

# La valeur des émotions : quel régime pour le « capitalisme mental » ?

par Judith ROCHFELD et Célia ZOLYNSKI

*Professeures de droit à l'École de droit de la Sorbonne,  
Université Paris 1 Panthéon Sorbonne, IRJS-DReDIS*

*Les potentialités de l'analyse des émotions.* – « Capitalisme cognitif<sup>1</sup> », capitalisme attentionnel<sup>2</sup> ou encore « capitalisme de surveillance<sup>3</sup> »... on ne compte plus les théories ou formulations évoquant une construction de valeur à partir d'éléments inédits. Avec l'exploitation monétaire d'informations, de connaissances, de données à caractère personnel, on assisterait à des mises en valeur qui composeraient de nouvelles formes, plus ou moins abouties et plus ou moins inédites, de capitalismes.

On peut donc ne pas éprouver de surprise devant l'hypothèse qui se diffuse aujourd'hui, selon laquelle il y aurait lieu d'en constater un naissant : un capitalisme des émotions, ou plus largement des velléités de valorisation d'une partie des activités mentales des individus. Cette hypothèse se fonde sur deux évolutions très contemporaines : l'exploitation de nos émotions, jusqu'aux plus intimes, serait aujourd'hui permise *via* divers procédés techniques, parfois réunis sous l'appellation d'« *emotion AI* » ou d'« *affective computing* » ; il en irait de même de nos images et activités mentales (nous nous concentrerons néanmoins, dans les lignes qui suivent, sur les émotions).

Cette exploitation est susceptible de prendre une ampleur inédite aujourd'hui, quantitative et qualitative, dès lors qu'elle peut s'appuyer sur des captations et analyses de plus en plus fines, appliquées à des éléments plus que variés : les mouvements et expressions de nos visages, les inflexions et textures de nos voix, nos gestes, notre écriture, nos battements cardiaques, etc. Les captations s'opèrent *via* des appareils photographiques, caméras, ordinateurs, objets connectés, capteurs biométriques, qu'ils soient ou non embarqués dans nos objets du quotidien (smartphones, montres...), tandis que leurs analyses reposent sur des algorithmes plus ou moins « intelligents ».

---

<sup>1</sup> Y. Moulier Boutang *et alii*, *Le capitalisme cognitif. La Nouvelle Grande Transformation*, broché, éd. Amsterdam, coll. « Multitudes/Idées », 2008, 315 p.

<sup>2</sup> T. Wu, *The Attention Merchants: the Epic Scramble to Get Inside our Heads*, Knopf, 2016. Adde N. Srnicek, *Le capitalisme de plateforme. L'hégémonie de l'économie numérique*, Montréal, Lux, 2018, 154 p.

<sup>3</sup> S. Zuboff, *L'Âge du capitalisme de surveillance*, trad. B. Formentelli et A.-S. Homassel, Zulma, 2020, 864 p.

Elles prennent également une ampleur temporelle et spatiale inédite – potentiellement continue et panoptique – tant elles s’opèrent pour l’ensemble de nos pans de vie (en ligne et hors ligne) et intègrent aussi notre environnement et les contextes dans lesquels les émotions collectées se développent. En conséquence, on nous promet de pouvoir détecter et interpréter les manifestations de nos émotions les plus subtiles, les plus intimes : rien n’est ou ne serait plus ignoré de nos craintes, impatiences, ennuis, désirs, que ceux-ci se manifestent au travers d’un clignement d’œil, d’un rythme de notre souffle, d’une parole, du léger affaissement de notre mâchoire... Là un bracelet qui serait à même de « lire » les émotions en écoutant la voix de son utilisateur pour en déduire exaltation, hésitation, lassitude, joie ou encore confusion<sup>4</sup>. Là encore un outil de « lecture » des expressions faciales, en temps réel, censé pouvoir détecter le niveau de stress de passagers à l’atterrissage et à l’arrivée en gare, de détecter des comportements suspects dans l’espace public ou un manque d’attention chez un conducteur en vue d’augmenter la sécurité de la conduite<sup>5</sup>. Ici, un système permettant de mesurer l’attention des élèves en situation de travail<sup>6</sup>, particulièrement promu à l’occasion de la crise du Covid (c’est en Chine, pour l’heure, que des logiciels analysent le visage des enfants pendant leurs cours à distance et évaluent leurs émotions pour permettre à leur professeur de mesurer leur motivation et de les noter<sup>7</sup>...). Songeons encore aux potentialités offertes par des dispositifs tels que les casques de réalité virtuelle qui permettront demain de naviguer au sein de divers « métavers » (de mondes virtuels parallèles) et ouvriront à des bénéfices juteux pour les fins limiers du marketing qui seront capables d’utiliser les émotions ressenties par les utilisateurs, émotions constatables par l’intermédiation des avatars<sup>8</sup>. Des domaines se trouvent particulièrement soumis à la pression de l’adoption de ce type de métriques, à l’instar des « *RH analytics* » censées assurer aux employeurs de pouvoir détecter si leurs salariés sont heureux dans leur travail ou risquent de quitter l’entreprise, frisent le surmenage ou

---

<sup>4</sup> V. le bracelet connecté baptisé « Halo » lancé par Amazon en août 2020 : [<https://press.aboutamazon.com/news-releases/news-release-details/introducing-amazon-halo-and-amazon-halo-band-new-service-helps>].

<sup>5</sup> L’entreprise Muvraline annonce développer cet outil de reconnaissance des émotions *via* l’image pour, plus généralement, déterminer automatiquement l’état de personnes indument nerveuses, agressives ou craintives ou encore pour évaluer, « d’une manière non invasive et fiable », l’attrait de produits ainsi que la capacité du contenu publicitaire à attirer efficacement l’attention et l’intérêt de spectateurs.

<sup>6</sup> V. la solution proposée par la société 4 Little Trees : [<https://www.findsolutionai.com/4littletrees>].

<sup>7</sup> P. Sirinelli et S. Prévost, « Reconnaissance émotionnelle, connaissance irrationnelle ? », *Dalloz IP/IT* 2021. 237.

<sup>8</sup> V. R. Chatellier, « Métavers : réalités virtuelles ou collectes augmentées ? », *LINC* nov. 2021, disponible à l’adresse suivante : [<https://linc.cnil.fr/fr/metavers-realites-virtuelles-ou-collectes-augmentees>].

nuisent à l'ambiance de groupe par une humeur maussade<sup>9</sup> ; de permettre aux aspirants à un emploi d'évaluer s'ils vont réussir à un entretien d'embauche sur la base de l'analyse de « leurs micro-expressions » ou du « ton de leur voix ». Un article récemment paru dans le journal *Nature* évaluait ainsi l'actuel développement de l'« intelligence artificielle » d'interprétation des émotions humaines à un marché de 37 milliards de dollars d'ici à 2026<sup>10</sup>. On mesure en conséquence les potentialités techniques et monétaires de ces analyses...

Certes, l'importance capitalistique des émotions peut ne pas paraître neuve : leur manipulation fondait déjà nombre de mécanismes publicitaires, assis entre autres sur le désir mimétique (au même titre que la propagande a toujours su flatter les désirs et les peurs pour pousser et diriger les votes). Certes, également, on peut n'y voir qu'un prolongement de techniques numériques antérieures. Ces collectes et analyses prennent en effet la suite d'une première génération de « *quantified self* » (quantification de soi) qui reposait déjà sur des captations diverses opérées par les objets connectés appliqués sur le corps : *via* différents capteurs, on mesure et analyse des données physiques et physiologiques « extérieures » (vitesse de battements du cœur, poids, position des yeux, du corps, etc.). Pour autant, des différences doivent être relevées. D'une part, on collecte et analyse désormais des données de sphères censément plus profondes qui n'étaient auparavant accessibles que par des dispositifs médicaux<sup>11</sup> : même en laissant de côté les images neuronales, on se saisit aujourd'hui des signaux extérieurs d'émotions – les mouvements du visage, les modulations de la voix, etc. – dont on infère une connaissance neuronale – celle d'un état d'esprit. D'autre part, répétons que ces captations pourraient bien se développer en continu et en intégrant les environnements. On parviendrait donc au traitement du « *digital mind* », c'est-à-dire à la quantification de l'esprit humain individuel *via* le traitement de ces données « mentales<sup>12</sup> », dont des définitions circulent déjà comme celle renvoyant à « toute donnée qui peut être organisée et traitée pour inférer l'état d'esprit d'une personne, soit ses états cognitif, affectif et conatif<sup>13</sup> ». Largement conçues, ces données embrasseraient ainsi les images

---

<sup>9</sup> Sur cette question spécifique, v. J. Rochfeld, « Les émotions du salarié. Traitements et décisions : quelles limites », in *Mélanges en l'honneur de P.-Y. Verkindt*, à paraître, 2022 et R. Chatellier, « *Affective computing* : les casques qui analysent le cerveau », *LINC* juin 2018, disponible à l'adresse suivante : [<https://linc.cnil.fr/fr/affective-computing-des-casques-qui-analysent-le-cerveau>].

