérée par nos moteurs de recherche ou nos

[36] Prenons un exemple en matière d'assurance-accidents de la route. Comment, en tant que client 'de bonne réputation', ne pas apprécier le fait que les assureurs automobiles, grâce à des systèmes de '*machine learning*', m'approchent en me proposant des polices calculées exactement sur les risques que je représente. Le calcul s'opère à la fois sur nombre de données relatives à l'environnement spatial et humain des conducteurs mais également sur les données de conduite (itinéraire, respect du code de conduite, état de nervosité du conducteur, etc.) ? Il permet de revoir la prime à la baisse et à ceux dont la prime serait revue à la hausse, il sera rétorqué qu'il est normal que chacun paie suivant le risque qu'il représente.

[37] Sur cette 'déshumanisation', lire A.ROUVROY, « *Des donnés et des hommes* » Droits et libertés fondamentaux dans un monde de données massives, 2016, https://rm.coe.int/16806b1659; G.NOTO LA DIEGA, "Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information", *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2018, n°.9-1, pp. 11–16

plateformes musicales) et qui répondent à nos injonctions (les robots, les enceintes connectées). Enfin, on souligne le sentiment de maîtrise d'un environnement désormais fixé aux limites de la terre, qu'il s'agisse de la culture, de l'information, de la consommation, etc. Nous sommes ubiquitaires, sentiment qui contraste avec celui d'une perte de la maîtrise du fonctionnement de l'outil technologique qui pourtant appuie ce sentiment de maîtrise de l'environnement.

.

II. Les risques liés à l'IA

15. Nombre de risques entourent tant le fonctionnement des systèmes que leurs applications. Le numérique se caractérise, aujourd'hui et chaque jour un peu plus, par diverses caractéristiques : la dimension ubiquitaire de son infrastructure, à l'heure de l"internet des objets' ; la capacité quasi illimitée de ses capacités de transmission, de traitement et de diffusion ; le caractère de plus en plus intrusif de ses applications et, enfin, la puissance de ceux qui, grâce à ces systèmes, détiennent un pouvoir informationnel sans précédent[38]. Ces caractéristiques permettent de comprendre la nécessité d'une protection non plus du seul individu et de ses libertés[39] (risques individuels) mais, au-delà, collectivement de catégories d'individus voire de l'ensemble des citoyens, en d'autres termes de la société et de sa démocratie (risques collectifs).

A. Les risques individuels[40]

16. Certes, la digitalisation de nos sociétés permet d'influer sur les capacités de développement de chaque citoyen, de mettre en cause notre dignité humaine et nos libertés et, enfin, de prévoir, d'influencer, manipuler voire de déterminer nos comportements. C'est à ces risques encourus par tout homme, au-delà des seuls internautes, que les législations de protection des données à caractère personnel entendent répondre. Cette réponse s'origine dans l'élargissement du concept de vie privée, consacré en particulier, par l'article 8 de la Convention des droits de l'Homme du Conseil de l'Europe. Au départ, les rédacteurs de la Convention le conçoivent de manière négative comme la protection d'un espace clos dont nous pourrions écarter autrui afin de pouvoir nous retirer en nous-mêmes. L'espace clos peut s'entendre au sens physique du terme : les quatre murs de la maison ; il peut s'entendre au

---

[38] « (Le Parlement) *insiste sur l'asymétrie entre ceux qui utilisent les technologies d'IA et ceux qui interagissent avec elles et qui y sont soumis ;souligne, dans ce contexte, que la confiance des citoyens à l'égard de l'IA ne peut être obtenue que grâce à un cadre réglementaire, éthique par défaut et éthique dès la conception» qui garantisse que tout système d'IA mis en service respecte et applique pleinement les traités, la Charte et le droit dérivé de l'Union; estime que cette approche devrait être suivie dans le respect du principe de précaution qui guide la législation de l'Union et constituer la pierre angulaire de tout cadre réglementaire relatif à l'IA; demande, à cet égard, l'élaboration d'un modèle de gouvernance clair et cohérent qui permette aux entreprises et aux innovateurs de développer davantage l'intelligence artificielle, la robotique et les technologies connexes. »* (Considérant n°3 de la résolution du Parlement européen déjà citée sur l'éthique de l'intelligence artificielle)

[39] Notre sentiment est que la notion de risques aux libertés individuelles ne doit pas être limité au seul droit à la protection des données au sens strict mais que ce droit à la protection des données doit se comprendre comme le moyen de protéger l'ensemble dfes libertés comme l'affirment à la fois le Conseil de l'Europe et l'Union européenne. "*As indicated in the Article 29 Data Protection Working Party Statement on the role of a risk-based approach in data protection legal frameworks, the reference to "the rights and freedoms" of data subjects primarily concerns the rights to data protection and privacy but may also involve other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion*." (Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "*likely to result in a high risk*" for the purposes of Regulation 2016/679 Adopted on 4 April 2017)".

[40] Sur cette analyse des risques, nous renvoyons à B. FRENAY et Y. POULLET, rapport sur le profilage pour le Conseil de l'Europe, déjà cité.

sens communicationnel : notre correspondance. Aujourd'hui, sans dénier cette conception négative de la notion mais, au contraire, en la voyant comme son élargissement à une dimension plus positive, la jurisprudence range sous cette notion toutes les garanties exigées afin que nous puissions réaliser notre capacité de développement. Ainsi, dans une société de l'information de plus en plus envahissante mais dont les infinies ressources peuvent contribuer à notre développement personnel, la notion de vie privée est désormais approchée comme unissant fondamentalement deux droits pareillement nécessaires au développement de notre personnalité, d'une part, celui traditionnel à la 'séclusion', au secret[41] et, d'autre part, le droit à la maîtrise de notre 'image informationnelle' ou, selon l'expression souvent retenue, le 'droit à l'auto-détermination informationnelle', c'est-à-dire à pouvoir contrôler : Qui détient des informations à notre propos ? Lesquelles ? Pour quelle (s) utilisation (s) ? Citons rapidement les risques individuels liés d'une part au fonctionnement des systèmes d'IA et, d'autre part aux applications utilisant ces systèmes.

17. En ce qui concerne les risques individuels, on épingle le fait que les systèmes IA réduisent les personnes à des données, qu'elles utilisent souvent de manière décontextualisé : si mon GSM note ma présence telle heure, tel jour dans une officine de pharmacie, cette donnée interprétée comme témoignant de mon état de santé reçoit, dans la réalité, une signification totalement différente lorsqu'on apprend que ma visite s'explique par la demande d'un voisin malade incapable de se déplacer. Aux risques de réductionisme et de décontextualisation, on ajoute les risques déjà dénoncés d'erreurs de programmation ou de biais qui peuvent fausser le résultat du fonctionnement du système. Ensuite, à l'occasion des données collectées depuis sa mise en fonctionnement, le système peut voir son raisonnement s'écarter du modèle qui fixait son fonctionnement initial. Enfin, on rappelle l'opacité des opérations effectuées par le système, qui empêche celui qui subit la décision à ne pas comprendre le pourquoi et ne pouvoir dès lors s'y opposer et l'incline à normaliser son comportement de manière à correspondre à ce qu'il croit être attendu du système, ainsi, exemple un peu simpliste mais évocateur, il n'utilisera pas le mot 'bombe' dans ses messages circulant dans les réseaux sociaux, afin d'échapper à la surveillance exercée par les systèmes de détection des messages terroristes par les opérateurs de ces réseaux

L'attention sur les applications de l'IA souligne d'autres risques. La capacité prédictive des systèmes IA déjà soulignée (*supra,* n° 7) permet à certains opérateurs de systèmes ou à leurs utilisateurs de connaître certes dans les limites probabilistes qui caractérisent l'IA les aptitudes, les goûts et les comportements futurs de l'individu et donc de lui adresser le message publicitaire adéquat, la communication politique conforme à son profil ou de lui suggérer telle

[41] Sans doute, serait-il important de revitaliser ce droit à la séclusion au regard des évolutions de la technologie et de leur ubiquité qui interdit dorénavant de pouvoir échapper au regard et à la sollicitation d'autrui. Nous reviendrons à cet égard sur un arrêt important de la Cour fédérale allemande de 2008 qui analyse l'ordinateur personnel (et au-delà tout terminal (mon mobile, mon robot, …) comme un 'domicile' dont je dois pouvoir exclure autrui qui ne pourra y pénétrer qu'avec mon consentement. Il est par ailleurs intéressant de souligner, en droit du travail, la reconnaissance récente par l'article 16 de la loi du 26 mars 2018 relative à la confiance dans l'économie et au renforcement de la cohésion sociale (*M.b.,* 30 mars 2018) d'un droit limité à la déconnexion, de manière à éviter le harcélement au travail en dehors des heures de bureau (pour le regret des limites mises à la loi en ce qui concerne ce droit à la déconnexion, lire K. Reyniers, 'Een recht op deconnectie, of hoe omgaan met technostress?', *TSR,* 2019, no. 1, 109 et s.).

démarche, bref de le manipuler[42] via notamment des *nudges*[43]. La mémoire de l'ordinateur maintiendra la figure positive tirée de ce profil mais également celle négative avec le risque de stigmatisation de l'individu et donc de sa stigmatisation. On ajoute que les applications de l'IA s'opèrent dans des domaines de plus en plus sensibles. Ainsi en matière de santé où le profil génétique peut déterminer la probabilité de telle ou telle maladie ; en matière de sécurité où la lutte contre le terrorisme a justifié des applications IA capables de déterminer la dangerosité de certains criminels voire de certains individus (par exemple lors de demande d'immigrants) ; en matière d'éducation, afin de déterminer les capacités de réussite futures de tel ou tel élève.

18. **Le risque d'éclatement de la sphère privée et le risque de surveillance sans limite** - L'ubiquité de la technologie conduit à la transparence de l'individu dans la mesure où celui-ci de plus en plus ne peut vivre en dehors de sa connexion aux outils et services de la société digitale, transparence totale ou, en tout cas, partielle notamment s'il renonce à certains des services qui lui sont offerts par la technologie, voire s'en déconnecte. Comme nous l'avons vu, la technologie n'enregistre pas seulement les traces laissées volontairement[44] par l'individu sur un réseau social ou sur des sites de services[45], pas seulement, les mouvements et déplacements tant de son corps que les expressions faciales, ses choix de services, de produits ou d'informations (son *surfing*). Cette technologie entend, à partir des données ainsi saisies grâce à des systèmes d'IA, mais également, avec *l'affective computing*, connaître ou plutôt deviner voire prédire nos émotions et nos sentiments et, à travers l'examen de nos données génétiques, de ce qui est notre 'identité'. L'abolition dans notre société technologique de la distinction entre **sphère publique et sphère privée** inquiète : la distinction entre les deux sphères, qui était un pilier de notre droit en garantissant l'inviolabilité du domicile à l'inverse des lieux publics est désormais abolie.

