Les dispositifs de reconnaissance faciale : une réalité socio-technique en développement, un enjeu pour les libertés publiques et privées, un défi pour l'éthique de l'IA

Thierry Ménissier

Les dispositifs de reconnaissance faciale : une réalité socio-technique en développement, un enjeu pour les libertés publiques et privées, un défi pour l'éthique de l'IA

Thierry Ménissier

IPhiG, Univ. Grenoble Alpes

Intervention au séminaire de la chaire « éthique & IA » MIAI Grenoble, 4/12/2019 : « La reconnaissance faciale est-elle une bonne ou une mauvaise chose (éthiquement parlant) ? »

On peut dresser le double constat qu'aujourd'hui les dispositifs de reconnaissance faciale (DRF) se développent rapidement dans les sociétés technologiquement avancées, et que rien ne semble devoir freiner ce développement, tant les organisations (privées et publiques) y trouvent de l'intérêt en termes de commodités (notamment d'efficacité et de sécurité). Parce qu'elles disposent désormais de technologies permettant d'effectuer l'identification et l'authentification des individus de manière à la fois efficace, permanente et invisible, elles peuvent offrir à leurs clients (pour les entreprises) ou à leurs administrés (pour les Etats et les collectivités) des services qui permettent à ceux-ci de vivre dans la quiétude. Du moins, c'est en cela que semblent consister à la fois la promesse et la tentation offertes par ces technologies.

Or, l'une comme l'autre apparaissent aussi séduisantes qu'inquiétantes, compte tenu de la nature des technologies engagées, mais également de leur mise en relation avec les bases de données et de la sous-qualification éthique dont elles sont aujourd'hui l'objet.

Cette contribution se donne un triple objectif : (1) caractériser les dispositifs de reconnaissance faciale, en tenant compte du contexte social de leur développement ; (2) produire une évaluation de leur impact possible sur les libertés privées et publiques à la lumière de la distinction entre « identité » et « personnalité » numériques – et de la fragilité contemporaine de cette distinction cruciale ; (3) questionner l'éthique de l'intelligence artificielle actuellement dominée par l'évaluation des dispositifs technologiques en termes utilitaristes, afin de souligner les dangers qu'il y aurait à s'en tenir seulement à une évaluation de ce type, alors qu'un élargissement de l'éthique de l'IA est salutaire et fécond. La contribution s'achève par des recommandations exprimées à propos des « lignes de défense » possibles pour contenir les dérives probables de tels dispositifs.

1/ Les dispositifs de reconnaissance faciale, des technologies biométriques, aisément connectables aux flux de méga-données :

Les dispositifs de reconnaissance faciale (DRF) relèvent des technologies biométriques : ils captent automatiquement la forme des visages (par captation directe ou à partir d'un artefact, par exemple une photo ou une vidéo) afin de l'authentifier ou de l'identifier, comme d'autres types de système le font à partir des empreintes digitales, des iris, de la voix humaine, etc. Schématiquement, les DRF fonctionnent en deux temps : premièrement, une

collecte initiale de données est nécessaire pour permettre l'établissement d'un « gabarit », à savoir, un modèle informatique représentant certaines caractéristiques de ce visage : deuxièmement, ce « gabarit » est utilisé pour authentifier ou identifier la personne dans les situations concrètes.

Parmi les dispositifs biométriques, la reconnaissance faciale occupe une place à part, car elle permet de capter des données issues du corps humain sans contact ni consentement préalable, tandis que pour d'autres systèmes, il faut prononcer des paroles de manière claire et distincte (reconnaissance vocale), approcher son œil d'un viseur (reconnaissance oculaire), imprimer les empreintes ses doigts (reconnaissance digitale), ou bien donner un échantillon de salive ou de sang (prélèvement d'ADN).

Les DRF qui permettent l'authentification des personnes à partir de la captation des images des visages sont d'ores et déjà largement utilisées dans des opérations de la vie ordinaire tel que le filtrage des accès et la sécurité, par exemple dans ces cas :

- Ouverture de sessions personnelles sur les terminaux électroniques (PC, tablettes, smartphones) et autorisation d'accès à des espaces, tels des lieux physiques ou à des sites numériques estimés sensibles : manifestations ou lieux publics, administrations et institutions (prison, enseignement, religions, etc.), banques, santé, sécurité sociale et assurances, etc.
- Autorisation d'accès à des lieux ou à des sites qui impliquent que le requérant soit accrédité et « en règle » : entreprises devant protéger leur activité contre le vol de leurs actifs, toutes les applications payantes…

Ils sont également employés pour l'identification générique des personnes ; par exemple lorsqu'il s'agit de surveiller une foule comme dans le cas des aéroports internationaux ou lors du récent Carnaval de Nice, ils permettent de retrouver des personnes estimées dangereuses en les distinguant automatiquement des autres à partir de leur gabarit.

Dans tous les cas, on pourrait dire, en s'inscrivant dans la perspective d'un travail de distinction conceptuelle classique, qu'il y a ici une **identification des individus**, mais que ce n'est pas une **identification des personnes, ou « personnification »**. En procédant de la sorte, on retrouve le type de travail philosophique que fit John Locke au XVIIème siècle dans l'*Essai sur l'entendement humain* (livre II, chapitre 27, voir Locke 2001, p. 505-542), qui a permis de construire une différence ontologique entre l'identité de l'individu (qualifiable extérieurement par son nom ou ses attributs apparents) et celle de la personne (caractérisée par ses états internes, corporels et mentaux).