<sup>10</sup> K. Crawford, « Time to regulation IA that interprets human emotions », *Nature* 8 avr. 2021, vol. 592, p. 167.

<sup>11</sup> Pour un historique plus détaillé, M. Ienca et G. Malgieri, « Mental Data protection and the GDPR », 5 mai 2021, disponible en version préparatoire : [<https://ssrn.com/abstract=3840403>], p. 3.

<sup>12</sup> M. Ienca et G. Malgieri, art. préc., p. 4.

<sup>13</sup> M. Ienca et G. Malgieri, art. préc., p. 4 (traduction de « *any data that can be organized and processed to infer the mental states of a person, including their cognitive, affective and conative states* »).

cérébrales, les émotions, les souvenirs, les intentions, que celles-ci soient captées par une observation directe du cerveau ou inférées d'analyse de manifestations extérieures<sup>14</sup>.

*Les mirages des analyses d'émotions.* – Le panorama peut évidemment faire frémir, tant on entrerait potentiellement dans une ère de transparence absolue où les personnes pourraient se trouver privées de toute sphère d'intimité. Il faut toutefois le nuancer quelque peu car les points faibles (pour le moment...) de ces technologies ne sont pas à ignorer et on doit se garder de se laisser submerger par les promesses des discours qui les portent et en enjolivent l'efficacité. *A minima*, et comme beaucoup de techniques de cet ordre, les analyses impliquées ne se déploient pas sans biais de lectures et d'inférences susceptibles de rendre les constats proposés inexacts. *A maxima*, on peut affirmer, avec le rapport rédigé en 2020 par le Centre *ad hoc* pour l'intelligence artificielle (CAHAI) sis auprès du Conseil de l'Europe, qu'« aucune preuve scientifique solide ne corrobore l'idée qu'il serait possible de "lire" les émotions ou l'état d'esprit d'une personne sur son visage ou dans sa démarche, son rythme cardiaque, le ton de sa voix ou sa température – sans parler de prédire ses comportements futurs<sup>15</sup> ». Dans le même sens, un groupe de scientifiques concluait dans une étude récente que la reconnaissance des émotions par une « intelligence artificielle » pouvait tout au plus interpréter, subjectivement, comment une autre personne réagissait<sup>16</sup>, cette « interprétation » ne coïncidant toutefois pas nécessairement avec les sentiments réels. En réalité, l'algorithme ne fait qu'« étiqueter » cette interprétation, par ailleurs fortement dépendante du contexte et de la culture dans lesquels l'analyse se déploie<sup>17</sup>. Il existe donc de forts risques d'erreurs (sans parler des risques de discriminations) et cette fiabilité insuffisante doit impérativement être prise en compte, notamment quand on prétend par exemple, ainsi qu'évoqué, étendre ces usages à la gestion de ressources humaines<sup>18</sup>. Il n'en demeure

---

<sup>14</sup> M. Ienca et G. Malgieri, art. préc., p. 18.

<sup>15</sup> C. Muller (dir.), *Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit*, rapport du Comité *ad hoc* sur l'IA (CAHAI), Conseil de l'Europe, 24 juin 2020, CAHAI(2020) 06 fin, n° 21.

<sup>16</sup> L. F. Barret *et al.*, « Corrigendum: Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements », *Psychological Science in the Public Interest* 20, p. 1-68, [doi:10.1177/1529100619832930] ; T. Le Mau, K. Hoemann, S. H. Lyons *et al.*, « Professional actors demonstrate variability, not stereotypical expressions, when portraying emotional states in photographs », *Nat. Commun* 2021, 12, p. 5037 [https://doi.org/10.1038/s41467-021-25352-6] ; et déjà R. E. Jack *et al.*, « Facial expressions are not culturally universal », *Proc. Nat. Acad. Sci. USA* 8 mai 2012, 109(19), p. 7241-7244.

<sup>17</sup> Rapport du haut commissaire aux Nations unies, *The Right to Privacy in the Digital Age*, sept. 2021, A/HRC/28/39, p. 28.

<sup>18</sup> Défenseur des droits, *Technologies biométriques : l'impératif respect des droits fondamentaux*, 2021, p. 6, spéc. p. 10, disponible à l'adresse suivante : [https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/rap-biometr-num-01.09.21.pdf], pointant le fait qu'« une technologie de reconnaissance des émotions dans le cadre d'une procédure de

pas moins que ces promesses sont émises, trouvent des destinataires à l'écoute et que les technologies qui les fondent se diffusent. Le tout oblige à prendre au sérieux les valorisations à l'œuvre et celles que l'on peut anticiper.

*Les finalités des analyses d'émotions.* – La question de l'encadrement de ces usages se fait d'autant plus complexe que ces techniques peuvent se déployer en servant plusieurs maîtres et qu'on ne doit pas raisonner de la même façon selon les finalités poursuivies<sup>19</sup>. On laissera ainsi de côté, dans les analyses qui suivent, les utilisations relatives à la surveillance étatique ou à des fins médicales (notamment des images cérébrales), tant celles-ci soulèvent des problématiques particulières<sup>20</sup>. On aurait également tendance à mettre à l'écart le « *quantified self* », même élargi à des signaux nouveaux : les captations et analyses sont censées, là, relever d'utilisations personnelles de connaissance de soi, plutôt inoffensives. Mais on retrouvera en réalité cette hypothèse très rapidement : ces données ne restent pas en général sous la seule coupe de la personne concernée ; le plus souvent, en effet, ce « *quantified self* » donnera lieu à des traitements de données par d'autres que cette dernière et à une monétisation, sur fond de consentement « faible » obtenu par accord à des conditions générales non lues (celles des divers objets connectés). Cette hypothèse tombe aussi, en conséquence, dans le domaine des modes de valorisation qui se mettent en place et prennent pour cible l'analyse des émotions à des fins commerciales, sur laquelle nous allons nous concentrer.

*Les limites aux analyses des émotions.* – Or, celle-ci soulève des problématiques spécifiques, dont le dédicataire de ces lignes s'est déjà fait l'écho<sup>21</sup>. C'est donc afin de lui rendre hommage, en tentant de prolonger ici sa réflexion, et pour lui témoigner notre émotion, non valorisée, d'avoir travaillé à ses côtés pendant de nombreuses années et d'avoir souvent été orientées, sous son impulsion moderne et visionnaire, vers les problématiques les plus novatrices, que l'on s'intéressera à ces modes de valorisation. On soulignera alors qu'ils se déploient selon au moins deux modalités, qui vont de captations et « interprétations » d'affects existants, à des jeux bien plus subtils et puissants consistant à induire et orienter les comportements (I). Sur les fondements de ces constats, on plaidera en conséquence pour l'établissement de limites à ce capitalisme

---

recrutement peut déterminer à tort qu'un candidat est « nerveux » et lui assigner une note lui retirant toute chance d'être embauché ».

<sup>19</sup> Pour une catégorisation plus précise et approfondir les finalités possibles – médicales et de recherche, d'intérêt public, d'intérêt légitime des responsables de traitement –, v. M. Ienca et G. Malgieri, art. préc., p. 12-13.

<sup>20</sup> Sur les imageries cérébrales, v. particulièrement L. Thomasset, *La neuroéthique saisie par le droit. Contribution à l'élaboration d'un droit des neurotechnologies*, Université Paris 1 Panthéon-Sorbonne, 2021.

<sup>21</sup> Avec S. Prévost, art. préc.

mental, limites fondées sur la consécration de nouveaux droits propres à asseoir la défense de l'intégrité mentale qu'il se révèle susceptible d'atteindre (II).

## I. - LES DIMENSIONS PLURIELLES DU CAPITALISME MENTAL

Pour approfondir les ressorts du capitalisme mental, et plus particulièrement celui des émotions, on démontrera qu'il peut se déployer selon différents degrés d'analyse des données. Pour rester schématiques, on peut en distinguer deux principaux : tout d'abord, il peut reposer sur des techniques de « captologies » tendant à saisir des émotions existantes chez les personnes, afin de monétiser ces dernières et aux fins de conserver leur attention, attention qui profite à des contenus publicitaires ou suscite de l'engagement envers diverses offres marchandes (A) ; ensuite, il peut en appeler à « l'*affective computing* » *stricto sensu*, soit le fait de susciter une émotion qui n'existait pas chez la personne concernée, afin de la projeter dans une interaction plus importante avec un message ou une machine (B).