    Enfin, la possibilité que le raisonnement automatisé soit entaché de ce qu'il est convenu d'appeler des biais, a été évoquée. Ces biais, qu'ils soient volontaires ou non, peuvent conduire à de possibles discriminations. Le poids trop important accordé à un critère, le fait que tel critère utilisé cache un autre critère 'discriminant une catégorie de populations (les noirs, les femmes, les étrangers, les handicapés, les pauvres, …) et ce, bien au-delà des critères liés aux catégories de données sensibles ou particulières énoncées par l'article 6 de la Convention 108+, le refus de prendre en considération un élément contextuel spécifique à la personne concernée, dès lors victime de l'automaticité de l'ordinateur, constitue un dernier risque tant vécu individuellement, que, le cas échéant, vécu par tout un groupe et donc collectif : ainsi,

---

[42] Comme l'affirmait le patron d'Amazon, « *avant même que vous passiez commande, nous avons déjà préparé votre colis.* » et le patron de Google renchérit : « *It will become very difficult for people to see or consume something that has not in some sense been tailored for them.* ».

[43] *La théorie du Nudge (ou théorie du paternalisme libéral*), nous explique Wikipedia, *est « un concept des sciences du comportement, de la théorie politique et d'économie issu des pratiques de design industriel, qui fait valoir que des suggestions indirectes peuvent, sans forcer, influencer les motivations, les incitations et la prise de décision des groupes et des individus, au moins de manière aussi efficace sinon plus efficacement que l'instruction directe, la législation ou l'exécution ».*«

[44] Volontairement certes mais souvent sans conscience des possibilités ouvertes d'utilisation des données ainsi confiées au réseau.

[45] Ainsi, récemment, une enquête de TECHCRUNCH, une société de média spécialisée en analyse de technologie digitale, révélait que « *la société israélienne* ***Glassbox*** *enregistre ce que vous faites sur votre téléphone, à chaque fois que vous êtes sur le site ou l'application de l'un de ses clients. Cette entreprise d'analyse de* ***données*** *tente de mieux comprendre les comportements des consommateurs et la manière dont ils naviguent dans certaines applications. Ainsi Hotels.com, Expedia, Abercombie & Fitch et bien d'autres encore, font appel à Glassbox pour enregistrer tout ce que font leurs clients lorsqu'ils sont sur leur* ***application*** *: saisir du texte, cliquer, zoomer, tous les gestes sont enregistrés ».*

l'analyse par un système d'IA des quartiers susceptibles de favoriser une population criminogène met en cause non seulement, à titre de personnes, les personnes physiques de ce quartier, mais également le quartier comme tel, son image et entraîner des conséquences sociales (personnes désertant le quartier ou refusant de s'y installer) voire la surveillance accrue de la police. Il s'agit donc d'un risque Ce risque de discrimination est d'autant plus important que le fonctionnement du mécanisme décisionnel apparaît comme neutre et objectif et que l'opacité de son fonctionnement empêche le décodage de la soi-disant 'logique' suivie.

B. Les risques dits collectifs

19. La réflexion sur le risque de discrimination témoignait déjà du fait que les risques liés au *machine learning* peuvent concerner au-delà de leur portée individuelle, une catégorie d'individus voire la société. L'affaire Cambridge Analytica en témoigne. Au-delà de l'analyse de messages individuels échangés sur Facebook et des réponses à un questionnaire apparemment anodin, ce qui était en cause ne se limitait pas aux choix politiques d'un individu et le cas échéant à sa manipulation mais avit une portée collective essentielle, par sa mise en danger du fonctionnement de nos démocraties.?. L'IA a une particularité, c'est qu'elle met en cause non seulement les **libertés et l'égalité de chaque individu pris en tant que tel mais, au-delà, celles des groupes** ethniques, philosophiques, à revenu modeste, les résidents d'un quartier, etc. Ces traitements affectent également d'autres valeurs que celles liées aux libertés individuelles, en particulier **la justice sociale ou la diversité culturelle** entre individus ou entre groupes[46] et, au-delà, de manière parfois importante, le fonctionnement de nos sociétés et en particulier de notre démocratie. Ainsi, par exemple, la manipulation abusive des individus met en cause tant les libertés que la dignité humaine au sens kantien du terme[47] mais si elle concerne en cela chacun de nous, elle peut également s'étendre à toute une population ou avoir des effets comme tels, y compris sur la qualité du débat public et l'opinion politique des citoyens. L'individualisation de l'offre de services ou encore l'exclusion de certaines personnes du bénéfice de ces services atteignent, au-delà des individus, des groupes de personnes et soulèvent des questions de justice sociale[48].

Prenons l'exemple déjà cité des assurances '*one to one*' qu'il s'agit d'assurances soins de santé ou de responsabilité civile. L'individualisation des primes se calcule au plus près des 'risques', que peut représenter chaque personne, risques définis instantanément par les systèmes d'IA. Cette individualisation soumet à dure épreuve le sacro-saint. Cette individualisation heurte le principe de la mutualisation des risques, pilier de notre système d'assurance et sa généralisation impacte l'ensemble de la population assurée. Autre exemple : des systèmes de

[46] Cf. La Convention de 2003 de l'UNESCO à propos de la bioéthique : « *Aucun individu ou groupe ne devrait être soumis, en violation de la dignité humaine, des droits de l'homme et des libertés fondamentales, à une discrimination ou à une stigmatisation pour quelque motif que ce soit*. ». Sur le bien-fondé de reprendre en matière de numérique, ces principes qualifiés d'universels par l'UNESCO, notre ouvrage, *Ethique et droits de l'Homme à l'heure du numérique,* Mémoire de l'Académie royale de Belgique, 2020, p. 85 et s.

[47] La dignité exige, selon la doctrine kantienne largement reçue dans nos pays européens, que la personne humaine ne soit jamais considérée comme moyen mais toujours comme fin.

[48] W. SCHULTZ et S. DREYER, *The General Data Protection Regulation and Automated Decision-making: Will it deliver? Potentials and limitations in ensuring the rights and freedoms of individuals, groups and society as a whole*, *Analysis commissioned by the Bertelsmann-Stiftung, disponible à l'adresse:* https://www.hiig.de/en/publication/the-general-data-protection-regulation-and-automated-decision-making-will-it-deliver-potentials-and-limitations-in-ensuring-the-rights-and-freedoms-of-individuals-groups-and-society-as-a-whole/

reconnaissance faciale détectent automatiquement l'"état de fatigue des conducteurs et certains peuvent imposer alors l'arrêt de la conduite. On conçoit que ces systèmes pourraient être prônés voire imposés par la loi ou, de manière plus insidieuse mais tout aussi efficace, par les compagnies d'assurance et ce, au nom de la sécurité publique de nos routes. Faut-il suivre cette voie sans qu'il y ait eu débat public en la matière ? Il est assez remarquable qu'on déborde ainsi les questions traditionnelles de protection des données ou de vie privée, au sens étroit du terme[49]. La même réflexion peut être adressée à un système d'intelligence artificielle qui aurait pour but de prédire les chances de réussite scolaire ou les risques familiaux d'enfants battus dans la population et en arriverait à identifier le poids de certaines données. Cette identification ne pose pas seulement un risque individuel mais bien collectif, dans la mesure où elle risque de stigmatiser certains types de population. Le rapport de la défunte Commission belge de la protection de la vie privée de 2016[50] distingue, de manière très convaincante, les risques individuels de ceux collectifs où, selon l'expression du rapport, des étiquettes sont appliquées consciemment ou inconsciemment à des catégories de personnes souvent sans lien entre elles et réunies par des critères non nécessairement prévisibles ni classiques comme ceux fondées sur la vulnérabilité physique, mentale, ethnique ou financière[51] : « *D'autres exemples sont la discrimination tarifaire et/ou les projets de big data qui compromettent l'accès égal à des opportunités sur le marché du travail ou la diversité sur le lieu de travail. Par exemple, un bureau de ressources humaines qui, à l'aide d'un logiciel de recrutement, appliquerait aveuglément une stratégie "hiring for culture fit" avec une discrimination inconsciente sur la base de critères protégés tels que la religion ou les convictions politiques ou philosophiques, l'origine ethnique, l'âge, le sexe ou la race, la grossesse, la maternité et le handicap. Dans cet exemple, il faudrait veiller à ce qu'un algorithme ne refuse pas la participation sociale à des groupes historiquement défavorisés ou vulnérables.* »[52]

20. Ce propos conduit à une réflexion complémentaire : les mégadonnées nous entraînent nécessairement dans une aventure solidaire. Lorsque les données de *surfing* d'un internaute sont collectées par son moteur de recherche préféré et englouties dans une vaste base de données où elles se retrouvent avec les données de millions d'autres internautes, il est clair que les algorithmes qui, à un moment donné, prendront une décision à propos de cet internaute, le feront en fonction de l'ensemble des données collectées et de la différenciation ou de l'approximation que les recherches algorithmiques induiront entre les données de cette personne et celles d'autrui. On note également l'effet dit 'domino' qui signifie que le choix d'un acteur, par exemple de se connecter à un réseau social, aura pour effet d'influer de manière majeure les choix de ses proches.

---

[49]Nous défendons en effet une conception large de la notion de vie privée qui prend en compte le développement de l'individu dans une société donnée. Sur notre conception de la vie privée au-delà de la protection des données, lire Y. POULLET, *La vie privée à l'heure du numérique - Essai,* Larcier, Cahier du Crids, n° 47, 2019.

[50] "*On a déjà évoqué ailleurs l'utilisation de certaines corrélations, ensembles de données ou algorithmes qui produisent une stratification sociale (ce qu'on appelle placer une population dans des catégories de "social sorting", impliquant un traitement potentiellement illégal et inégal de groupes sociaux. Il est certes possible que ce traitement inégal ne soit pas intentionnel mais se dissimule dans les données ou les algorithmes. Lorsque l'on aborde des problèmes sociaux à l'aide du big data, on peut voir apparaître, par des fausses corrélations, le "fondamentalisme de données" et la "distorsion de données", des effets néfastes pour une partie spécifique de la population. Des enquêtes sur certains algorithmes concernant le transport ou la circulation ont démontré le risque qu'une répartition d'avantages basée sur un algorithme de décision puisse donner lieu à une répartition inégale de moyens (publics) ou de ces avantages (réparation de routes, prestation ou non d'un service dans un temps de réaction donné, octroi d'une réduction sur des services).* » (CPVP, *Big Data Report,* 2016, p. 27 (document disponible à l'adresse : https://www.autoriteprotectiondonnees.be/sites/privacycommission/files/documents/Rapport_Big_Data_2017.pdf

[51] D. Le METAYER et J. Le CLAINCHE. « From the Protection of Data to the Protection of Individuals: Extending the Application of Non-Discrimination Principles". In S. Gutwirth, P. De Hert, & Y. Poullet (Eds.), *European Data Protection: In Good Health*. Springer Dordrecht, Heidelberg London New York, 2012, p. 315 à 329;

[52] Ibid., p. 28.

21. Nos considérations sur les risques collectifs trouvent écho dans la Résolution du Parlement européen du 20 octobre concernant les aspects éthiques de l'intelligence artificielle[53]. La proposition de règlement qui est annexée à cette Résolution parlementaire, appuyée par la Commission, élargit singulièrement la notion de « risques » au-delà des seuls risques identifiés par le droit traditionnellement, à savoir les risques « individuels » de dommages aux personnes, qu'ils soient physiques, financiers ou concernent leurs libertés, en particulier, leur vie privée (une attention particulière est accordée aux traitements de données biométriques) ou leur liberté d'expression. Ainsi, les dispositions mentionnent les risques à l'environnement causés par le caractère énergivore des technologies de l'IA et surtout ceux de discrimination de catégories d'individus avec une attention particulière à la lutte contre les biais de toute nature. Au-delà, le texte plaide pour que le développement de l'IA soit conduit « *d'une manière qui ne perturbe pas les élections ou ne contribue pas à la diffusion d'éléments de désinformation, qui respecte les droits des travailleurs, qui favorise une éducation de qualité et l'habileté numérique, qui ne creuse pas l'écart entre les genres en empêchant l'égalité des chances pour tous.* ». Sans doute, cette nécessité d'une gouvernance des systèmes d'IA contre les risques individuels et collectifs exige une réflexion sur la portée limitée des législations de protection des données et sur les interactions nécessaires entre les organes chargés de la protection des données et ceux qui pourraient être mis en place dans le cadre d'une réflexion plus large comprenant les risques collectifs[54].

