Strasbourg, 7 novembre 2019

T-PD(2019)07FIN

COMITÉ CONSULTATIF DE LA CONVENTION POUR LA PROTECTION

DES PERSONNES A L'ÉGARD DU TRAITEMENT AUTOMATISÉ

DES DONNÉES A CARACTÈRE PERSONNEL

CONVENTION 108

Profilage et la Convention 108+ : Rapport sur l'évolution de la situation après l'adoption de la Recommandation (2010) 13 sur le profilage

DG I – Droits de l'Homme et État de droit

Table des Matières

Résumé

Le profilage se répand dans tous les domaines d'activité. Ce peut être une administration qui souhaite mieux identifier les délinquants, y compris les délinquants potentiels ou offrir des services plus adaptés aux citoyens. Ce peut être des entreprises qui cherchent à avoir une connaissance plus ciblée de leurs ses clients actuels ou futurs, leur offrir de meilleurs services ou mieux sélectionner leurs employés. Ce peut être enfin des hôpitaux ou des chercheurs médicaux dans leur lutte contre des maladies ou pour utiliser des technologies qui leur permettront de meilleurs soins. Il y a dix ans, le Comité des Ministres du Conseil de l'Europe a adopté une recommandation qui prenait en compte les premiers systèmes de profilage basés sur des systèmes experts traditionnels aux capacités assez limitées et sur une logique intrinsèque assez transparente. Aujourd'hui, différentes technologies, internet des objets, megadonnées et intelligence artificielle (IA), appliquées ensemble, modifient radicalement les méthodes et l'impact du profilage. Les systèmes d'IA opèrent souvent de manière opaque – d'ailleurs avec biais et erreurs – en combinant le fonctionnement de réseaux neuraux. Ils s'appuient sur des combinaisons statistiques aléatoires englobant des millions de données et rendent maintenant possible un profilage d'une efficacité et avec un niveau de prévision des comportements individuels sans précédent. Cette réalité alliée à une possible absence de maîtrise des êtres humains sur la technologie augmente les risques que courent les personnes concernées, individuellement mai aussi collectivement. Ces raisons militent en faveur d'un cadre pour le profilage renouvelé et repensé.

Le présent rapport a été écrit à quatre mains, celles d'un informaticien et celles d'un juriste. Cette approche interdisciplinaire était une nécessité absolue pour couvrir le sujet et avait plusieurs objectifs.

- Il s'agissait d'abord de définir le contexte : que signifient IA, l'apprentissage automatique (*machine learning*), l'apprentissage profond (*deep learning*) ? Quelle typologie peut-on proposer en ce qui concerne les différentes méthodes utilisées dans les activités de profilage ? Comment fonctionne cette technologie en comparaison avec les systèmes experts traditionnels ? Qui en sont les acteurs ? Quels sont les principaux risques techniques liés à cette nouvelle technologie et comment les aborder ?
- Ensuite, il fallait effectuer une analyse soit des multiples activités de profilage en fonction des différents secteurs (public et privé) et de leurs finalités ultimes, soit sur la base de cette typologie des risques en lien avec les activités. En ce qui concerne les risques, nous avons particulièrement mis en avant les nouvelles menaces sur la vie privée que sont les risques de manipulation, de stigmatisation et de discrimination courus non seulement par des personnes individuellement mais aussi collectivement en groupes ainsi que par le développement de la démocratie. Le rapport propose de distinguer, parmi les activités de profilage, les « systèmes de profilage à haut risque », selon différents critères.

Enfin, il s'agissait de considérer comment les dispositions de la Convention 108+ récemment adoptée devaient être interprétées et parfois complétées à la lumière des développements décrits plus haut afin de répondre aux risques énumérés. C'est pourquoi le rapport suggère notamment de nouvelles définitions, une attention à porter au rôle d'intervenants particuliers, un élargissement du devoir d'information, du droit à recevoir des explications en cas de décisions (au sens large) fondées sur le profilage et du droit au recours, de nouveaux rôles pour les autorités de contrôle de la protection des données et, de manière générale, l'obligation de procéder à des évaluations en cas de profilage à haut risque.

1 Evolutions récentes, aspects techniques, acteurs et typologie des technologies pour le profilage[1&2]

1.1 Evolution du contexte et du profilage

Depuis la recommandation de 2010, le profilage a évolué. D'une part, l'explosion de technologies comme le *deep learning* a rendu possible des analyses précédemment impossibles. D'autre part, il y a eu une prise de conscience des opportunités et des risques de ces mêmes technologies pour la société et les individus. Ce rapport commence par un court exposé de cette évolution.

1.1.1 Facteurs expliquant l'évolution technologique

Les technologies qui révolutionnent actuellement le profilage ne sont pas apparues du jour au lendemain. L'intelligence artificielle est née en 1956 à la conférence de Dartmouth, tandis que le terme « *machine learning* » (apprentissage automatique en français) a été proposé par Arthur Samuel en 1959. Les racine du *deep learning* remontent quant à elles aux années 1980. Différents facteurs expliquent que ces technologies issues de la recherche ont connu un vif essor ces dernières années.

Premièrement, la recherche elle-même a énormément progressé dans ces domaines. Des algorithmes performants ont été développés pour analyser de grandes quantités de données. Ainsi, dans l'article « *ImageNet Classification with Deep Convolutional Neural Networks* »[3] de 2012 qui a relancé l'intérêt pour le *deep learning*, Krizhevsky, Sutskever et Hinton entraînent un réseau de 650.000 neurones possédant 60 millions de paramètres à partir d'ImageNet (1 200 000 images de 1000 classes différentes). A titre de comparaison, le réseau de neurones LeNet-5[4] proposé par Lecun, Bottou, Bengio et Haffner en 1998 ne comportait que 60 000 paramètres et ne pouvait reconnaître que des chiffres. Cette prouesse technique a permis de réduire de 11% le taux d'erreur en reconnaissance d'image. Depuis, de nombreuses architectures de réseaux de neurones ont été proposées.

Deuxièmement, les données sont actuellement disponibles à une échelle sans précédent. D'une part, des collections variées d'images, de textes, de sons, de données brutes, etc. ont été rendues publiques par des entités diverses (laboratoires de recherche, entreprises privées, organismes publics, organisations internationales, etc.). On pense ainsi à ImageNet[5], mais aussi aux *open data* de villes comme Paris[6], Londres[7],

[1] Cette première partie a été écrite par le professeur Benoît Frenay, informaticien, sous le contrôle et l'assistance du Professeur Yves Poullet.

[2] Dans le cadre de ce rapport, certaines entreprises ou logos sont utilisés afin de donner au lecteur une idée précise. Il ne s'agit nullement d'un jugement sur l'importance de celles-ci ou de la qualité des produits offerts.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. *ImageNet classification with deep convolutional neural networks*. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12), F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, 1097-1105.

[4] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 86(11), 2278-2323. https://doi.org/10.1109/5.726791

[5] http://www.image-net.org/

[6] https://opendata.paris.fr

[7] https://data.london.gov.uk/

New York[8] ou encore Namur[9], aux sites comme Wikipédia[10] qui fournissent d'énormes quantités de texte, au *dataset* YouTube-8M[11] de 237 000 segments vidéos publié par Google, à *l'International Collaboration on Cancer Reporting*[12] qui vise à produire des *datasets* uniformisés pour différents types de cancer et émet des recommandations dans ce sens[13], etc. D'autre part, les mêmes entités construisent en leur sein des collections de données pour leur usage propre. Il peut s'agir d'un organisme public qui souhaite améliorer les services qu'il offre aux citoyens, d'une société privée qui souhaite améliorer ou vendre ses produits, d'un service multimédia en ligne, etc. Nous reviendrons en seconde partie sur la diversité des usages possibles du profilage et des profileurs.

Troisièmement, les technologies matérielles ou logicielles ont évolué et rendent possible l'exploitation d'algorithmes gourmands en ressources sur de grandes quantités de données. Ainsi, en 2012, Krizhevsky, Sutskever et Hinton utilisaient déjà des GPUs, des cartes graphiques développées initialement pour équiper les ordinateurs personnels (jeu vidéo, rendu graphique, etc.). Les GPUs sont massivement utilisés pour accélérer de plusieurs ordres de grandeurs les calculs en *deep learning* et permettre d'entraîner rapidement de nouveaux réseaux de neurones. Bien que le stockage de grandes quantités de données ne soit pas neuf (citons par exemple Teradata fondé en 1979), d'autres technologies sont apparues dans les années 2000 comme MapReduce (2004) et son implémentation open source Hadoop (2006). Ces technologies sont indissociables du « phénomène big data ».

Quatrièmement, pour les acteurs privés, les données sont une source inestimable d'information et constituent pour certains le socle de leur activité. Il n'est donc pas étonnant qu'ils investissent en ce sens. Certaines sociétés participent activement à la recherche en intelligence artificielle et au développement de nouvelles technologies. Des budgets considérables sont consacrés à ces activités.

En conclusion, l'évolution rapide de la dernière quinzaine d'années s'explique par la convergence de plusieurs facteurs : des algorithmes efficaces peuvent désormais exploiter de grandes quantités de données grâce à des technologies matérielles et logicielles rendues possibles par la recherche et des investissements importants d'acteurs publics (notamment les bailleurs de fonds comme l'Union européenne avec Horizon 2020) et privés. Cette évolution ouvre la porte à des innovations bénéfiques pour les individus et la collectivité, mais comporte également des risques tant au niveau individuel que collectif (voir le titre 2).

1.1.2 Perception et impact de l'évolution technologique

L'évolution de l'intelligence artificielle, du *machine learning* et du *deep learning* a eu un impact considérable sur le profilage. Celui-ci ne nécessitait pas forcément l'utilisation de ces technologies, mais il devient possible d'exploiter des données à caractère personnel bien plus efficacement.

Les progrès rapides et l'engouement pour les technologies intelligentes ont naturellement eu pour résultat qu'elles ont été appliquées ces dernières années dans de nombreux contextes. D'une part, cela rend possible de s'attaquer à des problèmes jusqu'alors difficiles à résoudre efficacement et à grande échelle (diagnostique automatique de cancer de la peau via smartphone[14], détection de fraude, conseil personnalisé sur des plateformes de vente ou multimédia, etc.). D'autre part, l'utilisation à grande échelle de l'intelligence artificielle pour le profilage a attiré l'attention sur des problèmes relevant de l'informatique, de la sociologie,

[8] https://opendata.cityofnewyork.us/
[9] https://data.namur.be
[10] https://en.wikipedia.org/wiki/Academic_studies_about_Wikipedia
[11] https://research.google.com/youtube8m/
[12] http://www.iccr-cancer.org/
[13] https://www.ncbi.nlm.nih.gov/pubmed/27735079
[14] Esteva, Andre & Kuprel, Brett & Novoa, Roberto & Ko, Justin & M Swetter, Susan & M Blau, Helen & Thrun, Sebastian. (2017). *Dermatologist-level classification of skin cancer with deep neural networks.* Nature. 542. 10.1038/nature21056.

de l'éthique, du droit, etc. : risques encourus par les employés dans leur recrutement ou leur carrière, utilisation éthique du profilage, droits et devoirs liés à l'utilisation des données, développement de nouveaux algorithmes plus transparents et plus facilement interprétables, réduction des problèmes de biais et de discrimination, amélioration de la robustesse des algorithmes face à du bruit ou des attaques, etc.

Parallèlement, les acteurs publics et privés, notamment les médias, exposent le citoyen non-initié à ces technologies via leur utilisation et l'information qu'il reçoit. Le discours est souvent polarisé. D'un côté, le citoyen entend parler de voitures autonomes, d'intelligences artificielles qui jouent au go ou au poker et d'assistants personnels qui l'aident dans sa vie quotidienne. D'un autre côté, on lui parle de dérives, de menaces pour la démocratie, de discriminations et de pertes d'emplois. Il est difficile pour le citoyen de faire la part des choses, ce qui constitue une menace pour le débat public qui ne peut plus se dérouler dans de bonnes conditions. Exposé aux discours commerciaux, le citoyen a également difficile à déterminer quelles sont les capacités et les limites réelles des systèmes vendus. Plus largement, les décideurs au sein des organismes publics et privés sont parfois dans le même flou. De plus, il n'est pas aisé de trouver des profils techniques et non-techniques formés à ces technologies. Il y a un réel problème sociétal d'éducation à l'intelligence artificielle et, en particulier, aux spécificités du profilage.

La diffusion rapide et en profondeur dans la société des technologies d'intelligence artificielle pour le profilage met donc en lumière des potentialités inédites, accompagnées de questionnements techniques et non-techniques et d'un besoin urgent d'éducation à tous les niveaux de la société.

1.1.3 Réactions face à l'évolution technologique

Face à l'évolution de l'intelligence artificielle et son impact sur le profilage, différentes réactions ont marqué ces dernières années. D'une part, des règlementations sont applicables et des rapports ont été proposés à différents niveaux, tels que le règlement GDPR et le rapport « *Ethics guidelines for trustworthy AI* » à l'Europe, le rapport « *AI for humanity* » en France ou AI4Belgium en Belgique. D'autre part, certains financements ont été créés pour soutenir le développement de la recherche, tels que le récent « *H2020 Call on European Network of Artificial Intelligence Excellence Centres* ».

A différents niveaux (universités, hautes écoles, centres de formation, etc.), des formations ont été créées pour former les travailleurs aux défis du profilage et de l'intelligence artificielle. Des réflexions sont menées pour éduquer les enfants, les adolescents et le grand public, notamment en Belgique, en Finlande, en France et aux Pays-Bas. De manière duale, de nombreux laboratoires de recherche et entreprises s'intéressent à l'impact positif que peut avoir le profilage sur l'enseignement : parcours personnalisé pour l'apprenant avec des exercices et des cours adaptés à son profil, détection et prévention du décrochage scolaire sur base des résultats et des activités de l'apprenant, etc.

Dans le monde de la recherche, les problèmes évoqués ci-avant sont pris à bras-le-corps : de nombreuses conférences scientifiques comportent des sessions consacrées aux problèmes de biais, d'« interprétabilité », de fiabilité, d'éthique, de sécurité, de préservation de l'anonymat, etc. Dans le même temps, les entreprises ont été également sensibilisées à ces problèmes et font des efforts en ce sens. Des initiatives comme la plateforme « *AI for Good* » de l'ONU encourage l'utilisation bénéfique de l'intelligence artificielle. Malgré toutes ces initiatives, il est clair que les problèmes existants sont loin d'être résolus. Il est actuellement difficile d'assurer une parfaite « interprétabilité », fiabilité ou sécurité de nombreux systèmes d'intelligence artificielle utilisés en profilage. La recherche est plus que jamais nécessaire, en particulier en informatique d'où viendront les solutions techniques. Mais une approche interdisciplinaire est indispensable afin de mieux répondre à ces défis.

En conclusion, l'évolution technologique des dernières années a suscité de nombreuses réactions dans le monde politique, juridique, de l'éducation, de la recherche et de l'entreprise. Une tendance qui s'observe ces dernières années mérite une attention particulière. De nombreux chercheurs en intelligence artificielle,

machine learning et *deep learning* quittent les laboratoires de recherche pour les entreprises privées. Si ce mouvement est normal et souhaitable, son ampleur est inédite. En intelligence artificielle, on assiste à une réelle privatisation de la recherche. Il devient difficile pour les universités de garder leurs meilleurs éléments. Une politique ambitieuse de soutien à la recherche indépendante est plus que jamais nécessaire pour que l'Europe reste leader scientifique en la matière. Répondre aux défis du profilage nécessite une recherche fondamentale et appliquée forte et vivace, qui rend compte de ses résultats, de ses progrès et accepte en retour la discussion par les pairs.

1.2 Acteurs

L'utilisation de l'intelligence artificielle, du *machine learning* et du *deep learning* pour le profilage complexifient la mise en place de nouveaux projets. Désormais, il est rare qu'un projet implique un seul acteur pour créer les algorithmes nécessaires, les adapter, les paramétrer et les utiliser. Il faut également capter, stocker et organiser les données. De plus, les composants de profilage ne sont généralement qu'une partie d'un système bien plus large dans lequel ils doivent s'intégrer. Cette section part d'un exemple typique d'utilisation du *machine learning* et le décline selon différents scénarios où un nombre variable d'acteurs de différents types interviennent.

1.2.1 Scénario de base : un seul acteur

Les techniques d'intelligence artificielle utilisées pour le profilage relèvent principalement du *machine learning* et, dans certains cas, plus spécifiquement du *deep learning*, une sous-discipline du *machine learning* particulièrement indiquée lorsque des images, des vidéos, du son ou du texte sont traités. La figure 1 montre les phases principales du *machine learning* : l'entraînement et la prédiction.

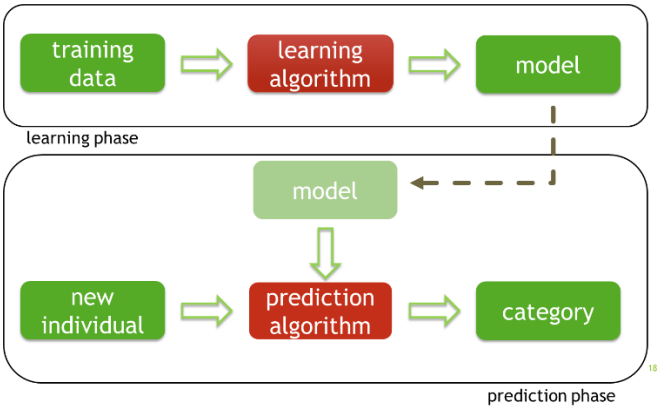


Figure 1 : entraînement (learning phase) et utilisation (prediction phase) d'un modèle de machine learning.

L'entraînement consiste à exploiter un ensemble de données pour apprendre un modèle de celles-ci. Le modèle est une abstraction mathématique qui fournit une description simplifiée des données pour résoudre la tâche à effectuer. Par exemple, une agence immobilière pourrait souhaiter prédire le prix de biens immobiliers à partir de leur surface, du nombre de chambres, de la présence de commerces à proximité, d'indicateurs de pollutions sonore, de l'âge du bien, etc. Dans ce cas, elle disposera de données relatives aux nombreux biens qu'elle a vendus par le passé, pour lesquels le prix de vente est connu : ce sont les données d'entraînement. Elle utilisera celles-ci pour trouver une formule telle que

$$\begin{aligned} prix = {} & paramètre_{surface} \times surface + paramètre_{chambres} \times chambres \\ & + paramètre_{commerces} \times commerces + paramètre_{\text{pollution}} \times \text{pollution} \\ & + pparamètre_{\text{age}} \times age \end{aligned}$$

où la valeur du poids de chaque caractéristique est inconnue *a priori*. L'entraînement va consister à chercher les meilleures valeurs de ces poids pour coller le mieux possible aux prix observés pour les biens déjà vendus. L'hypothèse est que le modèle (c'est-à-dire la formule avec les meilleures valeurs des poids) permettra d'estimer correctement le prix de nouveaux biens. Ce genre de modèle linéaire est trop simple pour prétendre expliquer parfaitement les prix du marché, mais il est probable qu'il permette une première estimation suffisamment précise pour les besoins de l'agence.

La phase de prédiction utilisera le modèle entraîné à partir des données pour faire des prédictions sur de nouveaux biens immobiliers. En pratique, un modèle peut prédire un nombre, mais également une catégorie, ce qui est plus fréquent en profilage. On peut ainsi partir de données disponibles sur un grand nombre de clients pour apprendre un modèle de rétention qui permet de prédire si un client risque de partir à la concurrence. Le modèle peut être une formule linéaire, comme au-dessus, mais il peut aussi se présenter sous la forme de règles logiques comme celui de la figure 2 ou être bien plus complexe comme le réseau de neurones Inception-v3 dont l'architecture est montrée en figure 3.

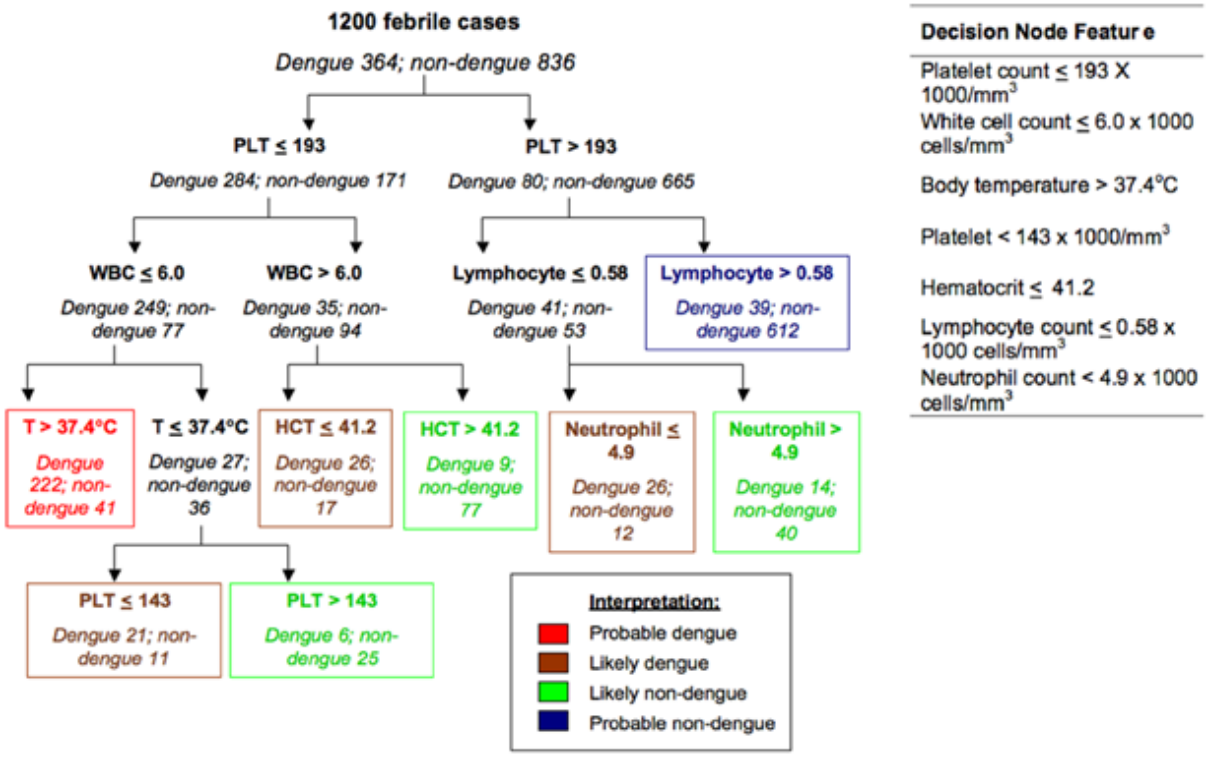


Figure 2 : exemple de modèle de catégories exprimé sous forme de règles logiques organisées en arbre de décision[15].

[15] Figure reproduite depuis Tanner, L; Schreiber, M; Low, JG; Ong, A; Tolfvenstam, T; Lai, YL; Ng, LC; Leo, YS; Thi Puong,L; Vasudevan, SG; +3 more... Simmons, CP; Hibberd, ML; Ooi, EE; (2008) *Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS neglected tropical diseases*, 2 (3). e196. ISSN 1935-2727 DOI: https://doi.org/10.1371/journal.pntd.0000196.

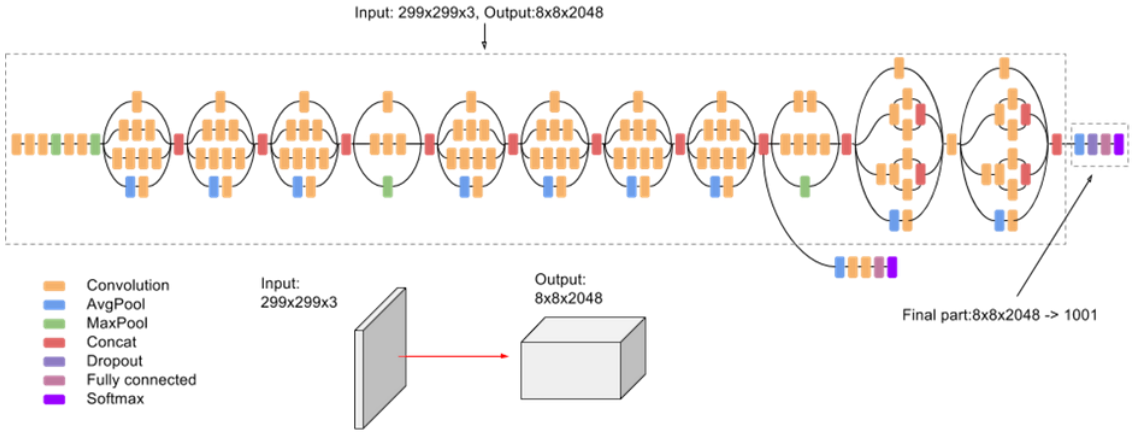


Figure 3 : exemple de réseau de neurones (deep learning) caractérisé par 23.885.392 poids (Inception-v3 de Google)[16].

Dans le scénario présenté ici, un seul acteur est impliqué : l'agence immobilière qui acquiert, stocke et gère les données, conçoit et utilise des algorithmes de *machine learning* et les déploie ensuite. En pratique, le profilage est rarement le fait d'un seul acteur et la situation est bien plus complexe.

Avant d'aller plus loin, une note sur la différence entre intelligence artificielle, *machine learning* et *deep learning*. L'intelligence artificielle est une discipline de l'informatique qui crée de nouveaux algorithmes capables de résoudre des problèmes qui requièrent normalement l'intelligence humaine. Cela recouvre des tâches variées comme la planification, les jeux, la satisfaction de contraintes, le raisonnement logique (notamment les systèmes experts) ou probabiliste, le traitement de la parole, du texte ou des images, etc. 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.

1.2.2 Utilisation de bibliothèques spécialisées

La figure 4 reprend le scénario développé ci-avant en se concentrant sur l'entraînement. Un seul acteur y est présent : celui qui fournit le service final intégrant du profilage. Même pour un spécialiste en *machine learning*, il est rare de ne pas au moins se reposer sur des bibliothèques spécialisées, comme le montre la figure 5. Une bibliothèque est un ensemble de fonctionnalités (autrement dit d'implémentations d'algorithmes) prêtes à l'emploi pour réaliser facilement de nouveaux programmes. Lorsqu'une équipe de développement utilise une bibliothèque pour un nouveau projet, elle évite ainsi de devoir réinventer la roue et de (re)développer toute une série d'algorithmes usuels. Par exemple, aucun expert n'implémente lui-même un modèle tel que les support *vector machines*, mais utilise plutôt la bibliothèque LIBSVM qui propose une implémentation efficace de ce modèle. Ce modèle a fait l'objet de milliers de publications scientifiques et n'est pas trivial à implémenter.

De nombreux algorithmes de *machine learning* sont déjà implémentés (c'est-à-dire rendus disponibles) dans des bibliothèques *open source* gratuites telles que LIBSVM, LIBLINEAR, scikit-learn, Weka, Keras, TensorFlow, PyTorch ou leurs équivalents commerciaux. Ces bibliothèques permettent un gain de temps extraordinaire et d'obtenir des résultats compétitifs. En effet, l'implémentation efficace et fiable de nombreux algorithmes de *machine learning* nécessite une importante expertise et un temps considérable, ceux-ci reposant souvent sur des dizaines d'années de recherche.

[16] Figure reproduite depuis https://cloud.google.com/tpu/docs/inception-v3-advanced.

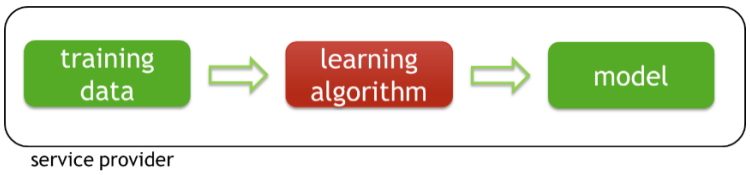


Figure 4 : phase d'entraînement avec un seul acteur.