### A. - ÉMOTIONS ET « CAPTOLOGIE » : EXPLOITER LES EMOTIONS

« *Captologie* » et exploitation des émotions – *Les éléments de permanence*. – Un premier niveau de valorisation des émotions réside ainsi dans le recours à des techniques de « captologie » d'émotions existantes de sorte à utiliser ces dernières comme des vecteurs d'attention et d'engagement, sur fond de connaissance fine des personnes et de ciblage de ces dernières avec des informations ou des offres de produits et services particulièrement adaptées à leurs goûts et besoins. Les émotions ne sont ici utilisées, comme pour les techniques publicitaires classiques, que pour mener l'exploitation de nos biais cognitifs ou nos besoins psychologiques et afin de limiter notre rationalité et de nous emporter vers des actions impulsives ou, à tout le moins, mues par les émotions actionnées<sup>22</sup>. Comme pour les techniques plus modernes également, les collectes et analyses d'émotions s'intègrent à des schémas d'enfermement de l'attention : depuis un moment, certaines grandes plateformes s'attachent à capter et retenir l'attention des individus grâce à une alliance de connaissance profonde (ou censée être profonde) de leurs goûts, besoins et envies – connaissance fondée sur l'exploitation de leurs données à caractère personnel – et de dispositifs de ciblage adaptés ; à partir de ce que l'on induit de chacun, on présente des recommandations de contenus personnalisés (recommandations fournies grâce aux analyses d'algorithmes de traitement des achats et consommation de la masse des utilisateurs), des notifications permanentes, des offres d'achat profilées, des

---

<sup>22</sup> V. not. G. Donadieu, *Les pratiques commerciales cognitives*, Université Montpellier, 2021, nos 162 s. Adde A.-L. Sibony *et al.*, *Nudge and the Law. A European Perspective*, Bloomsbury, 2015, 336 p.

informations... le tout sur fond de « design » d'interfaces sans cesse améliorés, conçus par des équipes d'ingénieurs, de graphistes et de spécialistes des sciences comportementales<sup>23</sup>. Tout juste pourrait-on noter que le maniement des émotions dans ce contexte serait à même d'accroître le risque déjà bien identifié d'utilisation de techniques pernicieuses et occultes, soit certains « design » manipulateurs dont on parle comme de « *dark patterns* » et que l'on définit comme « des éléments et mécanismes mis en place dans les interfaces pour influencer les décisions des utilisateurs dans un sens qu'ils ne choisiraient pas nécessairement s'ils étaient en face d'une interface loyale et transparente<sup>24</sup> » (la Commission nationale de l'informatique et des libertés (CNIL) les a déjà cartographiés<sup>25</sup>).

« Emotional machin » et interprétation des émotions – *Les éléments d'évolution*. – Néanmoins, loin de cette permanence et d'un simple ajout de « matériau » de connaissance des personnes, on peut considérer que l'on quitte l'effet des techniques publicitaires et attentionnelles « traditionnelles » quand on collecte et analyse aujourd'hui les émotions. Comme le souligne en effet Soshana Zuboff (auteure états-unienne qui a cependant davantage avancé l'hypothèse d'un « capitalisme de surveillance » que d'émotions), « ces réalisations ne se comparent pas à l'application scientifique des pouvoirs historiques de la machine et à la restitution continue et micro-mesurée de nos sentiments plus ou moins réels<sup>26</sup> » : leur but est non seulement d'« en savoir plus sur vous que vous-même » mais surtout de coder et de calculer vos émotions pour les restituer sous forme d'information comportementale particulièrement utile pour les secteurs du marketing et de la publicité et atteindre une « *emotion-driven advertising* » ou publicité conduite par la donnée d'émotion<sup>27</sup>. Celle-ci a alors pour particularité de se faire bien plus intrusive et manipulatrice que les précédentes d'autant que d'un point de vue quantitatif, ainsi qu'on l'a évoqué précédemment, « ce balayage émotionnel » pourrait demain devenir aussi

---

<sup>23</sup> M. Le Roy, F. Levin et C. Zolynski, « L'économie de l'attention saisie par le droit : plaidoyer pour un droit à la protection de l'attention », *Dalloz IP/IT* 2019. 614, spéc. p. 615. *Adde* CNIL, *La forme des choix. Données personnelles, design et frictions désirables*, *Cahiers IP Innovation et prospective*, 2019, n° 06.

<sup>24</sup> E. Harry, « *Dark patterns* : quelle grille de lecture pour les réguler ? », *LINC* sept. 2019 : [<https://linc.cnil.fr/fr/dark-patterns-quelle-grille-de-lecture-pour-les-reguler>].

<sup>25</sup> CNIL, *La forme des choix*, préc., p. 28-29. V. également ForbrukerRådet, « Deceived by design. How tech companies use dark patterns to discourage us from exercising our rights to privacy », 27 juin 2018 : [<https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf>].

<sup>26</sup> S. Zuboff, *op. cit.*, p. 384.

<sup>27</sup> R. Booth, « Facebook Reveals News Feed Experiment to Control Emotions », *The Guardian* 29 juin 2014, disponible à l'adresse suivante : [<https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds>].

*Adde*, R. Calo, « Digital Market Manipulation », *The George Washington Law Review*, vol. 82:995, p. 995.

étendu et banal que celui fondé sur les cookies aujourd'hui (ces petits espions qui tracent « seulement » nos navigations sur le réseau) et pourrait pister toutes nos émotions lors de nos navigations ou utilisations d'objets connectés ; que, d'un point de vue qualitatif, ces techniques de captologie, « bien » orientées et utilisées, permettraient d'augmenter bien davantage que les précédentes le « temps utile » que passent les utilisateurs et de les profiler encore plus finement. Pour toutes ces raisons, elles promettent d'accroître la rentabilité des modèles d'affaire et de monétiser en temps réel les émotions : on utilisera par exemple votre aversion « captée » à la perte ou à la peur de rater quelque chose pour vous faire parvenir en une seconde un message ou une information qui vous fera rester sur un site ou une application ; on vous présentera le produit ou l'information qui correspondra exactement à ce que votre état d'esprit réclame à un instant « t ». L'exploitation des émotions ne se limitera pas, par ailleurs et comme on l'a déjà montré, à quelques sites ou applications marchands et se développera de façon probable dans toutes les sphères de nos vies, engageant en outre de plus en plus nos corps. En témoigne le déploiement de la myriade d'objets biométriques à des fins d'interpréter les émotions et les projets portés par les diverses « jeunes pousses » contemporaines promettant des applications élargies. On aboutirait alors, dans cette version généralisée, à une monétisation en temps réel des individus par leurs émotions, monétisation reposant sur une classification et un « étiquetage » dont, rappelons-le, la fiabilité n'est rien moins qu'assurée...

## B. - *EMOTIONS ET AFFECTIVE COMPUTING : GENERER L'EMOTION*

*Affective computing et émotions générées par les machines.* – À un second degré de valorisation des émotions, ensuite, on trouve les techniques connues sous l'expression d'« *affective computing* ». Plutôt que d'une technique, il s'agit en réalité de l'intrication d'une diversité de techniques au service de trois finalités : non seulement détecter des émotions (1) et « raisonner » sur le fondement de ces informations affectives – comme précédemment – (2), mais également générer des expressions affectives nouvelles (3). On suscite donc ici des interactions humain-machine et, pour ce faire, on se fonde « sur la notion de théorie de l'esprit, qui désigne l'aptitude cognitive d'un individu à attribuer des états mentaux inobservables (intention, désir, croyance, émotion) à soi-même ou à d'autres individus », c'est-à-dire sur le fait que « [l]orsqu'un être humain perçoit une émotion chez autrui, il s'agit de perception subjective qui lui est propre<sup>28</sup> » (c'est l'objet de « l'informatique affective », puisque c'est désormais ainsi que certains auteurs la

---

<sup>28</sup> Comité national pilote d'éthique du numérique, CNPEN, *Agents conversationnels : enjeux d'éthique*, avis n° 3, nov. 2021, p. 20. V. également CNIL, *À votre écoute. Exploration des enjeux éthiques, techniques et juridiques des agents vocaux*, Livre blanc n° 1, 2020, p. 10 s.