Chapitre III : Et la législation de protection des données (le RGPD) face à l'IA : une réponse adéquate ?

[53]Résolution du Parlement européen du 20 octobre 2020 déjà citée, contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes (2020/2012(INL)). Il est à noter que trois textes du Conseil de l'Europe (Conseil de l'Europe, « *Lignes directrices sur les mégadonnées* », disponibles à l'adresse : https://rm.coe.int/CoERMPublicCommnSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a . et Conseil de l'Europe, « *Rapport sur l'intelligence artificielle* », rapport établi par A. MANTELERO, T-PD(2018)09Rev., Strasbourg, le 25 janvier 2019, https://rm.coe.int/intelligence-artificielle-et-protection-des-donnees-enjeux-et-solution/168091f8a5 et « Lignes directrices », T-PD(2019)01, https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40)) réclament une identification préalable et complète de ceux-ci, risques tant pour les individus que pour la société, de même qu'une prise en compte des différents types d'algorithmes utilisés : *« Les risques d'impact négatif sur les personnes et la société inhérents aux données décontextualisées et aux modèles algorithmiques décontextualisés devraient être dûment pris en compte lors du développement et de l'utilisation d'applications de l'IA. »* soulignent les lignes directrices sur l'IA. Les lignes directrices sur les mégadonnées vont dans le même sens, réclamant des mesures préventives et prévoyant diverses mesures relatives à l'évaluation des risques. Cette approche rejoint l'application du principe de précaution, bien connu en matière d'environnement.

[54] A cet égard, le règlement proposé par la Résolution déjà citée prévoit en son article 18 : « *Chaque État membre désigne un organisme public indépendant chargé de contrôler l'application du présent règlement (ci-après «organisme de surveillance») et de réaliser les évaluations des risques et de la conformité et la certification prévues aux articles 14,15 et 16 sans préjudice de la législation sectorielle » « Chaque organisme de surveillance national sert de premier point de contact en cas d'atteinte présumée aux principes éthiques et aux obligations juridiques énoncés dans le présent règlement, et notamment en cas de traitement discriminatoire ou de violation d'autres droits, à la suite du développement, du déploiement ou de l'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes. Dans ces cas, l'organisme national de surveillance concerné procède à une évaluation de la conformité en vue de soutenir le droit des citoyens à la contestation et au recours* »

22. Comme déclaré d'emblée, notre propos n'est pas de reprendre ici une analyse de toutes les dispositions du RGPD mais simplement d'en relever quelques-unes qui nous semblent poser difficulté lorsqu'il s'agit de les appliquer aux traitements qui utilisent le *machine learning*. Par ailleurs, nous nous concentrerons sur les traitements de profilage utilisant des systèmes d'IA, qui constituent les applications les plus fréquemment rencontrées et les plus problématiques, lorsqu'on envisage les risques d'atteinte à la protection des données à caractère personnel. A cette fin, nous nous servirons des travaux actuels entrepris avec et pour le bureau sur l'actualisation de la recommandation de 2010 sur le profilage et des guidelines fournis par rité même l'EDPB en la matière[55].

Le plan du chapitre suit la structure du RGPD. Ainsi, l'adéquation des définitions y compris de la liste des acteurs sera d'abord examinée ; le deuxième point concernera les principes de légitimité des traitements décrits aux articles 5 et 6 ; le troisième, final, s'attardera à quelques obligations significatives du responsable du traitement dont la lecture devra être interprétée à la lumière des particularités des systèmes d'IA.

A. Définitions et acteurs

a. Les données

23. Le RGPD limite sa protection aux seules données à caractère personnel. Certes, cette notion doit s'entendre de manière large puisque au critère d'identifiabilité qui d'une manière ou d'une autre continue à faire référence à la notion d'identité civile et aux éléments de celle-ci, se substitue progressivement[56] celui d' »individualisation »[57], c'est-à-dire cette possibilité de saisir

---

[55] Voir EDPB, Lignes directrices sur la prise de décisions individuelles automatisées et le profilage aux fins du Règlement 2016/679 (wp251rev.01), 22 Aout 2018, accessible sur le site de l'EDPB déjà cité. On citera également l'étude comparative réalisée dans le cadre du Fundamental Rights and Citizenship Programme of the European Commission: Protecting citizens' rights against illicit profiling- *Comparative report on automated profiling in 28 EU member states and Switzerland,* accessible à l'adresse: www.statewatch.org/media/documents/news/2014/nov/profiling-project-ws.pdf. (consulté pour la dernière fois le 14 janvier 2021)

[56] Vor en particulier l'avis du Groupe de l'article 29 sur la notion de données à caractère personnel : « Trois critères ont ainsi été affirmés par le Groupe de travail « article 29 » (dont la fonction de coordination des autorités de protection nationales des données a été transférée par le RGPD à l'European Data Protection Board) : l'individualisation (est-il toujours possible d'isoler un individu ?), la corrélation (est-il toujours possible de relier entre eux les enregistrements relatifs à un individu ?) et l'inférence (peut-on déduire des informations concernant un individu ?). Si le traitement répond à ces trois critères, il sera considéré comme anonyme ; il pourra également l'être s'il ne respecte pas l'un des trois critères, mais seulement après une analyse détaillée des risques de ré-identification (dans son avis du 10 avril 2014 (Avis n° 05/2014 du 10 avril 2014 sur les Techniques d'anonymisation du Groupe de travail « article 29 » sur la protection des données)

[57] Il s'agit de passer, selon le terme heureux de C. de TERWANGNE (C. de TERWANGNE , « Définitions clé et champ d'application du RGPDi », in DE TERWANGNE C. et ROSIER K. (dir.), *Le Règlement général sur la protection des données (RGPD/GDPR), Analyse approfondie, Cahiers du CRIDS*, n° 44, Bruxelles, Larcier, 2018, p. 64 et sde l'identification d'une personne, c'est-à-dire d'une possibilité de retrouver les éléments de son identité légale (nom, prénom, adresse, etc.), à l'individualisation c'est-à-dire la capacité de rapporter à un individu singulier des données, sans que les éléments de son identité légale soit connaissable. Ainsi, le tag RFID porté par une personne X se baladant dans un supermarché, ne permettra sans doute pas l'identité de son porteur mais permettra de le localiser au sein de ce supermarché, de tracer, le cas échéant, ses précédents achats voire de connecter de telles informations à celles résultant de ses habitudes de *surfing*.

la personne dans sa singularité , même si l'identité civile ne peut être connue[58]. Notre propos ne s'arrête pas là. Il s'agit de mettre en évidence le fait que la réglementation de l'IA doit également s'appliquer aux données à caractère non personnel, c'est-à-dire anonymes prises en compte par le système de *machine learning*. L'argument est double. Premièrement, il note que la puissance de certains systèmes d'IA est telle que des données, pourtant considérées comme rendues anonymes, peuvent être re- personnalisées[59]. Deuxièmement, on souligne que la plupart des systèmes d'IA, en particulier de profilage, utilisent des données anonymes, ainsi dans le cadre d'un profilage permettant de sélectionner les candidats à un logement, l'opérateur d'un système se référera à des données telles que le revenu moyen des personnes originaires de tel quartier, le niveau scolaire des habitants ou leur niveau d'endettement. Limiter les dispositions du RGPD aux seules données à caractère personnel représente dès lors un risque en matière de transparence pour la personne concernée : il est requis d'envisager également les **données anonymes** qui, dans bien des cas, peuvent également servir à la constitution du profil. Ainsi, conformément aux articles 13, 14 et ³15 du RGPD, limiter l'information et l'accès à la personne concernée aux seules données à caractère personnel intervenues dans la fabrication de son profil de candidat à l'emploi et exclure de cette information les données anonymes apparaîtrait comme une information incomplète voire biaisée. Le récent projet de recommandation du Conseil de l'Europe sur le profilage reconnaît, en ce qui concerne ce type d'opérations, ce besoin d'extension du champ d'application de la réglementation de protection des données : « *Dans le cadre de l'utilisation croissante de méga données (« big data »), des données à la fois personnelles et non personnelles sont traitées. Par ailleurs, avec des traitements automatisés, basés notamment sur l'utilisation de systèmes d'apprentissage automatique, il est difficile de savoir a priori quelles données permettront des corrélations ou des prédictions relatives à une personne concernée. Dans de tels cas, pour que les données à caractère personnel soient traitées de façon loyale, les organisations devraient garantir la pertinence et la qualité de toutes les données, y compris les données non personnelles, qui pourraient permettre les corrélations ou prédictions relatives à une personne concernée.* ».

24. Juste un ajout sur la notion de données sensibles. Le scandale 'Cambridge Analytica' témoigne du fait que des communications triviales sur Facebook, les réponses à des questions, dont la teneur était tout sauf politique, peuvent, une fois traitées par des systèmes d'intelligence artificielle, révéler ou, en tout cas, prétendre révéler les 'préférences' politiques des individus. L'exemple relevé ci-dessus (*supra*, n°7) attestent que les mots choisis pour l'interrogation d'un moteur de recherche peuvent révéler l'état de santé de l'internaute. En d'autres termes, la qualification de « sensibles » des données ne tient pas nécessairement à la nature en soi sensible des données traitées mais au résultat de leur traitement, en tenant compte de la finalité de celui-ci, ce que l'EDPB[60] appelle 'les données inférées'. Cette constatation rejoint l'argument en faveur d'une interprétation constructive de l'article 9 du RGPD qui, rattache le caractère sensible non à la nature des données mais au traitement dans lequel ces données sont considérées. Le texte de l'article 9.1.[61] énonce en effet : « *le traitement des données à caractère personnel qui révèle (et non qui révèlent, nous soulignons) l'origine raciale, les opinions politiques, ….* » Ainsi, c'est le traitement et non la donnée elle-même qui 'révèle' l'origine raciale ou ethnique (exemple: sélectionner tous les noms terminant en 'SKI' de façon à retrouver l'origine

[58] Sur cette évolution, lire les actes du colloque de Toulouse de 2019 tenu sur l'identité numérique, J. EYNARD (sous la direction de), *L'identité numérique – Quelle définition pour quelle protection ?,* Larcier, Collection Création et communication, Bruxelles , 2020

[59] A cet égard, à propos des données de communications téléphoniques rendues anonymes et la possibilité de res de tel quartier« On the privacy-conscientious use of mobile phone data », *Scientific Data, No 5*, 11 Decembre 2018, (https://www.nature.com/articles/sdata2018286.pdf.).

[60] Voir à ce propos, EDPB, Guidelines 8/2020 on the targeting of social media usersVersion 1.0, Adopted on 2 September 2020, en particulier p. 29 et 30.