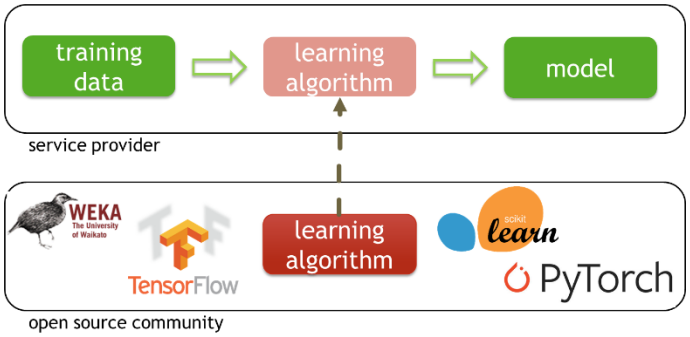


Figure 5 : phase d'entraînement reposant sur l'utilisation de bibliothèques spécialisées.

Les « bibliothèques » mentionnées ci-avant sont conçues grâce au travail considérable de larges communautés à but non-lucratif ou d'entreprises. Dans le cas des bibliothèques *open source* et gratuites, celles-ci sont disponibles sur internet et téléchargeables par n'importe quel acteur qui souhaite les utiliser. Elles sont généralement accompagnées d'un avertissement signalant à l'utilisateur qu'elles n'offrent pas de garanties (au sens légal) et doivent être utilisées avec toutes les précautions nécessaires. Il faut ici insister sur l'importance des bibliothèques *open source* gratuites en *machine learning*, utilisées très largement, sans lesquelles la plupart des développements actuels seraient impossible, notamment en *deep learning*. Malgré leur caractère non-commercial, le fait qu'elles soient développées par de grandes communautés est souvent un gage de qualité. Nos recommandations devront tenir compte des nécessités de développement de ces bibliothèques, facteur incontestable d'innovations et chercher dans l'autorégulation de la recherche scientifique et des communautés de chercheurs des garanties de protection contre les risques liés à l'utilisation de celles-ci à des fins de profilage.

1.2.3 Sous-traitance de l'apprentissage

Dans cas, l'acteur qui souhaite déployer un service de profilage ne dispose pas de l'expertise pour le faire. Il peut alors faire appel à un ou plusieurs acteurs qui vont concevoir des algorithmes adaptés à ses besoins. Il s'agit dans ce cas de déléguer la phase d'entraînement comme dans la figure 6. Dans ce cas, évidemment, les acteurs seront attentifs aux dispositions en matière de sous-traitance.

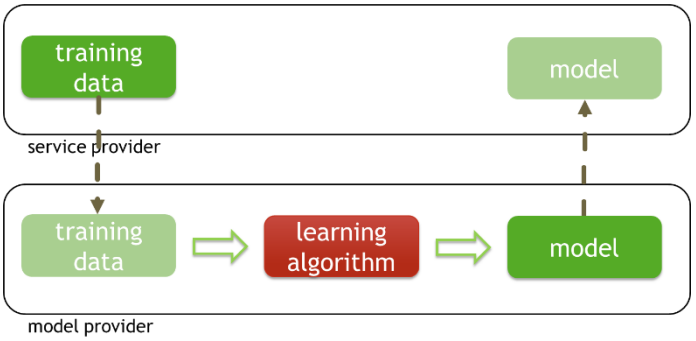


Figure 6 : délégation de la phase d'entrainement à un prestataire externe.

La délégation de la phase d'entraînement peut avoir plusieurs explications. Même lorsque l'expertise technique ne fait pas défaut, la quantité de données utilisée dans le cadre d'activité de profilage peut être telle (on parle souvent de « big data », même si le terme ne s'applique réellement qu'à quelques gros acteurs) qu'il est nécessaire de mettre en place ou d'utiliser une infrastructure spécifique, que ce soit pour le stockage de l'information ou son analyse. Cela nécessite l'utilisation de ressources matérielles et logicielles qui requièrent des compétences externes au *machine learning*. Un cas extrême (mais courant) est l'utilisation de plateformes de type « cloud » où un acteur paie pour avoir accès à d'importants moyens de stockage et de calcul pour y exécuter ses propres algorithmes. Enfin, parfois, il peut simplement être plus efficace de faire appel à un spécialiste d'un type de profilage.

1.2.4 Le cas particulier du *deep learning*

Dans le cas du *deep learning*, les problèmes liés à la conception et l'exécution des algorithmes sont encore plus importants. Pour entraîner le réseau Inception-v3 présenté dans la figure 3, les chercheurs de Google ont dû utiliser 50 GPUs. Chaque GPU peut coûter plusieurs milliers d'euros, une dépense qui n'est pas à la portée d'une petite ou moyenne entreprise. De plus, concevoir une architecture de réseau de neurones adaptée requiert une solide expérience en *deep learning*. Pour ces raisons, il est difficile pour un petit acteur de développer son propre modèle *deep learning*.

En pratique, pour les activités de profilage qui implique la reconnaissance d'images similaires à celles qu'on trouve dans les collections d'images, il n'est pas nécessaire de créer son propre modèle. En effet, certains grands acteurs ont rendu disponibles librement et gratuitement les modèles qu'ils ont entraînés sur de grandes collections d'images. Un plus petit acteur peut donc, comme dans la figure 7, le télécharger et l'utiliser immédiatement sans le coût de l'entraînement. C'est une pratique courante dans le monde de la recherche et de l'enseignement, particulièrement lorsque quelqu'un se forme au *deep learning* ou souhaite tester rapidement une idée à coût réduit. Evidemment, le prix à payer est que seules les catégories prévues à la conception du modèle seront reconnues. Ainsi, un réseau comme DeepLoc[17] entraîné sur des images de levures au microscope sera inutilisable pour reconnaître un chat d'un chien. Inversement, Inception-v3 ne pourra pas être utilisé à la place de DeepLoc puisqu'il ne reconnait que des images « naturelles » (chien, chat, arbre, etc.).

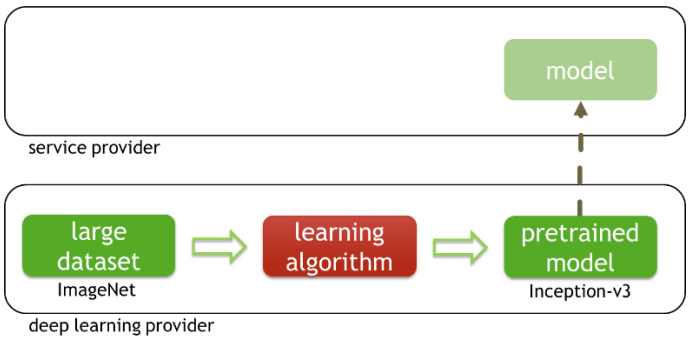


Figure 7 : utilisation d'un modèle pré-entraîné.

[17] Kraus, Oren & T Grys, Ben & Ba, Jimmy & Chong, Yolanda & J Frey, Brendan & Boone, Charles & J Andrews, Brenda. (2017). *Automated analysis of high-content microscopy data with deep learning*. Molecular Systems Biology. 13. 924. 10.15252/msb.20177551.

De nombreux réseaux libres d'accès ont été entraînés sur la même collection d'image faisant autorité dans la communauté scientifique : ImageNet. Comme le montre la figure 8, cela ajoute donc un nouvel acteur : celui qui conçoit et rend disponible la collection d'images, sans y associer de modèles.

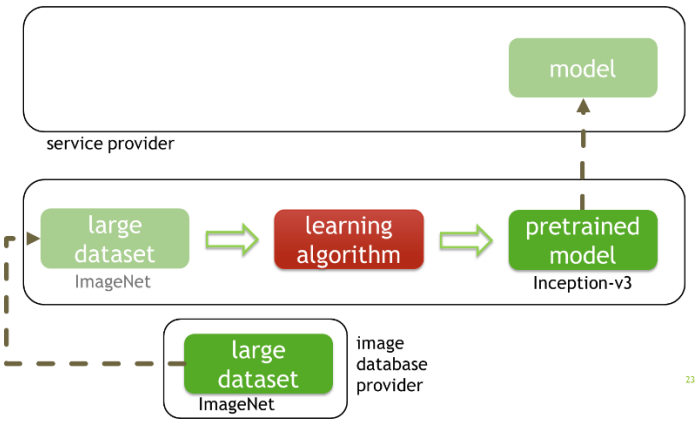


Figure 8 : ajout du fournisseur de la collection d'images.

Pour que le *machine learning* puisse fonctionner, il faut que les collections d'images contiennent des images suffisamment variées et associées à des catégories prédéfinies. Ainsi, ImageNet contient 1.200.000 images appartenant à 1000 classes différentes. 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 figure 9 montre la répartition de l'origine géographique des images : il y a un clair biais de représentativité[18] 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[19]. A nouveau, les collections d'images comme ImageNet sont rendues disponibles sans garantie (légale) et il incombe à l'utilisateur de ces images de vérifier qu'elles conviennent à l'application qu'il envisage. Ainsi, ImageNet est détaillé dans « ImageNet : A Large-Scale Hierarchical Image Database »[20] et « Construction and Analysis of a Large Scale Image Ontology »[21] et est utilisé dans de nombreuses études. Dans le domaine médical, les collections de données peuvent faire l'objet de recommandations, telles que celles de l'International Collaboration on Cancer Reportings[22].

[18] Shankar, Shreya, et al. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world." arXiv:1711.08536 (2017).

[19] Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proc. FAT in PMLR 81:77-91

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, *ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.*

[21] J. Deng, K. Li, M. Do, H. Su, L. Fei-Fei, *Construction and Analysis of a Large Scale Image Ontology. In Vision Sciences Society (VSS), 2009*

[22] Voir la liste de publications présente sur le site http://www.iccr-cancer.org

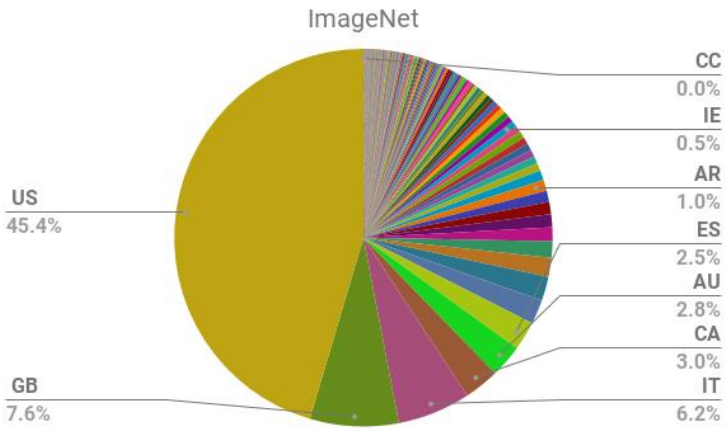


Figure 9 : répartition de l'origine géographique des images dans ImageNet[23]. Il est à noter que la mesure provient d'un échantillon éventuellement non uniforme et peut être imparfaite, bien que cela ne devrait pas avoir d'incidence sur les tendances générales.

1.2.5 La révolution des modèles pré-entraînés

Dans de nombreux cas, il n'est pas suffisant d'utiliser tel quel un modèle de *deep learning* existant. Toutefois, il est rare de disposer de la quantité d'images nécessaire à l'entraînement d'un réseau de neurones de type « *deep learning* » : une collection de quelques centaines ou de quelques milliers d'images est nettement insuffisante. Or, dans de nombreux contextes, c'est l'ordre de grandeur des collections à disposition de l'acteur qui souhaite développer un service utilisant du *deep learning*. Il faut en effet non seulement acquérir ces images, mais en plus les étiqueter manuellement une à une. Dans un contexte médical, par exemple, le processus est long et coûteux à cause de l'expertise et du matériel nécessaire. De plus, le nombre de patients disponibles pour une telle recherche sera limité.

Lorsqu'un acteur souhaite développer un modèle *deep learning* sans disposer de suffisamment d'images, une solution simple et souvent efficace existe. Il suffit de ré-entraîner, sur la petite collection d'images dont il dispose, un modèle pré-entraîné sur une bien plus grande collection d'images (voir figure 10). Par exemple, dans une étude[24] de 2017, des chercheurs de Stanford ont ré-entraîné sur 129.450 photos de lésions cutanées provenant de 18 collections d'image et réparties en 2.032 classes le réseau Inception-v3 pré-entraîné sur ImageNet. Ils ont ainsi exploité ce que Inception-v3 avait déjà appris pour résoudre un problème complexe à partir d'un nombre plus limité d'images. La littérature scientifique regorge d'autres exemples de techniques de « *transfer learning* » (on transfère à un problème ce que le réseau a appris sur un autre problème pour mieux résoudre celui-ci).

[23] Figure reproduite de Shankar, Shreya, et al. "*No classification without representation: Assessing geodiversity issues in open data sets for the developing world*." arXiv:1711.08536 (2017).

[24] https://cs.stanford.edu/people/esteva/nature/

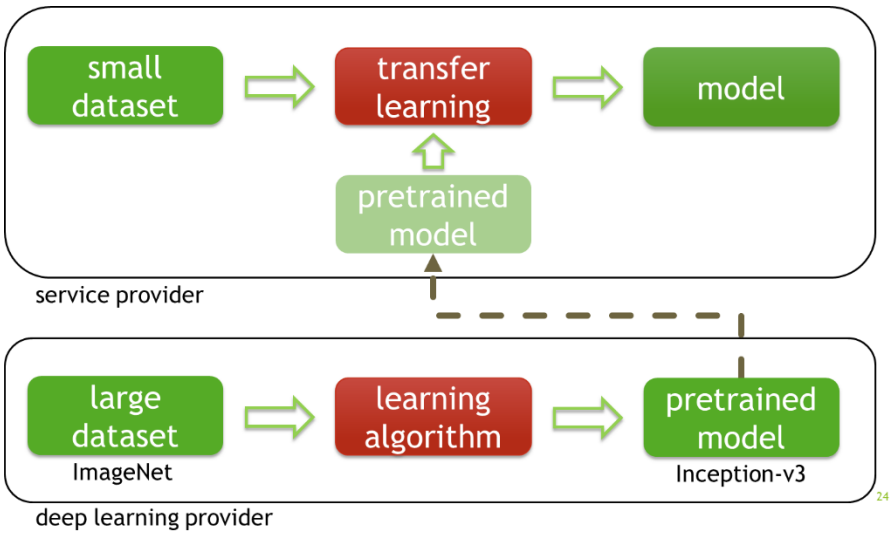


Figure 10 : ré-entraînement et utilisation d'un réseau de neurones pré-entraîné par transfer learning.

Lorsque le *transfer learning est* utilisé, les biais évoqués ci-avant peuvent provenir de plusieurs sources de données : celles utilisées pour pré-entraîner le modèle et celles utilisées pour le ré-entraîner. Toutefois, il faut souligner que le *transfer learning est* souvent incontournable en *deep learning*.

1.2.6 Complément d'information : l'ouverture pluridisciplinaire

Dans les scénarios évoqués ci-avant, le propos s'est concentré sur les aspects techniques du machine *learning*. En plus des spécialistes en intelligence artificielle, il est évident que de nombreux autres acteurs entrent en jeu : le profilage est par nature multidisciplinaire puisqu'il requiert des spécialistes en bases de données, en génie logiciel, en interfaces homme-machine, en droit, en éthique, etc. En effet, il faut concevoir les infrastructures pour stocker les données, construire avec rigueur les systèmes logiciels de profilage depuis l'analyse des besoins jusqu'à la mise en production, concevoir des interfaces adaptées à l'exploitation des résultats du profilage (par exemple pour présenter efficacement des recommandations de produits à un client), s'assurer que le système de profilage est conforme à la législation sur les données à caractère personnel (entre autres), vérifier que le système de profilage ne comporte aucun biais ou ne discrimine aucune catégorie d'individus, etc. Il est nécessaire de tenir compte de ces acteurs dans une réflexion sur le profilage car leur contribution est indispensable.

1.2.7 Conclusion

Il existe de nombreuses configurations possibles d'acteurs pour le profilage. Dans ces configurations, chaque acteur a une responsabilité et un impact différent sur le résultat final. Le cas des acteurs open source est particulier : ils n'ont pas de contrôle sur les acteurs qui utilisent les données, algorithmes et modèles qu'ils mettent à disposition. Pourtant, leur rôle est primordial dans l'innovation en profilage. Il est donc indispensable, dans une réflexion juridique sur le profilage de tenir compte des spécificités et d'attribuer à chacun une responsabilité adéquate. Dans le cas des outils et des collections de données open source, il faut noter qu'un effort de documentation est déjà entrepris et que par définition, les projets en open source sont soumis à la critique et à l'amélioration par leurs utilisateurs. La recherche en intelligence artificielle est une source d'information sur ces outils et leurs limites.

1.3 Typologie du profilage, des solutions techniques et des finalités

De nombreux types de profilages sont possibles, qui se distinguent par le genre de technologie utilisée et la finalité du profilage lui-même. Nous passons ici en revue quelques-unes de ces possibilités.

1.3.1 Types de solutions techniques

Le profilage utilise les données disponibles sur une personne pour mieux la comprendre ou inférer d'autres informations (risque de développer une maladie, comportement de consommateur, etc.). Lorsque le *machine learning* doit être utilisé pour construire un modèle à des fins de profilage, il est nécessaire de préciser la forme de profilage en termes techniques pour choisir la bonne technologie.

Lorsqu'on sait d'avance ce que le profilage est censé faire et qu'on possède des exemples de réponses correctes pour un suffisamment grand nombre de personnes, on utilisera du *machine learning* supervisé. Un **algorithme supervisé** cherche à apprendre le lien entre les données disponibles sur une personne et ce que le responsable du profilage souhaite pouvoir dire à propos de la personne. Dans le cadre d'une étude psychologique réalisée sur plusieurs volontaires, on pourrait ainsi par exemple chercher le lien entre les données collectées via des questionnaires et leur niveau de stress professionnel. Pour un tel profilage, la finalité est clairement établie. Un algorithme de *machine learning* supervisé pourra utiliser tous les questionnaires collectés et voir comment prédire au mieux le niveau de stress professionnel à partir des seules réponses données par chaque individu.

Parmi les 'problèmes' supervisés, on peut distinguer plusieurs variantes, notamment la classification, la régression ou la régression ordinale. La **classificatio**n associe à chaque individu une catégorie parmi un ensemble prédéfini de catégories possibles. Il peut s'agit de maladies, de types de clients, etc. Lorsqu'on cherche à associer un nombre à une personne, il s'agit plutôt de **régression**, par exemple pour prédire l'âge sur base d'une photographie, typiquement en utilisant du *deep learning*. Quant à elle, la **régression ordinale** cherche à prédire un niveau de satisfaction ou de préférence.

Les systèmes de « recommandations » sur base de profilage peuvent également utiliser du *machine learning*. Il s'agit ici de prédire les préférences d'une personne sur base des préférences d'autres personnes, comme c'est le cas pour des plateformes de commerce en ligne ou la publicité ciblée. Le filtrage collaboratif est une technique de recommandation qui compare les historiques de consommation des différents utilisateurs d'un service, faisant l'hypothèse qu'on pourra proposer à des personnes aux habitudes similaires des produits similaires qu'ils n'ont pas encore consommés. De nombreuses variantes existent.

Dans certains scénarios de profilage, il est parfois difficile, voire impossible de spécifier complètement la finalité à l'avance, dans le sens où la réponse attendue par le système de profilage est inconnue. C'est le cas typiquement lorsqu'on souhaite segmenter la population d'un pays, la patientèle d'un hôpital ou la clientèle d'une entreprise. Des groupes de personnes seront trouvés automatiquement par les algorithmes : l'intérêt est précisément d'apprendre de nouvelles choses sur ses citoyens, ses patients ou ses clients. On peut ainsi mieux les comprendre, développer des services plus adaptés, identifier des sous-populations à risque, etc. Ce **profilage non-supervisé** est très courant et pose la question de jusqu'où il est possible ou souhaitable de définir précisément la finalité d'un profilage. Un tel profilage a généralement pour but d'explorer les données et de dégager de nouvelles connaissances qui seront ensuite exploitées par des humains : il s'agit en fait souvent d'une étape préliminaire.

Une autre forme de profilage où le comportement attendu est difficile à prédire est la **détection d'anomalies**. Il s'agit de détecter, dans une population, des personnes qui sont « anormales », dans le sens où elles sont

significativement différentes du reste de la population. On peut ainsi détecter si une personne utilise un service d'une façon anormale (cas de fraude) ou plus simplement s'il s'agit d'une personne dont il ne faut pas tenir compte pour d'autres analyses car elle fausserait les résultats.

Il existe **des situations intermédiaires** dans le profilage. Par exemple, les algorithmes **semi-supervisés** permettent de faire du profilage supervisé, même si la réponse attendue n'est connue que pour un nombre limité de personnes. Ainsi, il est possible d'utiliser une grande population d'individus pour construire un modèle de classification, même si la catégorie correcte n'est connue que pour un faible pourcentage d'entre eux. Ce scénario est courant quand les données sont faciles à acquérir, mais la réponse (catégorie, nombre, préférence) est coûteuse et difficile à obtenir, particulièrement si elle nécessite l'intervention de spécialistes humains. C'est le cas par exemple en traitement d'images.

En conclusion, il existe de nombreux types d'algorithmes pour le profilage. Dans certains cas, ces algorithmes vont eux-mêmes aider leurs concepteurs à mieux comprendre les personnes qu'ils étudient et à définir plus précisément le profilage qu'ils souhaitent mettre en place. Il semble important de tenir compte de cette variété et du fait qu'il est particulièrement difficile de prédire quelles données à caractère personnel seront utiles dans le cas des algorithmes non-supervisé. Il faut également noter que les algorithmes susmentionnés sont capables de traiter non-seulement des informations chiffrées, mais également des images, du son, du texte, des séquences, etc.

1.3.2 Types de profilages et de finalités

Une analyse plus complète des finalités de profilage est proposée dans la partie II de ce document. Toutefois, sur base de la discussion technique ci-avant, il est intéressant de faire plusieurs remarques.

En fonction du caractère clairement défini ou exploratoire du profilage, la finalité du profilage sera plus ou moins précise. A cela s'ajoute également une distinction importante en *machine learning* : un **modèle** (et donc un profilage) **peut être à visée descriptive ou prédictive**. Un modèle descriptif décrit sous une forme compréhensible la relation (c'est-à-dire la formule mathématique ou les règles logiques) entre les données à caractère personnel d'une personne et la réponse recherchée. Il a pour but de mieux comprendre cette relation, par exemple lorsqu'un médecin cherche à comprendre comment et pourquoi une maladie apparaît chez certains patients et pas les autres. Par opposition, un modèle prédictif a pour seul but de prédire la réponse recherchée pour un individu spécifique. Il n'est donc pas strictement nécessaire que le modèle soit transparent, puisqu'il doit surtout être précis. En pratique, certains modèles peuvent convenir dans une certaine mesure pour la description et la prédiction, comme les modèles linéaires (voir l'exemple de l'agence immobilière) ou les arbres de décision (voir figure 2). Les réseaux de neurones sont un contre-exemple : leur puissance de prédiction se paie en opacité accrue. En particulier, le *deep learning* permet de traiter des images, du son ou du texte, mais sans qu'on puisse faire sens des calculs effectués. Ce problème ne vient pas du fait que les calculs sont inconnus (ils sont connus précisément, sans quoi un ordinateur serait incapable de les exécuter), mais du fait qu'ils sont bien trop complexes pour être « décortiqués » par l'esprit humain et qu'il puisse en tirer une interprétation précise. Il existe bien des outils comme les cartes de saillance (« *saliency maps* ») qui donnent une vague idée de la zone de l'image utilisée par le *deep learning* pour prendre une décision, mais elles sont insuffisantes et leur fiabilité est discutée[25].

Lors de la conception d'un système de profilage, un choix important sera le compromis entre transparence et la précision. La plupart des modèles transparents sont moins précis pour résoudre des problèmes complexes. A contrario, les modèles plus précis sont souvent difficiles à interpréter et peu transparents. **Le risque encouru par les personnes profilées semble un élément important pour déterminer le compromis**.

[25] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim. 2018. *Sanity checks for saliency maps*. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, and N. Cesa-Bianchi (Eds.). Curran Associates Inc., USA, 9525-9536.

Il est vraisemblablement moins grave de ne pas comprendre le fonctionnement d'un algorithme de recommandation musicale que de ne pas avoir d'explication sur celui d'un algorithme de refus de crédit. Mais dans certains cas, malgré un risque élevé, il peut être difficile de justifier l'utilisation d'un système de profilage moins précis, mais transparent, alors qu'il est possible de mettre en place un système bien plus précis. Ce problème se pose notamment en médecine : vaut-il mieux un modèle puissant mais opaque ou un modèle moins fiable mais explicable ?

Terminons cette partie en insistant sur une spécificité des technologies d'intelligence artificielle, de *machine learning* et de *deep learning* qui ont permis au profilage de progresser si rapidement. Ces algorithmes sont issus de la recherche scientifique, qui se fait aussi dans certains cas dans les laboratoires de grandes entreprises, qui a créé de nombreuses et d'importantes bibliothèques open source, ainsi que des collections de données (« *datasets* ») disponibles en accès public. L'utilisation à grande échelle du *machine learning* est rendue possible par ces algorithmes et ces collections qui sont documentés et évalués dans des centaines de milliers de publications scientifiques.

Les algorithmes présents dans les bibliothèques sont souvent basés sur les derniers développements scientifiques. Dans une réflexion juridique, il est important de tenir compte de cet écosystème pour le préserver.

2 Le profilage et la vie privée : de quelques considérations juridiques à quelques recommandations [26]

2.1 Enjeux et définition du profilage.

Introduction : Sous le terme 'profilage', on rassemble diverses opérations qui peuvent poursuivre nombre de finalités et présenter des degrés de risques bien différents. L'opération de profilage n'est pas nécessairement liée à l'utilisation de systèmes d'information. De tout temps, chacun de nous « profile » autrui. Nous cherchons tous à catégoriser ceux qui nous entourent à partir de leurs caractéristiques personnelles qu'elles soient pertinentes, objectives, permanentes ou non. Bref, nous utilisons les données subjectives ou objectives en notre possession pour catégoriser autrui et inférer, sans doute avec une marge d'erreur dans notre appréciation, d'autres traits ou manières d'être, qui nous sont inconnus. Le profilage est donc une opération propre à tout être humain, dans la mesure où nous cherchons tous à pouvoir nommer le réel qui nous entoure ou, en d'autres mots, à faire entrer autrui dans des catégories qui permettent de mieux le cerner et agir vis-à-vis de lui. Cependant, l'utilisation de systèmes d'information modifie les modalités et la portée du profilage pour diverses raisons.

La première tient au fait que les systèmes d'information actuels par leur interactivité et leur ubiquité permettent d'élargir - et ce de manière exponentielle – le nombre de données collectées. Si hier, les capacités de stockage et de communication des données étaient limitées. Aujourd'hui, d'une part, ces capacités sont devenues quasiment infinies comme le démontre le phénomène des mégadonnées (le *big data)* et, d'autre part, l'internet des objets et la multiplication ces services disponibles à portée d'un clic permettent la capture de moments de plus en plus triviaux de la vie quotidienne. Le détenteur de ces données pourra 'profiler' de manière de plus en plus fine et de plus en plus proche l'individu auquel ces données sont rapportées.

La seconde est l'utilisation d'algorithmes toujours plus puissants pour analyser cette quantité de données. Il y a vingt ans, les 'profileurs' s'aidaient d'algorithmes fondés sur une mise en forme du raisonnement humain. Ces systèmes dits experts permettaient de substituer (ou en tout cas d'aider) le responsable de traitement dans la mesure où ils traduisent et appliquent automatiquement les 'règles' mises en forme par des experts humains sur base de leurs expériences. Ces systèmes guidaient ainsi le raisonnement et évitaient la subjectivité et les risques de discrimination propres à tout décideur humain. A ces systèmes experts dont l'algorithme est totalement transparent, succèdent aujourd'hui des systèmes dits de '*machine learning',* voire de *deep learning*, capables de travailler sur bien plus de données que celles traitables par les experts. Ces systèmes opèrent des corrélations entre des données de plus en plus nombreuses suivant des algorithmes qui se nourrissent et s'affinent progressivement au fur et à mesure des données rencontrées. La variété et la complexité des 'modèles' suivis et développés par ces algorithmes sont telles que leur fonctionnement devient partiellement non transparent y compris pour ceux qui les ont développés et/ou les utilisent.