nomment). On agira alors sur cette perception par l'humain de la machine, mais également sur « [l]a perception de la machine [qui] est, quant à elle, une estimation probabiliste obtenue par un calcul dans un réseau stochastique, construit à partir de données émotionnelles produites par de nombreux sujets humains<sup>29</sup> ». Pour mieux comprendre cette direction, on peut par exemple considérer les agents conversationnels (dont les Siri, Alexa ou Google largement diffusés) qui composent de bons archétypes de ce qui pourrait être à l'œuvre très bientôt avec tous nos objets connectés du quotidien. Ces systèmes, que l'on peut définir comme des « système[s] informatique[s] qui dialogue[nt] de façon autonome avec des utilisateurs en langage naturel », peuvent en effet être « dotés de modules de détection des émotions, à partir d'indices linguistiques (on parle alors de détection des sentiments), mais aussi à partir des indices paralinguistiques (modulation dans la voix) et d'indices corporels (faciaux, gestuels, posturaux ou encore tactiles quand il s'agit d'un robot ou d'un objet incarné)<sup>30</sup> ». Certains même « intègrent un module de prédiction du comportement émotionnel, attentionnel ou intentionnel des humains<sup>31</sup> ». Certains « peuvent également simuler une expression affective dans leurs répliques écrites ou synthétisées<sup>32</sup> ». Dans ce cas alors, non seulement l'analyse des émotions et des sentiments permet une meilleure interprétation de la situation et une stratégie de dialogue plus adéquate<sup>33</sup> mais le développement de ces nouvelles interactions humain-machine entraînera des répercussions sur les émotions de l'humain du binôme en question : elles généreront chez lui des émotions nouvelles. Par exemple, la startup Emoshape se propose de « déterminer en temps réel les émotions des utilisateurs et [de] permettre aux robots et autres applications de répondre avec un état émotionnel en phase avec celui de l'utilisateur », en se fondant sur des « informations de bas niveau issues de diverses sources d'informations comme les analyses du visage réalisées par Affectiva, des analyses de la voix réalisées par d'autres outils et d'autres informations issues de capteurs divers (pouls...) » ; de permettre à une intelligence artificielle interagissant avec

---

<sup>29</sup> CNPEN, avis préc.

<sup>30</sup> L. Devillers, « Le dialogue humain-machine. Intelligence artificielle/intelligence humaine : manipulation et évaluation », *Futuribles* 2019, n° 433, p. 51, spéc. p. 52.

<sup>31</sup> CNPEN, avis. préc.

<sup>32</sup> CNPEN, avis préc.

<sup>33</sup> Pour une présentation de ces enjeux et une liste d'exemples de systèmes développées en ce sens, v. O. Ezratty, « Où en est l'IA émotionnelle », 2 mars 2018, disponible à l'adresse suivante : [<http://www.oezratty.net/wordpress/2018/ia-emotionnelle/>]. *Adde*, pour d'autres exemples, R. Chatellier, « Captation des émotions : comment vous le direz pourra être retenu contre vous... », *LINC* avr. 2018, disponible à l'adresse suivante : [<https://linc.cnil.fr/fr/captation-des-emotions-comment-vous-le-direz-pourra-etre-retenu-contre-vous>] ; Conseil de l'Europe, Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme, 8 avr. 2020, CM/Rec(2020)1, p. 192.

l'utilisateur « d'adopter son propre état émotionnel, sur une palette riche de 64 trillions d'émotions différentes, que ce soit par de la parole de synthèse comme avec WaveNet de DeepMind, de la génération d'avatars ou même la gestuelle dans le cas d'un robot humanoïde<sup>34</sup> ». Nul doute alors qu'il y aura des émotions humaines générées par la réaction-émotion de la machine...

*Affective computing et émotions perturbées par les machines.* – On franchit en outre une étape supplémentaire lorsque l'interaction décrite est susceptible d'encourager la personne concernée à changer d'émotion et de comportement et que c'est ce changement qu'une interface est capable d'exploiter sagement : dans la « conversation », en se fondant sur les émotions générées ou en en générant d'autres, on introduit des *nudges* ou incitations douces<sup>35</sup>. Le risque est d'autant plus présent que les techniques d'apprentissage qui sous-tendent le fonctionnement de ces interfaces permettent leur adaptation « au fur et à mesure aux habitudes, émotions et goût » de leurs utilisateurs<sup>36</sup> et, partant, une connaissance extrêmement fine de ces derniers et des voies d'une influence possible... Et l'on n'abordera que peu ici les effets d'ores et déjà dénoncés de la « co-adaptation humain-machine », soit le fait que la machine s'adapte à l'humain et l'humain à la machine, avec des risques d'anthropomorphisation certains au sens de « projections des réactions affectives et probablement emphatiques envers des entités artificielles<sup>37</sup> » et de perturbations encore plus poussées des émotions des humains. Si l'on ajoute la considération de contextes particulièrement sensibles où s'effectuera la production de ces émotions nouvelles et de jeux possibles sur ces dernières, comme celui de « *dead bots* » par exemple, c'est-à-dire d'agents conversationnels susceptibles de figurer une conversation avec une personne morte (conversation qui résultera d'un apprentissage à partir des données langagières ou textuelles de la personne défunte<sup>38</sup>), on mesure les questions éthiques et juridiques d'ores et déjà devant nous.

---

<sup>34</sup> Elle ajoute que « Le système s'enrichit de plus par l'apprentissage pour développer des états émotionnels associés aux utilisateurs qui interagissent avec lui. Il peut par exemple être associé à un générateur de langage naturel pour lui permettre d'accentuer son intonation en fonction des interactions émotionnelles avec l'utilisateur et des textes générés par l'IA ».

<sup>35</sup> V. sur ce point les travaux menés dans le cadre de la chaire HUMAINE (*Human-machine affective interaction & ethics*).

<sup>36</sup> L. Devillers, art. préc., p. 55.

<sup>37</sup> L. Devillers, « Des interfaces traditionnelles hommes-machines aux machines empathiques. Vers une coadaptation humain-machine », *Enjeux numériques. Annales des mines* déc. 2020, n° 12, p. 78, spéc. p. 80.

<sup>38</sup> CNPEN, avis préc., p. 14. V. également C. Buégin-Faynel, « Transcender la mort avec l'intelligence artificielle », in M. Bouteille-Brigant (dir.), *La personne face à l'intelligence artificielle*, 2021, p. 77 ; J. Rochfeld, « Les avatars d'éternité. Vers de nouvelles personnes résiduelles compassionnelles ? », in *Mélanges en l'honneur de C. Labrusse*, à paraître en 2022.

## II. - PENSER LES LIMITES DU CAPITALISME MENTAL

Pour y répondre, le droit positif ou prospectif n'est pas complètement désarmé. Pour autant, les solutions qu'il offre s'avèrent largement insuffisantes à ce jour et les justifications en termes de protection de droits fondamentaux qui pourraient venir les porter, imparfaitement dessinées. Cela conduit à les explorer plus avant et à soutenir la consécration de nouvelles limites, que ce soit à la licéité même de ces traitements des émotions et d'inférences d'émotions (A) ou aux manipulations auxquels ils peuvent aboutir dans les hypothèses où ces traitements seraient admis (B).