[61] Le texte de la directive 95/47 se référait non au traitement mais aux données. L'article 8.1 stipulait ; « les Etats-membres interdisent les traitements des données qui révèle**nt** … »

polonaise des personnes concernées), les opinions politiques (présence dans la rue à une manifestation d'un parti politique), l'appartenance syndicale, les convictions religieuses, la santé ou la vie sexuelle. A propos des données sensibles, on ajoute l'intérêt de l'ajout à la liste des catégories particulières de données, des données biométriques et génétiques [62], qui sont bien souvent l'objet d'applications de système de *machine learning,* comme les systèmes de reconnaissance faciale ou de recherche médicale fondé sur les analyses génétiques des personnes concernées.

b. le traitement

25. Qu'il soit clair que la simple utilisation de l'IA n'est pas en soi un traitement au sens du RGPD sauf à considérer que toute opération nouvelle, en l'occurrence la fouille de données grâce à un algorithme de *machine learning,* est un traitement, ce qui impliquerait, pour le responsable à chaque nouvelle opération, les obligations liées à l'existence d'un nouveau traitement Certes, le profilage défini par le RGPD en son article 4. 4) comme « *toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique,* … » est soumis à divers prescrits comme nous le verrons plus loin (*infra,* n°37) mais cette évaluation de la personnalité ne s'entend pas nécessairement de l'utilisation d'un système d'IA et peut se réaliser par la fois d'un système expert classique. Dans la mesure où l'utilisation d'algorithmes savants et d'une *big data* sert généralement à une meilleure réalisation d'une finalité de profilage au sens le plus large du terme déjà présente dans l'entreprise ou l'administration, l'utilisation d'un système d'IA ne constituera pas un traitement nouveau. Prenons l'exemple de la mise sur pied d'un système IA permettant le ciblage de personnes qui pourraient être suspectées de fraude fiscale, l'administration poursuivait déjà certes avec des moyens moins performants cette finalité. Sans doute, la réponse ne sera pas la même si l'utilisation de l'IA permet de poursuivre une finalité nouvelle, ainsi si une plateforme comme SPOTIFY utilise l'IA aux fins de développer un nouveau service, de conseil à ses utilisateurs qui paient ou non pour l'obtenir.

Le fait que l'IA soit souvent une simple technologie au service d'un traitement déjà existant prive en fait la personne de toute information a priori de son utilisation. Sans doute, en prendra-t-elle connaissance au moment de l'exercice de son droit d'accès ou par le fait qu'une décision est prise à son égard (voir *infra*, n°35) ; sans doute, si le système IA est un système qualifiable de système à haut risque (voir infra, n°39) , le responsable sera –t-)il tenu en interne de procéder à un « *Privacy Impact Assessment »*. Ne serait-il cependant pas utile en outre que toute utilisation d'un système de *machine learning* ayant un impact sur des personnes physiques ou sur une collectivité soit automatiquement l'objet au moins d'une information préalable du public suivant le principe de loyauté (*infra*, n°27) ?

[62] Les données tant génétiques que biométriques présentent des caractéristiques communes : elles s'adressent à une réalité biologique propre à la personne concernée (même si parfois partagée avec d'autres membres de la famille, comme dans le cas des données génétiques) et en tant que telles, elles ne peuvent être modifiées et collent de manière définitive à notre peau. On ne peut s'en affranchir. Sur la nécessité d'une prise en considération des risques particuliers liés aux traitements des données génétiques, lire *Recommandation CM/Rec(2019)2 du Comité des Ministres aux États membres en matière de protection des données relatives à la santé*, adoptée par le Comité des Ministres le 27 mars 2019, lors de la 1342e réunion des Délégués des Ministres. Comparer avec la définition du RGPD, art. 4.13) : « *données génétiques», les données à caractère personnel relatives aux caractéristiques génétiques héréditaires ou acquises d'une personne physique qui donnent des informations uniques sur la physiologie ou l'état de santé de cette personne physique et qui résultent, notamment, d'une analyse d'un échantillon biologique de la personne physique en question;* «

c. les acteurs

26. L'introduction insistait sur la qualité de leurs fournitu. Cette extension peut prendre deux directions.

27. Pour certains fournisseurs d'éléments nécessaires au fonctionnement d'un système de profilage et qui les offrent commercialement sur le marché (un algorithme de base, un jeu de données nécessaires au *testing*, une base de données), devraient être requis des engagements quant à la qualité du produit, la description des limites de celui-ci et, le cas échéant, la collaboration avec le responsable dans le cadre de l'évaluation des risques et lors de la phase de tests. Même si les qualifications de responsable ou sous-traitant ne peuvent être retenues, les textes cités plaident en effet pour une obligation à charge d'acteurs, intervenant dans la fourniture d'algorithmes, de bases de données servant de jeux de tests ou de mégadonnées, de collaborer avec le responsable du traitement pour diminuer les risques, en particulier en donnant une information sur la qualité et les limites de l'élément fourni au regard de l'objectif poursuivi par ce dernier à travers l'exploitation du système IA (par exemple : pour un logiciel IA de reconnaissance faciale destiné à des pays africains, le fournisseur d'une base de données d'images devrait avertir des risques de biais liés au fait que les images sur lesquelles l'algorithme a été testé sont pour 90% des femmes blanches européennes). On peut imaginer que cette collaboration pourrait l'obliger à participer à la phase de tests pour détecter les éventuels biais présentés par son produit.

La notion de « responsable conjoint » introduit par l'article 26 du RGPD semble pouvoir être utilisée également dans le contexte des montages nécessaires à l'exploitation des systèmes IA. Dans l'affaire Fashion ID (entreprise de vêtements en ligne)[63], la Cour a de même estimé que « *l'insertion par Fashion ID du bouton « j'aime » de Facebook sur son site Internet lui permet d'optimiser la publicité pour ses produits en les rendant plus visibles sur le réseau social Facebook lorsqu'un visiteur de son site Internet clique sur ledit bouton. C'est afin de pouvoir bénéficier de cet avantage commercial consistant en une telle publicité accrue pour ses produits que Fashion ID, en insérant un tel bouton sur son site Internet, semble avoir consenti, à tout le moins implicitement, à la collecte et à la communication par transmission des données à caractère personnel des visiteurs de son site, ces opérations de traitement étant effectuées dans l'intérêt économique tant de Fashion ID que de Facebook Ireland, pour qui le fait de pouvoir disposer de ces données à ses propres fins commerciales constitue la contrepartie de l'avantage offert à Fashion ID. …Dans de telles circonstances, il peut être considéré, sous réserve des vérifications auxquelles il incombe à la juridiction de renvoi de procéder, que Fashion ID et Facebook Ireland déterminent, conjointement, les finalités des opérations de collecte et de communication par transmission des données à caractère personnel en cause au principal. ».* Les récentes lignes directrices sur le 'ciblage des utilisateurs des réseaux sociaux énoncées par l'EDPB[64] confirment la jurisprudence de la Cour et examinent en particulier les différentes méthodes utilisées par les entreprises et les plateformes de réseaux sociaux pour cibler leur clientèle, pour conclure à l'existence d'une responsabilité conjointe de la plateforme qui, soit, fournit les profils demandés par l'entreprise, soit, aide à la sélection de ceux-ci sur base des critères retenus et permet l'accès électif à sa base de données. L'EDPB en déduit la nécessité selon l'article 26.2 d'une convention entre eux à propos de la répartition ces obligations mises à charge des responsables de traitement: " *In terms of scope, the EDPB considers that the arrangement between*

[63] CJUE, 29 juillet 2019, C.40/17, Fashion ID c. Verbraucherzentrale**,** points 79 et s.
[64] EDPB, Guidelines 8/2020 on the targeting of social media usersVersion 1.0, Adopted on 2 September 2020, en particulier p. 17.

targeters and social media providers should encompass all processing operations for which they are jointly responsible (i.e. which are under their joint control). By concluding an arrangement that is only superficial and incomplete, targeters and social media providers would be breach of non-compliance with their obligations under Article 26 of the GDPR". On ajoute que si cette qualification de responsable conjoint n'était pas retenue, celle de sous-traitant pourrait alors l'être. Ainsi, la plupart des promoteurs d'expérimentations portant sur des données hospitalières confient à des sociétés tierces, le plus souvent informatiques, le soin de mettre à disposition les outils d'IA qui analyseront les données médicales récoltées souvent en grand nombre. Dans le cas de sous-traitance, il importera, suivant les prescrits du RGPD (article 28) que le choix du sous-traitant soit entouré de garanties, qu'un contrat soit conclu entre le responsable et ce ou ces sous-traitants quant à l'utilisation des données, quant à leur sécurité[65]. Par ailleurs, le RGPD impose aux sous-traitants certaines obligations, comme celles de notification en cas de brèches de sécurité, de *risk assessment*, etc.

.

B. Les principes

28. L'article 5 du RGPD énonce nombre de principes applicables à tout traitement. Certains trouvent difficulté à s'appliquer lorsqu'il s'agit de traitements utilisant les technologies de l'intelligence artificielle. Nous serons loin d'être exhaustifs dans cette analyse, nous contentant d'examiner les points majeurs de difficulté : ainsi, que devient le principe de loyauté dans le cadre de systèmes aussi opaques que ceux dits de *deep learning ?* Les difficultés liées au principe de finalité (en particulier pour les opérateurs de plateforme seront ensuite évoqués et en particulier, pour certains traitements, le consentement peut-il encore être une base de légitimité. Que deviennent les principes de proportionnalité rt de minimisation ? Enfin, le principe de sécurité n'appelle-t-il pas des mesures supplémentaires, lorsqu'on considère les risques d'erreur, de biais ou dévolution des systèmes d'IA,

a. le principe de loyauté

29. La loyauté s'entend notamment d'une information des personnes concernées sur toute une série de caractéristiques du traitement. Cette information s'avère difficile vis-à-vis de tiers dont les données sont incidemment collectées66. Par ailleurs, les types de données collectées peuvent être nombreux (voir par exemple, en cas de voiture intelligente) et venir de sources diverses qu'il importera de lister, lors de l'information due que la donnée ait été collectée auprès de la personne concernée ou non67. Surtout, le fait que le système, par les qualités caractéristiques des systèmes d'IA, puisse prédire et décider vis-à-vis d'une personne entraîne des devoirs d'information complémentaires. Ainsi, on souhaitera que la personne soit informée, de la possibilité de refuser le consentement ou de le retirer, et les

---

[65] Article 29 Working Party. *Opinion 1/2010 on the concepts of controller and processor*. 2010. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf.

[66] Ainsi, dans le cadre d'un robot aide-soignant, les images prises de visiteurs du patient à domicile ou hospitalisé.