Un individu humain, c'est un nom et un état civil ; une personne c'est un ensemble de vécus, d'opinions, d'expressions diverses, variées et potentiellement contradictoires entre eux. L'individu concerne l'assignation à des éléments extérieurs et contingents, la personne renvoie à l'authenticité d'un vécu multiple. « Identifier » et « personnifier » un individu constituent normalement deux actions différentes, la première assigne l'individu à des déterminants contingents ou extérieurs à ce qu'il est au fond de lui (tels que son nom, son âge et son lieu de naissance, son sexe biologique – toutes choses non choisies et parfois subies) ; la seconde

représente une caractérisation plus intime (ce qu'il ressent et ce qu'il pense, ses états d'âme et ses convictions profondes, ses choix intimes de culture et de vie, ses options éthiques personnelles).

Or, il nous faut ici avancer d'un pas, et remarquer qu'en dépit des apparences, les DRF comme tous les autres systèmes biométriques, non seulement identifient les individus, mais également sont susceptible de les « personnifier », c'est-à-dire d'intégrer pleinement la vie des personnes.

Au point où nous en sommes déjà aujourd'hui, en effet, les DRF et les autres dispositifs biométriques *facilitent l'exercice de l'identification à partir de la captation des données personnelles*, et de ce fait, en étant susceptibles d'intégrer des éléments bien plus riches que l'identité stricto sensu, ils sont en mesure d'interférer plus intimement avec la personnalité individuelle. Ces données sont variées et peuvent concerner, outre la validation de l'identité des individus,

- le relevé de la localisation de ces derniers dans l'espace,
- le suivi de la chronologie de leurs activités,
- l'identification de la nature leurs activités.

Si l'on n'y prête un soin extrême, ces données de base, très aisément captées, peuvent se trouver corrélées à d'autres données, confidentielles ou secrètes, détenues par des opérateurs d'IA publics et privés :

- celles liées à l'activité bancaire et à la situation fiscale des personnes,
- celles contenues dans leur casier judiciaire,
- celles contenues dans leur dossier médical,
- celles détenues par les assureurs et les mutuelles,
- celles détenues par les services de sécurité (polices nationales et municipales, services secrets intérieurs et extérieurs),
- celles détenues par les employeurs (publics ou privés) des personnes,
- celles détenues par les compagnies privées tels que celles qui utilisent les réseaux (opérateurs de téléphonie et de services numériques, réseaux sociaux, commerçants variés) – ces dernières sont susceptibles de de révéler les orientations des personnes et de donner accès à leurs opinions ordinairement tues ou tenues secrètes.

Ces deux séries de données peuvent enfin se trouver associées à celles produites en temps réel par les terminaux utilisés (smartphones, PC, tablettes) mais également, dans le contexte du déploiement de l'IoT, par les objets connectés qui sont en contact avec les personnes et leur activité : transports privés (automobile, moto, vélo, etc.) ou publics (trams, trains, avions…), appareils électroménagers (réfrigérateurs, système d'éclairage et de chauffage, dispositifs de remise en forme personnels…), bâtiments connectés, etc.

En reprenant la distinction lockéenne, on pourrait dire que ces nouveaux ensembles offrent un cadre inouï pour l'identification des individus, mais également pour leur « personnification », concept qu'il faudrait entendre d'un double point de vue : d'une part, les

espaces (physiques et numériques) offrent aux individus des accès et des services, de l'autre, ils monitorent leur activité et enregistrent des données qui, agrégées en flux de mégadonnées (*big data*), ont une forte pertinence en termes de renseignement sur l'intimité et par conséquent une grande valeur en termes policiers et marchands. Or, du point de vue du déploiement de l'IoT, n'importe quel lieu (privé ou public), n'importe quel dispositif socio-technique (tels un véhicule pour l'action de se déplacer, un fauteuil ou un matelas pour celle de se reposer) sont susceptibles de devenir un outil pour l'identification et la personnification.

Au-delà de la captation des données par les DRF et autres dispositifs biométriques actuels, s'ajoutent aujourd'hui la collecte d'un type d'informations qui les rendent potentiellement encore plus efficients : celles qui touchent à la captation des données corporelles ou biologiques. Ce type de données concernent, d'une part, les comportements perceptibles par des capteurs plus sensibles et objectifs que ne peuvent être les sens humains normaux (ainsi, la perception et la mesure du pouls qui s'accélère, celles de la sudation accrue, celles de la dilatation des pupilles, etc.) ; et de l'autre, les informations véhiculées par le sang ou l'ADN des personnes, deux substances qu'il est éventuellement possible de recueillir et d'analyser à l'insu des personnes.

Il serait difficile de soutenir que la mise en relation de toutes ces *data* ne représente pas, au-delà de la simple authentification, une possibilité permettant de constituer des bases de données permettant de connaître une personne à travers ses comportements. Les DRF fournissent en tout cas à ces bases de données l'occasion d'être actives pour toute sorte de situations, déjà bien réelles en France comme à l'étranger, telle que la reconnaissance des personnes participant à une manifestation de type politique[1].

2/ Les DRF augmentent considérablement le risque de mise en œuvre d'une société de surveillance intégrale et invisible :

Dans ces conditions, l'inquiétude grandit légitimement pour la préservation de la sphère privée : l'intimité aussi bien que le secret personnel courent évidemment de grands risques, d'autant plus que ces technologies se trouvent sinon déjà connectées, du moins connectables d'une part à des réseaux denses de caméras de surveillance urbaine, et de l'autre aux flux de *data* engendrées à propos des personnes ou par les personnes elles-mêmes.