Avantages du profilage 'automatisé' - Ainsi, le 'profilage' numérique se distingue du profilage humain. Il présente à la fois des avantages mais également des risques pour l'individu[27]. Ces risques justifient une

[26] Ce titre a été rédigé par Yves Poullet sous le contrôle et l'assistance de Benoit Frenay.

[27] On connaît le reproche adressé au profilage du chef d'entreprise qui doit recruter un employé ou décider de sa promotion. Tant, la subjectivité du sujet humain, la mauvaise humeur, le sentiment de connivence que le manque de qualité et de quantité des données qui serviront de base à sa décision, expliquent les doutes que d'aucuns

réglementation qui permettent la confiance de ceux auxquels sont opposés voire imposés les résultats de ces traitements et la maximisation des bénéfices liés à l'utilisation de ces systèmes de décision ou d'aides à la décision.

Les **avantages** du profilage sont en effet évidents pour les **entreprises et administrations** qui utilisent de tels systèmes.

- Il s'agit pour les entreprises d'**optimise**r leur action et leurs investissements. Par exemple, le profilage permettra aux entreprises de cibler leur clientèle, de définir leurs stratégies en fonction d'une infinité de paramètres, de mieux comprendre les réactions de telle ou telle population, etc. il s'agira également de sécuriser le choix de leur implantation et d'aider au recrutement des employés voire leur avancement. Il s'agira, enfin, de détecter des possibilités de cyberattaques, de fraudes, etc.
- Pour les pouvoirs publics[28], les avantages résident d'abord dans une meilleure appréhension de la réalité et ces systèmes aident dès lors à mieux définir les stratégies d'action des pouvoirs publics que ce soit en matière de politique d'emploi, de lutte contre la criminalité ou de politique dans le domaine de l'éducation. Ensuite, les administrations y voient une manière d'appliquer de façon plus effective leurs réglementations.
- Pour la **personne concernée** profilée, les avantages sont tout aussi notoires. Prenons quelques exemples : le premier dans le domaine de la santé, le cas du patient dont l'analyse par intelligence artificielle des antécédents cliniques et des tissus tumoraux confrontés à celles de milliers d'autres patients permet en quelques secondes d'orienter le médecin vers tel ou tel type d'intervention et de prédire les risques de succès de l'opération. Un deuxième est tiré du domaine de la consommation, la possibilité pour une personne qui désire acheter la tondeuse la plus adéquate à ses besoins, de pouvoir, dans un marché où l'offre est à ce point diversifiée, être orientée progressivement, grâce aux vertus d'un système d'aide à la décision interactif, vers le ou les engins qui répondront à ces nécessités : il s'agit de permettre l'optimisation des choix du consommateur; le troisième concerne le service offert par les plateformes musicales ou de films : combien d'entre nous se félicitent de la façon dont les opérateurs de plateformes peuvent guider nos choix vers des musiques dont nous ne soupçonnions pas l'existence et que nous découvrons correspondre à nos goûts ; enfin, le quatrième, dans le domaine d'une politique de l'emploi, la possibilité pour les pouvoirs publics de définir en fonction de multiples critères relatifs tant à la population, aux besoins actuels et futurs de l'économie locale et bien d'autres données, les profils d'emploi souhaitables, les orientations d'étude à proposer, en même temps que les déficits ou trop plein de personnes formées dans les différentes disciplines. On ajoute que la mise à disposition des résultats de ces profilages intéresse également les citoyens qui pourront orienter leurs formations et leurs candidatures à l'emploi de manière adéquate.

… et les risques - Ces avantages du profilage automatisé sont à contrebalancer par des **risques** dont la gravité est à mesurer en fonction des conséquences et de l'impact des décisions de profilage confiées au numérique. Les dangers du profilage se mesure en effet à la finalité poursuivie ou découverte par celui qui le met en place. Soulignons en effet que le profilage n'est pas une finalité mais peut correspondre à une multitude de finalités : recherche médicale, ciblage de clientèle à des fins de marketing, définition de stratégies publiques, lutte contre la fraude ou prévention de la criminalité, etc. Ceci dit, de manière générale, les opérations de

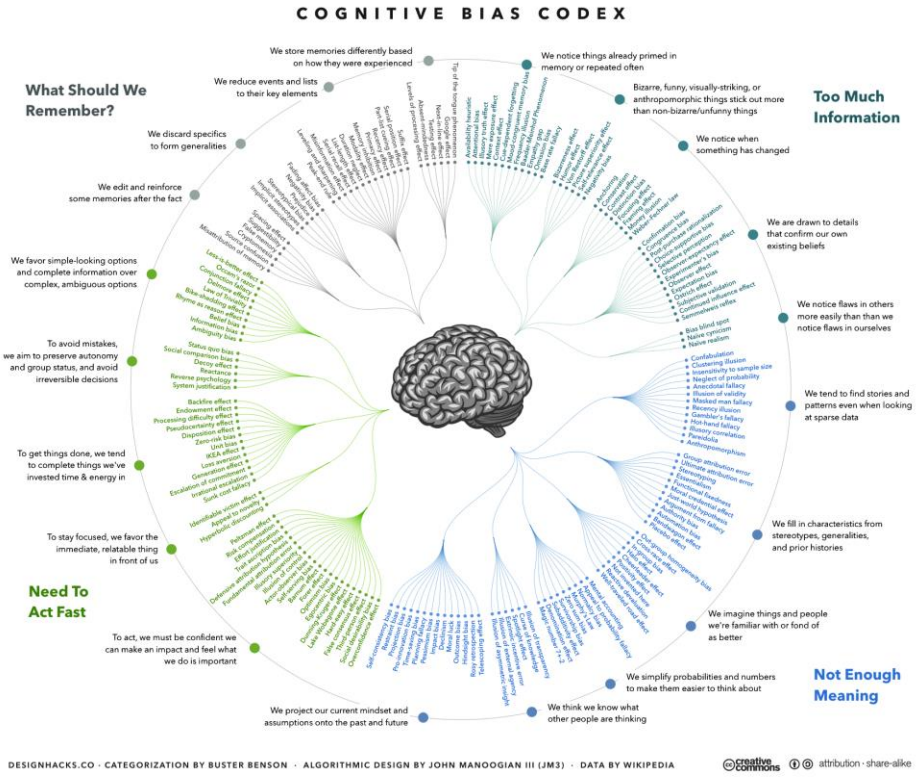
porteront à sa décision. A l'inverse, on louera le fait que la décision proposée ou fournie par l'ordinateur a travaillé sur de nombreuses données qui nous apparaissent objectives (il s'agit de spécimens d'écriture, de statistiques relatives à telle ou telle catégorie de candidats en fonction de leurs études et curricula vitae, de leurs comportements lors de l'interview analysés par des systèmes de reconnaissance faciales, …) et appliquées sans discrimination à tous les candidats. La neutralité du fonctionnement du système d'information, le volume des données traitées, son apparente objectivité expliquent les avantages de voir le jugement humain remplacé par celui numérique.

[28] Sur l'utilisation de systèmes automatisés de profilage dans 11 pays de l'Union européenne, lire le rapport: *Automating Society Taking Stock of Automated Decision-Making in the EU: A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations*, janvier 2019 disponible à l'adresse: www.algorithmwatch.org/automating-society.

profilage menées dans le cadre de l'utilisation des systèmes d'information de type *machine learning* présentent par leurs caractéristiques des risques inhérents à ces méthodes.

- Le premier est que nombre de données collectées sont traitées hors contexte. Prenons un exemple à propos de l'engagement d'un employé, le système exclut tout universitaire ayant réalisé son master en plus de 5 ans, tel candidat a réalisé ses études en 7 ans, vu un problème de santé et se trouve donc exclu !
- Le deuxième est que l'algorithme peut présenter des erreurs soit dans leur conception, soit dans les données reprises, fausses ou de mauvaise qualité, soit dans leur adéquation au problème à résoudre. Pire, les données ou les algorithmes choisis peuvent présenter certains biais[29] qui faussent les résultats ou ont pour conséquence de discriminer certaines personnes[30] ou parfois certains groupes.
- Le troisième est certes la crainte du '*Deus ex machina*', c'est-à-dire la confiance *a priori* accordée aux résultats de l'algorithme. Un jugement humain, tout subjectif qu'il soit, peut se réviser et surtout peut-être contredit par d'autres jugements humains. Sans doute, objectera-t-on, il suffit de réserver au jugement humain la note finale et donc, la possibilité de révision de la 'vérité sortie de l'ordinateur'. Nous reviendrons sur ce point mais notons que cette possibilité de révision n'est pas toujours existante et surtout que la décision 'proposée' par l'ordinateur présente une forte présomption de vérité, de par les qualités d'objectivité et de neutralité qui sont reconnues ou plutôt prêtées au fonctionnement du système d'information qui les produit et que partant le non-suivi de la 'proposition' de l'ordinateur risque d'être analysé comme une faute et, souvent, comme la preuve inacceptable de la subjectivité de celui qui s'écarte de cette 'proposition'. L'utilisation de tels systèmes, comme le note le RGPD dans ses considérants, incite, dès lors, à une véritable **déresponsabilisation** de la personne décideur. Enfin, ce mode de prise de décision[31], sans possibilité pour la personne à la fois d'être entendue et, parfois, de

29



30 Toujours à propos de l'engagement d'un employé (voir la note précédente), en l'occurrence informaticien, tenir compte du sexe des informaticiens performants est un biais dans la mesure où le nombre de femmes 'informaticiennes' expliquent facilement le peu de personnes présentant les performances attendues.

31 Comme le révèlent les travaux préparatoires du RGPD, le législateur européen en est venu à s'inquiéter d'une telle automatisation tant elle diminue le rôle joué par les personnes dans les processus de décision : « Cette disposition vise à protéger l'intérêt de la personne concernée à participer à la prise de décisions qui sont importantes pour elle. L'utilisation de profils de données exhaustifs de personnes par de puissantes institutions publiques et privées prive

comprendre les raisons de la décision prise vis à vis d'elles, peut parfois être considérée comme une atteinte à la dignité de la personne, réduite au rang de simple objet d'un calcul.

Une première réflexion sur le rôle du Droit - On conçoit dès lors les points sur lesquels porte l'intervention réglementaire de manière à corriger les risques d'un profilage numérique. Il s'agit d'abord de reconnaître le traitement de profilage comme tel, de le définir. Il s'agit ensuite, lors de la conception de systèmes automatisés de profilage, d'exiger une correcte appréciation des risques pour les personnes concernées, risques liés à ce traitement en les comparant (mise en balance) aux avantages que ces systèmes peuvent présenter tant à la personne concernée, qu'au responsable du traitement. Cette opération de balance peut dans certains cas nécessiter une véritable discussion multidisciplinaire et ouverte à une prise en compte des différents intérêts en cause. Il s'agit, enfin, de réintroduire une possibilité pour les personnes concernées de discuter la 'vérité sortie des ordinateurs'. Cette réglementation suggère dès lors que l'on distingue différents types de profilage tant les risques diffèrent suivant les finalités de ces opérations. Nous y reviendrons après avoir dit un mot des définitions présentes dans divers textes européens.

Des définitions - Le RGPD définit le profilage à l'article 4 (4) comme suit : « «profilage», « 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, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique. » La Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation des données reprend la même définition en son article 3(4).

Cette définition ne s'écarte pas de celle proposée dans la recommandation CM Rec (2010) 13. Notons cependant que cette dernière distinguait la notion de profil, résultat de l'algorithme susceptible de servir à de multiples finalités et de s'appliquer à de nombreuses personnes. Par exemple, la 'catégorie' 'personne suspecte de fraude fiscale' reprend les caractéristiques abstraites que doivent présenter les personnes, susceptibles d'avoir commis tel crime) et celle de profilage qui désigne dans le cadre d'une application poursuivant une finalité déterminée, l'application du profil. Ainsi, « le terme « profil » désigne un ensemble de données qui caractérise une catégorie d'individus et qui est destiné à être appliqué à un individu ». Le terme 'profilage' dans la recommandation du Conseil de l'Europe s'appuyait sur le terme 'profil', « *le* profilage » est une technique de traitement automatisé des données, qui consiste à appliquer un « profil » à une personne physique, notamment afin de prendre des décisions à son sujet ou d'analyser ou de prévoir ses préférences, comportements et attitudes personnels ». La notion de profil doit - elle être maintenue alors qu'elle est évacuée par les textes européens plus récents ?

Le terme 'profil' avait et garde tout son sens dans des systèmes qui distinguent les opérations de création de profils, de celles qui les appliquent, ainsi dans le cas d'une définition d'un profil d'employé idéal ou d'un criminel potentiel. Elle n'est cependant plus de mise lorsque le fonctionnement de l'algorithme ne permet plus de distinguer ces deux temps dans la mesure où il aboutit directement à une 'action' par rapport à un individu (par exemple, l'envoi d'une publicité *one-to-one* dans le cadre de la vente d'un produit ou d'un service). En d'autres termes, la mise en évidence du 'profil' permet une transparence des critères qui sont appliqués dans un second temps par l'opération de profilage. Par ailleurs, elle permet de mettre en évidence des discriminations, qui ne sont plus individuelles mais de groupe, dans la mesure où c'est l'ensemble des personnes répondant à telle catégorie qui risquent d'être soumises à telle évaluation négative. Ainsi, on peut imaginer que tel quartier présentant un profil de quartier dangereux soit systématiquement stigmatisé lors

la personne de la capacité d'influencer les processus décisionnels au sein de ces institutions, si les décisions sont prises sur la seule base de son "ombre de données »

d'opérations de recherche d'un criminel ou que présentant un profil de quartier d'illettrés, les résidents de ces quartiers soient automatiquement affectés d'un coefficient négatif lors de l'évaluation de candidats à l'emploi.

Notre analyse du fonctionnement des systèmes d'intelligence artificielle nous amène à préférer à propos de tels systèmes le terme de 'modèle' à celui de 'profil'. Le **modèle** est une abstraction mathématique qui fournit une description simplifiée des données pour résoudre la tâche à effectuer, c'est-à-dire la formule avec les meilleures valeurs des paramètres de solution. Le modèle n'est pas figé, il évolue au gré des données auxquelles il est confronté. Il rend compte du fait que le profilage qui utilise des méthodes d'apprentissage automatisé ne fonctionne pas sur le mode causaliste qui déduit ou prétend déduire d'une règle formulée par les experts, l'appartenance de telle ou telle personne à telle catégorie mais bien sur le mode de la corrélation purement statistique et évolutive entre des données. Dans le cas de l'intelligence artificielle, on n'est pas dans l'ordre de l'explication causale mais de la constatation purement statistique.

Les finalités génériques du profilage : détecter et prévoir - Mais au-delà, que retenir de ces définitions ? Les définitions distinguent deux finalités génériques du profilage : l'analyse et la prévention. Il s'agit tantôt de décrire afin de comprendre le passé, tantôt de prédire le comportement futur d'une personne. Ces deux finalités ne sont pas exclusives l'une de l'autre. La distinction est évidente par exemple, lorsqu'on pense aux traitements policiers : tout autre apparaît le profilage où à partir d'un fait passé, par hypothèse, un crime, on cherche un coupable potentiel à partir d'éléments (le lieu du crime, la présence de tel ou tel indice, les modalités du crime, etc.), il s'agira alors d'un profilage *réactif* qui en fonction de toute une série de critères y compris l'analyse de précédents plus ou moins semblables, le profil du criminel peut être établi. Par contre, s'il s'agit d'analyser les risques de récidive d'un prisonnier ou d'anticiper une attaque terroriste et ses éventuels auteurs, une analyse prospective *proactive* débouchera sur la prédiction du comportement de personnes identifiées ou identifiables. La distinction entre les deux types de profilage est loin d'être tranchée. Il est clair que l'analyse du passé d'une personne peut conduire à imaginer les comportements futurs de cette dernière.

Profilage au-delà des données à caractère personnel - La définition du profilage dans les textes de l'Union européenne envisage des traitements dans la seule mesure où ils portent sur des données à caractère personnel et les obligations du responsable du traitement n'envisage que ces données. On note que les Lignes directrices du Comité de la Convention 108 en matière de **mégadonnées** élargit sa préoccupation. Elles parlent d'opérations portant sur des données qu'elles soient à caractère personnel ou non. Ce point est important dans la mesure où la plupart des mégadonnées, sur lesquelles fonctionnent des systèmes d'intelligence artificielle de profilage, rassemblent tant des données anonymes que des données à caractère personnel. On pense ainsi aux données statistiques (par exemple, dans une base de données utilisées par les autorités policières, on reprendra des statistiques relatives aux divers types de criminalités par secteur urbain). Par ailleurs, cette distinction entre données anonymes et données à caractère personnel a-t-elle encore un sens au moment où il est parfois possible de 'désanonymiser' des données qualifiées d'anonymes ? En outre, ces données « anonymes » sont importantes dans la plupart des opérations de profilage et la limitation des obligations du responsable du traitement, par exemple l'obligation d'information, aux seules données à caractère personnel crée, nous semble-t-il, un risque de vue incomplète du fonctionnement du traitement de profilage. Nous reviendrons sur ce souci majeur.

La recommandation de 2010 : un texte d'avant-garde – La recommandation du Conseil de l'Europe date de 2010, soit un peu moins d'une dizaine d'années. Son contenu a été, à l'époque, salué comme un texte d'avant-garde. La recommandation explique l'importance donnée aux traitements de profilage par le Règlement général européen de protection des données (RGPD) et il est à noter, en particulier, la reprise par ce Règlement, ainsi l'obligation du responsable du traitement d'informer la personne concernée de la 'logique sous-jacente', selon les termes de la Recommandation ou la nécessité d'un ***Risk assessment***.

Pourquoi dès lors, réfléchir à une nouvelle recommandation. Divers arguments peuvent être proposés. **L'intelligence ambiante (IoT) et artificielle (IA)**, qui constituent les outils du profilage actuel et plus encore futur étaient encore peu utilisés il y a dix ans par les 'profileurs'. Ces deux innovations ont été jugées disruptives dans la mesure où elles modifient profondément notre environnement et notre rapport à celui-ci, créant des risques nouveaux non seulement pour l'individu mais également pour le fonctionnement de notre société en tant que telle. Le Comité de la Convention 108 a pris la mesure de ces risques par l'élaboration de deux textes clefs. Ainsi, le phénomène des 'mégadonnées' a justifié depuis des lignes directrices du Comité, de même que celui de l'IA. Il est donc important que nous tenions compte de ces risques nouveaux (I) mais également des idées nouvelles apportées par les lignes directrices que nous venons de citer (II).

2.2 Les risques encourus par nos libertés individuelles et les autres risques, liés au profilage dans le contexte des innovations disruptives de l'IA et de l'IoT.

La notion de 'risque' est centrale dans la Convention du Conseil de l'Europe et dans le Rapport qui le suit (voir par exemple et en particulier, le point 90). Encore faut-il savoir de quels risques on parle ? Notre propos distinguera d'abord ceux relatifs aux dangers encourus par nos libertés individuelles pour ensuite aborder d'autres risques dits collectifs, encourus cette fois par les notre société et son fonctionnement démocratiques, risques soulignés par les deux séries de lignes directrices récentes du Comité de la Convention 108.

2.2.1 Les risques encourus par nos libertés individuelles

Libertés individuelles – au-delà de la protection des données - Notons en préambule que la notion de 'libertés individuelles' ne s'entend pas uniquement du droit à la protection des données mais devrait viser l'ensemble des libertés individuelles susceptibles d'être mises en cause par le profilage. Sur ce point, tant le Conseil de l'Europe que l'Union européenne s'accordent: « Comme indiqué dans la déclaration du groupe de travail "Article 29" sur le rôle d'une approche fondée sur les risques dans les cadres juridiques de protection des données, la référence aux "droits et libertés" des personnes concernées concerne principalement les droits à la protection des données et à la vie privée, mais peut également impliquer d'autres droits fondamentaux tels que la liberté d'expression, la liberté de pensée, la liberté de mouvement, l'interdiction de discrimination, le droit à la liberté, de conscience et de 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*)".

Des risques pour nos libertés individuelles - Sans prétendre être complet, relevons les risques suivants pour nos libertés individuelles :

- **Le risque de réductionnisme –**. La collecte des données s'avère de plus en plus aisée et la technologie ubiquitaire en multiplie les sources, les coûts de stockage se sont effondrés et les capacités de transmission facilitent l'accès aux réservoirs de données ainsi constitués voire leur concentration. Tout cela explique la multiplication des *big data* et la possibilité d'utilisation par les systèmes d'intelligence artificielle. Au sein de ces entrepôts de données, les individus sont approchés non plus en tant que personnes mais à travers, d'une part, l'agrégation d'un certain nombre de données les concernant, considérées d'autant plus comme 'vérité' de chacun de nous qu'elles représentent des instantanés de vie (ma présence à tel endroit, mon *surfing*, mes heures et mes 'objets' d'écoute, mon achat de tel produit, ma consommation d'énergie) et, d'autre part, la mise en corrélation de telles données avec des données de même type collectées auprès d'autres individus ou des données anonymes propres à des

entités ou groupes auxquels j'appartiens. A la vérité des données saisies sur le vif, s'ajoute celle que confère la statistique. Comment nier que telle personne puisse être fraudeur, le candidat idéal ou la personne intéressée par telle publicité ou tel parti politique puisque son 'profil' ou plutôt le 'modèle' démontre que 95% des personnes ayant le même profil révèlent cette même capacité ou ce même choix. Bref, l'individu, dans cette alchimie des algorithmes, devient un agrégat de données décontextualisées et s'y réduit[32].

Cette réalité conduit à deux constatations : la première est que les systèmes d'information qui nous profilent et, le cas échéant, décident[33] de telle ou telle action vis-à-vis de nous, ne nous perçoivent et jugent nos personnalités qu'à travers de tels données et des corrélations pas toujours maîtrisées. Nous sommes reconnus à travers des « profils » ou des « modèles » créés en fonction de nos systèmes experts et d'intelligence 'artificielle' et en vue de finalités définies par ceux qui utilisent ces données, sans doute, au mépris du respect dû à la dignité humaine La seconde préoccupation concerne le fait que l'automatisation galopante des processus de décision engendre une acceptation quasi-automatique de la validité et de la pertinence de ces décisions et, corrélativement, un désinvestissement et une déresponsabilisation de décideurs « humains ».

- **Le risque de décontextualisation** - Le respect des « contextes », c'est-à-dire des zones de confiance dans lesquelles une donnée à caractère personnel est transmise par la personne concernée est fondamental dans nos sociétés. L'utilisation pendant mes occupations diverses des mêmes plateformes (par exemple, le même réseau social, le même outil de recherches, …), la présence de certains opérateurs de plateformes dans des activités diverses à travers des filiales en particulier, les GAFAM. On notera que ces acteurs opèrent non point en concurrence mais sur des marchés complémentaires même si tous vivent de l'exploitation des données qu'ils collectent à travers leurs plateformes et les services associés[34], en

[32] Sur le danger en particulier, de réduire la personne humaine aux seules données génétiques, lire l'article 3. 2 de la Déclaration universelle sur le génome humain et les droits de l'homme, de l'UNESCO (11 novembre 1997) : « Cette dignité impose de ne pas réduire les individus à leurs caractéristiques génétiques et de respecter le caractère unique de chacun et leur diversité. »

[33] Il va de soi que le système ne décide pas en tant que tel mais est l'agent auquel des humains (une entreprise, un organisme public) délègue par choix une telle compétence.

34

Puissance économique des GAFAM (Source : Wikipedia, v° GAFAM, 2018)

Société	Création	Produits phares	Source de revenu principale (en2017)[15]	Utilisateurs[16](Milliards)	Capitalisation boursière[17] (Milliards USD en mars 2018)	Acquisitions notoires
Google(Alphabet)	1998	Moteur de recherche, Régie publicitaire, Intelligence artificielle	Publicité (86 %)	1,42	719	reCAPTCHA, Waze, DoubleClick, YouTube, Android
Apple	1976	Ordinateur personnel	Matériel (81 %)	0,85	851[Note 1]	Beats Electronics
Facebook	2005	Réseau social, Publicité, Intelligence artificielle	Publicité (98 %)	2,13	464	Instagram, WhatsApp, Oculus
Amazon	1994	Commerce en ligne, informatique en nuage	Vente en ligne (82 %)	0,244	701	Whole Foods Market

leur offrant ainsi la possibilité de croisement de données émises dans des contextes différents abolissent les frontières que l'individu souhaite ou pourrait souhaiter mettre entre les différents 'domaines de sa vie'.

- **Le risque de stigmatisation -** La mémoire de l'ordinateur est ou du moins apparaît sans limite, bien au-delà de la mémoire humaine. Elle conserve une trace de notre passé ou de certains événements de notre passé que parfois nous souhaiterions voir oublier. Cette mémoire de l'ordinateur utile dans les traitements de profilage risque de stigmatiser pour sa vie entière l'individu. Ainsi, l'assureur pourrait garder indéfiniment la trace d'une maladie d'un client ; la banque de l'émission d'une traite sans provision, une association d'employeur, du vol commis par un candidat à l'emploi ou le renvoi ou l'échec d'un établissement scolaire. Considérer la capacité de la personne d'évoluer et ne pas le figer par la considération de son passé, n'est-il pas une exigence éthique ?

- **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[35] par l'individu sur un réseau social ou sur des sites de services[36], 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 qui garantissait l'inviolabilité du domicile à l'inverse des lieux publics est désormais abolie. La protection du domicile physique, lieu inviolable, apparaissait traditionnellement et, aux yeux du droit, comme quelque chose de fondamental pour la construction de la personnalité de l'individu. La notion de domicile, lieu de liberté à l'abri de tout regard d'autrui, se trouve, elle aussi, bouleversée par les développements technologiques qui conduisent à l'abolition de la distinction entre **sphère publique et sphère privée**.
- **Le risque de 'normalisation' lié à l'opacité des systèmes d'information –** A la transparence des individus, s'oppose souvent l'opacité des systèmes d'information. Cette opacité est d'abord celle du fonctionnement tant des **terminaux** (notamment, les cookies, les RFID présents dans l'Internet des objets) que des **infrastructures** (voir les « agents distribués » localisés tout au long de systèmes d'information comme ceux dits d'intelligence ambiante). Cette opacité entraîne la crainte de traitements non sollicités, non voulus et la volonté dès lors de se conformer à un comportement qui est celui que nous pensons être attendu par ces nouveaux « lieux » invisibles de surveillance. Les dangers de l'opacité de nos sociétés de l'information comme menace pour nos sociétés de l'information, où les citoyens ne peuvent connaître, de manière exacte, le fonctionnement des systèmes d'information, les données collectées, les lieux de traitement, les finalités poursuivies par ceux qui traitent les données, sont mis en

Microsoft	1975	Système d'exploitation, informatique en nuage	Logiciels (62 %)	1	703	Hotmail, Nokia, Skype, LinkedIn, GitHub

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

[36] 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 ».

évidence dès 1983 par le fameux jugement constitutionnel dans l'affaire du recensement (*Bundesverfassungsgerichtshof* 15 Décembre 1983, *EuGRZ,* 1983, p.171 et s.). La tentation des citoyens, face à cette opacité, est d'adopter le comportement, qu'ils croient attendu par la société et de ne point oser s'exprimer librement, ce qui est dommageable pour le fonctionnement de nos démocraties. On ajoute que notre vie sur les réseaux est médiatisée par le fonctionnement d'outils qui, d'une manière ou d'une autre, formate notre connaissance et approche du réel, nos actions et interactions avec autrui. Ainsi, nos engins de recherche nous propose - fort heureusement sans doute encore faut-il que nous en soyons conscients – un classement des sites en réponse à nos recherches et Facebook décide des informations prioritaires que nous recevons.

En bref, les systèmes d'IA contribuent à fixer insidieusement la 'norme' de nos comportements[37] non en nous les imposant mais, de manière plus subtile, en vous les proposant comme une évidence qui vous rend la vie facile : *« il suffit de cliquer* ». Ces systèmes opèrent à la manière de ce que d'aucuns qualifient de « capitalisme libertarien ». Dans ces systèmes, la norme n'est pas impérative mais son respect est suggéré; elle n'opère pas de manière transparente mais bien de façon masquée comme un conseil présenté comme répondant à vos besoins[38] et dont vous ne connaissez pas le mécanisme de production, si ce n'est qu'elle est induite du fonctionnement des systèmes que vous 'consentez' à utiliser… et d'autres données, celles d'autrui mais également des données inconnues jugées pertinentes par le concepteur et en tout cas par le système.