### *A. - DES LIMITES AUX TRAITEMENTS DES DONNEES EMOTIONNELLES : VERS UN DROIT AU RESPECT DE L'INTEGRITE PSYCHIQUE ?*

*La qualification en données à caractère personnel.* – Si l'on entend faire état des limites aux traitements mêmes des émotions, ou plus exactement aux signaux physiques d'émotions (modulation de la voix, froncement des sourcils, etc.), il faut se tourner vers le droit des données à caractère personnel pour trouver quelques pistes de réactions. Il faut toutefois immédiatement évoquer ce que d'aucuns nomment « *the inception problem* », c'est-à-dire le fait que l'approche par les seules données ne dit rien de la protection ou des limites à dresser face à une intrusion éventuelle dans la source de ces données, à savoir le corps ou le cerveau eux-mêmes<sup>39</sup>. On peut donc voir là un premier facteur d'insuffisance ou, comme le formulent certains, de « gap » de protection. Mais si, comme nous l'avons fait jusqu'à présent, on resserre l'analyse sur les traitements des seules données d'émotions captées ou analysées *via* des manifestations extérieures au cerveau et sans atteinte physique au corps, on peut effectivement penser que le droit des données érige des remparts à analyser. Pour le confirmer, il faut tout d'abord vérifier que ces signaux traités puissent bien être qualifiés de données à caractère personnel et tomber sous l'empire de ce pan du droit. Cela ne fait pas beaucoup de doutes à la lecture de l'article 4, 1, du RGPD, qui définit les données à caractère personnel comme « toute information se rapportant à une personne physique identifiée ou identifiable » et qui répute « être une "personne physique identifiable" une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ». En effet, et d'une part, les

---

<sup>39</sup> M. Ienca et R. Andorno, « Towards New Human Rights in the Age of Neuroscience and Neurotechnology », *Life Sciences, Society and Policy* 2017, 13:5, p. 14.

données d'émotions permettent bien de singulariser une personne (parmi les autres). D'autre part, elles composent bien des données propres à « son identité physiologique » (manifestations physiques des émotions) et « psychique » (analyse et inférence d'états psychologiques)<sup>40</sup>. L'avis de l'ancien groupe des autorités européennes de contrôle des données, dit de l'article 29 (G29), ne s'y était pas trompé qui, en 2007, considérait que la définition des données à caractère personnel devait être large et pouvait embrasser « les informations relatives [aux] activités [de la personne], quelles qu'elles soient, tout comme celles concernant [...] son comportement [économique ou social à cette époque] », comportement dans lequel on peut aujourd'hui faire entrer les signaux d'émotions<sup>41</sup>. Dans un secteur voisin qui nous intéresse également, la Cour de justice de l'Union européenne, dans ses arrêts *Breyer* du 19 octobre 2016<sup>42</sup> et *Nowak* du 20 décembre 2017<sup>43</sup>, a non seulement retenu une approche large des données à caractère personnel mais l'a fait à l'égard de données en lien avec les capacités mentales de l'individu. Notamment, dans la seconde affaire, des réponses écrites fournies par le candidat lors d'un examen professionnel furent qualifiées de données à caractère personnel (ainsi que les annotations de l'examineur) au sens où elles « reflète[nt] le niveau de connaissance et de compétence » (du candidat, tandis que celles de l'examineur renvoient à « l'avis ou l'appréciation [de ce dernier] sur les performances individuelles » du candidat). Pour la Cour surtout, conformément aux critères émis par le G29, ces données permettent de singulariser la personne et composent des informations qui, « en raison de [leur] contenu, [leur] finalité ou [leur effet,] [...] sont] liée[s] à une personne déterminée<sup>44</sup> ». Des compétences cérébrales aux états d'esprit et émotions, il n'y a qu'un pas qu'il est aisé de franchir et l'on peut donc estimer que les données d'émotions (signaux extérieurs et inférences) se classeront sans mal parmi les données à caractère personnel.

*La qualification en données de catégories particulières.* – En revanche, il est bien moins aisé de déterminer si les données d'émotions relèvent de la qualification des données sensibles

---

<sup>40</sup> On peut seulement mettre à l'écart les hypothèses où de telles données seraient dissociées de la personne qui en est à l'origine et, en conséquence, pourrait être non identifiantes. Ce sont les hypothèses d'anonymisation par exemple. Dans ce cas, elles ne permettraient pas de singulariser la personne. Sur ce point v. M. Ienca et G. Malgieri, art. préc., p. 8.

<sup>41</sup> G29, avis 4/2007 sur le concept de données à caractère personnel, 20 juin 2007, WP136, p. 7-8.

<sup>42</sup> CJUE 19 oct. 2016, *Patrick Breyer c/ Bundesrepublik Deutschland*, aff. C-582/14, D. 2016. 2215, *Dalloz* IP/IT 2017. 120, obs. G. Peronne et E. Daoud, *RLDI* 2017, n° 4944, note B. Pautrot, *CCE* 2016, comm. 104, note N. Metallinos, *Rev. int. compliance* 2016, comm. 130, note M. Griger et J. Schwartz.

<sup>43</sup> CJUE 20 déc. 2017, *Peter Nowak c/ Data Protection commissioner*, aff. C-434/16, *CCE* mars 2018, comm. n° 23, obs. N. Metallinos.

<sup>44</sup> *Ibid.*, pts 37, 43 et 35.

(appellation française) ou de catégories particulières (appellation européenne)<sup>45</sup>. La réponse est pourtant importante puisque, de cette qualification, est censé découler un régime d'interdiction des traitements (certes faible puisqu'assorti de multiples exceptions dont celle, très attaquée à juste titre, du consentement donné par la personne concernée au traitement)<sup>46</sup>. Or, au soutien d'une réponse négative, il faut constater que l'état d'esprit d'une personne, son état affectif ou ses émotions, ne figurent pas dans la liste des données considérées comme de catégories particulières par l'article 9 du RGPD<sup>47</sup> : celui-ci vise en effet toute donnée à caractère personnel « 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 ». Seules trois composantes pourraient embrasser quelques données relevant d'émotions, d'état d'esprit ou encore de situations mentales, et encore à certaines conditions. En premier lieu, les émotions ou données mentales qui révéleraient un état de santé de la personne – qui permettraient par exemple d'identifier une pathologie – pourraient tomber sous la qualification de « données concernant la santé ». En deuxième lieu et dans un sens voisin, les données qui relèveraient de la « vie sexuelle », comme celles traduisant une excitation par exemple, le pourraient également<sup>48</sup>. En troisième lieu, on peut s'interroger sur la catégorie des « données biométriques » : la qualification est ici rien moins qu'évidente. Pour qu'elles puissent relever de la définition de l'article 9, il faudrait que ces données résultent non seulement « d'un traitement technique spécifique » et soient « relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique<sup>49</sup> », mais également qu'elles soient traitées « aux fins d'identifier une personne physique de manière unique ». Or, les traitements des émotions ou état d'esprit dont il a été question dans les cas d'usage évoqués ne poursuivent pas en général cette finalité, mais celle de

---

<sup>45</sup> Sur les difficultés en général, v. C. Koumpli, *Les données personnelles sensibles. Contribution à l'évolution du droit fondamental à la protection des données personnelles. Étude comparée : Union européenne, Allemagne, France, Grèce, Royaume-Uni*, préf. O. Pfersmann, avant-propos J. Rochfeld, Pedone, à paraître en 2022 ; P. Quinn et G. Malgieri, « The Difficulty of Defining Sensitive Data. The Concept of Sensitive Data in the EU Data Protection Framework », 2021, *German Law Journal*, à venir, version préparatoire disponible à l'adresse suivante : [<https://papers.ssrn.com/abstract=3713134>], dernière consultation le 15 déc. 2020.

<sup>46</sup> Article 9, 2, a, RGPD. Sur la faiblesse voire l'inefficacité de ce régime, v. particulièrement C. Koumpli, *op. cit.*

<sup>47</sup> Dans ce sens, M. Ienca et R. Andorno, art. préc., p. 8.

<sup>48</sup> M. Ienca et R. Andorno, art. préc., qui ajoutent qu'un état d'esprit ou des émotions pourraient parfois permettre d'inférer une orientation politique, voire des convictions religieuses.

<sup>49</sup> RGPD, art. 4, 14. *Adde* RGPD, cons. 35.

profilage, d'évaluation ou de catégorisation des individus ; en conséquence, elles n'intégreraient pas la catégorie et le régime de l'article 9. En définitive, les données d'émotions ou d'état d'esprit qui ne seraient pas « pathologiques », ne relèveraient pas de la vie sexuelle ou qui ne se trouveraient pas traitées à dessein d'identification resteraient en dehors de ce champ<sup>50</sup>. Comme telles, elles ne justifieraient pas d'une protection particulière, c'est-à-dire d'une interdiction de principe de leurs traitements (interdiction dont on rappelle qu'elle est faible et peut être facilement contournée). Il y aurait là ce que l'on peut identifier, avec d'autres à nouveau, comme un second « gap » de protection<sup>51</sup>.

*Des données sensibles aux « traitements sensibles ».* – Pour y réagir, on pourrait alors, dans une première direction, avancer une demande d'extension de la liste de l'article 9 à ces données bien particulières<sup>52</sup>. Ce n'est pourtant pas l'orientation qui semble la plus judicieuse, tant la catégorie s'attire un régime de protection faible, notamment du fait de cette exception rappelée du consentement, largement utilisée. C'est pourquoi une autre direction pourrait être prônée.