Avec les DRF, on pourrait dire que l'espace social tend à devenir transparent ; de manière tendancielle, ils transforment les espaces qui leur sont offerts en espaces de visibilité

[1] Ainsi, le site de lanceurs d'alerte La Quadrature du Net révélait-il récemment que le gouvernement français a progressivement mis en œuvre un système capable d'identifier et de fournir des données sur les manifestants, qui met en relation la reconnaissance faciale et les bases de données des fichiers TAJ (« traitement des antécédents judiciaires » qui comprend les éléments contenus dans les casiers judiciaires individuels) et TES (« titres électroniques sécurisés », qui sert à l'établissement des papiers d'identité), et ceux collectés dans le cadre de la loi du 24 juillet 2015 relative au renseignement. Voir La Quadrature du Net, 18/11/2019, « La reconnaissance faciale des manifestant.e.s est déjà autorisée », https://www.laquadrature.net/2019/11/18/la-reconnaissance-faciale-des-manifestants-est-deja-autorisee/?fbclid=IwAR36yoXz-1iASeaX7lzwXjetdmcwguobjGfa0JCPjWKS-MepS0q5vkZ0ORA, consulté le 4/12/2019.

intégrale. Ici, il apparaît très difficile de séparer la mise en œuvre des DFR et la problématique du pouvoir. Mais comment entendre plus précisément ce dernier terme ?

Les DRF concernent plus précisément trois formes de pouvoir, qu'ils actualisent.

Premièrement, il y a un risque de surveillance généralisée des individus, comme le ferait un grand panoptique, selon le dispositif autrefois inventé par Jeremy Bentham, puis analysé par Michel Foucault dans son livre sur la naissance de la prison et dont il fait un symbole du pouvoir à l'époque contemporaine (Foucault, 1975). En termes foucaldiens, par référence à l'analyse du panoptique, les DRF pourraient être rattachés à la volonté de « discipliner » les agents sociaux, en les assignant, via l'identification et la « personnification » auxquelles ils contribuent, à normes implicites et explicites quant à leurs comportements.

Deuxièmement, ils peuvent également être l'expression de la volonté d'agir sur les conduites de ceux-ci en tant que population *via* une action sur les milieux dans lequel ils évoluent : ils sont de la sorte susceptible d'être analysés à l'aune des notions foucaldiennes de biopouvoir et de biopolitique (Foucault, 2004 a et b).

Troisièmement, on peut également les soupçonner, en reprenant les travaux de Jean-Gabriel Ganascia à propos de la manière dont, dans un réseau numérique, chacun contribue à apporter spontanément, volontairement ou involontairement, des données d'information, de nourrir le système « catoptique » ce qui engendre de la « sousveillance » à la fois auto-contributive, généralisée, permanente, et invisible (Ganascia, 2009).

Et cela d'autant plus qu'on voit en ce moment apparaître à la fois des projets de système et des systèmes dans lesquels les DRF (et tout autre dispositif biométrique) peuvent pleinement se déployer. Ces systèmes, ce sont les « smart cities », les ensembles urbains et sociaux intégralement outillés dès leur conception pour capter les données afin de mieux distribuer et d'optimiser les flux (d'énergie, d'eau, de l'air des climatiseurs, de transport, financiers et touchant alimentation). Si ces ensembles urbains, dans les projets des entreprises en bâtiment et selon les planificateurs urbanistes contemporains, ont été conçus comme des *ensembles sociaux*, ils n'ont encore jamais été conçu comme des *ensembles politiques*, c'est-à-dire qu'ils favorisent une idée d'autonomie plutôt liée au rapport à l'environnement que tournée vers la demande humaine d'autonomie (qui, elle, se décline notamment en termes d'imputation de responsabilité civique, d'aspiration à l'équité, et de participation politique en vue de l'autodétermination). Responsable du point de vue environnementale, la ville intelligente n'est pas encore « intelligible » [Caccamo et *alii* 2019], c'est-à-dire réflexive, critique et démocratique. Une smart city « intelligible » verrait les dispositifs techniques permettre à chacun de comprendre son action et d'exprimer ses choix de vie à parité avec les autres.

Les DRF renvoient aux trois niveaux de l'analyse du pouvoir. D'abord le pouvoir comme contrôle tendant à discipliner les individus ; de ce point de vue, tout individu peut être considéré comme suspect à un titre ou à un autre. Les méga-données sont en effet susceptibles de révéler tous les excès (petits et grands) des individus en termes de conduite, et d'attirer l'attention sur les petites ou grandes déviances par rapport à la norme statistique ou aux contraintes exprimées par les autorités en place dans les organisations publiques et privées.

Dans les smart cities, où les espaces physiques et numériques se rejoignent sans cesse et composent un seul et même espace, tout individu pourra donc être à la fois identifié et « personnifié » non seulement par les services de sécurité et de police, mais également par tous les services afférents aux conditions de son existence, concernant la rectitude de son action au sein de son « milieu de vie » (physique et numérique). Il convient ici de mobiliser la deuxième définition du pouvoir en intégrant à la première le caractère collectif du rapport aux milieux via la distribution des flux. Parce qu'ils agissent sur la relation des populations à leurs milieux, ces dispositifs sont pleinement ouverts à la gestion des populations par le biopouvoir. Du fait de l'intrication des milieux physiques et numériques ainsi que de l'auto-contribution des usagers du numérique, ils le sont également à leur identification/personnification en termes de sous-veillance, et renvoient donc à la troisième définition du pouvoir.

Dans les smart cities, le contrôle peut s'effectuer automatiquement : le système autorise les accès aux individus sans problèmes et les verrouille pour ceux qui sont suspects. Dans ces conditions la meilleure image possible pour une analogie avec l'architecture, pourrait être celle d'un gigantesque labyrinthe bidimensionnel. C'est en un tel labyrinthe que se transforme l'espace physique aussi bien que numérique régi par le système global dans lequel les dispositifs automatiques peuvent *distinguer* donc *discriminer* n'importe quel individu pour de « bonnes » ou de « mauvaises » raisons qui, parce qu'elles sont effectuées « en temps réel » ne seront pas examinées ni discutées au moment où il conviendra de le faire, c'est-à-dire au moment du blocage de l'accès.