[67] Voir l'article 4.2 de la proposition de recommandation sur le profilage : « *Lorsque les données personnelles ne sont pas obtenues auprès de la personne concernée, celle-ci devrait être informée par le responsable du traitement, au minimum au moyen d'une information générale, des éléments visés au Principe 4.1 dès le traitement des données personnelles ou, si une communication des données à caractère personnel à un tiers est envisagée, au plus tard lors de la première communication des données. Outre les informations énumérées au Principe 4.1, les informations devraient inclure l'origine des données collectées, leur responsable du traitement des données, la base juridique de la transmission ou du partage de données et la possibilité de s'opposer à cette transmission ou ce partage*. »

conséquences d'un retrait, des personnes, du caractère obligatoire ou facultatif de la réponse aux questions utilisées pour collecter les données personnelles et des conséquences pour les personnes concernées d'un défaut de réponse, de l'impact potentiel du profilage sur la personne concernée et enfin, des informations utiles sur le raisonnement qui sous-tend le profilage ou le modèle utilisé par le responsable du traitement des données. Le texte de l'article 13.2. f) du RGPD, réclame qu'au minimum, la personne concernée doit recevoir des 'informations utiles' ('*meaningful information*') concernant cette « logique sous-jacente »[68]. De manière générale, doivent être prises les mesures qui permettent de faciliter la compréhension du fonctionnement du traitement de profilage, les ressources sur lesquelles ils travaillent, l'impact sur la personne les voies de contestation des résultats qui découlent de ce profilage. L'OCDE[69] décrit comme suit l'objectif de telles mesures lorsqu'elles concernent un système de profilage utilisant l'IA (Recommandation 1.3 : 'transparence et explicabilité') : « *Les acteurs de l'IA devraient s'engager à assurer la transparence et une divulgation responsable des informations liées aux systèmes d'IA. A cet effet, ils devraient fournir des informations pertinentes, adaptées au contexte et à l'état de l'art, afin:*

i.de favoriser une compréhension générale des systèmes d'IA,

ii. d'informer les parties prenantes de leurs interactions avec les systèmes d'IA, y compris dans la sphère professionnelle,

iii.de permettre aux personnes concernées par un système d'IA d'en appréhender le résultat, et,

iv.de permettre aux personnes subissant les effets néfastes d'un système d'IA de contester les résultats sur la base d'informations claires et facilement compréhensibles sur les facteurs, et sur la logique ayant servi à la formulation de prévisions, recommandations ou décisions. ».

Il serait par ailleurs utile que l'utilisation d'un système de traitement automatisé traitant des données à caractère personnel soit signalé par la présence d'une icône. Enfin, toujours en vertu du principe de loyauté, des textes récents[70] ont affirmé l'obligation des organisations, qui utilisent des robots humanoïdes, de révéler leur qualité non humaine et la nécessité de pouvoir toujours distinguer leur nature au premier coup d'œil, par exemple par la présence d'une icône de signalement.

b. le principe de détermination et de légalité des finalités

[68] Sur le commentaire des termes 'logique sous-jacente', utilisés par le RGPD, on lira les '*Guidelines on Automated individual decision-making and Profiling*, publiées par le Groupe de travail de l'article 29 et depuis reprises par l'EDPB, le 3 octobre 2017 et revues le 6 février 2018, WP251rev.01 (texte disponible en anglais sur le site de l'EDPB: https://edpb.europa.eu/edpb_en) : « *Data controllers should find simple ways to tell the data subject[s] about the rationale behind, or the criteria relied on in reaching the decision[s] […but] not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision*"

[69] OCDE, *Recommendation on Artificial Intelligence (AI) – the first intergovernmental standard on AI* –, adopté par le Conseil des Ministres de l'OCDE, le 22 mai 2019

[70] Ainsi, le rapport du HLGE on Artificial Intelligence déjà cité, p. 34 : « *Human beings should always know if they are directly interacting with another human being or a machine, and it is the responsibility of AI practitioners that this is reliably achieved. AI practitioners should therefore ensure that humans are made aware of – or able to request and validate the fact that – they interact with an AI system (for instance, by issuing clear and transparent disclaimers). Note that borderline cases exist and complicate the matter (e.g. an AI-filtered voice spoken by a human). It should be borne in mind that the confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human.73 The development of human-like robots74 should therefore undergo careful ethical assessment."* Ce principe de 'distinction' interdit toute utilisation non transparente des technologies pour faire croire artificiellement à l'action d'une personne, telle que stigmatisée par la Déclaration de Montréal (Principe d'intimité et de vie privée, n° 8) : *« L'intégrité de l'identité personnelle doit être garantie. Les SIA (Systèmes d'intelligence artificielle) ne doivent pas être utilisés pour imiter ni modifier l'apparence physique, la voix et d'autres caractéristiques individuelles dans le but de nuire à la réputation d'une personne ou pour manipuler d'autres personnes*. » (*Déclaration de Montréal. Pour une IA responsable*, Déclaration développée sous les auspices de l'Université de Montréal, 2017 (www.declarationmontreal-iaresponsable.com/la-declaration).

30. L'affirmation de ce principe clé de la réglementation des données soulève dans son application aux systèmes d'IA de nombreuses difficultés. La première concerne la détermination des finalités. La richesse possible des agrégations permise par les algorithmes au sein des *big data* voir l'évolution de celles-ci en fonction de nouvelles données collectées permet à l'exploitant du système d'entrevoir de nouvelles applications possibles. Prenons quelques exemples : le profilage de la clientèle peut, dans un premier temps, avoir pour seule finalité la sélection de la clientèle, l'affinement du système peut amener à avoir une meilleure connaissance des potentiels clients ainsi sélectionnables et sur cette base de leur proposer des prix différenciés en fonction d'indices de leur demande potentielle pour tel ou tel produit. Autre exemple, si Facebook développe pour son propre intérêt un système d'analyse intelligente des données de l'utilisation de son réseau social, elle offre également à des 'partenaires', par son système d'IA, des accès ciblés à ses blogs et leur permet ainsi de mieux connaître leurs potentiels 'clients' ou utilisateurs[71]. Les données de géolocalisation reprises par Google servent à des finalités multiples, en dehors de celles poursuivies par le responsable lui-même, soit à des clients professionnels de Google, intéressés par l'envoi de publicités ciblées à des personnes proches des lieux où ils offrent leurs biens ou produits[72], soit demain à des autorités publiques municipales préoccupées par un meilleur planning urbain dans le cadre de leur lutte contre la pollution ou pour signaler des alternatives de mobilité à leurs citoyens[73]. On conçoit dès lors que la notion de finalités compatibles sera souvent invoquée pour permettre l'élargissement des finalités. Il sera utile de maintenir une interprétation stricte des critères retenus par l'article 6.4 du RGPD.

31. Les risques liés à l'impact dans certains secteurs de certains traitements utilisant l'IA peuvent amener à interdire *a priori* leur utilisation à certaines finalités. On cite les craintes d'utilisation de la reconnaissance faciale et dès lors la réglementation stricte qui s'y applique[74]. Récemment, la loi belge sur les assurances a été modifiée le 4 décembre 2020. L'article 4.2 prévoit : « *Lors de la conclusion du contrat visé à l'article 46/1, le refus du candidat assuré d'acquérir ou d'utiliser un objet connecté qui récolte des données à caractère personnel concernant son mode de vie ou sa santé ne peut en aucun cas conduire à un refus d'assurance ni à une augmentation du coût du produit d'assurance.*" Et l'article 5 introduit un article 46.3 qui énonce : «*Aucune segmentation ne peut être opérée sur le plan de l'acceptation, de la tarification et/ou de l'étendue de la garantie sur la base de la condition que le candidat assuré accepte d'acquérir ou d'utiliser un objet connecté*

---

[71] Le rapport de l'ICO (ICO, *Big Data, artificial intelligence, machine learning and data protection,* p. 11, disponible à l'adresse suivante: *https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.) Big Data, artificial intelligence, machine learning and data protection* (dernière consulytation, 17 janvier2021), l'autorité de protection des données, donne l'exemple de la société Data Sift. Cette société à partir de données venant de Twitter, Facebook et autres médias sociaux qu'elles analysent et revend pour des finalités marketing ou autres.

[72] Ces exemples sont tirés du rapport de l'ICO cités note précédente..

[73] Autre exemple : il est possible d'utiliser les nombreuses données collectées à propos de mes employés dans le cadre de leur travail, pour des finalités traditionnelles : sécurité, calcul des salaires et primes, suivi des dossiers mais dans le même temps, grâce à un système d'IA travaillant sur toutes ces données, d'instaurer un profilage de ses employés, de calculer automatiquement leur rentabilité et, en fonction de cela, leur évolution au sein de l'entreprise ou des primes liées à leur rentabilité.

[74] Ainsi, en France, la reconnaissance faciale pour le compte de l'Etat peut être justifiée par l'intérêt public (*article 6 III de l'Ordonnance de 2018*). Elle doit être autorisée par décret en Conseil d'Etat après avis de la CNIL, lorsqu'elle: 1. Intéresse la sûreté de l'Etat, la défense ou la sécurité publique ; 2. A pour objet la prévention, la recherche, la constatation ou la poursuite des infractions pénales ou l'exécution des condamnations pénales ou des mesures de sûreté. (*Article 31 II de l'Ordonnance*) ; 3. Est nécessaire à l'authentification ou au contrôle de l'identité des personnes, et que l'Etat agit dans l'exercice de ses prérogatives de puissance publique. (*Article 32 de l'Ordonnance*)

qui récolte des données à caractère personnel concernant son mode de vie ou sa santé, accepte de partager des informations récoltées par un tel objet connecté, ni sur la base de l'utilisation par l'assureur de telles informations." On doit s'attendre à une multiplication de telles réglementations spécifiques au vu des risques importants de discrimination que représentent les capacités prédictives et décisionnelles de l'IA.

32. La légitimité des traitements dans le secteur privé est souvent fondée sur le consentement, consentement parfois obtenu dès la visite d'un site web par l'acceptation de cookies, selon la formule souvent trompeuse à moins qu'elle ne soit ironique, des opérateurs de ces sites : « *Nous sommes soucieux de votre vie privée* ». Ce consentement via cookies interposés sera obtenu de manière globale ou à travers un paramétrage fastidieux et souvent aux critères incompréhensibles de vos préférences. Quelques questions se posent à propos de la validité d'un consentement obtenu par de tels moyens. Le RGPD exige que ce consentement soit libre, éclairé et spécifique[75]. Tout cela dit, soulignons que les qualités requises du consentement seront rarement remplies. Elles supposent, sous réserve d'autres exigences encore, un consentement libre, c'est-à-dire non manipulé et sur base d'un choix réel qui puisse être autre que le seul : *'j'accepte*'. La complexité des montages des systèmes supportés par les technologies de l'IA, la diversité des sources utilisées, l'impossibilité de prévoir les corrélations qui seront à la base des décisions du responsable et, dans le contexte de l'accès à des services gratuits et à portée d'un clic, la difficulté de prendre le recul nécessaire au moment où on pousse sur le 'j'accepte', tous ces facteurs rendent les conditions mises par le RGPD complétement illusoires d'autant plus que le refus conduit à un non accès au service. Dans de telles conditions, le consentement ne peut être, sauf exceptions, la condition d'accès à des services souvent ressentis comme nécessaires à l'exercice de la vie sociale. Il devrait, sauf nécessité lié au service lui-même[76] (par exemple, un service de confection de vêtement sur mesure offert à distance ou d'éducation à distance), être toujours accompagné d'une offre d'un service non profilé, qui devrait, par défaut, s'appliquer[77]. Que proposer dès lors ? Sans doute, et ces solutions ont notre préférence, faut-il, prescrire, là où c'est possible et suivant l'exemple du droit de la consommation, un consentement collectif négocié entre le responsable du traitement et les représentants des usagers (avec ou non la médiation des autorités de protection des données) ou, dans ces cas interdire en principe le consentement, et réclamer que seules les autres causes de validité soient invocables, qu'il s'agisse tantôt de la nécessité d'exécution d'un contrat ou des mesures précontractuelles, tantôt de l'intérêt légitime de l'opérateur qui trouvera, dans certains cas, dans les trois apports de l'IA déjà cités : la sécurisation, l'optimisation et l'objectivation, des justifications supplémentaires au traitement.