En effet, face à de tels défis, on peut se demander s'il ne faudrait pas basculer vers une appréhension de ce type de situation en termes de « traitements sensibles » plutôt que de données « sensibles » : le glissement permettrait de saisir des traitements particulièrement intrusifs, présentant des dangers particuliers, quand bien même ils ne porteraient pas sur des données dites « sensibles<sup>53</sup> ». Or, le RGPD les aborde déjà par endroits. En premier lieu, en effet, les exigences fondamentales du droit des données à caractère personnel, notamment celles synthétisées dans l'article 5, 1 du RGPD – selon lequel, entre autres, les données « doivent être : a) traitées de manière licite, loyale et transparente au regard de la personne concernée (licéité, loyauté, transparence) ; b) collectées pour des finalités déterminées, explicites et légitimes », tout en étant « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées » – peuvent être interprétés en ce sens : comme imposant une limitation des traitements en cas de risques disproportionnés – contrôle de proportionnalité – mais également une minimisation des données traitées ainsi qu'une

---

<sup>50</sup> Dans le même sens, C. Muller (dir.), *Impact de l'intelligence artificielle sur les droits de l'homme, la démocratie et l'État de droit*, rapport du Comité *ad hoc* sur l'IA (CAHAI), Conseil de l'Europe, 24 juin 2020, CAHAI(2020) 06 fin, n° 23.

<sup>51</sup> S. Rainey *et al.*, « Is the European Data Protection Regulation Sufficient to Deal with Emerging Data Concerns Relating to Neurotechnology? », *Journal of Law and the Biosciences*, 2020, n° 28, 11.

<sup>52</sup> En ce sens, M. Ienca et G. Malgieri, art. préc., p. 8.

<sup>53</sup> V. également en ce sens, M. Ienca et G. Malgieri, art. préc., p. 1 ; C. Koumpli, *op. cit.* ; CNIL, *Cahiers IP. Innovation et prospective. Vie privée à l'horizon 2020*, n° 1, p. 34 s., not. Y. Pouillet, évoquant un « quadrillage de données triviales » désormais.

véritable légitimité dans les finalités du traitement. À ces égards, on pourrait donc estimer que des traitements trop invasifs ne seraient pas proportionnés, minimiseraient insuffisamment les données, ou ne relèveraient pas de finalités légitimes.

En deuxième lieu, en son article 35, le règlement européen préconise une analyse d'impact (une évaluation précise des risques ainsi que l'exposition de mesures adaptées), et s'il le faut, un renoncement aux traitements<sup>54</sup> en matière de traitements particulièrement dangereux définis comme ceux présentant un « risque élevé pour les droits et libertés des personnes physiques ». Selon le texte, ceux-ci embrassent plus précisément tout « 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, [qui] est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques » – définition générale – et, « en particulier », ceux consistant en une « évaluation systématique et approfondie d'aspects personnels concernant des personnes physiques [...] 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<sup>55</sup> ». Par ailleurs, le G29 avait posé 9 critères pour apprécier ce risque (deux devant être réunis pour imposer l'analyse d'impact), parmi lesquels on trouvait le fait de procéder à une « évaluation ou notation » à partir « de profilage et de prédiction », notamment sur le fondement de données de « comportement » de la personne, ou encore tout traitement portant sur des données non pas sensibles *stricto sensu* mais « sensibles » dans le sens courant. Parmi ces dernières, il classait « certaines catégories de données [pouvant] être considérées comme augmentant le risque possible pour les

---

<sup>54</sup> Au préalable, en cas de risques élevés, une consultation de l'autorité de contrôle s'impose, v. cons. 84 : « Afin de mieux garantir le respect du présent règlement lorsque les opérations de traitement sont susceptibles d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement devrait assumer la responsabilité d'effectuer une analyse d'impact relative à la protection des données pour évaluer, en particulier, l'origine, la nature, la particularité et la gravité de ce risque. Il convient de tenir compte du résultat de cette analyse pour déterminer les mesures appropriées à prendre afin de démontrer que le traitement des données à caractère personnel respecte le présent règlement. Lorsqu'il ressort de l'analyse d'impact relative à la protection des données que les opérations de traitement des données comportent un risque élevé que le responsable du traitement ne peut atténuer en prenant des mesures appropriées compte tenu des techniques disponibles et des coûts liés à leur mise en œuvre, il convient que l'autorité de contrôle soit consultée avant que le traitement n'ait lieu. »

<sup>55</sup> Art. 35, 3, a. On mentionne également les situations visées par l'article 35, 3, b et c dans le cas où ces traitements auraient lieu « à grande échelle » et que l'on admettrait que l'on soit face à des « catégories particulières de données visées à l'article 9, paragraphe 1 » (b) ou pour une « surveillance systématique à grande échelle d'une zone accessible au public » (c).

droits et libertés des personnes<sup>56</sup> » et proposait de réserver un sort particulier aux « utilisation[s] innovante[s] ou application[s] de nouvelles solutions technologiques », qui peuvent « impliquer de nouvelles formes de collecte et d'utilisation des données, présentant potentiellement un risque élevé pour les droits et libertés des personnes<sup>57</sup> ». Dans cette lignée et appliquées aux données mentales (au sens large), d'aucuns plaident alors actuellement pour imposer une évaluation et une documentation spéciales : il y aurait lieu d'établir pour elles une « analyse d'impact spéciale<sup>58</sup> ».

En troisième lieu et enfin, les concepteurs de ce type de traitement pourraient être soumis à une obligation rehaussée de « *privacy by design* », c'est-à-dire à une obligation renforcée de minimiser au maximum, dès la conception des techniques de traitement, les impacts du traitement sur les personnes concernées.

*Consécration et protection du droit à l'intégrité psychique.* – Mais, on le comprend, il est nécessaire, pour aller dans ces directions et renforcer les exigences, d'identifier un « risque élevé » pour des « droits et libertés de personnes physiques ». Or, quels sont ces droits et libertés qui seraient ici en cause et pourraient être brandis aux fondements de ces conciliations ? C'est alors que l'on croise les réflexions très actuelles sur la nécessité de faire émerger et de consacrer de nouveaux droits fondamentaux<sup>59</sup>, qu'on les évoque comme ceux « au respect de l'intégrité psychique<sup>60</sup> », « à une liberté cognitive », « à une vie privée mentale », « à une intégrité mentale » ou encore à une « continuité psychologique<sup>61</sup> », voire « à une autodétermination mentale ». En Europe, ils pourraient se fonder sur l'article 3 de la Charte des droits fondamentaux selon lequel « chacun a droit à son intégrité physique et mentale ». Ils embrasseraient un droit à une protection de toute personne contre l'utilisation par d'autres, de façon coercitive ou non consciente, de neurotechnologies ou reconnaissance des émotions<sup>62</sup>. Dans ces conditions, un traitement serait dit sensible s'il est susceptible d'atteindre l'un de ces droits<sup>63</sup>.

---

<sup>56</sup> G29, Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est « susceptible d'engendrer un risque élevé » aux fins du règlement (UE) 2016/679, révisées le 4 oct. 2017, WP 248 rév. 01, p. 11, critères 1 et 4.

<sup>57</sup> G29, préc., p. 12, critère 8.

<sup>58</sup> M. Ienca et G. Malgieri, art. préc., p. 14 s., en anglais « *Mental Data Protection Impact Assessment* » (MDPIA).

<sup>59</sup> M. Ienca et R. Andorno, art. préc.

<sup>60</sup> J.-C. Bublitz, « My Mine is Mine! Cognitive Liberty as Legal Concept », in E. Hildt et A. Francke (eds.), *Cognitive Enhancement*, Springer Spring 2013, Chapter 19, p. 233-264. Pour plus de détails sur chacun des droits, v. M. Ienca et R. Andorno, art. préc., p. 12 s.

<sup>61</sup> M. Ienca et R. Andorno, art. préc., p. 9 ; C. Muller (dir.), Rapport préc., n° 77, p. 21.

<sup>62</sup> C. Muller (dir.), Rapport préc., n° 77, p. 21.

<sup>63</sup> Comp. Conseil de l'Europe, *Vers une régulation des systèmes d'IA. Perspectives internationales sur l'élaboration d'un cadre juridique fondé sur les normes du Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie*



Cette consécration permettrait aussi de réagir et de poser des limites propres non plus seulement aux traitements des émotions mais également à leurs manipulations.