En tant que tel, ainsi que l'a dénoncé Woodrow Hartzog, Professeur à la Faculté de droit de la Northeastern University à Boston, les SFR représentent « le parfait outil pour l'oppression » [Hartzog 2018], validant ainsi les analyses inquiètes exprimées sur le plan de la sociologie et dans le champ des « Surveillance Studies » par David Lyon [Lyon 1994 et 2009 ; Lyon & Bennet 2008].

Si nous analysons d'abord les DRF du point de vue de la philosophie politique, c'est bien entendu parce que les risques engendrés par leur déploiement concernent directement ou immédiatement le plan de la vie sociale, et mettent en danger la vie privée en violant notamment l'anonymat auquel chacun a droit dans une démocratie et qui permet d'élaborer une pensée réellement personnelle et originale.

C'est d'ailleurs bien dans cet esprit que les agences chargées de la préservation des droits fondamentaux des personnes, publiques ou non-gouvernementales, ont récemment réagi : ainsi, l'Agence Européenne pour les Droits fondamentaux a rendu un avis qui s'alarme de l'usage de technologies pouvant échapper à tout contrôle et comprenant de nombreux biais [FRA 2019] ; tout récemment, une organisation fameuse pour la défense des libertés démocratiques, l'ACLU (American Civil Liberties Union), a solennellement demandé à la police de Détroit, Michigan, de ne pas déployer les SRF [ACLU 2019] car ils font peser sur les populations des risques accrus de discrimination en fonction de la couleur de peau et, globalement, ils mettent en danger le droit à l'anonymat dont se nourrit la vie privée ; enfin le site La Quadrature du Net a également souligné les multiples dangers de l'usage massif de DRF [Quadrature du Net 2019].

A l'instar de cette alerte lancée par La Quadrature du Net, c'est du point de vue de leur inanité en termes réellement politiques, voire de l'illusion sur laquelle ils sont construits que nous voulons souligner le danger profond que représente l'usage des DRF : on ne peut pas remplacer « l'esprit de la cité » (à savoir, l'intention politique, celle qui anime le débat collectif dans le projet démocratique) par des systèmes techniques.

Vouloir remplacer l'intention politique par des systèmes techniques, cela évoque les tentatives totalitaires, lorsque, d'abord dans l'Italie fasciste, puis dans la Russie stalinienne et en Allemagne nazie dans le premier tiers du XXème siècle, des organisations autoritaires ont voulu maîtriser les oppositions qui leur étaient faites au niveau d'Etats-nations, et cela par tous les moyens techniques possibles et grâce à l'autocontrainte engendrée par l'idéologie, donc selon un processus conjuguant le niveau technologique et des dispositifs psycho-sociaux produisant des effet d'assujettissement extrêmement efficaces ainsi que l'a bien montré Hannah Arendt dans le troisième tome des *Origines du totalitarisme* [tome III, chapitre 13 : « Idéologie et terreur »].

Les DRF d'aujourd'hui font courir le même risque que les techniques d'encadrement déployés dans les Etats totalitaires d'autrefois, mais ils le font au sein de sociétés apparemment démocratiques, avec l'alibi des séductions de l'innovation technologique : la performance technologique au service du confort des individus qui effectuent des choix de consommation de manière apparemment libre – mais non réflexive, non critique. Dans les deux cas, le moyen est comparable et l'intention est la même. Le moyen est de dépolitiser la société au profit du confort « bourgeois », et l'intention est de faire abandonner l'exigence imposée par « l'esprit de la cité », à savoir, entrer dans le débat démocratique pour favoriser les disputes fécondes pour la liberté.

On peut donc dénoncer l'implantation des DRF, de même que tout usage des technologies biométriques, à l'aune de la dénonciation de l'illusion techniciste comme l'avaient fait en leur temps Ellul [1986] et Mumford [1956 et 1963]. Pour autant, il convient de remarquer que les DRF, comme les autres dispositifs biométriques, bénéficient de la complicité des usagers qui, par désir de confort ou par les séductions du narcissisme, les adoptent pour les systèmes d'accès de leurs terminaux (il est plus facile de se laisser identifier par le système technique que de mémoriser et de saisir les codes d'accès à ce système). Position d'autant plus curieuse qu'elle configure un mode étroit de l'identité, qui se superpose paradoxalement à des postures où le moi « digitalement étendu », se nourrit des fictions que sont ses propres avatars dans la sphère numérique [Belk 2013 ; Barry 2015].

En renonçant de la sorte à ce qui reste de leur droit à demeurer anonymes, les usagers-consommateurs de biométrie domestique renoncent au bienfait d'une pensée pleinement originale car, élaborée dans une forme bienfaisante de solitude, elle n'est suspecte d'être ni conformiste ni manipulée. Tel est ce parti qu'avait pris Descartes, qui en témoigne dans le passage fameux du *Discours de la méthode* (1637) où il écrit avoir « pu vivre aussi solitaire et retiré que dans les déserts les plus écartés ». Le philosophe exprimait par-là comment, afin d'élaborer une philosophie qui bouleversait tout ce qu'on savait à son époque, il lui avait semblé commode de bénéficier d'un droit à l'anonymat, en évoluant inconnu au milieu de la foule

d'Amsterdam, « grand peuple actif, et plus soigneux de ses propres affaires que curieux de celles d'autrui » [Descartes 1963, p. 601].