- **Le risque de manipulation -** L'opacité du fonctionnement présente une autre conséquence : le risque de **manipulation**. Cette possibilité de manipulation existe d'autant plus que l'intelligence artificielle permet ce que notre collègue A. ROUVROY (2014) appelle la « gouvernementalité algorithmique ». Comme nous l'avons déjà souligné, les profils créés constituent des outils non seulement d'analyse du passé mais du fait de la 'vérité' que ces profils prétendent refléter, une vérité, certes, purement statistique et non exempte de biais. Il est donc intéressant d'utiliser ces profils comme un instrument de prévision de nos comportements futurs[39].

Une première manifestation de cette manipulation réside certainement dans ce qu'il est convenu d'appeler les *nudges*[40] *:* les systèmes vous proposent à vous conducteur, la meilleure route à suivre ; à vous chercheur, la façon dont votre indice H pourra évoluer ; à vous responsable d'une commune, les zones d'insécurité ou d'abandon, où votre police doit intervenir ; à vous ministre de l'éducation ou enseignant, les critères selon

---

[37] « Ils séduisent par leur attrait. Les citoyens-consommateurs individuels participent volontairement et activement aux processus de modulation, en recherchant les avantages que peut apporter une personnalisation accrue. Pour les consommateurs privilégiés, il peut s'agir de réductions de prix, de produits et de services améliorés, d'un accès plus pratique aux ressources et d'un statut social plus élevé. Au sein des assemblages de surveillants, les modèles de flux d'informations sont accompagnés de discours sur les raisons pour lesquelles ces modèles sont naturels et bénéfiques, et ces discours favorisent une large internalisation des nouvelles normes de flux d'informations. Pour toutes ces raisons, une critique de la surveillance en tant qu'atteinte à la vie privée "ne rend pas justice au caractère productif de la surveillance des consommateurs". La modulation est un mode d'atteinte à la vie privée, mais c'est aussi un mode de production de connaissances conçu pour produire un mode de connaissance particulier et un mode de gouvernance conçu pour produire un type de sujet particulier. Son but est de produire des citoyens-consommateurs tractables et prévisibles dont les modes d'autodétermination préférés suivent des trajectoires prévisibles et génératrices de profits. Pourtant, parler de processus de surveillance et de modulation en réseau dans le langage courant de l'ère industrielle, comme systèmes de "consentement de fabrication" serait trop grossier ». (J.Cohen, "What Privacy is For?", 126 *Harvard Law Journal,* 2013, (draft du 11 mai 2012, p. 12))

[38] Ainsi, les algorithmes de 'recommandation' utilisés par exemple par Facebook en ce qui concerne les faits d'actualité dont la lecture est à privilégier par l'internaute.

[39] 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 : « Il deviendra très difficile pour les gens de voir ou de consommer quelque chose qui n'a pas été, dans un certain sens, adapté pour eux. »

[40] **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 ».«

lesquels *a priori,* les enfants ont des chances de réussir leur parcours scolaire ; à vous juges, les risques de récidive d'une personne auteur d'une infraction ou la décision la plus conforme au droit ou plutôt ce qui a déjà été jugé comme conforme au droit ; à vous lecteur, les ouvrages qui doivent correspondre à vos goûts.

L'ordinateur vous adresse à vous, consommateur, la 'publicité' ciblée du produit ou service, qui est supposé répondre le plus étroitement à vos goûts. Cette manipulation est-elle répréhensible ? Certes, non. Le commerçant a pratiqué de tout temps le '*bonus dolus',* sans que cette pratique ne soit considérée comme répréhensible. Elle peut apparaître même comme un bénéfice pour le 'client' futur dans la mesure où elle ajoute à son information, lui fait découvrir de nouveaux produits ou services voire répond à sa demande d'être guidé dans un marché de plus en plus complexe et offrant des produits de plus en plus diversifiés. La manipulation n'est répréhensible que si elle représente un '**abus de circonstances'** pour reprendre l'expression du projet de loi belge[41] : elle définit la notion comme suit : « déséquilibre manifeste entre les prestations par suite de l'abus par l'une des parties des circonstances liées à la position de faiblesse de l'autre partie »[42]. Cette manipulation est par ailleurs punissable, si elle constitue un « **abus de la faiblesse d'autrui** », selon l'expression de la loi pénale belge du 26 novembre[43]. La nécessité de prise en compte de la vulnérabilité de chacun dans notre société moderne exige cette extension. Même si le risque de manipulation 'abusive' est plus grand lorsqu'il s'agit de mineurs, de personnes âgées ou d'handicapés, la nécessité de prise en compte de la vulnérabilité de chacun dans notre société moderne exige cette extension nonobstant son caractère flou. Quant à savoir dans quelles hypothèses ce risque de manipulation sera retenu dépendra, on le pressent, d'un *Risk assessment*' qui balancera les intérêts de chaque protagoniste.

La manipulation peut encore avoir une toute autre portée lorsque le système de profilage est utilisé à des fins politiques, de manière à adresser le 'bon message' à l'interlocuteur qui achèvera de la convaincre de voter dans tel ou tel sens. Le scandale 'Cambridge Analytica' témoigne de cet usage possible. Ici, le risque n'est pas tant individuel que collectif dans la mesure où il touche à notre conception de la démocratie.

- **Le risque de déshumanisation –** L'autonomie de raisonnement et décisionnelle que l'homme pourrait confier à la machine peut conduire à substituer à un raisonnement humain fait de dialogue et d'attention à l'autre, un mécanisme automatisé et engendre une préoccupation éthique majeure. Comme le révèlent les travaux préparatoires, le législateur européen en est venu en effet à s'inquiéter d'une telle automatisation tant elle diminue le rôle joué par les personnes et, au-delà, par l'humain. Il s'agissait de balancer, d'une part les risques d'une décision humaine avec tout ce que cela comporte certes de subjectivité, d'empathie et de mauvaise appréciation dans les processus de décision, mais également les avantages d'une telle décision avec ses possibilités de correction, de dialogue et de motivation. Une autre préoccupation concerne le fait que l'automatisation galopante des processus de décision engendre une acceptation quasi-automatique de la validité et de la pertinence de ces décisions et, corrélativement, un désinvestissement et une déresponsabilisation de décideurs « humains »[44].

[41] La réforme actuelle menée en Belgique en matière de droit des contrats consacre ce concept, en son article 5.41 du projet de Code des obligations.

[42] Cf. sur cette disposition, son origine et ses commentaires, les réflexions de H. Jacquemin, « Protection du consommateur et numérique en droits européen et belge », in *Vulnérabilités et droits dans l'environnement numérique*, Actes du colloque tenu à Namur le 14 octobre 2018, ouvrage sous la coordination de H. Jacquemin et M. Nihoul, Larcier, Collection de la faculté de droit de Namur, p. 241 et s. ; dans le même ouvrage, lire également F. George et J.B. Hubin, « La protection de la personne en droit des obligations », p. 67 et s.

[43] La loi du 26 novembre 2011 introduit, dans notre Code pénal, la notion d'abus de la situation de faiblesse d'autrui. Au recours adressé par certains contre le caractère flou de cette législation peu respectueuse à leurs yeux du principe de 'prévisibilité' de la loi pénale, la Cour constitutionnelle, le 7 novembre 2013, justifie l'extension de la manière suivante : « dans une société démocratique, la protection des personnes en situation de faiblesse constitue une condition essentielle pour protéger les droits fondamentaux de chacun ».

[44] A cet égard, les experts IA de la Commission relèvent que « les résultats produits par la machine, utilisant des logiciels de plus en plus sophistiqués, voire un système expert, ont un caractère apparemment objectif et incontestable auquel un décideur humain peut accorder trop de poids, renonçant ainsi à ses propres responsabilités ».

- **Le risque de discrimination -** Enfin, la possibilité que le raisonnement automatisé soit entaché de ce qu'il est convenu d'appeler des biais a été longuement 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, 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.

2.2.2 Les risques dits collectifs et ceux sociétaux

Les risques individuels sont-ils les seuls risques ? On note que nombre des risques abordés ci-dessus (risques de discrimination, risques de stigmatisation, etc.) mettent en cause non seulement les **libertés de chaque individu pris en tant que tel mais au-delà des groupes** ethniques, philosophiques, à revenu modeste, de résidents d'un quartier, etc. Ces traitements affectent d'autres valeurs, en particulier **la justice sociale ou la diversité culturelle** entre individus ou entre groupes[45] 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[46] 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 l'opinion politique des citoyens. L'individualisation de l'offre de services, 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.

Autre réflexion : les mégadonnées nous entraînent nécessairement dans une aventure solidaire. Lorsque mes données de *surfing* sont collectées par mon moteur de recherche préféré et englouties dans une vaste base de données où elles se retrouvent avec les données de millions d'autres données relatives aux choix d'autres internautes, il est clair que le modèle qui, à un moment donné, prendra telle décision sur moi, le fera en fonction de l'ensemble des données collectées et de la différenciation que les recherches algorithmiques induiront entre mes données 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 sur les choix de ses proches.

Prenons l'exemple des assurances '*one to one'* qu'il s'agit d'assurances soins de santé ou de responsabilité civile, l'individualisation des primes au plus près des 'risques' que peut représenter chaque personne, risques calculés en fonction de leur profil, soumet à dure épreuve le sacro-saint principe de la mutualisation des risques pilier de notre système d'assurance. Il est assez remarquable qu'on déborde ainsi les questions traditionnelles de protection des données ou de vie privée, au sens étroit du terme. 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

[45] 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. »

[46] 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.

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. Cette nécessité d'élargir les préoccupations de la Convention 108+ est soulignée à juste titre dans les Lignes directrices récentes déjà citées, celles à propos des mégadonnées et de l'intelligence artificielle. On ajoute que cet élargissement des risques à prendre en compte ne se retrouve pas dans les textes de l'Union Européenne mais qu'elle est également présente dans les recommandations de l'OCDE qui viennent d'être adoptées en mai 2019 : « Les acteurs de l'IA devraient respecter l'état de droit, les droits de l'homme et les valeurs démocratiques tout au long du cycle de vie des systèmes d'IA. Ces droits et valeurs comprennent la liberté, la dignité et l'autonomie, la protection de la vie privée et des données, la non-discrimination et l'égalité, la diversité, l'équité, la justice sociale, ainsi que les droits des travailleurs reconnus à l'échelle internationale ». Ce point est majeur et soulève la question de la possibilité de lutter contre ces risques collectifs et sociétaux via les textes de protection des données et les organes créés par ces législations. La portée de ces textes et la compétence de ces organes ne sont-ils pas limités à la seule protection des risques individuels ? Nous reviendrons sur ce point important que nos recommandations doivent prendre en compte[47].

2.3 Les textes de l'Union européenne et du Conseil de l'Europe.

2.3.1 Introduction

Notre propos se limite à lister les dispositions des textes européens et d'ajouter un bref commentaire sur la façon dont les instruments récents du Conseil de l'Europe (Lignes directrices adoptées par le Comité de la Convention 108) indiquent la nécessité d'un élargissement du débat, particulièrement de mise en ce qui concerne les traitements de profilage.

2.3.2 Le RGPD et la Directive 206/680

En ce qui concerne les **textes de l'Union européenne**, la question du profilage est abordée par diverses dispositions. Nous avons déjà cité la définition donnée par l'article 4 (4) du RGPD ou l'article 3 (4) de la directive 2016/680 sur les traitements policiers. Au-delà, on souligne que les articles 13. 2 (f) et 14. 2 (g) du RGPD[48] mentionnent parmi les informations à communiquer à la personne concernée : « l'existence d'une prise de décision automatisée, y compris un profilage … et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée », L'article 15.1 (h) prévoit que l'accès s'étendra aux mêmes informations. L'article 22. 1. intitulé : « décision individuelle automatisée, y compris le profilage » octroie à la personne concernée le droit de s'opposer à être soumise à ce type de traitement si celui-ci produit des effets juridiques le concernant ou l'affecte 'de manière significative' sauf les hypothèses prévues au point 2, à savoir les hypothèse s de nécessité du contrat, l'autorisation par la loi ou le consentement explicite. Lorsque de telles exceptions s'appliquent, L'article 22.3 ; oblige le responsable du traitement « à mettre 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 ». L'article 35 concerne l'obligation du responsable de traitement de procéder à une 'analyse d'impact' du traitement projeté lorsque celui

[47] A notre connaissance, seul le rapport récent de S. Dreyer et W. Schultz («*The GDPR and automated decision making : Will it deliver ?* », *Bertelsman Foundation Report*, janvier 2019*,* disponible sur le site de la fondation Bertelsmann*)* aborde cette question*.*

[48] Nous ne détaillons pas ici les dispositions du texte de la directive sur les traitements policiers. Elles suivent pour l'essentiel celles du Règlement. Nous y reviendrons lorsque nous parlerons des profilages réalisés par les autorités policières ou des juridictions criminelles (i*nfra*, n° 24)

comporte un 'risque élevé' pour les personnes concernées. Le point 3 de l'article oblige à cette analyse en particulier « en cas d'évaluation systématique et approfondie d'aspects personnels concernant des personnes physiques, qui est fondée sur un traitement automatisé y compris le profilage », lorsque des décisions sont prises sur base de cette évaluation ayant des effets juridiques sur les personnes concernées ou l'affectant de manière significative. On ajoute que le 3 octobre 2017, le Groupe dit de l'article 29 émettait, dans la perspective de l'application alors prochaine du RGPD, des *Guidelines on automated decision making and Profiling*, destinées à interpréter les diverses dispositions du Règlement. Ces *Guidelines* particulièrement étoffées ont été reprises par *l'European Data Protection Board (EDPB)* créé depuis par le Règlement et qui hérite notamment des compétences du Groupe dit de l'article 29. Nous aurons l'occasion de mentionner quelques points d'interprétation du texte du RGPD, présents dans ces *Guidelines.*

On ajoute que la Directive 2016/680 cette fois applicable aux traitements de données en matière d'infractions ou de sanctions pénales outre qu'elle reprend en son article 11, le principe de non-suffisance d'une décision individuelle automatisée consacrée par l'article 22 ajoute en son alinéa 3 que « Tout profilage qui entraîne une discrimination à l'égard des personnes physiques sur la base des catégories particulières de données à caractère personnel … est interdit, … »

2.3.3 Et la Convention 108+ ?

Le texte de la Convention 108+ ne mentionne ni n'envisage comme tels les traitements de profilage. On souligne cependant l'article 9.1. à propos des décisions automatisées (Comparer avec l'article 22 du RGPD) : « Toute personne a le droit: a. de ne pas être soumise à une décision l'affectant de manière significative, qui serait prise uniquement sur le fondement d'un traitement automatisé de données, sans que son point de vue soit pris en compte *».* L'importance d'une approche fondée sur les risques est par ailleurs soulignée par l'article 10. Le point 2 exige : « Chaque Partie prévoit que les responsables du traitement, ainsi que, le cas échéant, les sous-traitants, doivent procéder, préalablement au commencement de tout traitement, à l'examen de l'impact potentiel du traitement de données envisagé sur les droits et libertés fondamentales des personnes concernées, et qu'ils doivent concevoir le traitement de données de manière à **prévenir ou à minimiser les risques** d'atteinte à ces droits et libertés fondamentales*. »* et le point 4 accorde à chaque Etat la possibilité de mesures complémentaires (voir aussi l'article 13) *: «* Chaque Partie peut, eu égard aux risques encourus pour les intérêts, droits et libertés fondamentales des personnes concernées, adapter l'application des dispositions des paragraphes 1, 2 et 3 dans la loi donnant effet aux dispositions de la présente Convention, en fonction de la nature et du volume des données, de la nature, de la portée et de la finalité du traitement et, le cas échéant de la taille des responsables du traitement et des sous-traitants ».

L'élargissement par les Lignes directrices sur les mégadonnées et l'IA -Nous apparaissent plus significatifs les élargissements affirmés par les lignes directrices déjà citées. Ils sont de divers ordres et doivent être pris en compte lors de l'écriture de tout nouvel instrument sur le profilage, d'autant plus que l'on sait que le profilage s'appuie désormais de plus en plus à la fois sur les ressources offertes par les 'mégadonnées' et sur les potentialités de l'Intelligence artificielle :

a. La première a déjà été mentionnée : les deux textes des Lignes directrices ouvrent singulièrement le débat : elles prennent en compte d'autres dimensions éthiques comme la dignité, la justice sociale ou non-discrimination, la diversité culturelle et les impératifs sociétaux de démocratie et non simplement les seules préoccupations autour de la protection des données à caractère individuel[49]. Ainsi, les "Lignes

[49] Voir dans le même sens, « Les systèmes d'IA ne doivent pas nuire aux êtres humains. De par leur conception, les systèmes d'IA doivent protéger la dignité, l'intégrité, la liberté, la vie privée, la sûreté et la sécurité des êtres humains dans la société et au travail. Les systèmes d'IA ne doivent pas menacer le processus démocratique, la liberté d'expression, la liberté d'identification ou la possibilité de refuser les services d'IA. À tout le moins, les systèmes d'IA ne doivent pas être conçus de manière à renforcer les préjudices existants ou à en créer de nouveaux pour les individus. Les préjudices peuvent être physiques, psychologiques, financiers ou sociaux. Les préjudices spécifiques à

directrices sur les Mégadonnées" du 23 janvier 2017[50] n'hésitent pas à l'affirmer explicitement: « L'utilisation des mégadonnées pouvant porter atteinte non seulement à la vie privée et à la protection des données de façon individuelle, mais également à la dimension collective de ces droits, les politiques préventives et l'évaluation des risques doivent tenir compte de l'impact juridique, social et éthique de cette utilisation, y compris au regard du droit à l'égalité de traitement et à la non-discrimination. » Et l'orientation de principe en exergue des Lignes directrices sur l'IA suit la même voie : « La protection de la dignité humaine et la sauvegarde des droits de l'Homme et des libertés fondamentales, en particulier le droit à la protection des données à caractère personnel, sont essentiels au développement et à l'adoption d'applications basées sur l'IA qui sont susceptibles de produire des effets sur les personnes et la société. Ceci est particulièrement important lorsque l'IA est utilisée dans les processus décisionnels. ».

b. La deuxième est d'élargir le cercle des acteurs dont le rôle implique certaines obligations. Là où le RGPD et la Convention 108+ ne mentionnent que les obligations du « responsable » du traitement, les deux Lignes directrices ajoutent la responsabilité d'autres acteurs intervenant soit dans la fourniture des mégadonnées (*big data*) ou bibliothèques (*libraries*) qui serviront aux systèmes de profilage soit dans la conception et la mise en œuvre des algorithmes de base ou l'adaptation de ces algorithmes aux besoins particuliers d'un secteur ou d'un système particulier[51]. On ne s'étonne pas de ce souci tant l'analyse des acteurs montre bien que les qualités du traitement et l'étendue des risques y liés sont loin de dépendre du seul responsable du traitement et trouvent leurs explications dans le choix des données souvent acquises à l'extérieur (le fournisseur de données), le choix de l'algorithme de base ou adapté par un tiers aux besoins de l'application spécifique du responsable (les développeurs et fabricants, selon l'expression des lignes directrices IA)[52]. Enfin, ne peut-on exiger, comme le préconisait la Recommandation de 2010

---

l'IA peuvent découler du traitement des données sur les individus (c'est-à-dire la manière dont elles sont collectées, stockées, utilisées, etc.) Pour éviter tout préjudice, les données collectées et utilisées pour la formation des algorithmes d'IA doivent être faites de manière à éviter toute discrimination, manipulation ou profilage négatif. Tout aussi important, les systèmes d'IA doivent être développés et mis en œuvre de manière à protéger les sociétés contre la polarisation idéologique et le déterminisme algorithmique. » High Level Expert Group on Artificial Intelligence dans ses *Ethical Guidelines for a Trustworthy AI* (2019).

[50] Cf. également : « Le traitement des données à caractère personnel ne devrait pas aller à l'encontre des valeurs éthiques communément acceptées dans la communauté ou les communautés pertinentes, et ne devrait pas porter atteinte à des intérêts, des valeurs et des normes sociétaux, y compris la protection des droits de l'Homme. …»

[51] Lignes directrices sur l'IA : *« 1*es développeurs, fabricants et prestataires de service en IA devraient adopter une approche de conception des produits et services centrée sur les valeurs, conformément à la Convention 108 +, notamment son article 10.2, et aux autres instruments pertinents du Conseil de l'Europe.

2. Les développeurs, fabricants et prestataires de service en IA devraient évaluer les éventuelles conséquences négatives des applications d'IA sur les droits de l'Homme et libertés fondamentales des personnes concernées et au regard de ces conséquences, adopter une approche de précaution basée sur des mesures de prévention et de réduction des risques appropriées.

3. Les développeurs, fabricants et prestataires de service en IA devraient, à tous les stades du traitement des données, y compris lors de la collecte, adopter une approche des droits de l'Homme dès la conception (by-design) et éviter tout biais potentiel, y compris les biais non intentionnels ou cachés, ainsi que les risques de discrimination ou d'autres effets négatifs sur les droits de l'Homme et libertés fondamentales des personnes concernées.

4. Les développeurs d'IA devraient évaluer de manière critique la qualité, la nature, l'origine et la quantité des données à caractère personnel utilisées, en réduisant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage, puis en vérifiant l'exactitude du modèle lorsqu'il est alimenté par de nouvelles données. Le recours à des données synthétiques pourrait être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA. *»*

[52] Cette même réflexion est partagée par le Groupe de l'article 29 qui, dans le contexte de ses *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* (adoptées, le 4 avril 2017), écrit: « Une analyse d'impact sur la vie privée peut également être utile pour évaluer l'impact sur la protection des données d'un produit technologique, par exemple un élément de matériel ou de logiciel, lorsque celui-ci est susceptible d'être utilisé par différents responsables de traitement pour effectuer différentes opérations de traitement. Bien entendu, le responsable du traitement qui déploie le produit reste tenu de réaliser sa propre analyse d'impact sur la protection des données en ce qui concerne la mise en œuvre spécifique, mais il peut en être informé par une analyse d'impact sur la protection des données préparée par le fournisseur du produit, le cas échéant. Un exemple pourrait être la relation entre les fabricants de compteurs intelligents et les entreprises de services publics. Chaque fournisseur ou responsable de traitement doit

(Point 3.8.) que : « La diffusion et l'utilisation, à l'insu des personnes concernées, de logiciels visant à l'observation ou à la surveillance dans le cadre du profilage de l'usage d'un terminal donné ou de réseaux de communications électroniques ne devraient être autorisées que si elles sont expressément prévues par le droit interne et assorties de garanties appropriées » ?

c. Le troisième point d'attention est l'amplification de l'approche par les risques. Les deux Lignes directrices 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.
d. La quatrième prolonge l'approche droit de l'environnement et réclame une évaluation participative, constructive, '*multistakeholders*' et interdisciplinaire des risques et solutions à trouver. Elle prône la création, le cas échéant et ce en fonction du caractère ou non élevé des risques envisagés, d'un Comité d'éthique (Lignes directrices sur les mégadonnées) *: «* Si l'évaluation de l'impact potentiel d'un traitement de données envisagé, telle que décrite à la section IV.2, révèle un fort impact de l'utilisation des mégadonnées sur les valeurs éthiques, les responsables du traitement des données peuvent établir un comité d'éthique ad hoc, ou s'appuyer sur les existants, afin d'identifier les valeurs éthiques spécifiques qu'il convient de protéger dans le cadre de l'utilisation de ces données. Le comité d'éthique devrait être un organe indépendant composé de membres choisis pour leurs compétences, leur expérience et leurs qualités professionnelles et accomplissant leur mission de façon impartiale et objective*. »* En écho, on lit dans les Lignes directrices sur l'IA : « Les développeurs, fabricants et prestataires de service en IA sont encouragés à recourir à des comités d'experts issus de différents domaines ainsi qu'à des institutions universitaires indépendantes qui peuvent contribuer à concevoir des applications de l'IA fondées sur les droits de l'Homme et orientées de façon éthique et sociale, et à détecter des biais potentiels. Le rôle de ces comités peut être particulièrement important dans les domaines où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles en raison d'intérêts et de droits concurrents, comme par exemple dans les domaines de la justice prédictive, de la prévention et de la détection des infractions. » et « Des démarches participatives d'évaluation des risques, reposant sur l'engagement actif des personnes et groupes potentiellement affectés par les applications de l'IA, devraient être encouragées. »

2.4 Classification des profilages

Quels critères retenir ? - L'approche par 'risque' soulignée par la Convention 108+ du Conseil de l'Europe et les deux Lignes directrices suggère qu'il importe de préconiser une régulation différenciée, fondée sur le degré de risque présenté par les différents types de traitement. Ce degré de risque s'apprécie en tout cas en fonction premièrement de la finalité poursuivie par le traitement de profilage, deuxièmement des conséquences que le traitement peut avoir sur la personne concernée ou les groupes visés et finalement en fonction, nous l'avons souligné, de la méthode de traitement suivie. Ainsi, on ne peut en effet, traiter avec la même sévérité des profilages où le 'profileur utilise les seules données volontairement adressées par le consommateur et ce, en fonction des besoins mêmes du service qu'il entend obtenir et ceux où des plateformes utilisent à usage multiple des données de plus en plus nombreuses obtenues par l'utilisation de leurs services.

Le profilage est donc loin de constituer une catégorie homogène de traitements. Notre propos est de mettre en évidence l'existence de nombreuses sous-catégories présentant des risques de nature et de portée

partager les informations utiles sans compromettre les secrets ni entraîner de risques pour la sécurité en révélant les vulnérabilités. »

différentes. L'examen des finalités poursuivies par les applications concrètes de profilage conduit à entrevoir ces sous-catégories (I). Le second critère distingue les traitements suivant le degré de risques y liés : nous parlerons de « traitements de profilage à risque élevé » que nous distinguerons quant à leur besoin de réglementation des autres (II). Nous reviendrons enfin sur les conséquences que nous pourrions tirer de cette double classification (III).

2.4.1 Classification en fonction des finalités

Quelles finalités ? - Nous listons ici une série de finalités suivant le contexte de l'opération de profilage et prendrons en compte, la classification des finalités de traitement, qu'opère les textes de protection des données. Ainsi, le profilage peut se concevoir dans une phase précontractuelle de la relation entre le responsable de traitement et la personne concernée ; il peut intervenir dans le cadre de leur relation contractuelle. Le profilage peut être un service en soi qui alimenté de l'extérieur ou non permet de mettre des profils à destination de tiers. La recherche, en particulier mais non uniquement médicale peut également procéder à, voire nécessiter, des profilages. Dans le cadre de traitements opérés par les pouvoirs publics, on distinguera également divers types de profilage suivant, d'une part, qu'ils aident à la prise de décisions générales ou suivant, d'autre part, qu'ils s'opèrent dans le cadre de l'application d'une réglementation à des cas individuels. Un point particulier sera réservé aux profilages des autorités policières ou juridictionnelles, dans le cadre de la prévention ou de la détection d'infractions pénales. La question de la justice prédictive sera également abordée. Sans doute d'autres classifications eurent été possibles : on pense à des classifications par secteurs : marketing direct, administration, emploi, banques, assurances, police et justice.

2.4.2 Profilage et phase précontractuelle dans la relation responsable de traitement- personne concernée.

De la publicité ciblée à l'exclusion et à *l'adaptative pricing* - La publicité ciblée est sans doute la première opération, à laquelle on songe en matière de profilage. Le profilage substitue à un adressage sans grandes nuances adressée à un public et qui reste relativement neutre (par exemple, il est clair qu'aux lecteurs d'une revue *people* de luxe, on n'adressera pas le même message publicitaire que celui à destination des lecteurs d'une presse populaire à scandales), l'envoi de messages, une publicité *one-to one,* qui cherchent à épouser au mieux la personnalité et les besoins du destinataire afin de renforcer l'efficacité de l'action publicitaire. Bien souvent, le destinataire du message y trouvera intérêt et peut même sollicitera-t-il cette publicité ciblée. La manipulation même légitime que représente toute action publicitaire est ici accrue de manière exponentielle, d'autant que le message peut apparaître de manière cachée (sous la forme d'une information) ou trompeuse (vous ignorez que le message provient d'un tiers et non du site avec lequel vous êtes entré en dialogue). Par ailleurs, la manipulation sera d'autant plus importante que votre profil vous cernera 'au plus profond', jouant sur votre affectif, vos penchants y compris sexuels, votre race, voire vos handicaps ou opinions[53]. On note que ce risque est d'autant plus grand que, dans les systèmes d'IA complexes (tels que le *deep learning*), les modèles sont difficiles à rendre transparents et le risque de manipulation est difficile à détecter. On parle d'ailleurs de modèles « back box ».