---

[75] Groupe de travail de l'article 29, « *Lignes directrices sur le consentement au sens du règlement 2016/679* », Adoptées le 28 novembre 2017,Version révisée et adoptée le 10 avril 2018.

[76] Mais on est alors dans une hypothèse de licéité du traitement à savoir le traitement nécessaire à l'exécution du contrat ou de mesures précontractuelles (article 6. 1. b))

[77] C'est en tout cas dans ce sens que s'oriente la proposition de règlement e-Privacy. Les utilisateurs doivent être informés de toute collecte de données et de la finalité de l'utilisation de celle-ci. Par conséquent, le consentement ne doit pas être caché dans les conditions générales ou lié à d'autres services. Ainsi, un achat en ligne nécessite le transfert des données par l'acheteur pour mener à bien la transaction. Par contre, il n'est pas permis d'utiliser les données transmises à des fins publicitaires. Cela nécessiterait en effet soit un nouvel accord spécifique et, dans ce cas, un consentement clair du client, soit dans le cadre où l'intérêt légitime du responsable du traitement est invoqué, la preuve que la balance d'intérêts a été opérée (*supra*, n° 21).

33. A défaut, ne faut-il pas laisser le choix à la personne concernée entre un accès non profilé et un accès profilé, voire entre un accès anonyme ou au contraire identifié ? On reprendra volontiers sur ces deux points (droit à l'anonymat et droit à ne pas être profilé), la recommandation 3.8 du projet du Conseil de l'Europe relatif au profilage, qui affirme : « *Dans toute la mesure du possible, les prestataires de services et les plateformes devraient offrir différents services plus ou moins personnalisés, voire non personnalisés, en fonction du service offert, afin de garantir que la personne concernée ait le choix en ce qui concerne l'intensité du profilage. Pour qu'il soit libre, le consentement suppose pour le moins, pour la personne concernée, la possibilité d'un choix informé. Le consentement au profilage ne devrait pas pouvoir être exigé comme condition de la prestation d'un service. Quand le consentement est requis, il incombe au responsable du traitement de prouver que la personne concernée a accepté explicitement le profilage au-delà de ce qui était nécessaire à l'exécution de la prestation et ce après avoir été informée,* … » Au-delà, il importe que la personne concernée puisse dans certains cas de traitements fondés sur des systèmes d'intelligence artificielle, en particulier lorsqu'il s'agit de profilage à des fins publicitaires[78], définir la finalité du profilage qu'elle souhaite et, dès lors, réduire le champ des données qui seront exploitées. Prenons un exemple, l'accès à un service de musique en ligne ne suppose pas que vous soyez d'accord avec le profilage de vos goûts musicaux lequel, par contre, est nécessaire si vous souhaitez que le fournisseur vous conseille ou vous propose des musiques adaptées à vos goûts. Votre choix devrait pouvoir porter sur chacune des diverses finalités et, le cas échéant, sur les destinataires tiers qui permettront de réaliser les finalités. Ainsi, pour reprendre toujours l'exemple du service de musique en ligne, peut-être, le souhait de l'internaute est-il de recevoir l'annonce publicitaire émanant de tiers à propos de la sortie d'une chanson de son interprète favori ? A l'inverse, le choix de l'internaute peut s'exprimer en sens contraire. En d'autres termes, admettre le profilage à des fins publicitaires par le responsable de traitement ne signifie pas nécessairement admettre le profilage par des tiers ou la cession de données ou de mon profil à des tiers.

c. les principes de minimisation des données et de proportionnalité de leur durée de conservation

34. Les principes de minimisation et de proportionnalité de la durée rencontrent également des difficultés d'application dans le cadre des traitements utilisant des systèmes d'IA. Ces principes présupposent que l'on puisse *a priori* déduire de la finalité de l'application les données nécessaires à son obtention et la durée de leur conservation. Or les systèmes dits de *machine learning* fonctionnent grâce à des corrélations statistiques établies sur base de rapprochements souvent non prévisibles de données et exigent donc que les réservoirs de données brassent très largement[79]. Quelle solution trouver ? La pseudonymisation de certaines données prônées dans des lignes directrices du Conseil de l'Europe en matière

[78] Autre exemple, l'utilisation de systèmes d'IA dans le cas de voitures connectées devrait distinguer les hypothèses où le possesseur de la voiture connectée souhaite connaître son profil de conducteur à des fins personnelles (respect des limites, analyse de la consommation, risques pris (par exemple : conduite en état de somnolence ou d'alcoolisme, …) sans que ces 'profils' ne soient accessibles à son garagiste hormis pour des raisons de sécurité, de l'hypothèse où ce possesseur refuse l'utilisation de tout système d'IA, …

[79] Ainsi, hypothèse purement fictive, il pourrait apparaître, aux yeux de l'administration fiscale lors de l'utilisation de vastes banques de données, que les dirigeants d'entreprise de plus de 200 employés et moins de 400, disposant d'une voiture rouge immatriculée entre telle et telle année, ayant l'habitude de voyages '*all inclusive*' dans les pays méditerranéens, habitant tel type de quartier dans des villes de plus de 50.000 habitants, avec un enfant et un chien, constituent des fraudeurs potentiels. Cet exemple témoigne du fait qu'il est difficile, *a priori* du moins, de fixer les éléments qui serviront à établir le profil.

de mégadonnées n'est pas une solution répondant aux principes évoqués[80]. Le projet de recommandation du Conseil de l'Europe sur le profilage, souvent cité, en son point 3.2. au nom de ces principes, recommande au moins la balise des « *legitimate expectations* » *« Les données personnelles utilisées dans le cadre du profilage devraient être adéquates, pertinentes et non excessives par rapport aux finalités pour lesquelles elles seront traitées. Dans les systèmes de 'machine learning' il est difficile de connaître a priori quelles données permettront des corrélations significatives. Par ailleurs, il est important de limiter le traitement de profilage à des catégories de données dont la personne concernée peut raisonnablement s'attendre (légitimement s'attendre) à ce qu'elles soient prises en considération au vu des finalités du profilage. ».* Sans doute cette balise est-elle insuffisante et faudra-t-il, secteur par secteur avec le cas échéant une intervention législative fixant elle-même certaines limites aux données utilisées, répondre à des questions, qui sont loin d'être triviales comme les suivantes : « Jusqu'où, une compagnie d'assurances peut-elle utiliser des données relatives aux personnes assurées dans le cadre de l'offre de services individualisés ? » ; « Jusqu'où une banque ou un organisme de crédit peut-il au nom de sa responsabilité de donneur de crédit et suivant les exigences du principe '*Know your customer*' profiler ses clients ? » ; « Dans quelle mesure, un employeur peut-il utiliser, vis-à-vis des employés ou candidats employés, des systèmes *d'affective computing* dans le cadre de leur sélection, gestion de carrière, etc. ? ».

d. le principe de sécurité des traitements

35. Le chapitre I de l'article énonçait les multiples risques attachés au fonctionnement des systèmes d'IA : mauvaise qualité, non pertinence ou non actualisation des données collectées et traitées par le système ; biais ou erreur dans la programmation, évolution imprévisible et opacité du système, sans omettre les risques d'intrusion et d'atteinte au fonctionnement du système. La sécurité des systèmes est donc un principe majeur au vu des conséquences que l'avènement de ces risques peut entraîner. Les principes dits éthiques du rapport d'experts de la Commission européenne[81] y attachent une attention particulière en particulier aux besoins d'une évaluation régulière des mesures de sécurité. On se contentera de se référer au projet dc recommandations à propos du profilage du Conseil de l'Europe[82] soulignent en son point. « *Les responsables du traitement et, le cas échéant,*

[80] Voilà comment les Lignes directrices (Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679) adoptees le 6 février 2018 par le Groupe de l'article 29, s'expriment à ce propos (p. 11), sans résoudre le problème: "*The business opportunities created by profiling, cheaper storage costs and the ability to process large amounts of information can encourage organisations to collect more personal data than they actually need, in case it proves useful in the future. Controllers must make sure they are complying with the data minimisation principle, as well as the requirements of the purpose limitation and storage limitation principles. Controllers should be able to clearly explain and justify the need to collect and hold personal data, or consider using aggregated, anonymised or (when this provides sufficient protection) pseudonymised data for profiling.*"

[81] Comme le note la recommandation n° 5 de l'OCDE (OCDE, Recommandation du Conseil sur l'intelligence artificielle, OECD/LEGAL/0449 *le*, adopté par le Conseil des Ministres de l'OCDE, le 22 mai 2019, disponible à l'adresse : https://legalinstruments.oecd.org/api/print?ids=648&lang=fr (consultée pour la dernière fois, le 22 janvier 2021) à propos de la sécurité des traitements utilisant l'IA: *'AI systems should be* ***robust, secure and safe*** *throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.'*[81].

[82] Le texte du Règlement repris en annexe de la Résolution du Parlement européen sur les aspects éthiques de l'IA est plus complet encore (article 8. 1) : « *L'intelligence artificielle, la robotique et les technologies connexes à haut risque, y compris les logiciels, les données et les algorithmes utilisés ou produits par ces technologies sont développées, déployées et utilisées de manière à garantir qu'elles sont:*
a) développées, déployées et utilisées de manière résiliente afin d'assurer un niveau de sécurité adéquat en se conformant aux références minimales en matière de cybersécurité proportionnées par rapport aux risques repérés, et d'empêcher toute vulnérabilité technique d'être exploitée à des fins malveillantes ou illicites;

les sous-traitants veillent à évaluer de manière critique la qualité, la nature et la quantité des données utilisées en éliminant les données inutiles et toutes celles qui pourraient biaiser les résultats. En particulier, certains seuils minimaux d'exactitude, d'exactitude des résultats doivent être respectés. Ils s'assurent de la robustesse du modèle en cas d'apport de nouvelles données… » et les points 7.7 et 7.8 « *Afin d'assurer la confiance dans les systèmes d'IA, les responsables du traitement et, le cas échéant les sous-traitants, veillent à l'utilisation de systèmes fiables et sûrs, notamment en ce qui concerne la mise sur pied de procédures en cas de non-fonctionnement, d'erreurs ou d'incohérences pendant toute la durée de vie du système. Ils s'assurent de manière régulière tout au long de la vie du système que celui-ci est fiable et que ses résultats sont conformes au modèle et sont reproductibles. Le système devrait être robuste pour résister aux attaques ou à d'autres manipulation des données ou des algorithmes. » « Les responsables du traitement et, le cas échéant, les sous-traitants veillent à évaluer de manière critique la qualité, la nature et la quantité des données utilisées en éliminant les données inutiles et toutes celles qui pourraient biaiser les résultats. En particulier, certains seuils minimaux d'exactitude, d'exactitude des résultats doivent être respectés. Ils s'assurent de la robustesse du modèle en cas d'apport de nouvelles données. Les résultats eux-mêmes sont évalués pour évaluer leur impact sur la personne concernée, y compris le droit à la non-discrimination. Les applications d'intelligence artificielle devraient permettre le contrôle effectif par les personnes comme par les groupes concernés des effets de ses applications autant sur les personnes, sur les groupes que sur la société*. ». On ajoute que tant les textes du Parlement européen, du Conseil de l'Europe que de l'OCDE[83] insistent sur les obligations des fournisseurs de données et des algorithmes de collaborer à la sécurité du système en particulier par une documentation adéquate.