B. - *DES LIMITES AUX MANIPULATIONS*  
*DES DONNEES EMOTIONNELLES : VERS LA CONSECRATION*  
*D'UNE AUTODETERMINATION COGNITIVE ?*

*L'obligation de loyauté et l'interdiction des « dark patterns ».* – Ces traitements des émotions seraient-ils toutefois admis, c'est alors les manipulations de ces dernières qu'il faudrait encadrer. On peut ici raisonner sur l'exemple d'une interface de montre connectée qui serait utilisée pour exploiter les émotions que l'objet permettrait de capter, en orientant les comportements de son utilisateur (pour lui faire prendre, par exemple, à des fins commerciales, un chemin plutôt qu'un autre lors d'une course à pied, pour le mettre en relation avec telle ou telle information, etc.). Or, plusieurs voix s'élèvent désormais pour interdire ce type « d'orientations », au moins quand celles-ci se développent de façon déloyale et en ayant recours à des *dark patterns*. Certes, on pourrait ici considérer que les limitations propres au droit des données à caractère personnel et au RGPD sont suffisantes : les principes évoqués de transparence, de garantie d'un consentement éclairé et de mise en œuvre du *privacy by design* ne pourraient pas être considérés comme respectés dans ces situations et ces types de traitements déloyaux des émotions ne devraient pas s'effectuer<sup>64</sup>. Mais il paraît plus sage et pertinent d'ajouter une sanction des manipulations cognitives pour elles-mêmes. Celles-ci pourraient alors se fonder, en premier lieu, sur les mécanismes de lutte contre les pratiques commerciales déloyales<sup>65</sup> : ils trouveraient à s'appliquer aux techniques dites de « *bad nudge* », incitant l'utilisateur à prendre une décision allant à l'encontre de ses intérêts, ou de « *bad sludge* » qui visent à l'empêcher d'agir selon son intérêt en créant une friction, une difficulté artificielle,

---

*et de l'État de droit*, DGI 2020(16), p. 42, plaidant pour la reconnaissance de cinq nouveaux principes « éthiques génériques » en matière d'utilisation d'intelligence artificielle, principes issus d'une comparaison des lignes directrices diverses qui se mettent en place partout dans le monde, soit les principes de transparence, de justice, de « non-malfaisance », de responsabilité et de respect de la vie privée.

<sup>64</sup> CNIL, *La forme des choix*, Cahier préc., p. 39 s.

<sup>65</sup> En ce sens, G. Donadieu, *Les pratiques commerciales cognitives*, *op. cit.*, not. n° 217 pour une définition de ces pratiques commerciales cognitives déloyales. Egalement, S. Bernheim-Desvaux, « Protéger le neuro-consommateur ? », CCC 2021. Repère 11. *Adde* Commission européenne, *Orientations concernant la mise en œuvre/application de la directive 2005/29/CE relatives aux pratiques commerciales déloyales*, 25 mai 2016, SWD (2016) 163 final, pt 3.3, p. 65.

laquelle entrave sa liberté de choix<sup>66</sup>. La loi californienne adoptée le 15 mars 2021 et portant réforme du *Consumer Privacy Act Regulations* montre le chemin, qui a renforcé les exigences relatives aux données à caractère personnel pour prohiber le recours à des procédés visant à compromettre ou entraver le choix d'un consommateur de se retirer des systèmes exploitant ses données<sup>67</sup>. Au niveau fédéral, on compte également une proposition de loi « S. 1084 », dite « *Deceptive Experiences to Online Users Reduction Act* », déposée le 9 avril 2019 devant le Sénat, qui entend « interdire de concevoir, modifier ou manipuler l'interface utilisateur dans le but ou avec l'effet substantiel d'obscurcir, de subvertir ou d'altérer l'autonomie, la prise de décision ou le choix de l'utilisateur pour obtenir son consentement ou ses données, ou encore de créer un usage compulsif, notamment dans la “consommation” de contenus et à l'égard de mineurs de moins de 13 ans<sup>68</sup> ». La France n'est pas en reste puisqu'une proposition de loi visant à garantir le libre choix du consommateur dans le cyberspace<sup>69</sup> proposait de modifier le Code de la consommation pour imposer, en un nouvel article L. 111-7-3, que « [l]es opérateurs de plateforme en ligne dont l'activité dépasse un seuil de nombre de connexions défini par décret s'abstiennent de concevoir, de modifier ou de manipuler une interface utilisateur ayant pour objet ou effet de subvertir ou d'altérer l'autonomie du consommateur dans sa prise de décision ou d'obtenir son consentement ». Mais cette proposition, qui aurait permis d'imposer la loyauté des interfaces utilisateurs et de saisir les nouvelles interfaces humains-machines (comme les agents conversationnels), n'a pas été adoptée. Il faudra donc attendre les avancées européennes de l'interdiction d'interfaces truquées visant à fausser ou entraver la liberté de choix des utilisateurs de services numériques, débattue dans le cadre des réformes dites *Digital Services Act* et *Digital Markets Act*<sup>70</sup>. Cette interdiction paraît d'autant plus essentielle que se développent des « *augmented dark patterns* » permettant, par un traitement algorithmique, d'agir dynamiquement sur les *stimuli* des utilisateurs pour exploiter leur vulnérabilité<sup>71</sup>.

---

<sup>66</sup> Sur cette distinction, v. not. N. de Marcellins-Warin, F. Marty, E. Thelisson et T. Warin, « Intelligence artificielle et manipulations des comportements de marché : l'évaluation *ex ante* dans l'arsenal du législateur », *RIDE* 2020/2, t. XXXIV, p. 203, spéc. p. 215 s.

<sup>67</sup> Chapitre 20, § 999.315, h : « *A business shall not use a method that is designed with the purpose or has the substantial effect of subverting or impairing a consumer's choice to opt-out.* »

<sup>68</sup> Bill S. 1084, 9 avr. 2019, disponible à l'adresse suivante : [<https://www.govinfo.gov/app/details/BILLS-116s1084is>]. On relèvera également l'objectif général poursuivi par ce texte : « *prohibit the usage of exploitation and deceptive practices by large online operators ant to promote consumer welfare in the use of behaviorial research by such providers* ».

<sup>69</sup> Proposition de loi n° 49 (2019-2020) visant à garantir le libre choix du consommateur dans le cyberspace et l'amendement n° COM-15, 3 févr. 2020.

<sup>70</sup> V. M. Le Roy, *La loyauté des plateformes à l'égard des consommateurs*, Université Paris Saclay, 2021, n° 354.

<sup>71</sup> Sur ce point, v. N. de Marcellins-Warin, F. Marty, E. Thelisson et T. Warin, art. préc., p. 220.

*L'obligation de loyauté et la transparence sur « l'interlocuteur machine ».* – En second lieu, il serait nécessaire de consacrer diverses obligations de transparence, à la charge des concepteurs et utilisateurs d'algorithmes voire de systèmes d'« intelligence artificielle » interagissant avec des personnes physiques : dès lors que les individus seraient susceptibles d'être manipulés par jeux sur leurs émotions en pensant interagir avec un humain et non un avec un agent virtuel, il y aurait lieu de les détromper et de contourner le risque de création d'émotions inappropriées. On plaide ainsi pour que les personnes puissent, dans tous les cas, « connaître l'origine et la nature des messages qu'elles reçoivent<sup>72</sup> » et la qualité de leur interlocuteur (humain ou machine). Tel est le sens de l'article 52 de la proposition de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle, en date du 21 avril 2021<sup>73</sup>, qui, dans le prolongement des préconisations de lignes directrices en la matière<sup>74</sup>, enjoignent les fournisseurs de systèmes d'IA destinés à interagir avec des personnes physiques de veiller à ce que ceux-ci soient conçus et développés de manière à ce que les personnes physiques soient informées qu'elles interagissent avec de tels systèmes. On relèvera que la Californie a déjà consacré une disposition visant à prévenir ce type de tromperie à l'égard des utilisateurs de logiciels robots (*bots*), ce par l'adoption du *Boltering Online Transparency Act of California Senate Bill 1001* (dit « Bot Act »)<sup>75</sup> : depuis le 1<sup>er</sup> juillet 2019, il « est illégal

---

<sup>72</sup> N. Nevejeans, « Les problématiques juridiques et éthiques posées par les robots en santé mentale », in S. Tisseron et al. (dir.), *Robots, de nouveaux partenaires de soins psychiques*, Érès, 2018, p. 43.

<sup>73</sup> Proposition de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union européenne, COM(2021), 206 final.