3/ Quelle évaluation éthique de la reconnaissance faciale ? (une critique de l'utilitarisme)

Du point de vue de l'analyse en termes éthiques, notre hypothèse de travail au sein de la chaire « éthique & IA » est de contribuer à l'approche éthique de l'IA en enrichissant la *computer ethics* : celle-ci apparaît dominée par une inspiration utilitariste, et le projet est notamment de l'enrichir grâce à celle des autres « familles » de l'éthique que nous identifions au nombre de quatre : déontologiste, arétaïque et axiologique. La notion de « famille » de l'éthique signifie, métaphoriquement, le regroupement des différentes manières de configurer l'intention éthique qui, quant à elle, coordonne toujours quelle que soit la doctrine dont on parle, la définition d'un sens pour l'action ou l'existence humaines et l'établissement de règles de conduite.

L'approche utilitariste renvoie à l'établissement (d'abord individuel ensuite collectif) du rapport coûts-bénéfices d'une décision ou d'une action. La déontologiste, à l'expérience du devoir, c'est-à-dire d'une part à la détermination de la responsabilité (et de ses différents niveaux ou de ses nuances variées) via l'examen de conscience individuelle, et de l'autre à la résolution de suivre une ligne de conduite réglée par la responsabilisation individuelle. L'arétaïque, à toutes les doctrines qui visent la définition des vertus et des vices, ces dispositions personnelles qui paraissent (sur le plan social-culturel comme sur le plan réflexif et rationnel) intéressantes à encourager ou à dénoncer. L'axiologique, aux doctrines qui procèdent en premier lieu à l'établissement ou à la reconnaissance de valeurs intrinsèquement bonnes, ce qui permet une priorisation des décisions et des actions, et une hiérarchie des genres de conduites.

Cela posé, nous voulons, dans les développements qui suivent, établir quatre choses différentes :

- l'analyse des DRF en termes utilitariste est à la fois aisée, efficace et nécessaire ;
- mais elle apparaît également limitée, voire intrinsèquement faussée, si bien qu'il convient de se méfier des fausses facilités qu'elle offre pour la décision ;
- l'analyse à l'aune des autres familles éthiques est peu intuitive, elle n'est ni intégrale (mais seulement partielle) ni totalement capable d'apporter une aide complète à la décision (mais seulement inspirante),
- elle est toutefois intéressante et importante.

a) L'analyse des DRF en termes utilitaristes est à la fois aisée, efficace et nécessaire :

L'approche utilitariste, qui évalue le coût ou le bénéfice d'une décision ou d'une action en fonction de ses conséquences, présente d'incontestables avantages pour la détermination de la valeur éthique des DRF. Elle permet en effet de construire, grâce à l'analyse objective de cas d'usage observés dans des situations réelles, ou simulées et réalistes, ainsi que par le

raisonnement logique, une réponse qui semble éthiquement satisfaisante, de manière toujours adéquate à la situation donnée et, chose très importante en matière de diffusion de l'innovation dans la société, sans jamais préjuger du résultat avant d'intégrer le point de vue des parties prenantes à la situation. En d'autres termes, son avantage vient qu'elle ne soumet l'évaluation éthique à aucune autre autorité (intellectuelle, morale, religieuse ou politique) qu'au raisonnement humain concentré sur le cas qu'il s'agit d'observer.

Aujourd'hui, dans les conditions concrètes du marché des innovations, l'approche utilitariste se voit très fréquemment associée à l'approche dite « centrée clients » (*Customer centric*) ou plus généralement à celle dite « expérience utilisateurs » (*User experience* ou *UX*). Et en tant que telle, elle est susceptible de produire des effets en termes d'acceptation ou acceptabilité des technologies, en tout cas elle y contribue efficacement.

Dans le cas de l'évaluation utilitariste des DRF, il y a lieu de penser que les parties prenantes trouveront aisément un accord sur leur validation : en dépit des différences entre leurs intérêt particulier un intérêt commun se dessinera rapidement. Et donc sur l'acceptation sociale. En effet tant les responsables des firmes qui conçoivent les composants des DRF que les responsables des organisations privées et publiques (centres commerciaux, aéroports, administrations, banques…) ont des intérêts à ce que les DRF se développent. Comme souvent, l'usager suivra, motivé d'ailleurs par ses propres intérêts en tant que consommateur de solutions ou d'applications électroniques personnalisées (augmentation de la proximité avec ses terminaux et les services auxquels il désire accéder, effets subjectifs de personnalisation et d'appropriation, promesse d'augmentation du confort par la montée de l'intimité avec ce que permet le dispositif).

Or les succès mêmes de l'approche utilitariste ne vont sans poser un certain nombre de problèmes qui pourraient rendre cette approche discutable si on se limitait à elle.

b) Elle apparaît également limitée, voire intrinsèquement faussée, si bien qu'il convient de se méfier des fausses facilités qu'elle offre pour la décision :

En dépit de l'aisance qu'il y a à mobiliser un tel outil d'évaluation, il convient de s'en méfier car, outre le caractère discutable de son aspect « sacrificiel » identifié depuis longtemps dans la littérature de théorie morale [Williams 1997 ; Baertschi 1998], il présente cet autre défaut de tendre à faire accepter certains usages technologiques (inutiles ou sur certains plans nocifs), au motif de bienfaits collectifs et individuels. Les séductions même exercées par une telle approche doivent être dénoncées comme fallacieuses et dangereuses si du moins on ne reconnait qu'elle seule comme susceptible de contribuer à la régulation éthique des technologies contemporaines liées au développement de l'IA. Cela, pour plusieurs types de raisons qui peuvent être ramenés à un ensemble que l'on peut nommer *les « angles morts de la théorie utilitariste* lorsqu'elle est appliquée à l'évaluation des technologies contemporaines :