Ceci dit, le ciblage de clientèle peut poursuivre des finalités complémentaires : le système d''*adaptative pricing'* couple au 'profil' une estimation de l'intensité de la demande pour tel ou tel produit et propose de

[53] Voir notamment, l'article de Kosinsky et all, *Private traits and attributes are predictable from digital records of human behavior* (2013) qui étudie ce que Facebook peut déduire des *likes*: « Nous montrons que des enregistrements numériques facilement accessibles du comportement, Facebook Likes, peuvent être utilisés pour prédire automatiquement et précisément une série d'attributs personnels très sensibles, notamment : l'orientation sexuelle, l'appartenance ethnique, les opinions religieuses et politiques, les traits de personnalité, l'intelligence, le bonheur, la consommation de substances addictives, la séparation des parents, l'âge et le sexe. »

manière différenciée aux internautes un prix en fonction de cette variable. Autre finalité, celle de sélectionner les potentiels contractants, en excluant la visite ou en écartant par des propositions transactionnelles non sérieuses, d'autres. On connaît le recours lancé contre Facebook aux États Unis à propos de l'utilisation par une société immobilière des services de profilage de Facebook, afin de filtrer toute demande d'acquisition ou de locations émanant de certaines catégories de personnes (les noirs, les personnes à revenu modeste, les LGBT, etc.). Enfin, en matière d'emploi, il n'est pas rare ainsi de sélectionner les candidats sur la base d'un certain nombre de données analysées automatiquement, selon des critères plus ou moins explicités lors de l'utilisation des algorithmes de profilage.

2.4.3 Profilage dans le cadre d'une relation contractuelle

Les nécessités du contrat - Le traitement de profilage aura alors pour but une évaluation de la personne contractante mais cette évaluation poursuit elle-même différents objectifs : soit il s'agit d'évaluer les performances contractuelles d'un employé par exemple dans le cadre d'une promotion ou d'un licenciement ou plus largement d'un client dans le cadre d'une banque ou d'une assurance, qui désirent fixer le montant d'une prime, d'un crédit ou la rentabilité d'un client. On peut également penser à l'évaluation des risques financiers ou autres, liés à l'évaluation d'un partenaire, personne morale ou physique et ce, afin de décider de la poursuite ou non des relations.

2.4.4 Profilage et activités relatives à la 'commercialisation' de ces profils.

La commercialisation des données ou des profils - Les détenteurs de mégadonnées peuvent être tentés d'offrir des services de profilage à des tiers soit en tant que simples fournisseurs de données, soit sur base des requêtes du 'client' de définir les algorithmes nécessaires et de les appliquer sur leurs propres données, mettant alors à disposition les personnes profilées ou s'offrant à opérer eux-mêmes l'opération souhaitée par le tiers. L'existence de 'collection de données' (*dataset)* disponibles en *open source'* ou non, la fourniture d'algorithmes gratuitement par des laboratoires de recherche ou à titre commercial par des entreprises offrant des services de profilage ont été développés dans le premier chapitre. Les cas Akademia et Cambridge Analytica, concernent de telles hypothèses. Dans le premier cas, il s'agissait pour Facebook de fournir au responsable d'un blog de fans les informations sur les profils des personnes se connectant au site de manière à mieux connaître le public intéressé par le blog et de mieux définir sa stratégie. L'affaire Cambridge Analytica est celle de l'utilisation des données générées par le réseau social et complétées par des questionnaires soi-disant de recherche proposés par la société Cambridge. Le scandale Facebook-Cambridge Analytica renvoie aux données personnelles de 87 millions d'utilisateurs du réseau Facebook que la société Cambridge Analytica (CA) a commencé à recueillir dès 2014. Ces informations ont servi à influencer les intentions de votes en faveur d'hommes politiques qui ont retenu les services de CA Il s'agissait de connaître les sensibilités politiques voire prédire les choix politiques des personnes et de commercialiser auprès d'acteurs politiques ces résultats, ce qui leur permettait d'adapter les stratégies de marketing politique. Tout récemment, le Financial Times (4 septembre 2019) révélait la façon dont Google partageait des données avec des entreprises commerciales précisément pour faciliter leurs activités de profilage[54].

[54] Le *Financial Time*s (4/09) révèle que **Google a violé sa propre politique en matière de confidentialité des données, ainsi les réglementations de l'UE en matière de vie privée exigeant le consentement et la transparence**. La firme attribuait un "tracker" pour identifier les internautes. Elle invitait ensuite ses clients annonceurs à se connecter à une page Web cachée contenant une adresse unique qui leur permettait de connaître l'activité de navigation de ces utilisateurs. Elles pouvaient ainsi leur adresser des publicités ciblées

2.4.5 Profilage et activités de recherche

Le profilage comme aide à la recherche - Dans le cadre de projets de recherche, les chercheurs peuvent à partir de données généralement collectées par les autorités publiques ou privées pour des finalités premières, développer des traitements de profilage pour mieux comprendre tel ou tel phénomène et ainsi progresser dans leurs connaissances. Le chercheur qui s'intéresse à la gestion de personnel utilisera des données venant d'entreprises, d'interviews d'acteurs, de données détenues par le ministère de l'emploi ou de l'économie pour essayer de comprendre en quoi certains facteurs peuvent expliquer la réussite ou non professionnelle. En matière médicale, l'intelligence artificielle est largement utilisée pour découvrir les profils y compris génétiques des personnes atteintes de telle ou telle maladie et ainsi pouvoir prédire l'évolution de cette maladie et, le cas échéant, intervenir préventivement. Les constructeurs de véhicules peuvent de même bénéficier de recherches utilisant des données enregistrant les détails d'utilisation de véhicules par leurs conducteurs ainsi que des données relatives au contexte de l'utilisation des véhicules.

L'activité de recherche scientifique fait l'objet de diverses mesures d'exception au motif de l'intérêt que présente celle-ci. Comme on le sait, le RGPD considère que moyennant des garanties appropriées (agrément des laboratoires, utilisation de pseudonymes, etc…), la finalité de recherche n'est pas incompatible avec les finalités premières des traitements de données. Plus récemment, nonobstant la directive 'Protection des bases de données', la directive '*Copyright in a Digital Society*' exempte les institutions scientifiques d'intérêt public de toute restriction en ce qui concerne l'exploitation par '*Data mining* ' des bases de données. En d'autres termes, les textes européens tout en soumettant le profilage dans le cadre d'activité de recherche à certaines conditions voient cependant cette activité de manière positive.

2.4.6 Profilage par les autorités publiques

Le profilage au service des gouvernements et administrations - Le profilage permet aux autorités publiques et aux administrations de poursuivre différents objectifs. Il permet tout d'abord de définir des **stratégies** qu'il s'agisse de politiques d'expansion économique, d'aide aux logements, de mobilité, d'éducation ou d'aide à l'emploi.[55] Pour ce faire, elles prennent ainsi en compte une multitude de facteurs et brassant des données nombreuses venant de sources publiques ou privées et tirant de là des 'modèles prédictifs' et permettant d'envisager l'impact de telle ou telle politique. On conçoit que ces traitements peuvent, s'ils sont mal programmés (biais, mauvaise qualité des données ou erreurs dans les algorithmes), affecter certains groupes et, en tout cas, que les décisions prises sur base de telles prévisions affecteront les membres de ces groupes.

L'**application** des réglementations trouve facilement une plus-value d'efficacité dans des systèmes d'IA et de profilage. Ces derniers permettent, dans le cadre d'une philosophie du '*benevolent government',* c''est à dire d'une politique proactive de l'État à l'égard de ses citoyens, de détecter voire sélectionner les personnes qui pourraient bénéficier d'une aide particulière, de conseiller au mieux tel étudiant sur son parcours d'étude… mais également de détecter les familles à problème (enfants battus) voire les éventuels fraudeurs à la sécurité sociale ou au fisc. La justice prédictive, censée remplacer les juges, peut également être évoquée ici dans la mesure où tout litige peut être 'profilé' en fonction des caractéristiques multiples et variées de l'affaire, analysées à la lumière des décisions précédentes.

[55] Sur différents cas de systèmes de '*decision-making'* dans le secteur public et à l'appui de stratégies gouvernementales, le rapport : *Automating Society Taking Stock of Automated Decision-Making in the EU : A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations*, janvier 2019 disponible à l'adresse : www.algorithmwatch.org/automating-society.

2.4.7 Profilage et autorités d'investigation et de poursuites en matière d'infractions pénales

L'IA au service des autorités policières et des parquets - L'Agence européenne des droits fondamentaux dans son rapport de 2017[56] distingue deux types de profilage en la matière : « Le profilage réalisé dans le contexte des services de police et de la gestion des frontières poursuit deux objectifs principaux : identifier des personnes connues sur la base de renseignements concernant une personne spécifique et, à titre de méthode prédictive, identifier des personnes 'inconnues ' susceptibles d'intéresser les autorités répressives et les autorités chargées de la gestion des frontières ». Il s'agit donc d'agir soit en réaction à une infraction déjà posée, soit de manière proactive de façon à pouvoir agir préventivement par rapport à la possibilité d'infractions pénales. Dans ce second cas, pourra se définir une stratégie globale qui mènera à des décisions soit relatives à des groupes (par exemple, surveiller telle communauté), soit individualisées.

Dans tous les cas, l'Agence souligne le risque de discrimination lié à ces profilages et rappelle que suivant la directive européenne : « Le profilage relevant du champ d'application de la directive « police » doit se conformer à l'article 11, paragraphe 3, de la directive « police ». Il dispose que « [t]out profilage qui entraîne une discrimination à l'égard des personnes physiques sur la base des catégories particulières de données à caractère personnel visées à l'article 10 est interdit, conformément au droit de l'Union ». Reste la question des critères fondés sur d'autres données que celles dites sensibles. Ainsi, un niveau d'éducation, la mobilité des personnes, le niveau d'éducation pourraient également être pris en considération pour détecter les milieux criminogènes. On le pressent ce n'est pas la nature des données mais la finalité du traitement qui crée le risque.

2.5 Classification en fonction des risques

L'importance des risques comme critère à prendre en considération dans le cadre de la recommandation - Il nous paraît que tous les profilages ne sont pas à réglementer de la même manière mais doivent se comprendre en fonction de la nature des risques créés et des conséquences du traitement sur les personnes concernées ou la société. Cette préoccupation est exprimée par la Convention 108+, tant en son article 10.2, traduite par le rapport explicatif (n°88) comme suit : « Le paragraphe 2 précise qu'avant d'effectuer une activité de traitement, le responsable du traitement doit examiner son impact potentiel sur les droits et libertés fondamentales des personnes concernées. » Elle l'est également par l'article 10.4 qui permet « eu égard aux risques encourus » d'ajouter des obligations complémentaires à celles prévues par les autres dispositions Par ailleurs, ces risques peuvent s'analyser soit à un niveau micro, c'est-à-dire de la personne concernée, soit à un niveau macro, c'est-à-dire de groupes sociaux organisés ou non, existant de la société, soit aux deux. Enfin, dans la mesure où le traitement de profilage met en cause différents acteurs dont l'intervention peut chacune être source de risques, il importe de préciser comment chacun supporte sa part de risque.

Notre réflexion part de la notion de traitement à risque élevé ou '*High Risk*' développé par les textes de l'Union européenne et suggéré par le Conseil de l'Europe en particulier dans ses Lignes directrices sur les mégadonnées et l'IA (A). Ensuite, sera développé le régime juridique lié à cette notion (B).

2.5.1 La notion de '*high risk*' dans les textes européens

Le RGPD

[56] FRA, Agence pour la protection des droits fondamentaux, « Guide pour la prévention du profilage illicite aujourd'hui et demain » Publications officielles, 2019.

L'obligation de '*Risk Assessment*'- L'article 35.1 du RGPD énonce: « Lorsqu'un type de traitement, en particulier par le recours à de nouvelles technologies, et compte tenu de la nature, de la portée, du contexte et des finalités du traitement, est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement effectue, avant le traitement, une analyse de l'impact des opérations de traitement envisagées sur la protection des données à caractère personnel. ». L'article 27.1. de la directive 'Police' reprend le même libellé. La notion de 'risque élevé ' est donc au centre de la réflexion européenne. Le même article du RGPD énonce trois hypothèses notamment constitutives de 'risque élevé'. Ces trois hypothèses sont particulièrement pertinentes lorsqu'il s'agit de traitements de profilage.

a) l'évaluation systématique et approfondie d'aspects personnels concernant des personnes physiques, qui est fondée sur un traitement automatisé, y compris le profilage, et sur la base de laquelle sont prises des décisions produisant des effets juridiques à l'égard d'une personne physique ou l'affectant de manière significative de façon similaire ;

b) le traitement à grande échelle de catégories particulières de données visées à l'article 9 paragraphe 1, ou de données à caractère personnel relatives à des condamnations pénales et à des infractions visées à l'article 10 ; ou

c) la surveillance systématique à grande échelle d'un zone accessible au public

Le point a) fait directement allusion au profilage et doit donc être pris en considération ; le second point vise la constitution de mégadonnées portant sur des données à caractère sensible. Lorsqu'est associé un traitement de profilage à ces mégadonnées, le *Risk Assessment* s'impose *a fortiori*. Enfin, la dernière hypothèse vise la collecte à des fins de surveillance d'une zone accessible au public (rues, cinémas, magasins, …) Il est certain que cette surveillance si elle souhaite être systématique utilisera des techniques de profilage basés par exemple sur la reconnaissance faciale, l'analyse des mouvements, l'internet des objets, etc.

Les traitements à haut risque selon le RGPD - L'analyse des textes de l'Union européenne conduit à considérer le profilage comme « traitement à haut risque », dans les cas suivants.

- L'évaluation d'une personne physique par un traitement de profilage est présumée à haut risque si elle est réalisée de manière systématique et approfondie. On suppose qu'il s'agit là d'écarter des évaluations d'une part, ponctuelles et non répétées mais également aboutissant à une analyse sommaire de la personnalité. L'interprétation de ce second point est ambigüe. Notre sentiment est que le critère tient non pas à la qualité plus ou moins sensible des données (les données peuvent être triviales) mais bien au résultat (le traitement de ces données triviales peut aboutir à détecter la nervosité de la personne dans des moments difficiles). En outre, ce traitement doit avoir des **conséquences de nature juridique ou des effets significatifs sur la personne**, il s'agit ici de la reprise de notions utilisées par la disposition sur les traitements automatisés, disposition présente dans le RGPD (article 22 du RGPD). Les Guidelines définissent longuement ces deux notions. La portée juridique de la décision née ou proposée par le traitement s'entend soit de la mise à terme d'un contrat, soit de l'octroi ou refus du bénéfice d'un avantage social soit, ajoutent les *Guidelines* de l'accès à un territoire ou à la nationalité (ou de leur refus). « Affecter significativement la personne concernée » suppose que l'atteinte perdure dans le temps, qu'elle conduise à l'exclusion d'une personne ou à sa discrimination, qu'elle concerne un service financier, de santé ou d'éducation ou un emploi. Les *Guidelines* n'excluent pas certains profilages publicitaires, en fonction de l'étendue des données collectées, des attentes raisonnables de la personne concernée, des techniques utilisées pour adresser la publicité et surtout de l'utilisation de la connaissance de la faiblesse des personnes concernées. Le document ajoute que des pratiques comme le '*dynamic pricing*' méritent également dans certains cas d'être considérées comme traitement à haut risque.

- Le traitement « à grande échelle » des données sensibles visées par l'article 9 du RGPD, soit « le traitement de données qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique » de même que les données relatives à des condamnations pénales est considéré à 'haut risque'. Ici, c'est principalement le risque de discrimination qui justifie une telle qualification. Notons que cette discrimination peut concerner certes des individus mais également des groupes ethniques et que cette possibilité doit également être prise en compte, nonobstant le texte du RGPD. On ajoute, autre critique du texte de l'Union européenne, que cette catégorie de traitement à 'haut risque' ne devrait pas nécessairement être liée au contenu sensible des données, comme le dit l'article 35, mais bien à sa finalité dans la mesure où la révélation du caractère sensible peut ressortir du traitement (voir le texte de l'article 9 : « le traitement de données à caractère personnel qui révèle… »). Ainsi dans l'affaire Cambridge Analytica, il a été démontré que les opinions politiques pouvaient être déduites de données non sensibles liées à une simple utilisation du réseau social. Le qualificatif 'à grande échelle' laisse supposer que seuls des traitements portant sur une large population, tenant compte du contexte géographique ou de la population susceptible d'être concernée sont retenus.
- La troisième hypothèse s'intéresse aux traitements dans les *zones*[57] *accessibles au public* et semble se concentrer sur les seuls traitements de surveillance systématique opérée, on peut l'imaginer, à des fins de sécurité (ex : pour un grand magasin, assurer la sécurité des lieux contre le vol ou autres infractions). La raison invoquée est l'impossibilité pour toute personne présente dans un lieu public (la rue, le grand magasin, l'hôtel, l'administration, …) de pouvoir échapper à cette surveillance. On ajoute que cette collecte dans les lieux publics peut poursuivre d'autres finalités que la seule « surveillance ». Ainsi, un grand magasin suggérera d'autres achats et, le cas échéant, cédera les données collectées à des entreprises d'un secteur différent comme le secteur du tourisme à des fins de profilage. On note que le texte n'exclut pas une interprétation large de la notion de surveillance et pourrait dès lors viser également de telles finalités commerciales.

Les textes du Conseil de l'Europe

Et le Conseil de l'Europe ? - Le rapport explicatif de la Convention 108+ lorsqu'il commente la possibilité pour les Etats-membres de prévoir des obligations complémentaires en cas de risques supérieurs donne quelques critères : « Une telle adaptation doit tenir compte de la nature et du volume des données traitées, de la nature, de la portée et de la finalité du traitement et, dans certains cas, de la taille de l'entité responsable du traitement ». On ajoutera que la qualité des personnes visées pourrait également être un indice tant, comme déjà souligné, la manipulation de certaines catégories comme les enfants mais peut-être d'autres groupes comme les patients ou les employés est plus facile. Il est clair que l'addition de nombre de ces critères justifiera certainement le risque élevé. Les Lignes directrices du sur les mégadonnées énonce un autre critère de prise en compte d'un 'haut risque' : « Exposer la personne concernée à des risques différents ou supérieurs à ceux envisagés pour les finalités initiales pourrait être considéré comme un traitement ultérieur inattendu ». Cette réflexion est importante tant nos données de plus en plus se mélangent à d'autres données et constituent ensemble des réservoirs dans lesquels les systèmes d'intelligence artificielle qui autorisent le profilage viennent puiser afin de poursuivre de multiples fins. La cession à des tiers de profils ou de données conduisant à du profilage que du partage de données nous paraît devoir également être considérées comme un traitement à 'risque élevé'.

Il est à noter que la recommandation sur le profilage de 2010 retenait déjà la notion de risques particuliers qui sans doute a inspiré le RGPD lorsqu'il parle de traitements à 'haut risque'. L'article 9.2 de la

[57] L'utilisation du mot 'zones' semble indiquer que seuls les lieux physiques et non virtuels sont visés. On peut s'interroger sur l'intérêt qu'il y aurait à élargir le propos vis-à-vis de plateformes de communication ou d'information dont les services sont largement ouverts au public.

recommandation s'exprimait comme suit : « Par ailleurs, dans le cas de traitements ayant recours au profilage et présentant des risques particuliers au regard de la protection de la vie privée et des données à caractère personnel, les Etats membres peuvent prévoir : a. que les responsables des traitements soient tenus de les notifier préalablement à l'autorité de contrôle ; ou b. que ces traitements fassent l'objet d'un contrôle préalable par l'autorité de contrôle. » On regrette que cette notion n'ait pas été mieux définie dans la recommandation, même si son emploi démontre l'importance accordée dès 2010 par le Conseil de l'Europe à définir un régime particulier de contrôle voire d'agrément par l'autorité de contrôle pour les traitements de profilage présentant des 'risques particuliers ou spéciaux'.

Un autre point des Lignes directrices 'mégadonnées' et 'Intelligence artificielle' a déjà été souligné à plusieurs reprises. Pour reprendre le point 2.3. des lignes directrices sur les mégadonnées : « L'utilisation des mégadonnées pouvant porter atteinte non seulement à la vie privée et à la protection des données de façon individuelle, mais également à la dimension collective de ces droits, les politiques préventives et l'évaluation des risques doivent tenir compte de l'impact juridique, social et éthique de cette utilisation, y compris au regard du droit à l'égalité de traitement et à la non-discrimination». Notre conviction, en accord avec les deux Lignes directrices du Comité de la Convention 108, est de considérer que, pour déterminer la gravité des risques liés à un traitement de profilage, il importe de prendre également en considération l'ensemble des risques sociétaux ou collectifs et ce, bien au-delà des risques encourus par les individus, comme les risques de discrimination, de justice sociale, de fonctionnement de nos démocraties, etc.

Quelle réglementation pour les traitements de profilage en particulier à haut risque ?

Vers de multiples régimes ? - Sans doute, est-il difficile de définir un régime juridique unique pour l'ensemble des traitements de profilage. La règle de proportionnalité, c'est-à-dire la nécessité d'une réglementation à la hauteur des risques créés et d'une réponse adaptée à ceux-ci, est à rappeler ici. Une réglementation qui impliquerait des coûts disproportionnés lors de sa mise en œuvre et manquerait sa cible est à éviter. Nous aurons l'occasion de le montrer lors du point III, lorsque nous repasserons en vue les différents traitements de profilage. A ce stade, nous reprenons ici différentes idées émises tant par les textes de l'Union européenne ou du Conseil de l'Europe déjà cités mais également d'autres documents, en particulier relatifs à l'éthique et l'intelligence artificielle.

La première règle est certes pour toute personne mettant en place un traitement de profilage ou, simplement, un outil logiciel (algorithme de profilage) ou une base de données destinés à un ou des traitements de profilage d'identifier les risques liés aux traitements projetés ou auxquels ils contribuent ou pourraient contribuer. Le risque est de double nature. Si les textes européens cités se concentrent sur les impacts possibles liés au traitement de profilage : risque de discrimination, de manipulation, de surveillance forcée voire les risques sociétaux liés à tel ou tel traitement, il faut également prendre en considération, comme pour tout traitement mais sans doute en tenant compte des caractéristiques du système, les risques propres à la sécurité du système : ainsi, ceux de confidentialité, d'absence d'intégrité des données de base ou une fois traitées voire d'indisponibilité du système. Les recommandations de l'OCDE adoptées récemment de même que les recommandations du Groupe européen dit '*High leve*l' sur l'intelligence artificielle (2018) insistent sur cette exigence de sécurité. Ainsi le point 1.4 des recommandations OCDE, intitulé : 'Robustesse, sûreté et sécurité'**,** énonce **:**

« a) Les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie, de sorte que, dans des conditions d'utilisation normales ou prévisibles, ou en cas d'utilisation abusive ou de conditions défavorables, ils soient à même de fonctionner convenablement, et ne fassent pas peser un risque de sécurité démesuré.
b) Pour ce faire, les acteurs de l'IA devraient veiller à la traçabilité, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA,

afin de permettre l'analyse des résultats produits par lesdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art de la technologie.
c) Les acteurs de l'IA devraient, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques y afférents, notamment ceux liés au respect de la vie privée, à la sécurité numérique, à la sûreté et aux biais. *»*

Une réflexion pluridisciplinaire - Cette réflexion doit être collective et pluridisciplinaire dans la mesure où le risque lié à un traitement dépendra bien évidemment de la qualité et du caractère approprié à la fois de la base de données choisie (cf. l'exemple des bases de données utilisées pour la reconnaissance automatique d'images de mariage, qui reprendraient de manière plus que majoritaire des images de mariage à l'occidental ou les bases de données identifiant les familles d'enfants battus dans un contexte historique et social différent) et à la fois de l'algorithme choisi. A cet égard, on rappellera le Rapport de la Convention 108+ (n°86) : « Le paragraphe 2 de l'article 10 précise qu'avant d'effectuer une activité de traitement, le responsable du traitement doit examiner son impact potentiel sur les droits et libertés fondamentales des personnes concernées. Cet examen peut être fait sans formalités excessives. Il évaluera également le respect du principe de proportionnalité, en s'appuyant sur une présentation détaillée du traitement envisagé. Dans certains cas, lorsqu'un sous-traitant intervient en plus du responsable du traitement, ce sous-traitant peut également avoir à procéder à un examen des risques. Des développeurs de systèmes d'information, et notamment des spécialistes de la sécurité ou des concepteurs ainsi que des usagers et des juristes peuvent prêter assistance dans l'examen des risques. ». Sans doute, un document écrit consignera les points essentiels de cette évaluation, des personnes réunies pour ce faire et les décisions prises.

Le RGPD impose un contenu particulier à cette procédure de « *Risk Assessment »* dans le cadre de traitements de profilage à 'haut risque'. L'article 35.7 précise : « L'analyse contient au moins :

a) une description systématique des opérations de traitement envisagées et des finalités du traitement, y compris, le cas échéant, l'intérêt légitime poursuivi par le responsable du traitement ;
b) une évaluation de la nécessité et de la proportionnalité des opérations de traitement au regard des finalités ;
c) une évaluation des risques pour les droits et libertés des personnes concernées conformément au paragraphe 1 ; et
d) les mesures envisagées pour faire face aux risques, y compris les garanties, mesures et mécanismes de sécurité visant à assurer la protection des données à caractère personnel et à apporter la preuve du respect du présent règlement, compte tenu des droits et des intérêts légitimes des personnes concernées et des autres personnes affectées. *»*

Pluridisciplinarité et ouverture de l'évaluation - Les Lignes directrices du Conseil de l'Europe à propos de l'IA précise : « Les développeurs, fabricants et prestataires de service en IA devraient, à tous les stades du traitement des données, y compris lors de la collecte, adopter une approche des droits de l'Homme dès la conception (by-design) et **éviter tout biais potentiel, y compris les biais non intentionnels** ou cachés, ainsi que les risques de discrimination ou d'autres effets négatifs sur les droits de l'Homme et libertés fondamentales des personnes concernées. Les développeurs d'IA devraient évaluer de manière critique **la qualité, la nature, l'origine et la quantité des données à caractère personnel utilisées, en réduisant les données inutiles, redondantes ou marginales lors des phases de conception et d'apprentissage, puis en vérifiant l'exactitude du modèle lorsqu'il est alimenté par de nouvelles données**. Le recours à des données synthétiques pourrait être considéré comme une solution possible pour minimiser la quantité de données personnelles traitées par des applications de l'IA. »

Par ailleurs, d'autres textes insistent sur la nécessité d'ouverture de cette procédure. Il s'agit notamment des Lignes directrices sur les mégadonnées (point 7.3). L'ouverture est double : premièrement, il s'agit d'une procédure menée par un **groupe interdisciplinaire** : « Le 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. » En second lieu, il d'agit d'associer à cette évaluation des représentants des personnes ou des groupes concernés. Les Lignes directrices plaident pour une **participation active des groupes concernés à cette évaluation ('une évaluation participative')** : « Les développeurs, fabricants et prestataires de service en IA sont encouragés à recourir à des comités d'experts issus de différents domaines ainsi qu'à des institutions universitaires indépendantes qui peuvent contribuer à concevoir des applications de l'IA fondées sur les droits de l'Homme et orientées de façon éthique et sociale, et à détecter des biais potentiels. Le rôle de ces comités peut être particulièrement important dans les domaines où la transparence et la mobilisation des parties prenantes peuvent être plus difficiles en raison d'intérêts et de droits concurrents, comme par exemple dans les domaines de la justice prédictive, de la prévention et de la détection des infractions. ». Ce point nous apparaît particulièrement important : la complexité, la difficulté de mesurer les impacts du traitement mais également le fait que le profilage concerne, au-delà de chaque individu profilé, l'ensemble des personnes profilées ou non plaident en ce sens, de plus, une approche individualiste de la protection est illusoire face à des responsables de traitement dont la puissance est sans commune mesure avec celle de chaque individu.
Il est important que le Conseil de l'Europe plaide en faveur d'une consultation voire d'une négociation avec des associations représentant les intérêts des personnes concernées. Les Lignes directrices sur les mégadonnées vont dans ce sens : « En ce qui concerne l'utilisation de mégadonnées susceptibles de porter atteinte aux droits fondamentaux, les Parties devraient encourager la participation des différents acteurs (par exemple des personnes ou groupes qui pourraient être concernées par l'utilisation des mégadonnées) au processus d'évaluation des risques et à la conception du traitement des données. »

Vers la création d'un organe multidisciplinaire d'évaluation 'éthique' des systèmes d'IA et en particulier des traitements de profilage fondés sur l'utilisation d'un système IA - Au-delà de cette procédure interne aux entreprises ou afin d'appuyer cette procédure interne, le Conseil de l'Europe pourrait recommander aux Etats-membres la création d'un organe multidisciplinaire d'évaluation 'éthique' des systèmes d'IA[58] et, dans le cadre qui nous intéresse, des traitements de profilage. Le rôle de cette « Autorité nationale pluridisciplinaire indépendante d'évaluation des risques liés à l'intelligence artificielle et en particulier aux traitements de profilage utilisant des procédés d'apprentissage automatique (***machine learning***) » serait multiple. Cette autorité indépendante serait en charge de l'audit, des tests, et de la labellisation des systèmes IA des secteurs privé ou public. L'intervention de cette autorité serait obligatoire en matière d'IA utilisée à des activités du secteur public et, sous réserve de ce qui pourrait être décidé par les Etats membres à propos des systèmes à risque élevé, volontaire pour les systèmes opérant dans le secteur privé.