<sup>74</sup> Celles du groupe d'experts de haut niveau nommé par la Commission européenne, v. *Lignes directrices en matière d'éthique pour une IA digne de confiance*, 2019, pt 78, qui embrassaient les préconisations suivantes, au nom d'un principe de transparence : « Les systèmes d'IA ne devraient pas se présenter comme des êtres humains auprès des utilisateurs ; lorsqu'ils interagissent avec un système d'IA, les êtres humains ont le droit d'en être informés. Cet aspect implique que les systèmes d'IA doivent être identifiables en tant que tels. Qui plus est, la possibilité de s'opposer à cette interaction humaine devrait être proposée le cas échéant, afin de garantir le respect des droits fondamentaux. Outre cet aspect, il convient de communiquer aux professionnels de l'IA ou aux utilisateurs finaux des informations appropriées sur les capacités et les limites des systèmes d'IA, selon des modalités adaptées au contexte d'utilisation concerné. Ces informations pourraient comprendre le degré de précision des systèmes d'IA, ainsi que ses limites. »

<sup>75</sup> Business and Professions Code, Division 7, Part 3, Chapter 6, Section 17940 [[https://leginfo.ca.gov/faces/billCompareClient.xhtml?bill\\_id=201720180SB1001&showamends=false](https://leginfo.ca.gov/faces/billCompareClient.xhtml?bill_id=201720180SB1001&showamends=false)].

pour toute personne d'utiliser un bot pour communiquer ou interagir avec une autre personne en Californie en ligne, avec l'intention de tromper l'autre personne sur son identité artificielle dans le but de tromper sciemment la personne sur le contenu de la communication afin d'inciter à l'achat ou à la vente de biens ou de services dans une transaction commerciale ou d'influencer un vote dans une élection ».

*Au-delà de la loyauté. L'interdiction per se de la manipulation des émotions et la reconnaissance de l'autonomie cognitive.* – Par ailleurs, au-delà de la prohibition des manipulations déloyales des émotions et des exigences de loyauté et de transparence, on pourrait aller beaucoup plus franchement vers l'interdiction de toute manipulation des émotions. Il paraît en effet nécessaire d'appréhender spécifiquement les risques que ce type de pratiques comportent, ce en raison de leurs conséquences potentielles hautement préjudiciables sur le plan tant individuel que collectif. La question d'aller dans cette direction demeure d'ailleurs ouverte en Europe, si l'on en juge aux négociations relatives à l'adoption d'un cadre réglementaire propre aux systèmes d'intelligence artificielle<sup>76</sup>. En réalité, deux directions paraissent discutées, dont seule la seconde nous semble acceptable.

Dans une première direction, on ne parle en effet que d'interdire le recours aux systèmes de reconnaissance émotionnelle en raison de l'insuffisante preuve, à ce jour, de leur fiabilité. D'aucuns s'interrogent là sur le déploiement de technologies fondées sur des méthodologies non éprouvées scientifiquement et appellent à déterminer le taux de fiabilité et de précision minimale des algorithmes qui pourraient être utilisés, notamment au regard du risque de discrimination pouvant en résulter<sup>77</sup>. Cette fiabilité serait-elle attestée que ces utilisations pourraient se déployer.

Cette solution nous semble insuffisante et c'est donc pour une seconde direction que l'on propose d'opter, consistant à consacrer une interdiction de principe de ces systèmes de reconnaissance et de manipulation émotionnelle. À nouveau, en conséquence, se pose la question du fondement de cette interdiction. À cet égard, dans le prolongement de la consécration d'une autodétermination informationnelle reconnu au bénéfice de

---

<sup>76</sup> Proposition de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle, préc.

<sup>77</sup> Défenseur des droits, *Technologies biométriques : l'impératif respect des droits fondamentaux*, 19 juill. 2021, p. 18, rappelant que l'article L. 122-1-8 du Code du travail précise que les « méthodes et techniques d'aide au recrutement ou d'évaluation des candidats à un emploi doivent être pertinentes au regard de la finalité poursuivie ». – Adde, K. Crawford, art. préc. : « Il est temps de mettre en place une protection législative contre les utilisations non prouvées de ces outils dans tous les domaines – éducation, soins de santé, emploi et justice pénale. Ces protections permettront de recentrer la science rigoureuse et de rejeter le mythe selon lequel les états internes ne sont qu'un autre ensemble de données que l'on pourrait effacer de notre visage. »

toute personne concernée par un traitement de ses données à caractère personnel, ainsi que du ou des droits évoqués précédemment (au respect de l'intégrité psychique ou autre appellation), il y aurait lieu à notre sens de consacrer celui à une autodétermination cognitive, notamment face aux risques de manipulation que ces systèmes comportent. L'idée fait d'ailleurs son chemin, qui apparaît par exemple dans les lignes directrices du groupe d'experts de haut niveau auprès de la Commission européenne : celles-ci rappellent que « les êtres humains qui interagissent avec des systèmes d'IA doivent être en mesure de conserver leur *autodétermination totale et effective* et prendre part au processus démocratique. En l'absence de justifications, les systèmes d'IA ne devraient pas subordonner, contraindre, tromper, manipuler, conditionner ni régenter les êtres humains<sup>78</sup> ». De son côté, le Conseil de l'Europe n'a pas manqué de souligner l'impact des systèmes de reconnaissance émotionnelle sur l'intégrité psychologique des individus<sup>79</sup>. En outre, dans sa Déclaration sur les capacités de manipulation des processus algorithmiques, il a relevé que « les outils d'apprentissage automatique (*machine learning*) actuels sont de plus en plus capables non seulement de prédire les choix, mais aussi d'influencer les émotions et les pensées et de modifier le déroulement d'une action, parfois de façon subliminale [...] » et que, en raison des « niveaux, très subtils, subconscients et personnalisés de la persuasion algorithmique », ils peuvent avoir des effets significatifs sur « *l'autonomie cognitive* des citoyens et leur droit à se forger une opinion et à prendre des décisions indépendantes<sup>80</sup> ». Enfin, le Contrôleur européen à la protection des données de l'Union européenne et le Comité européen de la protection des données soutiennent une interdiction générale des méthodes d'évaluation des émotions, sous réserve de certains usages spécifiques (à finalités de recherche ou de protection de la santé qu'il conviendrait néanmoins de soumettre à un strict encadrement)<sup>81</sup>.

---

<sup>78</sup> *Lignes directrices en matière d'éthique pour une IA digne de confiance*, préc., p. 15. Adde, Y. Meneceur, *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne*, Bruylant, 2020, p. 211 s.

<sup>79</sup> Conseil de l'Europe, *Vers une régulation des systèmes d'IA*, préc., p. 26, pt 20.

<sup>80</sup> Conseil de l'Europe, *Déclaration du Comité des ministres sur les capacités de manipulation des processus algorithmiques*, févr. 2019. Il ajoutait que « [c]es effets ne peuvent pas être sous-estimés même s'ils ne sont pas suffisamment étudiés. Non seulement ils peuvent nuire à l'exercice et à la jouissance des droits de l'homme individuels, mais ils peuvent également aboutir à l'érosion du fondement même du Conseil de l'Europe. Ses piliers centraux que sont les droits de l'homme, la démocratie et l'État de droit reposent sur la croyance fondamentale en l'égalité et la dignité de tous les êtres humains en tant qu'agents moraux indépendants ».

<sup>81</sup> CEPD-EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 18 juin 2021, pt 35.

Pour l'heure toutefois, la Commission européenne ne paraît pas concrétiser cette direction. Sa proposition de règlement IA se limite à poser, en son article 5 § 1, une interdiction des utilisations qui interviennent «à des fins d'influencer de manière subliminale le comportement d'une personne en vue de lui causer ou de causer à un tiers un dommage ou les systèmes qui exploitent la vulnérabilité d'un groupe de personnes en vue de fausser le comportement de l'une de ces personnes et de causer un dommage ». Plus précisément encore, s'agissant des systèmes de reconnaissance des émotions ou de catégorisation biométrique, l'article 52.2 se contente de prévoir que leurs utilisateurs seront tenus d'une obligation de transparence les enjoignant d'informer du fonctionnement du système les personnes qui y seront exposées... Elle s'en tient donc à la première direction et aux seules manipulations déloyales.

\*\*

\*

Tout reste ainsi à arbitrer dans cette matière complexe et émergente de la connaissance et de la manipulation des émotions, en passe de prendre une ampleur nouvelle et potentiellement panoptique au regard des technologies à disposition et des valorisations mirifiques attendues. Au même titre que la consécration du droit au respect de la vie privée en son temps, il semble que soit venu celui de l'identification de droits fondamentaux propres à répondre à ces défis : celui ou ceux au respect de l'intégrité psychique de chacun et à une autodétermination cognitive. Un marché juteux de monétisation des émotions se déploie et il revient aux juristes de poser les limites propres à contenir ces appétits... Le temps est compté et ils doivent s'y atteler.

*Janvier 2022*