Premièrement, dans les conditions concrètes de leur évaluation, toutes les parties prenantes n'ont pas la même importance dans le processus ; ces parties prenantes sont les investisseurs, les concepteurs, les responsables des firmes qui produisent, les responsables des organisations publiques où les DRF seront implémentées, le législateur, enfin les usagers-clients-citoyens. Il

convient de souligner le rôle particulier joué par ces derniers dans le processus d'évaluation : ce rôle est à la fois majeur et mineur car, du fait de la condition générale de la *mise en société et en marché* des DRF, il représente la source de validation ultime, soit que l'on parle en termes d'acceptation (les citoyens passifs, dont le comportement est discipliné dans l'espace social ou public, se satisfait de cette quiétude) ou en termes de création de valeur (la satisfaction du consommateur s'exprime par le consentement à payer). Le citoyen qui demande aux responsables des espaces sociaux et publics d'éprouver un sentiment de sécurité peut se trouver rassuré par l'implémentation de DRF dans ces espaces ; le consommateur qui utilise des DRF domestiques sur ses terminaux se montre séduit par eux, et ainsi la validation peut être obtenue pour le développement de ces systèmes de la part des firmes comme des espaces publics – à moins que le législateur ne nuance cette tentation, ainsi qu'il le fait régulièrement dans les sociétés technologiquement évoluées au nom du respect de la *privacy*.

A cet égard les matrices éthiques proposées par Castelluccia et Le Métayer [2019], inspirées par celles déployées en bioéthique, sont intéressantes car elles permettent d'exprimer la variété des points de vue qui entrent en jeu dans l'organisation et la validation des DRF. Cependant, leur limite vient du fait qu'en elles-mêmes, elles ne rendent pas compte du potentiel conflit entre les intérêts. Et elles n'interrogent pas les manipulations possibles du dernier terme de la chaine des parties prenantes, le citoyen-consommateur ; elles seraient en revanche valides et bien plus légitimes pour une approche éthique des DRF si on les utilisait dans un sens qui permette d'éveiller la conscience critique de ce dernier.

Deuxièmement, l'approche utilitariste ne permet nulle critique des valeurs impliquées dans l'évaluation. Plus exactement, dans ses fondements mêmes, elle repose sur des valeurs estimées indiscutables car c'est bien avec elle que le sujet de l'utilitarisme peut calculer ce qui est préférable pour son intérêt, et par suite éthiquement valable. Ces valeurs sont, au plan individuel, la recherche du bien-être et de la sécurité, au plan collectif, la recherche de l'efficacité sociale. Morale moderne essentiellement « bourgeoise », l'utilitarisme valorise le point de vue personnel de sujets autonomes dans la recherche de ce qui maximise d'abord leur intérêt individuel pragmatiquement défini, ensuite la composition sociale des intérêts individuels, ce qui transforme la société en marché (pour une critique de l'utilitarisme en ce sens, voir les analyses classiques de Rawls [Rawls 1997].

Troisièmement, l'utilitarisme, si pertinent pour déterminer ce qui est *bon* ou *mauvais*, (en fonction du calcul des préférences) a tendance à « éliminer » la vie morale, ainsi que le suggère, dans le champ de la bioéthique, Anne Maclean [Maclean 1993]. Dans son efficacité même, il simplifie des dilemmes profonds qu'il est intéressant, pour la vie morale, de laisser ouverts. En d'autres termes il ne dit rien du *bien* et du *mal*, ces valeurs qui sont, pour une subjectivité qui vise à une forme de responsabilisation morale, les objets de sa méditation fondamentale.

c) *L'analyse à l'aune des autres familles éthiques est peu intuitive, elle n'est ni intégrale (mais seulement partielle) ni totalement capable d'apporter une aide complète à la décision (mais seulement inspirante) :*

Certes, les autres familles éthiques ne bénéficient pas de la même proximité que l'utilitarisme avec les questions technologiques : l'expérience du devoir et la recherche des vertus, par exemple paraissent même fort éloignées des débats, et bien incapables de pouvoir trancher la question de savoir si les DRF constituent une bonne ou une mauvaise chose. Rassemblant des doctrines morales variées, elles constituent toutefois les indispensables relais pour une vie éthique riche. On peut certes admettre que l'éthique dont les développements contemporains de l'IA ont besoin relève d'une doctrine efficace, qui, parce qu'elle clarifie des situations d'usage, aide à la prise de décision rapide. Pour autant, perdre de vue que toute l'éthique n'est pas concernée par une telle finalité (à savoir, la prise de décision rapide) serait non pertinent et même particulièrement dangereux.

En effet, les limites mêmes de l'utilitarisme, qui constituent le revers de son efficacité pratique, rendent profitable pour les usagers-consommateurs-citoyens la fréquentation des autres familles de l'éthique. Par exemple, l'éthique déontologiste recentre les personnes sur l'expérience de leur intériorité, en les confrontant, via l'épreuve du devoir, à la dimension universelle que, selon Kant, chaque sujet rationnel retrouve au moment de savoir si la maxime de son action peut être valable pour d'autres personnes que pour lui-même. L'éthique arétaïque permet d'œuvrer à la constitution des vertus, ces dispositions acquises par l'exercice à partir du tempérament initial des personnes, qui sont un relai pour le développement et l'affirmation de leur propre caractère. L'éthique axiologique conduit la réflexion à la détermination des valeurs souhaitables pour orienter le jugement et l'action, les valeurs étant des principes pratiques estimés supérieurs et susceptibles de ce fait de hiérarchiser les préférences, de faciliter les choix de vie et de justifier les stratégies d'action.