Cette autorité émettrait des avis à propos, premièrement, de tout traitement de profilage individuel ou collectif envisagé par les administrations ou l'autorité réglementaire pour appuyer leurs stratégies ou appliquer les réglementations, secondement, à propos de l'évaluation des risques liés des politiques privées ou publiques en matière de 'partage des données' et d''open data' et de soutien à la définition et à l'implémentation de 'bonnes pratiques'. Cette autorité devrait émettre des recommandations relatives aux qualités des mégadonnées et des algorithmes de profilage afin d'assurer leur fiabilité, leur transparence et leur conformité aux législations applicables, notamment en matière de protection des données, de protection des consommateurs, de non-discrimination, de concurrence, etc. Elle travaillerait en étroite coopération avec les autorités de contrôle même si son évaluation déborde les compétences de ces autorités, dans la mesure où elle envisage également les risques collectifs. Ces avis et recommandations seraient publics.

Il est également proposé que cette autorité, sous réserve de l'existence d'autres organismes sectoriels ou non de labellisation ou certification, puisse également répondre aux demandes des entreprises d'évaluer et de 'certifier' leurs propres systèmes. Ce mécanisme de certification serait purement volontaire mais ; par le

[58] A ce propos, le Rapport sur l'intelligence artificielle remis par A. Mantelero au Conseil de l'Europe, qui a servi de base à l'établissement des Lignes directrices sur l'intelligence artificielle, en particulier les pages 16 et s. L'auteur souligne l'intérêt d'une approche à un niveau national en complément à la mise en place d'une procédure au niveau des entreprises.

label délivré, créerait aux yeux du public la confiance et l'acceptabilité sociétale nécessaires. L'évaluation interne ou externe proposée doit être particulièrement attentive aux points suivants :

- La lutte contre les erreurs, par l'examen à la fois des jeux de bases de données utilisées, jeux ou données non adéquats, non mis à jour ou contenant des erreurs et des algorithmes mal conçus ou inadaptés ou contenant des bugs qui peuvent mener à des résultats non désirables ;
- Le souci de respecter lors du design même du traitement de profilage, les principes du '*Privacy by design*' et du '*Privacy by default' ;*
- La nécessité de prévoir une période de test et l'évaluation des résultats de ce test avant de lancer effectivement le traitement de profilage ;
- La nécessité d'une information de la personne concernée sur l'existence du profilage en cours ou projeté avec une obligation d'accès facile (en cliquant sur une icône ?) à une description plus détaillée des caractéristiques des profilages exercés vis-à-vis de lui.

De manière générale, 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 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. À 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*. »*

2.6 Premières réflexions sur l'application des principes à quelques catégories recensées de traitements de profilage.

Rappel du principe de proportionnalité des mesures de régulation - Le but de l'exercice est de reprendre quelques réflexions à propos à la fois des risques et des mesures susceptibles d'être imaginées pour les catégories retenues. Sans doute, ces réflexions n'épuisent pas les suggestions qui pourraient être adressées et peut-être y trouvera-t-on à redire mais il nous semble absolument important de les conduire pour montrer combien une réglementation qui s'étendraient à l'ensemble des traitements de profilage est, à notre avis, dangereuse et risque d'être disproportionnée dans bien des cas.

2.6.1 Le profilage dans un cadre précontractuel

Le profilage publicitaire - Ainsi, parmi les traitements de profilage susceptibles d'être opérés dans la phase contractuelle, on distinguera ceux qui consistent en un adressage publicitaire élaboré sur base de catégories tantôt fixées par le responsable entièrement ou partiellement (systèmes d'IA supervisés), tantôt découvertes par le système lui-même (système d'IA non-supervisés). Dans de tels cas, on se limitera à rappeler le besoin d'un *risk assessment* interne portant sur la qualité des données utilisées et on veillera à ce que la personne concernée soit informée du profilage effectué (via une icône) et puisse par un moyen simple (cliquer sur l'icône avertissant du profilage), d'une part, connaître les caractéristiques du système et de son

fonctionnement voire son label de qualité et, d'autre part, s'y opposer. L'obligation de '*privacy by default*', suppose en tout cas pour certains prestataires de services opérant de manière dominante sur le marché ou gérant un service d'utilité sociale (comme les plateformes d'information ou les réseaux sociaux) d'offrir une alternative effective et non discriminante (par exemple, par l'obligation pour ceux qui refuseraient de devoir 'payer' l'accès) au profilage publicitaire ou en tout cas comme certains sites le proposent déjà de moduler les utilisations possibles des données collectées.

Des profilages à risque élevé - Cinq cas nécessitent à notre avis des mesures complémentaires propres aux traitements dits à haut risques Au-delà des mesures déjà signalées, chacun de ces cas mérite cependant quelques commentaires additionnels. Le premier est certes, le profilage concernant des **jeunes** afin de contrer les risques supérieurs de manipulation. On peut penser à un système d'autorisation parentale en ce qui les concerne et en tout cas à l'intervention d'un comité interne d'évaluation. Le deuxième est le profilage qui permet le ***'dynamic pricing'*** … information sur cette caractéristique du profilage doit être donnée aux personnes concernées et les marges d'évaluation du prix, de même que les critères et en particulier leur caractère non discriminatoire appellent l'intervention de représentants des consommateurs dans l'évaluation des risques. On ajoute, troisième figure, la question du profilage utilisant ou non des **données sensibles** mais ayant pour effet de manière directe ou indirecte de révéler celles-ci, ce type de profilage doit être interdit sauf exceptions, en particulier lorsque la personne concernée s'étant vue offrir une alternative de non profilage, a cependant consenti explicitement au traitement de cette donnée, en rapport direct avec la qualité du service offert[59]. Une attention particulière, à propos de la possibilité de biais favorisant des discriminations fondées sur des données sensibles ou l'exclusion de certaines personnes peu importe le critère, doit être prêtée lors de tout *risk assessment* et lors de l'évaluation des tests préalables à l'utilisation des traitements de profilage.

Les deux derniers cas appellent également des commentaires dans la mesure où ils sont programmés pour prendre des décisions ayant un 'impact significatif' sur les personnes concernées. La quatrième hypothèse, le profilage ayant pour but de **sélectionner la clientèle** doit être opéré à partir de **critères pertinents en lien direct avec la transaction envisagée.** On peut imaginer en ce sens que certains critères soient exclus d'office. Il est important de maintenir un juste équilibre entre l'intérêt de l'entreprise, en particulier de crédit, de bien connaître leurs futurs clients selon l'exigence du droit résumé par l'adage : '*Know your customer*' et la nécessité de ne pas juger les personnes sur base de critères de décision, qui sortent des légitimes attentes raisonnables des personnes concernées. Enfin, il serait utile dans ce cas de prévoir la nécessaire labellisation possible voire obligatoire du traitement de profilage par l'autorité évoquée ci-dessus (*supra*, n°33) ou par des laboratoires sectoriels (par exemple : le secteur banque-assurance, le secteur logement, …), agréés par l'autorité(?) qui en toute indépendance pourraient évaluer les systèmes décisionnels envisagés. En cas de label, il serait intéressant de prévoir que cliquer sur l'icône du label renvoie à la description des critères utilisés pour le profil. Par ailleurs, on rappelle le principe de non-suffisance (nécessité de permettre la contestation de la décision automatisée) et accès à son profil sous une forme compréhensible suite au traitement.

La cinquième hypothèse vise les traitements de profilage utilisés dans le cadre de recrutement d'employés. Nombre de remarques développées à propos de la sélection de clientèle pourraient s'appliquer : nécessité d'offrir un droit de discuter les vérités sorties de l'ordinateur ; utilisation de critères pertinents par rapport à la candidature, accès au profil, On ajoutera qu'une labellisation des systèmes de profilage utilisés serait recommandée si possible après consultation des représentants des employés, en particulier s'il s'appuie sur

[59] Prenons un exemple, je souhaite de l'opérateur de ma plateforme musicale qu'il me présélectionne des musiques en rapport avec mes goûts. Dans ce cas, ma culture africaine de telle région d'Afrique peut être utile à cette sélection.

des techniques de reconnaissance d'émotions[60], étant donné les risques de biais et d'interprétation indue liés à l'utilisation de telles techniques.

Enfin, le risque de manipulation abusive par exploitation de la faiblesse d'autrui (*supra* sur cette notion, n°11) exige une attention particulière à la position économique, intellectuelle et sociale des parties en présence et aux techniques utilisées par le responsable de traitement pour amener la décision attendue du second.

2.6.2 Le profilage dans le cadre de l'exécution d'un contrat

Le profilage des performances des clients - L'utilisation de traitements de profilage dans la cadre de l'évaluation des 'performances' d'un client (par exemple d'une banque) ou d'un employé sont des traitements à haut risque par les conséquences que leur utilisation entraîne. Ils peuvent décider de l'octroi d'un crédit ou d'une promotion. A nouveau, l'exigence d'une réflexion sur les risques liés à ces traitements, prenant en compte le point de vue des différents intérêts (les représentants des consommateurs ou des employés) s'impose ; la validité des algorithmes et l'absence de biais devront être analysées (procédure de '*risk assessment*') et on rappellera ici les réflexions à propos d'une labellisation sectorielle.

Une information minimale devra être donnée sur les données prises en compte, le poids approximatif accordé à chaque critère et les conséquences attachées à chaque profil et ces points, discutés. Les personnes concernées auront accès aux données entrant en ligne de compte, leurs sources, une indication sur le poids de chaque critère, soit le 'profil' obtenu et les conséquences de ce 'profil', soit une explication en termes intelligibles sur le modèle suivi par l'algorithme. Enfin, doit être organisé le droit à une possibilité de contestation de la décision proposée par l'ordinateur, sans conséquences négatives pour celui qui l'exerce, etc. Cette possibilité de contestation pour être effective nécessite la mise sur pied au sein de l'entreprise d'une cellule compétente et suffisamment qualifiée pour remettre en cause la 'vérité' sortie de l'ordinateur.

2.6.3 Le profilage opéré par les autorités publiques (hormis le cas des autorités policières et des juridictions en charge de la poursuite des infractions pénales)

Le profilage par les autorités publiques - Le profilage exercé par les autorités publiques poursuit ou peut poursuivre diverses finalités, en particulier dans le cadre d'une politique proactive de '*benevolent government*' ou aux fins d'assurer l'effectivité des lois en instaurant le contrôle de leur respect par l'utilisation de système experts ou d'intelligence artificielle. La mise sur pied de tels systèmes de profilage devrait, d'une part, faire l'objet d'une 'loi' claire, proportionnées et répondant aux nécessités dans une société démocratique et, d'autre part, être subordonnée à l'analyse préalable et régulière par l'organisme multidisciplinaire et indépendant décrit ci-dessus (voir *supra* n° 33) tant du point de vue des algorithmes à retenir ou utilisés (qualité des données, absence de biais, etc.. ;), que de la sécurité du système. **La transparence des algorithmes et des sources nous apparaît requise dans la mesure où elle permet de répondre tant aux obligations d'accès aux documents publics qu'à l'exigence de motivation des décisions des autorités publiques.** Sans doute, faut-il concevoir des exceptions quand la transparence pourrait desservir l'intérêt général (ainsi, rendre transparents les critères retenus pour lutter contre la fraude sociale aurait un effet contreproductif) ? Deux précautions apparaissent par ailleurs à rappeler :

- mettre en place du point de vue organisationnel et managérial, 1.une procédure de réelle supervision des propositions de décisions sortis des traitements de profilage ; 2. l'information *a priori* par les autorités publiques de l'application d'un système de profilage, en même temps que les détails de son fonctionnement

[60] On note que ces systèmes dits *d'*'*affective computing*' peuvent également être utilisés dans les traitements de sélection de clientèle (4ème hypothèse).

et des sources sur lesquelles il s'appuie ; 3. une procédure de réelle écoute du point de vue de la personne concernée et de son droit à une simulation de l'algorithme à son cas.

- lorsque le traitement de profilage a pour conséquence une décision négative vis-à-vis de la personne concernée (par exemple, refus d'une aide sociale ou déclenchement d'une mesure d'inspection de sécurité sociale pour risque de fraude), créer une catégorie permettant d'indiquer clairement que la personne est répertoriée come telle en fonction d'un traitement de profilage.

En particulier, cet organe interviendrait lorsque le traitement de profilage est développé pour les besoins de l'administration. Ainsi, En avril 2018, l'*AI Now Institute* a défini un cadre pour les organismes publics qui souhaitent mettre en place des outils de prise de décision algorithmique, de permettre aux citoyens affectés par ces outils de faire appel des décisions prises à leur sujet. Les recommandations s'adressent également aux concepteurs de ces systèmes utilisés par l'autorité publique et déjà largement répandus. L'initiative *AI Now*[61] appelle ainsi à la fin de l'utilisation de systèmes opaques pour les décisions publiques, afin d'assurer l'équité et la régularité des procédures et de prémunir la population contre toute discrimination. Il s'agit aussi pour ces systèmes de respecter le droit à l'information du public, consacré par les directives européennes sur l'accès aux documents du secteur public et, au-delà, de répondre au devoir de motivation qui incombe à l'autorité publique dans ses décisions.

Pour ce faire est préconisé[62] le recours à un audit régulier par l'autorité qui disposerait d'un centre d'évaluation des systèmes IA, centre de test et d'audit géré de manière indépendante de l'Etat. Ce Centre devrait également veiller à améliorer l'expertise des organismes qui les conçoivent les systèmes IA utilisés par l'autorité et son administration. Il fixerait les procédures et modalités tant organisationnelles que techniques des modalités offertes aux citoyens afin de contester les décisions prises sur base des systèmes IA. L'initiative recommande aux agences publiques de répertorier et décrire les systèmes de décision automatisés, y compris d'évaluer leur portée et impact. Elle recommande également de mettre en place des modalités d'accès afin que des chercheurs, des experts indépendants, des associations ou des journalistes puissent accéder et évaluer ces systèmes et pour cela doivent s'assurer notamment que leurs fournisseurs privés de systèmes acceptent ces vérifications. Elle souligne également que les agences doivent monter en compétences pour être expertes des systèmes qu'elles mettent en place, notamment pour mieux informer le public, et invite les fournisseurs de solutions à privilégier l'équité, la responsabilité et la transparence dans leurs offres. Cela permettrait également aux organismes publics de développer des procédures de médiation, d'appel ou de réfutation des décisions prises. Obliger les systèmes à publier des analyses d'impact de leurs outils de décision automatisée pourrait enfin permettre au public d'évaluer les outils et la transparence des services.

3 Analyse de l'application des différentes dispositions de la Convention 108 +

Objet du chapitre III - Le souci de ce chapitre est d'attirer l'attention sur le fait que les dispositions de la Convention 108+ appliquées aux traitements de profilage méritent quelques interprétations qui pourront servir de base aux recommandations.

[61] Comme l'écrivait dès 1971, G. BRAIBANT, « (il) faudrait imposer à l'autorité publique, chaque fois qu'elle se fonde sur les résultats d'un traitement par l'informatique, de faire connaître les données et les programmes à partir desquels ces résultats ont été obtenus ; ces données et programmes pourront ainsi faire l'objet de discussions susceptibles de remettre en cause leurs résultats. ». (G. Braibant, « La protection des droits individuels au regard du développement technologique », Revue internationale de droit comparé, 1971, 23, p. 812)

[62] Voir en particulier le rapport (anglaise), Select Committee on Communications, *Regulating in a digital world,* 2nd *Report of Session 2017-19*, House of lords, published 9 March 2019.

3.1 Objet et but de l'intervention en matière de traitements de profilage

Des risques individuels aux risques collectifs : où la protection des données s'avère insuffisante - Nous l'avons dit et redit, les Lignes directrices récentes du Comité de la Convention insistent lourdement sur ce point. Comme le note le Rapport Mantelero, qui introduit les Lignes directrices relatives aux mégadonnées, *« Bien sûr, la protection des données en soi ne couvre pas tous ces aspects qui requièrent une approche plus large englobant les droits de l'homme et les questions de société ».* Cette réflexion a de nombreuses conséquences en ce qui concerne les traitements de profilage. Sans doute, est-il absolument nécessaire de passer d'une conception de '*Data Protection Risk Assessment*' et d'une approche '*Data Protection by Design*' à une conception et approche plus globale *d''Ethical values Assessment* et *Design'* où les préoccupations de justice sociale, de diversité culturelle, de dignité humaine et de démocratie soient également présentes à côté de celles liées à la seule protection des données. A cet égard, il ne s'agit pas de tourner le dos à la Convention ni à la cause de la protection des données. Comme le note très finement, le rapporteur cité : « L'accent mis sur l'individu, la prise de conscience des conséquences sociales de l'usage des données et le lien avec les droits de la personnalité peuvent élargir l'approche du responsable de traitement : au-delà de la protection ces données, il s'agit de garantir les droits fondamentaux et les intérêts collectifs. La protection des données contribue ainsi à titre complémentaire à révéler la façon dont sont utilisées les données et les finalités du traitement, éléments essentiels pour mieux appréhender les conséquences potentielles sur divers droits et libertés*. ».* Cette prise de position justifie notamment l'élargissement de la portée des risques à envisager lors de l'évaluation d'un traitement de profilage et la présence à côté des représentants de la protection des données d'autres intérêts. Comme le notent Schultz et Dreyer dans le rapport déjà cité : *The GDPR and Automated Decision making : Will it deliver ? : « Pourtant*, le RGPD n'offre pas un grand potentiel lorsqu'il s'agit de protéger les intérêts des groupes et de la société tels que la non-discrimination, l'inclusion sociale ou le pluralisme ... il est nécessaire d'adopter une approche complémentaire*. »*

3.1.1 Définitions

Nous nous attarderons à deux définitions : celle de 'données à caractère personnel' et celle de 'responsable du traitement'.

Données à caractère personnel

La notion de données à caractère personnel - La première notion appelle les commentaires suivants. Il est important de noter que désormais la notion de données à caractère personnel s'élargit au-delà du critère de l'identifiabilité, qui font référence à la possibilité pour le profileur de relier les données collectées de manière directe ou indirecte à l'identité de la personne concernée, aux critères de la contractabilité et la possibilité d''impacter' la personne concernée. Il s'agit de passer de l'identification à l'individuation. Ainsi, le tag RFID porté par une personne X se baladant dans un supermarché, ne révélera sans doute pas l'identité de son porteur mais permettra de le localiser au sein de ce supermarché et de lui envoyer la publicité d'un produit placé à proximité. Cette réflexion soulève la question de l'exercice des droits par la personne concernée individuée certes mais non identifiée. Il est donc important que cette personne puisse exercer ses droits (par exemple celui d'opposition ou simplement d'accès) en ligne sans révéler son identité. Par ailleurs, les recommandations doivent veiller à ne pas limiter leur objet aux seules données à caractère personnel mais doivent envisager les **données anonymes** qui servent également à la définition du profil. Ainsi, donner une information à la personne concernée sur les seules données à caractère personnel intervenues dans la fabrication du profil en excluant les données anonymes constituerait une information incomplète voire sans signification. C'est au regard de la finalité de l'ensemble des opérations, qui mènent au profilage d'une personne ou d'un groupe, que le caractère anonyme ou non d'une donnée doit se juger. A cet égard, nous rappelons la nécessité de défendre une vision holistique des traitements de profilage (voir la

Recommandation de 2010). Il ne peut être question de séparer les diverses étapes du traitement de profilage (*data collection, data mining* et application à une personne concernée ou un groupe) et juger chacune d'elles pour elles-mêmes. C'est à partir de l'étape finale que doit être jugée chaque autre étape. Par ailleurs, il nous apparaît que le profilage (voir supra notre réflexion sur les objectifs et but) constitue un risque majeur de discrimination de groupes voire d'entreprises. Cette préoccupation ramène à des réflexions sans doute à raviver autour de la protection des groupes qu'il s'agisse de groupes ethniques ou de groupes de personnes caractérisés par un profil commun (les habitants de tel quartier sont des criminels en puissance, tel quartier est condamné à terme en ce qui concerne le développement industriel, etc.).

Qu'en est-il des personnes morales ? - Par ailleurs, cette même préoccupation prolonge une discussion toujours présente au sein du Conseil de l'Europe[63], en ce qui concerne la **protection des personnes morales**, vis-à-vis desquelles le profilage est une réalité de plus en plus importante et aux conséquences parfois désastreuses. La liberté d'entreprendre est mise à mal par la dissymétrie d'informations qui actuellement caractérise le marché de l'information. Certaines plateformes, certains opérateurs de services d'information, les banques, les assureurs disposent, grâce au numérique, d'outils puissants y compris d'IA capables de prédire l'avenir d'une entreprise ou d'une filiale voire de les condamner à travers des *rankings* négatifs. Protéger en particulier les entreprises petites et moyennes contre ces traitements de profilage m'apparaît nécessaire, en ce compris pour les travailleurs y occupés et les territoires où ces entreprises sont installées. Affirmant cela, **nous ne prônons pas l'extension de la Convention 108+ aux personnes morales- il y va d'une autre logique plus proche du droit de la concurrence que des droits du citoyen mais ceci dit, que des droits subjectifs d'accès, de rectification, d'effacement copiés de ceux conférés aux individus, la protection contre les biais, … soient octroyés aux personnes morales non dans le cadre de notre recommandation mais bien d'une réglementation à portée plus économique serait utile voire nécessaire à l'heure des prédictions de l'IA.** On évoquera à ce propos, les droits nouveaux accordés par le règlement européen récent 2019/1150 du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne (JOUE 11.7. 2019, L 186/57) qui entend protéger l'entreprise utilisatrice des services de la plateforme, notamment en ce qui concerne leur *ranking.*

La notion de responsable de traitement

Pluralité d'acteurs, diversité des montages et qualification des acteurs - Notre introduction insistait sur la diversité des montages qui préside à la mise sur pied d'un système de profilage, en particulier lorsqu'il s'appuie sur des technologies d'intelligence artificielle. Ainsi, les données peuvent provenir de sources diverses, ayant chacun leurs responsables de traitement, la fourniture de jeux de données, sur lesquels des algorithmes ont déjà tourné, peut être le fait d'un autre acteur ; certains algorithmes de base sont disponibles sur le web en open source, d'autres sont acquis auprès de leurs concepteurs sous forme de licences privées. L'adaptation de ces différents éléments aux besoins d'un secteur ou d'une entreprise déterminée est souvent le fait d'un acteur spécialisé. Enfin, l'application du traitement de profilage aux personnes ou aux groupes peut être confiée à un tiers, en particulier comme un service 'clé sur porte' offert par une plateforme. En effet, cette dernière dispose déjà de nombre de données relatives aux personnes concernées et souvent utilise déjà à son profit des algorithmes de profilage. Elle peut comme nous l'avons souligné (*supra,* n°21) mettre de tels algorithmes à disposition de tiers, voire les résultats de ces algorithmes. L'intervention de tous ces acteurs pose la question de la qualification de chacun. On connaît la position des deux Lignes directrices du Comité de la Convention 108. Elles vont nous l'avons dit dans le sens d'un élargissement des obligations de chaque acteur qui de près ou de loin participent à la conception, à la mise sur pied et à l'exploitation des traitements de profilage. Cette extension peut prendre deux directions.

Pour certains fournisseurs d'éléments nécessaires au fonctionnement d'un système de profilage et offert commercialement sur le marché (un algorithme de base, un jeu de données, une base de données), ce qui

[63] J.P. Walter, Le profilage des individus à l'heure du « cyberespace » : un défi pour le respect du droit à la protection des données

est requis c'est certes la qualité du produit, la description des limites de celui-ci et la collaboration, le cas échéant, dans le cadre de l'évaluation des risques et lors de la phase de tests. Peut-on aller plus loin et considérer vis-à-vis du traitement de profilage mis en cause, que certains de ces acteurs sont sous-traitants voire responsables conjoints de ce traitement ? La décision *Akademia* de la CJUE va dans ce sens d'une responsabilité parfois conjointe d'acteurs participant l'un et l'autre au profilage. En l'occurrence, une ASBL gérante du club des fans d'un chanteur avait souhaité profiler ses 'clients', en l'occurrence les visiteurs du blog d'un chanteur. Pour ce faire, elle avait obtenu (non gratuitement d'ailleurs) les données relatives aux visiteurs du blog de la part du gestionnaire de la plateforme par laquelle les visiteurs se connectaient au blog, à savoir Facebook. Bref, la première, l'ASBL, définissait la finalité mais la seconde : la plateforme fournissait les moyens. Cette dernière a donc été jugée co-responsable ou pour reprendre l'appellation de la Convention du Conseil de l'Europe, d'un co-décideur partageant avec l'ASBL le pouvoir de décision à l'égard du traitement : *«* Afin de déterminer si un organe ou une personne peut être qualifiés de responsable du traitement, une attention particulière doit être portée au fait de savoir s'il ou elle détermine les motifs justifiant le traitement, à savoir ses finalités, ainsi que les moyens utilisés. D'autres facteurs pertinents dans cet exercice de qualification comprennent le fait de contrôler ou non les méthodes du traitement, le choix des données à traiter et les personnes autorisées à y accéder. *»*. Si cette qualification de responsable conjoint n'est pas retenue dans certains cas, celle de sous-traitant pourrait alors l'être. Dans ce cas, il importera 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 et que certaines obligations lui soient imposées comme celle de notification en cas de brèches de sécurité, de *risk assessment*.