C. Les obligations du responsable

36. Dans le cadre de cet article nous nous limiterons à l'analyse de deux obligations, majeures lorsqu'il s'agit de systèmes IA : la première découle du principe de non suffisance du traitement automatisé et comprend en particulier la nécessité d'une information relative à la logique du système et l'explicabilité de la décision par une personne humaine 'compétente' ; le second consacre l'obligation d'évaluer les risques lorsque le système présente un degré élevé de risques.

a. Le principe de non suffisance du traitement automatisé en cas de décision vis-à-vis d'une personne concernée.

b) développées, déployées et utilisées de manière sécurisée afin d'assurer la mise en place de garanties, notamment d'un plan et de mesures de secours en cas de risque pour la sécurité ou la sûreté;
c) développées, déployées et utilisées de manière à garantir que les résultats relatifs à la réalisation des activités et des objectifs pour lesquels elles ont été conçues, sont fiables, comme l'utilisateur peut raisonnablement s'y attendre, notamment en veillant à ce que toutes les opérations soient reproductibles;
d) développées, déployées et utilisées de manière à garantir la précision des résultats relatifs aux objectifs et aux activités des technologies particulières; si la survenance d'inexactitudes occasionnelles ne peut être évitée, le système informe, dans la mesure du possible et par des moyens appropriés, les déployeurs et les utilisateurs de la probabilité que des erreurs et des inexactitudes seproduisent;
e) développées, déployées et utilisées de manière à pouvoir être facilement expliquées afin de garantir qu'une analyse desprocédés techniques associés aux technologies peut être réalisée;
f) développées, déployées et utilisées de manière à ce qu'elles informent les utilisateurs qu'ils interagissent avec des systèmes d'intelligence artificielle, en exposant correctement et de façon exhaustive leurs capacités, leur précision et leurs limites aux développeurs, aux déployeurs et aux utilisateurs d'intelligence artificielle;
g) conformément à l'article 6, développées, déployées et utilisées de manière à permettre, en cas de non-respect des dispositifs de sécurité énoncés aux points a) à g), une désactivation temporaire des fonctionnalités concernées et un retour à un état antérieur qui rétablit des fonctionnalités sûres. »

[83] OCDE, Recommandation du Conseil sur l'intelligence artificielle, OECD/LEGAL/0449 *le*, adopté par le Conseil des Ministres de l'OCDE, le 22 mai 2019, disponible à l'adresse : https://legalinstruments.oecd.org/api/print?ids=648&lang=fr (consultée pour la dernière fois, le 22 janvier 2021).

37. L'article 22 du RGPD consacre le droit de la personne concernée « *de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire* » appellent les remarques suivantes. L'analyse de cette disposition laisse apparaître nombre de lacunes ou, en tout cas, d'ambiguïtés. Que veulent dire les expressions : '*décision fondée exclusivement*' ; « *affecter de manière significative* » et le terme « *uniquement* » ? Enfin, la décision dont parle l'article 22 doit viser une « personne concernée ». C'est la conséquence certes d'une législation centrée sur la protection de personnes individuelles mais ne faudrait-il pas également prendre en compte le fait que des systèmes en particulier prédictifs visent des catégories de personnes : ainsi les personnes habitant tel quartier, ayant tel type de comportement sur le net, telle mobilité… ? Le risque est ici collectif et, de ce fait, mériterait *a fortiori* d'être pris en compte. Nous ne pourrons dans le cadre de cette contribution analyser tous ces ambiguités[84]. Par ailleurs, l'article 22.2 prévoit des exceptions qui s'appliqueront à la plupart des systèmes d'IA : besoins contractuels ou précontractuels, consentement de la personne concernée ou exécution d'une mission d'intérêt public autorisée par l'Etat. Manque l'hypothèse d'un système IA dont la licéité reposerait sur un intérêt légitime prépondérant du responsable du traitement.

*38.*L'article 22.3 réclame au cas où les exceptions s'appliqueraient, que le responsable mette « *en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement d'exprimer son point de vue et de contester la décision* ». Au-delà, la personne concernée peut-elle avoir accès à une explication par écrit ou, à défaut, orale des critères utilisés pour justifier la décision et de leur application à son cas concret ? L'article 22[85] ne le réclame pas. La disposition évoque simplement le droit à des 'mesures appropriées' et à l'obtention d'une intervention humaine. Certes, l'intervention humaine ne peut se limiter à une simple réaffirmation par oral de la 'vérité sortie de l'ordinateur' mais à partir de quand pourra-t-on considérer que l'humain a une réelle capacité de remise en cause de la présomption de vérité sortie des ordinateurs[86] ? Que recouvrent les termes '*garanties appropriées*' : le droit à une audience en face à face ? Un droit de contestation de la décision après explication ? Les '*garanties appropriées*' exigent-t-elles, par exemple, que l'autorité policière, sans doute une fois l'enquête bouclée (secret de l'instruction oblige !), explique le raisonnement suivi qui a présidé à la suspicion ou à la mesure prise par l'autorité[87]? Nonobstant le flou de l'article 22 et la difficulté due à l'opacité des systèmes complexes d'IA, le considérant n° 71 qui lui est lié exige que le bénéfice de ces exceptions de l'article 22.2 soit assorti par ceux qui s'en prévalent « *de garanties appropriées, qui devraient comprendre une information spécifique de la personne concernée ainsi que le droit d'obtenir une intervention humaine, d'exprimer son point de vue,* ***d'obtenir une explication quant à la décision*** *prise à l'issue de ce type d'évaluation et de contester la décision* » . Ainsi, le responsable se devra, suite à une décision prise, d'offrir, non seulement, une interface humaine capable de recevoir la

[84] Sur cette analyse, nous renvoyons à notre ouvrage : *Le RGPD face aux défis de l'intelligence artificielle,* Cahier du CRIDS, n° 48, Larcier, Bruxelles, 2020, p. 109 et s., n°35 et s ; et les différentes références y reprises

[85] Le considérant n° 71 semble par contre le réclamer comme nous le dirons dans l'alinéa qui suit.

[86] Sur cette 'incontestabilité' de la décision produite par la machine, lire entre autres, M. KAMINSKY : "*And where human decision-making can often be contested, algorithmic decision-making (…) is often taken at face value, and left unchallenged and unchallengeable.* ». ("Binary governance: lessons from the GDPR's approach to algorithmic accountability", *Southern California Law Review*, 2019, 76, p. 15. Il semble que les autorités singapouriennes exigent que les personnes charges de répondre aux demandes d'explication ou de contestation des décisions d'applications de '*machine learning*' disposent d'une réelle compétence en matière *de machine learning* et connaissent l'application

[87] On peut songer par exemple dans le cadre d'un profilage effectué aux fins de recherche d'auteurs d'infractions que les personnes ainsi profilées soient clairement classifiées de manière distincte des personnes suspectées suite aux méthodes d'enquête classiques.

personne concernée et de répondre à ses questions mais également de donner toutes les informations qu'il détient dans le cadre des exigences d''explicabilité' de la décision prise[88], c'est-à dire au minimum les bases suffisantes pour permettre la compréhension du processus, qui a mené à la décision prise, et la possibilité, dès lors, pour la personne concernée, de la contester en connaissance de cause. Cette possibilité de contester doit s'entendre de la possibilité d'un recours à l'autorité certes mais surtout d'un recours interne par une personne ayant compétence de revoir les décisions prises par ou à la suite de l'utilisation du système de *machine learning*. L'ensemble de ces obligations est repris aux points 5.6 et s. du projet de recommandation du Conseil de l'Europe sur le profilage déjà cité : *« .. Lorsque le système de traitement de profilage émet une décision ou un projet de décision, il est fortement recommandé que :*

a. *les responsables du traitement tiennent compte de toutes les particularités des données et ne se fondent pas simplement sur des informations ou des résultats du traitement pris hors de son contexte ;*
b. *en cas de traitement de profilage à risque élevé, le responsable du traitement informera la personne concernée des opérations algorithmiques qui sous-tendent le traitement de données, y compris les conséquences pour elle de ces opérations. Dans ce cas, l'information doit être telle qu'elle permette à la personne concernée de comprendre la justification des décisions ou propositions de décision prises à son encontre. Cette exigence dépend fortement des conséquences que peut avoir l'impact du résultat obtenu pour la personne concernée, conformément au principe d'explicabilité ;*
c. *dans ce cas, la personne nommée par le responsable du traitement doit pouvoir, sur la base d'arguments raisonnables, décider de ne pas se baser sur les résultats des recommandations découlant de l'utilisation du traitement de profilage ;*
d. *en présence d'indications permettant de penser qu'il y a eu discrimination directe ou indirecte fondée sur le fonctionnement du traitement de profilage, les responsables du traitement et les sous-traitants apportent la preuve de l'absence de discrimination. .*

Les personnes affectées par une décision fondée sur un traitement de profilage devraient avoir le droit de recevoir toute explication utile sur cette décision, ou la proposition de décision, afin d'en comprendre la justification. La propriété intellectuelle ou l'existence de secrets commerciaux ne peuvent être contestées que lorsque les informations à fournir affecteraient gravement ces droits. L'invocation de ces droits et intérêts par le responsable du traitement ne peut conduire à priver la personne concernée ou le groupe concerné de la capacité de comprendre les décisions ou les projets de décision adoptés. Nonobstant le recours devant l'autorité de contrôle ou le recours juridique, les personnes concernées devraient avoir le droit de contester le profilage devant une personne désignée par le responsable du traitement, ayant accès à toutes les informations sur le profilage et son fonctionnement et compétente pour modifier ou supprimer la décision ou le projet de décision. »

b. L'obligation de procéder à un Privacy ou Ethics Impact Assessment

39. Pour les traitements présentant un **risque élevé** pour les personnes concernées, l'article 35 du RGPD prescrit aux responsables de traitement l'obligation de procéder à une évaluation des risques (en abrégé *PIA*). Il s'agit de procéder à une évaluation interne des risques en matière de protection des données et la présentation des mesures prises pour atténuer le risque. L'article liste les critères qui conduisent à qualifier le traitement comme présentant ce risque élevé et laisse à l'autorité de contrôle le soin, le cas échéant d'établir et de lister les applications correspondant à ce niveau de risque. Sans nous livrer à une