Il n'est pas interdit de penser que les éthiques non utilitaristes peuvent contribuer à l'évaluation des technologies de l'IA. Par exemple, une analyse arétaïque pourrait établir que la paresse constituant un vice, et que le fait de déléguer à un DRF les capacités que requiert la mémorisation des codes d'identification à des terminaux personnels, il est éthiquement préférable de ne pas utiliser ce dispositif.

d) Elle est toutefois intéressante et importante :

Il est également permis d'affirmer que, dans le contexte de la perte des points de repère traditionnels sous l'effet de l'émergence des technologies de l'IA dans tous les secteurs de l'existence humaine, à une époque où presque tout fait question (rapports incertains à l'énergie disponible dans l'environnement en danger, à la santé « améliorée » ou « augmentée », au transport autonomisé, à la sécurité automatisée, etc.), la fréquentation des familles non utilitaristes de l'éthique, à défaut de permettre une assistance à la décision qui serait intégrale et rapide, fournit une ressource précieuse pour un jugement éthique approfondi.

C'est pourquoi, si elles ne sont pas intégralement déterminantes des choix d'usage technologiques, les familles de l'éthique sont fortement inspirantes pour la vie morale des personnes, laquelle constitue une dimension non superficielle de leur existence.

En particulier, l'éthique axiologique peut jouer un rôle important, en permettant d'établir pour toute question importante, au nom de quelle valeur on va agir. Ce type d'approche permet, par les débats contradictoires qu'elle peut engendrer, de voir plus clair dans le type de société qui peut aujourd'hui être désirable.

Conclusion :

La définition et la constitution d'une éthique de l'IA se trouvent mise au défi par la reconnaissance faciale. Celle-ci semble représenter un cas d'espèce qui apparait immédiatement bien plus ambigu que d'autres, tels que le véhicule autonome (terrestre, maritime ou aérien) ou la smart city ; du moins révèle-t-il mieux que les autres les dangers que la civilisation humaine encoure à privilégier de manière systématique les dispositifs automatiques à la place des actions humaines.

Le danger que font courir pour les libertés publiques et privées les DRF, pointe avancée des capteurs biométriques, est susceptible de s'accroitre rapidement, sous l'effet de trois types de facteurs. D'abord, il faut souligner l'effet du développement technologique (celui de l'IoT, celui des capteurs d'émotions). Ensuite, il convient de relever la corrélation entre ce développement et l'émergence de marchés ouverts aux organisations (privées et publiques) et aux consommateurs particuliers. Enfin, la faiblesse de l'évaluation éthique dominante, qui ramène la démarche éthique à la prise de décisions particulières sous la condition d'une acceptation globale des technologies efficaces mises en circulation sur les marchés. Ces facteurs conjugués induisent deux conséquences possibles : le déploiement massif des DRF et le fait que les capteurs biométriques non seulement permettront l'identification/authentification des individus, mais également délivreront une connaissance de leur personnalité.

L'évaluation éthique des DRF constitue donc un « problème méchant » (ainsi qu'on en a récemment proposé l'idée dans une communication : « […] ethical considerations in AI are dominated by 'Wicked Problems' » [Bennett 2019]). « *Wicked Problems* » est un terme inventé par Rittel et Webber (1973), qui désigne des problèmes difficiles à définir et caractérisés par une résistance à la résolution, basés sur des jugements de valeur et, au mieux, qui offrent des solutions nécessitant un redéveloppement constant et qui sont par nature politiques.

Si la situation apparaît d'une certaine gravité, c'est parce qu'avec le regard, le visage humain représente pour chacun d'entre nous la plus importante interface entre l'intime et le social. Toute une tradition éthique se rattache à devenir ou à rester pleinement humain en acceptant la demande que nous fait autrui par son visage, et en assumant l'altérité irréductible de l'autre par le biais de son visage, différents du mien, face au mien. Ainsi que l'écrivait Lévinas, rescapé de la Shoah, « Le visage s'impose à moi sans que je puisse rester sourd à son appel, ni l'oublier, je veux dire sans que je puisse cesser d'être responsable de sa misère. » [Lévinas 1972, p. 49]. On pourrait qu'en délégant leur sécurité et leur bien-être à des systèmes qui schématisent les visages humains par le travail des algorithmes, se profile aujourd'hui le risque d'oublier l'injonction à rester humain, sur le fond d'un déni techniciste du caractère profondément éthique de la relation interindividuelle, ce caractère que nous rappelle sans cesse,

dans sa singularité et son altérité, le visage d'autrui. Si la question de la reconnaissance faciale apparaît d'une certaine gravité, c'est, au-delà de ses aspects pragmatiques du point de vue éthique et politique, en raison du sentiment qu'une limite anthropologique est irrémédiablement franchie, les humains courent désormais un risque radical et se trouvent exposés.

Aussi, face à une telle menace, pourrions-nous émettre deux recommandations. D'abord, se dessine une ligne de défense nécessaire sur le plan de l'usage social des DRF, qui apparaît nécessaire importante bien que fondamentalement fragile. Il convient que les personnes identifiées car soumises aux DRF le soient en connaissance de cause, voire en expriment le consentement. On doit imposer aux autorités de surveillance/sécurité/police à la fois l'obligation de l'information et le respect du consentement. Seul un état d'urgence publiquement exprimé et politiquement justifié permettrait de passer outre cette double condition. Condition salutaire et minimale, mais qui apparaît impossible à garantir.

Ensuite, on peut préconiser une ligne de défense dans la relation personnelle aux DRF, à maintenir coûte que coûte : par le biais de l'éducation et de l'auto-éducation aux usages numériques, favoriser le refus personnel, toutes les fois que c'est possible, d'accepter l'identification/la personnification biométrique, notamment dans les usages personnels / séductions du marketing des commodités, car cela revient à sacrifier les conditions de sa liberté (et en premier lieu, le respect de l'anonymat) à son confort. En termes arendtiens, cela revient nier son pouvoir d'« action » au profit des conditions de sa « vie », ce qui constitue une forme de régression aussi tentante que dangereuse [Arendt 2012].