Légitimité du traitement et qualité des données (art. 5)

Des principes de base - Les principes de proportionnalité des traitements par rapport à la finalité légitime poursuivie et de juste équilibre entre tous les intérêts en présence sont affirmés comme principes de base. Soulignons d'abord que le profilage n'est pas une finalité en soi mais constitue une technique qui permet d'atteindre une des nombreuses finalités déjà décrites dans la classification proposée. C'est donc au regard de ces finalités que les deux principes de légitimité et de proportionnalité doivent guider la réflexion des personnes responsables de traitement de profilage et ce, insiste l'article 5.1., dès la conception du traitement et tout au long de celui-ci. Ces principes portent sur le traitement et renvoie donc à la question d'une motivation du recours au profilage envisagé et la possibilité de choix à ce propos pour la personne concernée : le profilage est-il bien nécessaire et ne faut-il pas laisser le choix à la personne concernée entre l'accès non profilé et celui 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 3.7 de 2010 qui réclame : « Dans la mesure du possible et à moins que le service requis nécessite de connaître l'identité de la personne concernée, toute personne devrait avoir accès aux informations relatives à un bien ou à un service ou avoir accès à ce bien ou à ce service sans devoir communiquer de données à caractère personnel au fournisseur du bien ou au prestataire du service. Aux fins d'assurer un consentement libre, spécifique et éclairé au profilage, les prestataires de services de la société de l'information devraient assurer, par défaut, un accès non profilé aux informations relatives à leurs services. » Au-delà de cette première réflexion, il importe que la personne concernée puisse dans certains cas, en particulier le profilage publicitaire, définir la finalité du profilage qu'il 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 qui, 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 les diverses finalités et, le cas échéant, sur les destinataires qui permettront de réaliser les finalités. Ainsi, pour reprendre l'exemple du service de musique en ligne, peut-être, mon souhait est que je puisse recevoir l'annonce publicitaire émanant de tiers à propos de la sortie d'une chanson de mon interprète favori. A l'inverse, il 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.

Le besoin d'une évaluation - Ces principes justifient également l'**évaluation participative, multidisciplinaire et si possible *multistakeholders* des risques** et exigent que cette évaluation soit continue. La continuité de l'évaluation est d'autant plus nécessaire que les systèmes d'IA qui appuient nombre de traitements de profilage peuvent, au gré des nouvelles données collectées et de la découverte de nouvelles corrélations possibles, voir leurs finalités évoluer par ceux qui les exploitent. Ainsi, un traitement de profilage au départ conçu pour des opérations de publicité *one-to–one* peut se doubler à partir de quelques données supplémentaires et de nouveaux algorithmes en logiciel d'exclusion ou d'évaluation des 'clients'. L'administration des prisons peut, à partir des données relatant la situation, le comportement et l'environnement des prisonniers, développer un traitement de profilage qui lui permettra d'aider aux décisions en ce qui concerne les libérations anticipées. La collecte d'informations de localisation réalisées à partir de nos mobiles permet certes à l'utilisateur de ce dernier de trouver aisément un restaurant ou un hôtel à proximité (finalité de base) mais peut ensuite dans le cadre d'une cession des données à des tiers ou par le même opérateur que celui du traitement de base être utilisé, par exemple, dans le cadre de publicités pour des grands magasins ou autres prestataires de produits ou de services.

Systèmes IA et principes de finalité - Le respect du principe de finalité déterminée pose donc une **difficulté face à des systèmes d'information, complexes sans cesse évolutifs** et capables, au gré de l'aléa du fonctionnement d'un système d'IA, de se voir assigner de nouvelles tâches liées aux opportunités du « *machine learning ».* Cette question se pose non seulement à propos des traitements de profilage et soulève la question de la façon dont un principe aussi fondamental que celui des finalités déterminées peut être respecté. L'article 5.4. permet d'assouplir l'application du principe grâce à la non interdiction de « finalités compatibles » avec celles de départ, à savoir celles poursuivies par le traitement initial. Les Guidelines de l'Union européenne de 2017 rappellent les critères d'examen du critère de compatibilité énoncés par l'article 6.4 du RGPD[64] : « le responsable du traitement, afin de déterminer si le traitement à une autre fin est compatible avec la finalité pour laquelle les données à caractère personnel ont été initialement collectées, tient compte, entre autres :

a) de l'existence éventuelle d'un lien entre les finalités pour lesquelles les données à caractère personnel ont été collectées et les finalités du traitement ultérieur envisagé ;
b) du contexte dans lequel les données à caractère personnel ont été collectées, en particulier en ce qui concerne la relation entre les personnes concernées et le responsable du traitement ;
c) de la nature des données à caractère personnel, en particulier si le traitement porte sur des catégories particulières de données à caractère personnel, en vertu de l'article 9, ou si des données à caractère personnel relatives à des condamnations pénales et à des infractions sont traitées, en vertu de l'article 10 ;
d) des conséquences possibles du traitement ultérieur envisagé pour les personnes concernées ;
e) de l'existence de garanties appropriées, qui peuvent comprendre le chiffrement ou la pseudonymisation. »

Cette longue liste de critères peut servir à éclairer le ou les responsables et doit être prise en considération lors de l'évaluation continue des traitements de profilage et, en tout cas, tout au long de la mise sur pied du traitement, c'est-à-dire dès la conception initiale, dans le choix des algorithmes, des jeux de données, de la définition des tests, etc. Les fournisseurs d'éléments comme des bases de données prêteront également attention à ces critères sur base de leurs devoirs d'information et de conseil. On soulignera que la finalité générique de certaines de ces bases de données (par exemple un jeu de photos permettant le

[64] Pour les traitements de l'autorité publique y compris les traitements policiers, il nous semble que seule l'exécution de missions publiques peut légitimer ce 'glissement de finalités', le tout sous contrôle de l'organe décrit précédemment. Par ailleurs, le Conseil de l'Europe autorise, moyennent des 'garanties complémentaires' (règles d'accès aux données, pseudonymisation, agrément des labos de recherche, principe de non commercialisation des résultats, …), la possibilité de traitements de données collectées à des fins primaires (par exemple pour les soins à donner à des patients) vers en l'occurrence de profilage à des finalités de recherche scientifique, historique ou statistique (dans l'exemple en matière de recherche contre la cancer).

fonctionnement d'algorithmes de reconnaissance faciale utilisables dans différents contextes[65]) exige cette attention particulière. C'est donc, comme le notent les lignes directrices sur les mégadonnées (point 2.5.), au regard de chaque application que le responsable des '*data sets'* mis sur le marché commercial doit « identifier et d'évaluer les risques de chaque activité de traitement de mégadonnées et de ses incidences potentiellement négatives sur les droits et libertés fondamentales des personnes, en particulier le droit à la protection des données à caractère personnel et le droit à la non-discrimination, en tenant compte des impacts sociaux et éthiques *»* .

Qu'en est-il des opérateurs de plateforme ? Un point particulier concerne les opérateurs de plateformes. L'activité liée à l'utilisation par les internautes des différents sites rendus accessibles ou en tout cas accédés par ces opérateurs génère des données que ces opérateurs considèrent comme leurs. Or le traitement de ces données d'utilisation des sites accédés (non les simples données relatives aux choix des sites consultés) excède en principe la finalité poursuivie par l'internaute qui certes emprunte la plateforme de l'opérateur mais n'entend pas nécessairement les voir utiliser par ce dernier à des fins de profilage ou de commercialisation vers des tiers. On peut imaginer certes que ces utilisations par l'opérateur de la plateforme soient acceptées voire souhaitées par l'internaute qui désire que l'opérateur tienne compte de son profil pour sélectionner les sites adéquats en fonction de sa personnalité voire mais ce sera plus rare, renseigne des tiers sur son comportement et les caractéristiques de sa personnalité mais ces utilisations ne sont pas légitimes *a priori* et ne devraient être permises que sur base du consentement de la personne concernée[66].

---

[65] Ces bases de données peuvent servir à des traitements pour des applications variées développées par le détenteur des données ou par des tiers (par exemple, la définition de stratégies, le profilage de clientèles, la recherche scientifique, …).

[66] A ce propos, mais avec une préoccupation différente, à savoir celle de protéger l'entreprise utilisatrice des services de la plateforme, le règlement européen récent 2019/1150 du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne (JOUE 11.7. 2019, L 186/57) dispose (Considérant n° 34) : « Dans le même esprit, il est important, pour les entreprises utilisatrices, de savoir si le fournisseur partage avec des tiers toute donnée qui a été générée par l'utilisation du service d'intermédiation par l'entreprise utilisatrice. Les entreprises utilisatrices devraient notamment être informées de tout partage de données avec des tiers qui répond à des finalités qui ne sont pas nécessaires au bon fonctionnement des services d'intermédiation en ligne, par exemple lorsque le fournisseur du service tire profit de ces données à des fins commerciales. Afin de permettre aux entreprises utilisatrices de faire pleinement valoir leur droit à avoir leur mot à dire sur ce partage de données, les fournisseurs de services d'intermédiation en ligne devraient également informer clairement les entreprises utilisatrices des possibilités de refuser ledit partage lorsqu'une telle possibilité est prévue par leur relation contractuelle avec l'entreprise utilisatrice. » L'article 9 prévoit ainsi explicitement une obligation de transparence de l'opérateur vis-à-vis des entreprises utilisatrices, en particulier : « Par la description visée au paragraphe 1, les fournisseurs de services d'intermédiation en ligne informent de manière appropriée les entreprises utilisatrices en particulier des éléments suivants: a) la question de savoir si le fournisseur de services d'intermédiation en ligne a accès aux données à caractère personnel ou à d'autres données, ou aux deux, que les entreprises utilisatrices ou les consommateurs transmettent pour l'utilisation de ces services, ou qui sont produites dans le cadre de ces services, et dans l'affirmative, les catégories de ces données qui sont accessibles et les conditions applicables; b) la question de savoir si une entreprise utilisatrice a accès aux données à caractère personnel ou à d'autres données, ou aux deux, qu'elle transmet dans le cadre de son utilisation des services d'intermédiation en ligne concernés, ou qui sont produites dans le cadre de la fourniture de ces services à ladite entreprise utilisatrice et aux consommateurs de ses biens ou services, et dans l'affirmative, les catégories de ces données qui sont accessibles et les conditions applicables; c) outre le point b), la question de savoir si une entreprise utilisatrice a accès aux données à caractère personnel ou à d'autres données, ou aux deux, y compris sous forme agrégée, qui sont transmises ou produites dans le cadre de la fourniture des services d'intermédiation en ligne à toutes les entreprises utilisatrices et à leurs consommateurs, et dans l'affirmative, les catégories de ces données qui sont accessibles et les conditions applicables; et d) la question de savoir si des données visées au point a) sont transmises à des tiers, ainsi que, lorsque la transmission de telles données à des tiers n'est pas nécessaire au bon fonctionnement des services d'intermédiation en ligne, des informations précisant le but d'un tel partage de données, ainsi que les possibilités dont disposent les entreprises utilisatrices de ne pas participer à ce partage de données. ». Sans doute faudrait-il que ces informations soient également transmises aux internautes qui doivent pouvoir connaître dans quelle mesure

Principe de minimisation ou de proportionnalité des données traitées - La question de la proportionnalité ou de minimisation des données au regard de la finalité pose d'autres difficultés. Les systèmes dits de *machine learning* fonctionnent sur base de corrélations statistiques sur base de rapprochements arbitraires de données. Ainsi, hypothèse purement fictive, il peut apparaître aux yeux de l'administration fiscale 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 sont 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. Réaffirmer le principe de pertinence voire de nécessité des données au regard de la finalité exigerait ce travail or, à défaut d'avoir une interprétation large du critère, les utilisateurs de systèmes IA opposeront à l'interprétation stricte, que les exigences mises freinent l'innovation et interdisent de donner sa pleine efficacité au profilage. L'exemple de la recherche de criminels appuie cet argument. C'est parfois à partir de corrélations inattendues, que les auteurs d'infraction peuvent être retrouvés Ce point est délicat. Les Guidelines maintiennent le principe : « Les responsables de traitement doivent s'assurer qu'ils appliquent le principe de minimisation de données » … et voient le principe de pseudonymisation comme une solution.

Sans reprendre toutes les exigences posées par l'article 5.4 (loyauté, finalités déterminées, pertinence des données, exactitude et mise à jour des données, durée de conservation, ...), pointons quelques interrogations qui devront être résolues dans un premier temps par une approche sectorielle et impliquant une discussion *multistakeholders,* dans un second temps, le cas échéant, par une intervention des autorités de protection des données voire le législateur.

-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 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. ?

De la durée du traitement et de la qualité des données - Deux réflexions supplémentaires en ce qui concerne la durée du traitement et la qualité des données s'ajoutent. En ce qui concerne la durée de conservation des données, n'est-il pas utile de prévoir pour certaines données, l'interdiction d'utilisation au-delà d'une certaine période (par exemple, une donnée relative à une opération chirurgicale ne devrait plus être utilisée après une durée X de rémission de la maladie), de même la trace d'un accident si d'autres accidents ou de risques d'accidents (par ex. : détection d'un taux d'alcoolémie,) ne sont plus signalés. Au-delà, le profilé ne doit-il pas pouvoir exiger que le responsable ne prenne plus en compte des données le concernant au-delà d'une certaine durée, cette durée pouvant varier en fonction de la nature des données et de la finalité poursuivie par le profilage.

La qualité des données lorsqu'elles proviennent d'une *big data* d'un tiers devrait pouvoir être documentée par ce tiers voire certifié par un organisme agréé. « Outre la protection de la vie privée et des données personnelles, des exigences doivent être remplies pour garantir des systèmes d'IA de haute qualité. La qualité des ensembles de données utilisés est primordiale pour la performance des systèmes d'IA. Lorsque des données sont recueillies, elles peuvent refléter des préjugés sociaux ou contenir des inexactitudes, des erreurs et des fautes. Il convient d'aborder ce problème avant d'entrainer un système d'IA avec tout ensemble de données. En outre, **l'intégrité** des données doit être garantie. Les processus et les ensembles de données utilisés doivent être testés et documentés à chaque étape comme la planification, l'entrainement, les tests et le déploiement. Cela doit également s'appliquer aux systèmes d'IA qui n'ont pas

leurs utilisations de tel ou tel site accédé par une plateforme sont utilisées par la plateforme elle-même voire commercialisées à des tiers !

été développés en interne mais acquis ailleurs. Enfin, l'accès aux données doit être régi et contrôlé de manière adéquate. »

La question cruciale du consentement au profilage - Certes, le consentement est une cause de légitimité et on peut estimer, comme déjà noté, que le traitement de profilage correspond à un réel avantage pour la personne concernée qui souhaite se voir conseiller dans le dédale des produits et services offerts. Notons cependant que dans ce cas, on voit mal en quoi le consentement se distinguera du consentement à un contrat dans la mesure où la personne concernée reconnaît que le traitement de ses données et autres est nécessaire au service de profilage souhaité. On objectera que le consentement est révocable mais le contrat qui permet au responsable de traitement de profiler son client peut être à durée indéterminée et être révocable *ad nutum.* Par ailleurs, l'idée d'une négociation collective des conditions contractuelles entre associations de consommateurs et responsable de traitement lorsqu'il s'agit de profilage proposé à grande échelle nous apparaît préférable en termes de protection de la vie privée des personnes concernées que le consentement individuel qui laisse un individu seul face à la puissance économique du responsable de traitement[67].

Cela dit, soulignons que les qualités du consentement requises par l'article 5.2. seront rarement remplies. Elles supposent, sous réserve d'autres exigences encore, un consentement libre, c'est-à-dire non manipulé (voir *supra)* et sur base d'un choix réel qui soit autre que le 'j'accepte' ; il ne peut être la condition d'accès à des services souvent ressentis comme nécessaires à l'exercice de la vie sociale et devraient être liés, sauf nécessité lié au service lui-même (par exemple, un service de confection de vêtement sur mesure offert à distance ou d'éducation à distance) à l'offre par défaut d'un service non profilé; les requis de l'article 5.2. nécessitent par ailleurs un consentement éclairé, ce qui soulève nombre de problèmes (voir notre commentaire de l'article 8). Par ailleurs, le consentement suppose et la condition est clairement exprimée par le rapport du Conseil de l'Europe[68] le respect des autres conditions de légitimité énoncées aux points 1, 3 et 4 de l'article 5. Ce qui suppose notamment pour les traitements de profilage à risque élevé, le respect des procédures et des conditions déjà énoncées et pour tous les traitements de profilage, l'évaluation des risques et le respect impératif même si interprétés largement des principes de proportionnalité, finalité et de sécurité.

D'autres questions se posent encore à propos de ce 'consentement' : à quelles conditions admettra -t-on le consentement au traitement de profilage contre 'rémunération' ? A cet égard, tout dogmatisme nous paraît à écarter. Tout dépend de la hauteur de cette 'rémunération' par le responsable, des données en question, des finalités du profilage envisagé, des limites mises aux possibilités de profilage et, surtout, d'une négociation avec non seulement les autorités de protection des données que des représentants des consommateurs.

Les catégories particulières de données (article 6)

Considérations liminaires - On se contente de rappeler que la sensibilité d'un traitement est certes souvent due à la nature en soi des données qui la constituent mais peut-être également, dans le cadre du dernier tiret de l'article 6.1., lié directement à la finalité du traitement de données non sensibles en soi mais dont le traitement révèle l'origine raciale ou ethnique (exemple sélectionner tous les noms terminant en 'SKI' de façon à retrouver l'origine 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. On insistera sur le fait que de nombreux biais conscients ou inconscients[69].peuvent ainsi

[67] Sur tous ces points, lire notamment, Y. Poullet, « Consentement et RGPD : des zones d'ombre ! », DCCR, 2018, p. 3 – 39.

[68] « Les paragraphes 1, 2, 3 et 4 de l'article 5 sont cumulatifs et doivent être respectés pour garantir la légitimité du traitement des données. » (Rapport explicatif, n°41 ; voir aussi, le n°44)

69 J. Grimmelmann and D. Westreich, "*Incomprehensible Discrimination*", California Law Review, 2017, vol. 7, pp. 170–176. A cet égard, les réflexions de L. Edwards and M. Veale, "*Enslaving the algorithm: from a 'right to an explanation' to a 'right to better decisions*'?", IEEE Security and Privacy Magazine, 2018, n.° 16, p. 52: Par exemple, les préjugés

affecter les outils de fonctionnement utilisés par le traitement de profilage tant dans la sélection des données de base que des algorithmes choisis. On ajoutera que ces traitements de profilage sont à haut risque dans la mesure où les risques de discrimination et de limitation de libertés religieuses, syndicales ou philosophiques attachées à ces données sont importants (voir l'article 6.2.). Enfin que le profilage de groupes (et non seulement d'individus) opéré sur de telles bases doit également être réglementé.

Sécurité des données (art. 7)

La notion de sécurité : une acceptation large - L'utilisation de systèmes d'IA dans le cadre de traitement de profilage souligne la nécessité de définir l'intégrité des traitements comme la recherche et l'élimination (obligations de moyens) des biais qui pourraient affecter le traitement. A nouveau le fait de recourir à des algorithmes de tiers[70] obligera le responsable, suivant le principe du '*reasonable coder*'[71] à réclamer le résultat de tests subis par le producteur d'algorithmes voire un label de qualité fourni par un organisme indépendant, si du moins le traitement entraîne des risques importants pour les individus ou la société[72]. On note que bien souvent ces algorithmes sont développés par des centres de recherche scientifique qu'ils soient d'entreprises privées ou d'universités ont fait l'objet de publications amplement commentées et souvent après avoir subi les critiques de tiers (voir Chapitre 1). On souligne en outre que dans certains secteurs, en particulier le secteur médical, les algorithmes sont labellisés par des experts du secteur.

L'évaluation de la 'sécurité' dans ces trois acceptions : intégrité, confidentialité, disponibilité doit être entamée dès la conception du système de profilage et être continue (voir déjà la Recommandation de 2010). Elle nécessite un rapport identifiant les risques tant d'impact négatif externe vis-à-vis des individus et de la société qu'une analyse des risques techniques et de biais que présentent le choix des données et du système de fonctionnement ou dus à l'environnement (risques d'attaques externes). Il s'agit également de justifier les choix. La question de la réversibilité des conséquences négatives ou illégitimes du fonctionnement du système est également soulevée. Le principe de sécurité implique l'application maximale des principes de '*ethical values by design*'. Comme le note la recommandation n° 1.4 de l'OCDE à propos de la sécurité des traitements utilisant l'IA : « Les systèmes d'IA devraient être robustes, sûrs et sécurisés tout au long de leur cycle de vie, de sorte que, dans des conditions d'utilisation normales ou prévisibles, ou en cas d'utilisation abusive ou de conditions défavorables, ils soient à même de fonctionner convenablement, et ne fassent pas peser un risque de sécurité démesuré. »

Pour ce faire, les acteurs de l'IA devraient veiller à la traçabilité, notamment pour ce qui est des ensembles de données, des processus et des décisions prises au cours du cycle de vie des systèmes d'IA, afin de

peuvent entraîner une discrimination dans la prise de décision automatisée sur la base de la couleur de la peau ou d'autres attributs protégés par la loi. Ces préjugés peuvent reposer directement sur ces attributs, ou indirectement. Dans le second cas, les outils d'IA détectent les corrélations entre divers attributs non protégés pour obtenir les mêmes résultats discriminatoires que s'ils prenaient en compte des attributs protégés par la loi (approximations). Par exemple, l'utilisation de l'adresse d'une personne (c'est-à-dire dans une région où il y a une forte concentration de personnes ayant la peau noire) et de son salaire moyen pour refuser un crédit, alors qu'une personne ayant les mêmes attributs mais vivant dans une région où la plupart des gens ont la peau blanche recevrait le crédit. Quelle serait votre attitude si le problème n'était pas la couleur de la peau des personnes mais simplement l'adresse ?

[70] La question se pose lorsque l'algorithme est '*open source*' fourni gratuitement par exemple par un organisme de recherche. De lourdes obligations mises à charge de tels concepteurs seraient contreproductives et empêcheraient la circulation de 'produits' susceptibles d'amélioration par ceux qui les utilisent et donc d'innovations

[71] P. Terzis, « *The reasonable coder* », article en projet disponible sur *Academia edu*, 2019

[72] Voir déjà la Recommandation Profilage de 2010, article 9.2 : « Par ailleurs, dans le cas de traitements ayant recours au profilage et présentant des risques particuliers au regard de la protection de la vie privée et des données à caractère personnel, les Etats membres peuvent prévoir : a. que les responsables des traitements soient tenus de les notifier préalablement à l'autorité de contrôle ; ou b. que ces traitements fassent l'objet d'un contrôle préalable par l'autorité de contrôle. »

permettre l'analyse des résultats produits par lesdits systèmes d'IA et le traitement des demandes d'information, compte tenu du contexte et de l'état de l'art de la technologie.

Les acteurs de l'IA devraient, selon leurs rôles respectifs, le contexte et leur capacité à agir, appliquer de manière continue une approche systématique de la gestion du risque, à chaque phase du cycle de vie des systèmes d'IA, afin de gérer les risques y afférents, notamment ceux liés au respect de la vie privée, à la sécurité numérique, à la sûreté et aux biais.'[73]

Transparence du système (article 8)

Information sur quoi : logique sous-jacente ? Les informations spécifiquement prévues dues sur base de l'article 8.1. méritent peu de commentaires. L'article prévoit cependant des obligations complémentaires d'information en son alinéa 2 lorsque cette information est nécessaire « pour garantir un traitement loyal et transparent des données à caractère personnel ». Il va de soi que toute personne soumise à un profilage doit être avertie de l'existence de ce profilage via une icône qui lui permet d'accéder aux informations de base et complémentaires exigées des responsables de traitement. Pour les traitements de profilage à haut risque, on ajoutera une information sur la procédure suivie en ce qui concerne l'évaluation participative (l'accès au rapport peut être requis en outre par l'autorité de protection des données) et une information sur l'impact du profilage sur la situation du profilé. C'est pour cette raison également qu'en outre, le RGPD réclame une information sur la 'logique sous-jacente' au système, ce qui était les termes mêmes de la Recommandation du Conseil de l'Europe de 2010.

Pour aller plus loin - Plusieurs réflexions à propos des termes : 'logique sous-jacente' :

> le terme « logique sous-jacente », s'il est adéquat en ce qui concerne des systèmes experts classiques, traduisant la logique des experts dans des types de situations ou de décisions donnés, est par contre inadéquat lorsqu'il est fait usage de systèmes de '*machine learning*' travaillant sur des corrélations statistiques et non plus sur des logiques causales. Dans de tels cas, il est question de réclamer au minimum l''explicabilité' (*explainability*) du fonctionnement du système. Ainsi, le High Level Group of Experts de la Commission européenne réclame une explicabilité très large visant non seulement l'explicabilité du fonctionnement de l'algorithme mais son rôle dans le processus organisationnel, la motivation du recours à l'IA : « En relation à cela, l'**explicabilité** du processus de prise de décision algorithmique, adaptée aux personnes impliquées, devrait être fournies autant que possible. La recherche en cours pour développer des mécanismes d'explicabilité devrait être poursuivie. En outre, les explications sur le degré dans lequel un système d'IA influence et modèle les processus de prise de décision, les choix pour la conception des

[73] Cf. également le « *Policy document* » élaboré par le High Level Group of Experts on AI, document établi à la demande de la Commission européenne : « Une IA digne de confiance exige que les algorithmes soient suffisamment sûrs, fiables et robustes pour traiter les erreurs ou les incohérences pendant toutes les phases du cycle de vie du système d'IA, et pour faire face de manière adéquate aux résultats erronés. Les systèmes d'IA doivent être **fiables,** suffisamment sûrs pour **résister** à la fois aux attaques manifestes et aux tentatives plus subtiles de manipulation des données ou des algorithmes eux-mêmes, et ils doivent prévoir **un plan de secours** en cas de problème. Leurs décisions doivent être **exactes**, ou du moins refléter correctement leur niveau de précision, et leurs résultats doivent être **reproductibles**. En outre, les systèmes d'IA doivent intégrer des mécanismes de sûreté et de sécurité par conception afin de garantir qu'ils sont **sûrs de manière vérifiable** à chaque étape, en prenant à cœur la sécurité physique et mentale de toutes les personnes concernées. Cela inclut la minimisation et, si possible, la réversibilité des conséquences involontaires ou des erreurs dans le fonctionnement du système. Des processus visant à clarifier et à évaluer les risques potentiels liés à l'utilisation des systèmes d'IA, dans les différents domaines d'application, doivent être mis en place. »

systèmes ainsi que les raisons de leur déploiement devraient être disponibles (d'où le besoin d'assurer la transparence non seulement des données et des systèmes mais aussi des modèles d'entreprises »[74]."

Quelles informations l'exigence d'informer sur la logique sous-jacente[75] ou, en cas de *machine learning*, sur l'explicabilité du fonctionnement du système implique-t-elle ? L'article 9.c de la Convention évoque le droit à obtenir le "**raisonnement qui sous-tend le traitement des données**". Cette formulation est vague et erronée dans la mesure où en matière d'IA, il est impossible de parler de raisonnement, ce qui supposerait une vue causaliste, là où le système fonctionne sur base de corrélations plus ou moins supervisées. Faut-il faire un cours de *machine learning* pour expliquer quels genres de calculs sont faits ? Faut-il expliquer le calcul pour une décision particulière ? Faut-il donner des détails très précis ou une vague indication des poids de chaque information à caractère personnel dans la décision ? Y- a - t-il une proportionnalité dans l'exigence de transparence en fonction de la nature de la décision et de ses impacts sur la personne concernée ? Il s'agit en tout cas de mettre la personne concernée en position de comprendre comment fonctionne le système et pourquoi une décision est prise à son encontre dont la recommandation, la décision ou la prédiction lui sont opposées[76]. Il est clair que cela s'entend des types

---

[74] La recommandation de l'OCDE parle également d'explicabilité en ce qui concerne la transparence due pour les systèmes d'IA et la conçoit de façon assez large : 'Les acteurs de l'IA devraient s'engager à assurer la transparence et une divulgation responsable des informations liées aux systèmes d'IA. À 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. '

[75] 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 repris 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 : « Les responsables de traitement devraient trouver des moyens simples de donner des information aux personnes concernées sur le raisonnement et les critères appliqués pour parvenir aux décisions prises [… mais] pas nécessairement fournir une explication complexe sur les algorithmes utilisés ni révéler complètement l'algorithme. Toutefois, l'information fournie devrait être suffisamment de nature à permettre à la personne concernée de comprendre les raisons de la décision ».