[88] Comme l'affirme le Groupe de l'article 29 dans ses Lignes directrices reprises et confirmées par l'European Data Protection Board (Groupe de l'article 29, *Lignes directrice relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679*, WP251 rev.01, adoptées le 3 octobre 2017 et révisées le 6 février 2018, p.18), « *compte tenu du principe fondamental de transparence qui sous-tend le RGPD, les responsables du traitement doivent veiller à expliquer clairement et simplement aux personnes concernées la manière dont fonctionne le profilage ou le processus décisionnel automatisé* », ce qui ne signifie pas « *une explication complexe des algorithmes utilisés* »

analyse détaillée de cet article et de son application aux applications de profilage[89], relevons que ce concept de traitement à « risque élevé » ou à « haut risque » est repris par le projet de règlement du Parlement européen en matière d'IA. La notion[90] y est entendue non vis-à-vis des seuls risques individuels d'atteinte à la protection des données mais également collectifs comme ceux d'atteinte à l'environnement, à la justice sociale ou à la démocratie. L'article 4 e) définit en effet le 'haut risque' comme suit : *« «haut risque», un risque important associé au développement, au déploiement et à l'utilisation de l'intelligence artificielle, de la robotique et des technologies connexes de causer du tort ou un préjudice aux individus ou à la société en violation des droits fondamentaux et des règles de sécurité établis par le droit de l'Union, en tenant compte de leur utilisation ou finalité spécifique, du secteur dans lequel elles sont développées, déployées ou utilisées et de la gravité des torts ou des préjudices susceptibles de se produire; »* On ajoute que le projet du Parlement liste les traitements en fonction à la fois des secteurs et de la finalité des applications les traitements IA à haut risque mais surtout alors que le RGPD se contente d'un contrôle interne avec le cas échéant une obligation de consultation de l'autorité de protection des données, le contrôle imaginé par le projet du Parlement européen est confié préventivement à un organe externe multidisciplinaire et *multistakeholder*s ou à un organisme accrédité par cette Autorité. On note que ce contrôle préventif suppose l'auditabilité des systèmes d'IA[91] et qu'il est prévu à des intervalles réguliers. Reste à s'interroger sur la manière dont les deux dispositifs et les deux autorités pourront collaborer. Enfin, il est intéressant de noter l'appui tant par le RGPD que celui du Parlement européen à des mécanismes de

---

[89] A cet égard, l'article 1.1. k) qui définit les applications de profilage à haut risque comme suit : « L'expression « traitements de profilage à risque élevé » peut notamment désigner :

i. *le profilage dont le fonctionnement entraîne des effets juridiques ou qui ont un impact significatif pour la personne concernée ou pour le groupe de personnes identifié par le traitement de profilage ;*
ii. *le profilage qui, en raison du public visé ou du contexte ou de la finalité du traitement, en particulier en raison de la possibilité d'abus ou de détournement du déséquilibre dans le pouvoir d'information, notamment lorsqu'il s'agit de mineurs ou de personnes vulnérables, comporte un risque d'affecter ou d'influencer des personnes concernées ;*
iii. *le profilage ayant pour objet des données relevant des catégories particulières de données au sens de l'article 6 de la Convention 108+ ou ayant pour finalité de les détecter ou les prédire ;*
iv. *le profilage affectant un très grand nombre de personnes, notamment celui opéré par des services d'intermédiaires en ligne à leur bénéfice ou pour celui de tiers.*

[90] Pour distinguer les systèmes à haut risque des autres, le Livre Blanc de la Commission sur l'intelligence artificielle (*Livre blanc – Intelligence artificielle – Une approche européenne basée sur l'excellence et la confiance*, Bruxelles, le 19 février 2020, COM(2020)65 final, disponible sur le site : https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf (consulté pour la dernière fois, le 16 janvier 2021) prend dans le domaine médical (p. 20), l'exemple suivant : l'utilisation d'un système IA pour la fixation des agendas du corps médical d'un hôpital ne présente pas les mêmes risques que l'utilisation d'un système pour la détection préventive des personnes qui pourraient demain souffrir de la maladie d'Alzeihmer. *« Si une telle approche est importante pour garantir la proportionnalité de l'intervention réglementaire, elle requiert néanmoins des critères précis pour différencier les diverses applications de l'IA, et notamment pour déterminer si elles sont ou non «à haut risque» . Les éléments permettant d'établir qu'une application d'IA est à haut risque devraient être clairs, faciles à comprendre et applicables à toutes les parties concernées ».*

[91] Dans le même sens, le projet de recommandation du Conseil de l'Europe sur le profilage déjà cité : « *Aux fins d'une évaluation continue des risques tant individuels que collectifs, et en tout cas lorsqu'il s'agit de traitements de profilage à risque élevé, les responsables du traitement et, le cas échéant, les sous-traitants devraient documenter l'entraînement du modèle et effectuer des évaluations d'impact régulières en traitant des risques spécifiques du profilage basé sur des systèmes d'IA. Pour atteindre cet objectif, ils devraient s'entourer d'une équipe d'évaluation multidisciplinaire et consulter les représentants des intérêts concernés par le profilage, y compris les personnes profilées. Ce processus d'évaluation devrait être mené par des personnes dotées des qualifications professionnelles et des connaissances adéquates pour apprécier les différents impacts, y compris dans leurs dimensions juridique, sociale, éthique et technique. »*

certification et de labellisation des traitements et en particulier, pour le second texte, des traitements utilisant l'IA.

CONCLUSIONS

40. L'irruption de l'IA dans la conduite des entreprises, des administrations et, en définitive, de nous-mêmes mérite-t-elle une nouvelle réglementation spécifique à cette technologie particulièrement puissante et au fonctionnement peu transparent ? Sans doute, au risque de ne pas suffire[92], certaines interprétations « *ulta legem* » des dispositions du RGPD sont nécessaires pour rencontrer les besoins de protection nouveaux créés par le *machine learning*. L'exposé a surtout montré combien au nom du principe d'*accountability* consacré par l'article 5.2 du RGPD[93], les obligations du responsable se trouvent élargies. C'est au(x) responsable(x) du traitement de faire la preuve du respect des obligations liées aux traitements utilisant la technologie de l'IA, en tenant compte des risques liés à ces traitements. Ainsi, il vérifiera la qualité des données non seulement du point de vue de leur exactitude et de leur mise à jour, mais également du caractère non biaisé de leur utilisation par telle application. A cet égard, avant tout démarrage, il s'avère utile chaque fois qu'il est possible de tester le système sur des jeux de données. La même précaution s'entend en ce qui concerne le ou les algorithmes utilisés qu'ils soient développés ou non par le responsable lui-même. Il s'agira ensuite de documenter les diverses opérations menées au cours du traitement, de conserver les logs des décisions. Enfin, le responsable du traitement prévoira les procédures organisationnelles nécessaires pour respecter le droit à une intervention humaine dans la décision et à une réelle possibilité d'écoute et, le cas échéant, et de prise en compte du point de vue des personnes concernées Il est clair que le respect de la plupart de ces obligations peut faire l'objet d'une certification par l'organe de contrôle des systèmes d'IA mentionné ci avant (*supra,* n°39) ou par d'autres organismes agréés par celui-ci[94].

41. Deux lignes de force, déjà présentes mais seulement en filigrane et à peine dans le RGPD, semblent devoir être affirmées dorénavant pleinement lorsqu'on évoque la question des

---

[92] Cf. les points relevés en conclusion de notre ouvrage, Le RGPD face à l'intelligence artificielle », Cahier du CRIDS, n°48, Larcier, Bruxelles, 2020, p. 132 et s., n°42.

[93] Comme l'affirme C. CASTETS-RENARD, « Accountability of Algorithms in the GDPR and beyond: A European Legal Framework on Automated Decision-Making», (May 20, 2019), *Fordham Intellectual Property, Media & Entertainment Law Journal*, https://ssrn.com/abstract=3391266 or http://dx.doi.org/10.2139/ssrn.3391266.

[94] Le HLGE on artificial intelligence (op.cit., p. 26) résume, comme suit, ses recommandations au 'responsable de traitement' : « *Ensure that the AI system's entire life cycle meets the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability. ⟩ Consider technical and non-technical methods to ensure the implementation of those requirements. ⟩ Foster research and innovation to help assessing AI systems and to further the achievement of the requirements; disseminate results and open questions to the wider public, and systematically train a new generation of experts in AI ethics. ⟩ Communicate, in a clear and proactive manner, information to stakeholders about the AI system's capabilities and limitations, enabling realistic expectation setting, and about the manner in which the requirements are implemented. Be transparent about the fact that they are dealing with an AI system. ⟩ Facilitate the traceability and auditability of AI systems, particularly in critical contexts and situations. ⟩ Involve stakeholders throughout the AI system's life cycle. Foster training and education so that all stakeholders are aware of and trained in Trustworthy AI. ⟩ Be mindful that there might be fundamental tensions between different principles and requirements. Continuously identify, evaluate, document and communicate these trade-offs and their solutions.*"

traitements de données à caractère personnel utilisant l'IA. La première est le passage pour le ou les responsables d'une **obligation d'information à celle d'explication**. Comme le note J.EYNARD[95] « *On passe ainsi petit à petit d'une obligation d'information à une obligation d'explication ce qui va dans le sens d'une meilleure maîtrise par l'usager des processus mis en oeuvre*. » La seconde est le renforcement d'une **approche préventive** des risques qui semble systématiquement s'imposer, en ce qui concerne les systèmes IA à haut risque par l'obligation de PIA, par l'appel à une régulation douce par la certification et, enfin, par la réglementation légale *a priori* de certaines applications comme la reconnaissance faciale ou l'utilisation de l'IA dans certains secteurs comme l'assurance ou l'octroi de crédit.

42. Le dernier point de nos conclusions est plus essentiel encore. Ce qui caractérise les enjeux de l'IA, c'est qu'ils débordent de loin les aspects de protection d'intérêts et de libertés individuels, ceux pour la protection desquels ont été adoptées les législations de protection des données. Les enjeux environnementaux, de justice sociale, de démocratie sont désormais au cœur des textes éthiques qui se succèdent dans les instances internationales. L'IA certes peut être un moyen de prévention et de lutte contre ces risques collectifs, il est également cause de leur aggravation. Les applications IA soulèvent des problématiques éthiques bien au-delà de la seule protection des données. Elles représentent un défi pour la justice sociale, réservant certains avantages aux seules personnes capables d'acquérir les outils IA et pour la démocratie, dans la mesure où elles normalisent parfois à l'excès les comportements et favorisent la manipulation des masses. La protection des consommateurs, des SME et l'étiquetage collectif de personnes constituent d' autres facettes de la réflexion. Enfin, on note que le développement des systèmes d'IA est de plus en plus le fait des *Tech Giants* (ou GAFAM) et des plateformes et pose des questions au regard du droit de la concurrence et du rôle de l'Etat. Ces questions doivent être abordées de manière concertée entre les autorités en charge de ces diverses thématiques. Le respect des libertés et des droits fondamentaux, et notamment du droit à la vie privée et à la dignité humaine, mais aussi à la liberté d'expression et du principe de non-discrimination, mais également des impératifs de justice sociale, de diversité culturelle et de démocratie, doivent être garantis lors du traitement de profilage visé par la présente recommandation. Nous prônons l'existence, à l'instar de pays proches et conformément aux vœux de la Commission européenne, d'un organe de coordination chargé d'une approche éthique de l'innovation que représentent les applications du '*machine learning'*. Il s'agit de construire, comme y appelle l'UIT, une *« AI for good*[96].

[95] J. EYNARD, « Réflexions pour une intelligence artificielle digne de confiance », article à paraître, p.6

[96] L'ITU a organisé en 2018 l'"*AI for Good Global Summit*". Le prochain sera organisé à Genève en septembre de cette année. Voir le site de l'ITU : https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx. « *As the UN specialized agency for information and communication technologies, ITU is well placed to guide AI innovation towards the achievement of the UN Sustainable Development Goals. We are providing a neutral platform for international dialogue aimed at building a common understanding of the capabilities of emerging AI technologies."*