Bibliographie

- ACLU 2019 : « ACLU Condemns Detroit Board of Police Commissioners' Vote to Approve Detroit Police Department's Facial Recognition Technology Policy », 19/09/2019 : https://www.aclu.org/press-releases/aclu-condemns-detroit-board-police-commissioners-vote-approve-detroit-police
- ARENDT Hannah 2002 : *Les Origines du totalitarisme* (1951), éd. sous la dir. de P. Bouretz, Paris, Gallimard.
- ARENDT Hannah 2012 : *Condition de l'homme moderne* (1958), trad. G. Fradier, in L'Humaine condition, éd. Sous la dir. de Ph. Raynaud, Paris, Gallimard.
- BENNET Sarah Joy 2019 : « Investigating the Role of Moral Decision-Making in Emerging Artificial Intelligence Technologies », CSCW'19 (Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, p. 28-32), November 9-13, 2019, Austin.
- BAERTSCHI Bernard 1998 : « Ombres et lumières de l'utilitarisme », *Revue de théologie et de philosophie*, volume 130, 1998, cahier IV, p. 357-380.
- BARRY David M. 2015 : « Subjectivités computationnelles », *Multitudes*, 2015/2 n°59, p. 196-205.
- BELK Russel W. 2013 : « Extended Self in a Digital World », *Journal of Consumer Research*, Vol. 40, n°3 (Octobre 2013), p. 477-500.
- CACCAMO Emmanuelle, WALZBERG Julien, REIGELUTH Tyler & MERVEILLE Nicolas 2019 (dir.) : *De la ville intelligente à la ville intelligible*, Québec, Presses de l'Université du Québec, Cahiers du GERSE.

- CASTELLUCCIA Claude, LE METAYER Daniel 2019 : *Analyse des impacts de la reconnaissance faciale. Quelques éléments de méthode*, rapport 20/11/2019, accessible à l'URL : https://hal.inria.fr/hal-02373093.
- CNIL 2019 : *Reconnaissance faciale. Pour un débat à la hauteur des enjeux*, 15/11/2019, accessible à l'URL : https://www.cnil.fr/fr/reconnaissance-faciale-pour-un-debat-la-hauteur-des-enjeux
- DESCARTES René 1963 : *Discours de la méthode* (1637), in *Œuvres philosophiques*, tome I, édition de Ferdinand Alquié, Paris.
- ELLUL Jacques 1986 : *Le Bluff technologique*, Paris, Hachette Littératures.
- FRA (European Union Agency for Fundamental Rights) 2019 : « Facial recognition technology: fundamental rights considerations in the context of law enforcement », 27/11/2019, accessible à l'URL : https://fra.europa.eu/en/publication/2019/facial-recognition
- FOUCAULT Michel 1975 : *Surveiller et punir. Naissance de la prison*, Paris, Gallimard.
- FOUCAULT Michel 2004 a : *Sécurité, Population, Territoire, Cours au Collège de France, 1977-1978*, Édition établie sous la direction de François Ewald, Alessandro Fontana et Michel Senellart, Paris, Gallimard-Le Seuil.
- FOUCAULT Michel 2004 b : *Naissance de la biopolitique, Cours au Collègue de France*, 1978-1979, Édition établie sous la direction de François Ewald, Alessandro Fontana et Michel Senellart, Paris, Gallimard-Le Seuil.
- GANASCIA Jean-Gabriel 2009 : *Voir et pouvoir : qui nous surveille ?*, Paris, Editions Le Pommier.
- HARTZOG Woodrow 2018 : « Facial Recognition Is the Perfect Tool for Oppression », *Medium*, 2/08/2018, accessible à l'URL : https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66
- LEVINAS Emmanuel 1972 : *Humanisme de l'autre homme*, Editions Fata Morgana.
- LOCKE John 2001 : *Essai sur l'entendement humain* (1689), Livres I-II, trad. J.-M. Vienne, Paris Editions philosophiques J. Vrin.
- La Quadrature du Net 2019 : « Le vrai visage de la reconnaissance faciale », 21/06/2019, accessible à l'URL : https://www.laquadrature.net/2019/06/21/le-vrai-visage-de-la-reconnaissance-faciale/
- LYON David 1994 : *The Electronic Eye: The Rise of Surveillance Society* - Computers and Social Control in Context, Minneapolis, University of Minnesota Press.
- LYON David, BENNET Colin 2008 (eds.) : *Playing the Identity Card: Surveillance, Security and Identification in Global Perspective*, Londres, Routledge.
- LYON David 2009 : *Identifying Citizens: ID Cards as Surveillance*, Cambridge, Polity Press.
- MACLEAN Anne 1993 : *The Elimination of Morality. Reflections on Utilitarianism and Bioethics*, Londres, Routledge.
- MUMFORD Lewis 2008 : *La Transformation de l'Homme* (1956), trad. B. Pêcheur, Paris, Editions de l'Encyclopédie des Nuisances.
- MUMFORD Lewis : « Technique autoritaire et technique démocratique » (1963), trad. A. Gouilleux, in *Notes et morceaux choisis. Bulletin critique des sciences, des technologies et de la société industrielle*, n°11, 2014, p. 109-121.
- RAWLS John 1997 : *Théorie de la justice* (1971), trad. C. Audard, Paris, Editions du Seuil.

- RITTEL Horst W. J., WEBBER Melvin M., 1973 : « Dilemmas in a General Theory of Planning », Policy Sciences, Vol. 4, n° 2, p. 155-169, acessible à l'URL : http://www.jstor.org/stable/4531523
- WILLIAMS Bernard 1997 : « Une critique de l'utilitarisme » in Smart J.J.C. & Williams B., *Utilitarisme. Le pour et le contre* (1973), trad. H. Poltier, Genève, Labor et Fides, 1997.