[76] A titre de comparaison, en matière de classement des sites par les services d'intermédiation en ligne, l'article 5 du règlement européen 2019/1150 du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne (JOUE 11.7. 2019, L 186/57) prévoit également des obligations de transparence des prestataires de ces services vis-à-vis des entreprises utilisatrices de leurs services, en ce qui concerne le classement des sites par le prestataire des services d'intermédiation en ligne (les opérateurs de plateforme) : «1. Les fournisseurs de services d'intermédiation en ligne indiquent dans leurs conditions générales les principaux paramètres déterminant le classement, et les raisons justifiant l'importance relative de ces principaux paramètres par rapport aux autres paramètres. 2. Les fournisseurs de moteurs de recherche en ligne indiquent les principaux paramètres qui, individuellement ou collectivement, sont les plus importants pour déterminer le classement ainsi que l'importance relative de ces principaux paramètres, en fournissant une description facilement et publiquement accessible, énoncée dans une formulation claire et compréhensible, sur les moteurs de recherche en ligne de ces fournisseurs. Ils tiennent cette description à jour. 3. Lorsque les principaux paramètres incluent la possibilité d'influer sur le classement contre toute rémunération directe ou indirecte versée par les entreprises utilisatrices ou les utilisateurs de sites internet d'entreprise au fournisseur concerné, ce fournisseur présente également une description de ces possibilités et des effets de cette rémunération sur le classement, conformément aux exigences énoncées aux paragraphes 1 et 2. 4.. 5. Les descriptions visées aux paragraphes 1, 2 et 3 sont suffisantes pour que les entreprises utilisatrices ou les utilisateurs de sites internet d'entreprise puissent acquérir une compréhension suffisante pour déterminer si, et dans l'affirmative, comment et dans quelle mesure, le mécanisme de classement tient compte des éléments suivants: a) les caractéristiques des biens et services proposés aux consommateurs par le biais des services d'intermédiation en ligne ou des moteurs de recherche en ligne; b) la

de données anonymes ou non traitées, de leurs sources, du mode de fonctionnement de l'algorithme et sans doute sans devoir donner les détails, des poids accordés à chaque type de données dans l'algorithme de base. Ne peut-on estimer que la 'logique sous-jacente' à une décision oblige à fournir à la personne concernée tous les éléments utilisés pour produire le résultat, non seulement les données mais également les combinaisons et relations entre elles opérées par le système. Cette information peut se trouver sur un site web accessible via l'icône qui avertit la personne concernée de sa soumission à un profil. Par ailleurs, le besoin de transparence nous paraît devoir s'imposer et l'emporter sur les arguments du responsable qui, sur base du droit d'auteur, du brevet ou du secret d'affaires, s'opposerait de manière abusive à **toute** communication sur la logique ou l'explicabilité du système de profilage retenu[77]. Doit être affirmé le principe que le secret des affaires et les droits de propriété intellectuelle ne peuvent être opposés aux exigences de la transparence que de manière limitée et strictement nécessaire à la protection des intérêts de leurs titulaires.

l'utilisation de traitements de profilage par les administrations doit obéir à des règles plus strictes en matière de transparence, sauf lorsque des raisons d'intérêt général l'emportent, ainsi lorsqu'il s'agit de traitements de recherche de fraudeurs ou de délinquants. L'idée de faire évaluer les systèmes en particulier d'IA par un organe de contrôle, de manière à éviter en particulier tout biais et discrimination, celle de publier les algorithmes de fonctionnement, corollaire de l'obligation de motivation des décideurs publics nous paraît s'imposer. Cette obligation a des répercussions en matière de marchés publics en particulier pour les fournisseurs ou plutôt concepteurs d'algorithmes.

enfin, nous insistons : suite à cette transparence, des associations de libertés ou civiles, voire le parquet, puissent au nom de personnes exclues ou lésées en tant que membres de groupes discriminés, de ces groupes eux-mêmes ou de l'intérêt général, agir en justice. Dans le cadre de l'assurance *one to one* qui met en cause le principe de mutualisation, par exemple, on peut imaginer l'intervention de représentants de l'intérêt général ou d'associations de consommateurs au nom des personnes affectées par le système sans être nécessairement des personnes concernées. Quoiqu'il en soit, l'exigence de transparence reste une des exigences majeures nécessaires à assurer la confiance dans le fonctionnement de nos systèmes. C'est la raison pour laquelle nous estimons que le Conseil de l'Europe doit encourager la recherche en la matière et favoriser la publication en '*open source'* des algorithmes.

55. Les droits de la personne concernée en cas de décision automatisée - Les droits de la personne concernée « de ne pas être soumise à une décision l'affectant de manière significative qui serait prise uniquement sur le fondement d'un traitement automatisé de données sans que son point de vue soit pris en compte » appellent les remarques suivantes :

les termes « **affecter de manière significative »** ont été longuement commentés sur base de termes similaires par le Groupe de l'article 29 de l'Union européenne. Il s'agit en tenant compte du contexte, des personnes concernées et de la durée des conséquences (par exemple : on ne met pas sur le même pied le refus de visa suite à une demande d'immigration et la fixation d'un prix par profilage).

---

pertinence de ces caractéristiques pour ces consommateurs; c) en ce qui concerne les moteurs de recherche en ligne, les caractéristiques de conception du site internet utilisé par les utilisateurs de sites internet d'entreprise.»
L'article 6 a) du projet de directive récemment approuvé par le Parlement européen sur la meilleure mise en œuvre et modernisation des règles de protection des consommateurs (*Position of the European Parliament adopted at first reading on 17 April 2019 with a view to the adoption of Directive (EU) 2019/… of the European Parliament and of the Council …as regards better enforcement and modernisation of EU consumer protection rules*) reprend les mêmes idées en ce qui concerne cette fois l'information que l'opérateur doit au consommateur qui recourt à sa plateforme en ce qui concerne le *ranking* des sites.

[77] Comparer le considérant 63 du RGPD qui affirme : « Ce droit (d'accès à la logique sous-jacente) ne devrait pas porter atteinte aux droits ou libertés d'autrui, y compris au secret des affaires ou à la propriété intellectuelle, notamment au droit d'auteur protégeant le logiciel. Cependant, ces considérations ne devraient pas aboutir à refuser toute communication d'informations à la personne concernée.

le terme ***'uniquement'*** pose difficulté : il importera d'analyser dans quelle mesure la personne humaine chargée d'intervenir à la suite de la proposition sortant de l'ordinateur dispose dans le cadre de son organisation et de la procédure y prévue d'une réelle capacité en temps et en compétence, de s'écarter de la proposition et l'exerce effectivement.

les termes '**sans que son point de vue ne soit pris en compte'** exigent que la personne soit dûment informée de cette possibilité d'exprimer son point de vue, que l'organisation interne prévoit explicitement la procédure qui sera suivie en la matière et que, dans toute la mesure du possible, elle puisse se faire accompagner.

Cette dernière exigence, pour être effectivement rencontrée, exige suivant l'article 9.1.c) que la personne puisse avoir accès à une explication par écrit des critères utilisés pour justifier la décision et de leur application à son cas concret. Certes, il faut admettre des exceptions[78] lorsque l'accès à ce raisonnement va à l'encontre d'intérêts supérieurs du responsable du traitement consacrés par la loi[79] (par exemple en cas de lutte contre la fraude) (voir l'article 9.2., qui mentionne cependant la nécessité « de mesures appropriées pour la sauvegarde des droits, des libertés et des intérêts légitimes de la personne concernée[80] ») ou que cet accès est impossible vu la complexité des corrélations effectuées par la machine et qui rendent incompréhensibles y compris pour le développeur du système le 'modèle' de fonctionnement de la machine (système de '*deep learning*'). Dans ce dernier cas, le responsable donnera cependant toutes les informations qu'il détient dans le cadre des exigences d''explicabilité' (voir *supra*, nos réflexions à propos de l'article 8). A cet égard le Groupe d'experts de haut niveau en matière d'IA nommé par la Commission européenne réclame la 'traçabilité' des différentes décisions prises par le système : « La **traçabilité** des systèmes d'IA devrait être assurée. Il est important d'enregistrer et de documenter à la fois les décisions prises par les systèmes et les processus complets qui y ont conduit (y compris une description de la collecte de donnés destination et de leur qualification ainsi qu'une description de l'algorithme utilisé). »

Les droits de la personne concernée - On reprendra les recommandations (points 5.2 et 5.3.) suivantes du texte émis en 2010 : « Les personnes concernées devraient pouvoir obtenir, selon le cas, la rectification, l'effacement ou le verrouillage de leurs données, lorsque le profilage dans le cadre du traitement de données à caractère personnel s'effectue en méconnaissance des dispositions du droit interne donnant effet aux

[78] La recommandation 5.4 de 2010 nous paraît devoir être reprise : « S'il existe des motifs de restreindre les droits énoncés dans le présent paragraphe en application du chapitre 6, cette décision devrait être communiquée à la personne concernée par tout moyen permettant d'en garder la trace, avec mention des raisons juridiques et matérielles d'une telle restriction. Il est possible d'omettre cette mention pour une raison nuisant au but de la restriction. Dans ce cas, la personne concernée devrait être informée des modalités de contestation de cette décision devant l'autorité de contrôle nationale compétente, une autorité judiciaire ou un tribunal. »

[79] Il va de soi que cette loi doit être spécifique, claire et poursuivre des objectifs conformes aux exigences d'une société démocratique. Il nous apparaît difficile de renvoyer sans nuances à des exceptions comme celles du 'consentement' ou 'de l'existence d'une relation contractuelle'. Le point 5.5 de la recommandation de 2010 proposait : « Dans le cas où une personne concernée est soumise à une décision produisant des effets juridiques à son égard ou l'affectant de manière significative, prise sur la seule base d'un profilage, elle devrait pouvoir s'opposer à cette décision, à moins :
a. que la loi l'autorise et précise les mesures garantissant la sauvegarde de l'intérêt légitime de la personne concernée, notamment en lui permettant de faire valoir son point de vue ;
b. que la décision ait été prise dans le cadre de l'exécution d'un contrat auquel la personne concernée est partie ou en application des mesures précontractuelles prises à la demande de celle-ci et que les mesures garantissant la sauvegarde de l'intérêt légitime de la personne concernée soient mises en place. »
Cf. également la recommandation 6 qui élargit le propos aux principes de légitimité et de transparence des traitements de profilage : « Lorsque cela est nécessaire dans une société démocratique pour des raisons de sécurité nationale, de sûreté publique, de défense des intérêts monétaires du pays, de prévention ou de répression des infractions pénales, ou à la protection des personnes concernées ou des droits et libertés d'autrui, les Etats membres n'appliquent pas les dispositions des chapitres 3, 4 et 5 de la présente recommandation, pour autant que cela soit prévu par la loi. »

[80] 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.

principes énoncés dans la présente recommandation. Sauf si une loi prévoit le profilage dans le cadre du traitement de données à caractère personnel, la personne concernée devrait avoir le droit de s'opposer, pour des raisons prépondérantes et légitimes tenant à sa situation particulière, à l'utilisation de ses données dans le cadre du profilage. En cas d'opposition justifiée, le profilage ne devrait plus impliquer l'utilisation des données personnelles de la personne concernée. Quand le but du traitement est la prospection, la personne concernée n'est pas tenue de formuler un justificatif *».*

Obligations complémentaires en fonction des caractéristiques du profilage (article 10)

Le principe d'« *accountability* [81]» - Les articles 10.1 et 10.4 consacrent le principe d' '*accountability'.* C'est au(x) responsable(s) du traitement de faire la preuve de leur respect des obligations liées aux traitements de profilage qu'ils développent et ce, en tenant compte des risques liés à ces traitements. Il sera donc important de vérifier la qualité des données non seulement du point de vue de leur exactitude, de leur mise à jour mais également du caractère non biaisé de leur utilisation dans telle application. 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 par le 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 l'ensemble de ces obligations peuvent faire l'objet d'une certification par l'organe de contrôle des systèmes d'IA mentionné ci avant ou par d'autres organismes agréés par celui-ci.

Cette obligation du responsable n'évacue pas la responsabilité des autres acteurs qui interviennent sur le marché commercial en offrant l'un ou plusieurs éléments (data sets, algorithmes). Ainsi, en fonction des risques 'prévisibles' liés à leur fonctionnement, le fournisseur d'algorithme documentera le produit qu'il met sur le marché, en décrivant les applications pour lesquelles il est conçu ou au contraire proscrit ou déconseillé, le fait qu'il ait déjà fait l'objet d'applications ou ait été testé, il collaborera lors de la période de test, etc. On peut concevoir bien évidemment que ces produits fassent l'objet également de certifications au regard d'un secteur particulier ou de manière générale.

L'article 10. 2 prescrit au responsable voire au sous-traitant le devoir pour tout traitement de procéder à une évaluation des risques et l'utilisation de mesures organisationnelles et techniques permettant la diminution de ces risques. Il va de soi que relève de la responsabilité interne de l'organisation qui recourt au profilage ou développe des outils qui permettront celui-ci, de former les personnes impliquées dans la réalisation de tels traitements ou outils à la détection et à la lutte contre les risques liés au profilage. Au-delà, ces organismes favoriseront, le cas échéant, la discussion interdisciplinaire et ouverte nécessaire à cette détection et cette prise en compte. Par ailleurs, nous avons longuement développé la signification de cette obligation, pour les traitements de profilage à haut risque. Il importe de rappeler que cette notion devrait être approfondie non seulement par des approches sectorielles (secteur de la médecine, grande distribution, plateformes, banques et assurances, police, …) mais également en fonction de certains types de relation (voir la question des profilages dans le cadre des relations de travail). Ce '*Risk assessment'* nécessite la concertation avec des groupes et associations externes. Les autorités de protection des données doivent jouer également un rôle important en la matière.

Rôle des autorités de contrôle (article 15)

De quelques rôles de l'autorité de contrôle au regard des spécificités de l'utilisation de l'IA en matière de profilage - De multiples rôles doivent être dévolus aux autorités de contrôle si possible travaillant

[81] Pour information : Le cadre de l'audit de l'ICO (*AI auditing framework')* constitue une étape en vue d'améliorer la responsabilisation des organisations. Il porte sur de nombreux risques soulevés ici et donne, par exemple aux organisations des outils pour les identifier et les réduire elles-mêmes.

conjointement. La collaboration avec l'autorité interdisciplinaire et indépendante chargée de tester, d'évaluer (évaluation obligatoire au moins pour les traitements de profilage IA des autorités publiques, volontaires pour les autres mais avec obligation des responsables pour les traitements à risque élevé de remettre à l'autorité le rapport de *Risk Assessment* et les mesures prises pour diminuer ce risque), de définir, le cas échéant, des *'best practices*' et de délivrer des labels de conformité aux exigences de la Convention. En outre, nous rappelons la Recommandation de 2010 qui en son point 10 soulignait : « Par ailleurs, dans le cas de traitements ayant recours au profilage et présentant des risques particuliers au regard de la protection de la vie privée et des données à caractère personnel, les Etats membres peuvent prévoir : a. que les responsables des traitements soient tenus de les notifier préalablement à l'autorité de contrôle ; ou b. que ces traitements fassent l'objet d'un contrôle préalable par l'autorité de contrôle. » Cette compétence nécessite à notre avis un certain degré d'expertise des **autorités de contrôle** vis-à-vis des techniques parfois très avancées de traitement des données ou, à tout le moins, une connaissance de base pour en saisir le fonctionnement, les limites et les possibilités. Il s'agira de renforcer les liens entre les universitaires spécialisés en IA et les autorités de contrôle. On ajoutera que les universitaires peuvent également jouer un rôle dans le conseil au Comité conventionnel de la Convention 108+. Il faut absolument que les universitaires spécialisés en IA soient impliqués dans toute standardisation pour éviter qu'elle soit "artificielle" (sic) ou "déconnectée" de toute réalité technique et/ou technologique.

Les Etats membres devraient charger les autorités de contrôle exerçant leurs fonctions en toute indépendance de veiller au respect du droit interne mettant en œuvre les principes énoncés dans la présente recommandation et disposant à cet effet des moyens d'investigation et d'intervention nécessaires, en particulier la compétence d'examiner les recours déposés par des individus. Par ailleurs, dans le cas de traitements ayant recours au profilage et présentant des risques particuliers au regard de la protection de la vie privée et des données à caractère personnel, les Etats membres peuvent prévoir, en particulier en cas de traitement à risque élevé et de traitements de profilage mis en place par les autorités publiques soit que les responsables du traitement soient tenus de les notifier préalablement à l'autorité de contrôle, soit que ces traitements fassent l'objet d'un contrôle préalable par l'autorité de contrôle.

Deux points délicats - Deux points méritent l'attention. Le premier concerne le **devoir de coopération**, en ce qui concerne l'évaluation et le contrôle des traitements de profilage, entre, d'une part, les autorités de contrôle et, d'autre part, les autorités de protection des consommateurs et de la concurrence de même qu'avec les institutions en charge de l'égalité des chances ou de promotion de la démocratie. Le second est plus délicat. Nous avons noté que les traitements de profilage mettaient en cause, au-delà des questions qui indéniablement sont de l'ordre de la protection des données, des problématiques plus collectives de justice sociale, de protection des groupes, de mise en danger de la démocratie. L'approche protection des individus sans considération des enjeux sociétaux du numérique transparaît en effet des textes même si dans la réalité, les autorités n'hésitent pas à, agir de manière plus large. Cette tendance des autorités de contrôle est à suivre. Il apparaitrait dangereux d'écarter les autorités de protection des données de tels débats alors même que la Convention met en place des instruments adéquats d'analyse des risques (voir la procédure de *Risk assessment*[82]), d'information des sujets concernés et surtout de contestation des décisions prises sur une

[82] Il est piquant de noter que les Guidelines du Groupe de l'article 29 repris par l'EDPB sur le risk assessment' : *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, wp248rev.01 (*disponible à l'adresse: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236) ne mentionnent que les risques encourus par la personne concernée par le traitement sans analyser l'impact que ce traitement peut avoir sur d'autres personnes, sur un groupe ou sur la société : « Le RGPD n'exige pas une évaluation des risques pour chaque opération de traitement qui pourrait entraîner des risques pour les droits et les libertés des personnes physique. Conduire une évaluation des risques n'est obligatoire que lorsque le traitement "susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques » (article 35(1), illustré par l'article 35(3) et complété par l'article 35(4)). Cela est particulièrement pertinent lorsqu'une nouvelle technologie pour le traitement des données est introduite ».

base automatisée. Nous suggérons que les autorités élargissent leur champ d'analyse aux risques collectifs et sociétaux. Elles veilleront à ce que leurs avis mentionnent de tels risques et que leurs décisions les prennent en compte. Le cas échéant, elles provoqueront des débats en la matière. Elles attireront l'attention des Etats membres sur l'intérêt d'élargir leurs compétences en la matière. Sans doute, leur manquera-t-il dans l'attente d'un élargissement de leurs compétences la possibilité d'agir par voie de sanctions en la matière mais rien n'empêche nos autorités de contrôler, d'instruire, de convoquer des débats publics en la matière et de stigmatiser des pratiques contraires, sinon aux législations de protection des données, aux valeurs de nos démocraties : dignité, justice sociale notamment. Dans ce contexte, les autorités de contrôle devraient accueillir et instruire les plaintes émanant d'associations et visant l'intérêt collectif d'un groupe ou l'intérêt général. Le cas échéant, les autorités émettront des recommandations à ce propos. Ces autorités devraient informer le public de l'application de la législation mettant en œuvre les principes concernés.

Un dernier rôle important à confier aux autorités de contrôle est l'information et la sensibilisation du public aux enjeux tant positifs que négatifs du profilage. Cette sensibilisation s'entend également d'une éducation des jeunes dans le cadre scolaire à cette réflexion.

BIBLIOGRAPHIE:

ACCESS NOW, "*The Toronto Declaration: Protecting the rights to equality and non-discrimination in Machine-Learning Systems, 2018*", https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

AI NOW Institute, « *Litigating Algorithms : Challenging Government Use of Algorithmic Decision Systems*", 2018, https://ainowinstitute.org/litigatingalgorithms.pdf

M. BRKAN, "Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond", *International Journal of Law and Information Technology*, 2019, eay017, p. 15S.

M. BRUNDLAGE, J. CLARK, G. C. ALLEN, G. C., FLYNN, S. FARQUILHAR, S., R. CROOTOF, et J. BRYSON, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.Information Society Project. Future of Human Institute*, 2018,. Disponible à l'adresse suivante: https://www.repository.cam.ac.uk/bitstream/handle/1810/275332/1802.07228.pdf?sequence=1

J. BURELL, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", *Big Data & Society*, 2016, vol. 3, pp. 3–4.

A.CAMBON-THOMSEN, "Acteurs et outils de la prédiction génétique: l'éthique au coeur de la gouvernance", *Journal international de bioéthique,* 2016, Vol. 25 -2, p. 159 à 168

C.CATH, S.WACHTER et al., "Artificial Intelligence and the 'GoodSociety': the US, EU, and UK approach", *Science and Engineering Ethics, 2018,* 24 (2), p. 505-528.

CNIL, « *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle.»*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une république numérique, https://www.cnil.fr/sites/default/files/atoms/.../cnil_rapport_garder_la_main_web.pdf

Comité économique et social européen, " L'éthique des mégadonnées (Big Data): Équilibrer les avantages économiques et les questions d'éthique liées aux données massives dans le contexte des politiques européennes », 2017, étude disponible sur le site du Comité à l'adresse suivante : https://www.eesc.europa.eu/sites/default/files/resources/docs/qe-04-17-306-fr-n.pdf

Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, "Artificial Intelligence for Europe", Brussels, 25 April 2018, COM(2018) 237 final, p. 17.

Conseil de l'Europe (CoE), *Algorithms and Human Rights*, Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, 2017, disponible à l'adresse suivante : https://rm.coe.int/study-hr-dimension-of-automated-dataprocessing-incl-algorithms/168075b94a

Conseil de l'Europe (CoE). *Technological convergence, artificial intelligence and human rights*. 2017, disponible à l'adresse : http://www.assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=24236&lang

Conseil de l'Europe, « *Lignes directrices sur les mégadonnées* », disponibles à l'adresse : https://rm.coe.int/CoERMPublicCommnSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a .

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

D.DESAI and J. KROLL, "Trust But Verify: A Guide to Algorithms and the Law", *Harvard Journal of Law & Technology*, 2017, vol. 31, pp. 46–47.

F. DOSHI-VELEZ and K. MASON, “Accountability of AI Under the Law: The Role of Explanation”, *Berkman Klein Centre Working Group on Explanation and the Law, Berkman Klein Centre for Internet & Society*, 2017, pp. 2-3.

R. DWORKIN « What is equality? » Part 2: equality of ressources, *Philos Public Aff*, 1981, 10, p. 283-345.

EDPS Ethics Advisory Group, “*Towards a digital Ethics”.* Le rapport est disponible sur le site: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

L. EDWARDS and M. VEALE, “Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For”, *Duke Law & Technology Review*, 2017, vol. 16, p. 38 et s.

European Group on Ethics (EGE). 2018. “Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems. European Group on Ethics in Science and New Technologies.” Brussels: European Commission. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

A.FERRETTI, M. SCHNIEIDER and A. BLASIMME, “Machine Learning in Medicine”, *European Data Protection Law Review*, 2018, vol. 4, p. 326.

L. FLORIDI et al. (2018), "AI4People —An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”, *Minds and Machines*, 2018, 28 (4) p. 689-707.

L. FLORIDI, ‘Tolerant Paternalism: Pro-Ethical Design as a Resolution of the Dilemma of Toleration.’, *Sci. Eng. Ethics*, 2016, 22(6): 1669-1688.

GOOGLE, “Perspectives in Issues in AI Governance”, January 2019, p. 8.

I.GRIMMELMANN and D. WESTREICH, “Incomprehensible Discrimination”, *California Law Review*, 2017, vol. 7, pp. 170–176.

Groupe européen d’éthique des sciences et des technologies nouvelles, *Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems*. 2018, Bruxelles, Commission européenne, disponible à l’adresse suivante : https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

M. HILDEBRANDT, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar Publishing, 2016.

E.HIRSCH et F. HIRSCH (sous la direction de), *(Les) Les nouveaux territoires de la bioéthique, Traité de bioéthique,* Vol. IV, Collection Espaces éthiques, érès, Toulouse, 2018, ISBN 978-2- 7492-6083-9

Th. HOEREN et M. NIEHOFF,”AI in Medical Diagnoses and the Right to Explanation”, *EdpL*, 2018, n°3, p. 308 et s.

IEEE *Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, version 1, IEEE, 2016, disponible à l’adresse suivante: https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead_v1.pdf.

IEEE, General Principles, 2nd version of Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, 2017, document disponible sur le site: https://ethicsinaction.ieee.org/

M.KOSINSKI, D.STIWELL et T.GRAEPEL, « Private traits and attributes are predictable from digital records of human behavior », *PNAS*, Avril 2013, Vol. 110, p. 5803 et s.

J. KROLL, J. HUEY, S. BAROCAS, E. FELTEN, J. REIDENBERG, D. ROBINSON, and H. YU, “Accountable Algorithms”, *University of Pennsylvania Law Review*, 2017 vol. 165, pp. 660 – 706.

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

T.LINNET, L. FLORIDI et B. van der SLOOT (dir.), *Group Privacy: New Challenges of Data Technologies*, Springer International Publishing, 2017

G. MALGIERI et G. COMANDE, « Why a Right to Legibility of Automated Decision – Making Exists in the GDPR?", *International Data Privacy Law*, 2017, Vol.7, n° 4, p. 243 et s.

A.MANTELERO, Artificial Intelligence and Data Protection: Challenges and Possible Remedies – Report, Comité consultatif de la convention n108 du Conseil de l'Europe, 3 décembre 2018, T-PD(2018)09Rev

A.MANTELERO « AI and Big Data: A blueprint for a human rights, social and ethical impact assessment », in *Computer Law & Security Review,* 2018, https://doi.org/10.1016/j.clsr.2018.05.017.

A.NAUDTS, "Fair or Unfair Algorithmic Differentiation? Luck Egalitarianism as a Lens for Evaluating Algorithmic Decision-Making", *article à paraître*

J. NEW and D. CASTRO, "How Policymakers Can Foster Algorithmic Accountability", *Centre For Data innovation*, May 2018, pp. 20-22.

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.

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.

PARLEMENT EUROPEEN, Résolution du Parlement européen du 14 mars 2017 sur les incidences des mégadonnées pour les droits fondamentaux : respect de la vie privée, protection des données, non-discrimination, sécurité et application de la loi (2016/2225(INI)), disponible : http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0076+0+DOC+XML+V0//FR

G.PASQUALE and D. CITRON, "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, 2014, vol. 91, pp. 5-6.

Y. POULLET, *La vie privée à l'heure du numérique - Essai,* Larcier, Cahier du Crids, n° 47, 2019.

F.ROSSI, « *Artificial Intelligence: Potential Benefits and Ethical Considerations* », Parlement européen, Département thématique C « Droit des citoyens et affaires constitutionnelles », 2016, note d'information PE 571.380, disponible à l'adresse suivante : http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_ EN.pdf..

A.ROUVROY, « *Des donnés et des hommes* » Droits et libertés fondamentaux dans un monde de données massives, 2016, https://rm.coe.int/16806b1659

A. ROUVROY, "Privacy, Data Protection, and *the* Unprecedented Challenges of Ambient Intelligence" (September 11, 2007). disponible sur le site SSRN: http://ssrn.com/abstract=1013984

A.ROUVROY, « Face à la gouvernementalité algorithmique, repenser le sujet de droit comme puissance », 2012, p.42.

A.ROUVROY et Y. POULLET, "The right to informational Self-determination and the value of Self-development: Reassessing the importance of Privacy for Democracy", in *Reinventing Data Protection?,* Springer, Dordrecht, 2009, p. 159 et s.

D.SADIN, La vie algorithmique : critique de la raison numérique, Paris, l'Echappée, 2015

C.SANDVIG, K. HAMILTON, K. KARAHALIOS, and C. LANGBORT, "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms", presentation given at the conference *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, Washington, 22 mars 2018

A.SELBST and S. BAROCAS, "Big Data's Disparate impact", California Law Review, 2016, vol. 104, pp. 691-692.

P. TERZIS, « The reasonable coder », article en projet disponible sur *Academia edu*, 2019

J. Van DIJK, "Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology". *Surveillance & Society*, 2014, 12 (2), p.197-208.

C.VILLANI, "Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne", French public report, March 2018, p. 145.

M. WACHTER, B. MITTELSTADT and C. RUSSEL, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law and Technology*, 2018, vol. 31, pp. 845–846

J.P. WALTER, Le profilage des individus à l'heure du « cyberespace » : un défi pour le respect du droit à la protection des données