1. *Le symbiote et le mécanisme de délégation*

Dans son livre *Uomo e tecnologia*, Giuseppe Longo écrit : « Le caractère essentiel du rapport entre la technologie et l'homme est révélé par la rétroaction que les innovations techniques exercent sur les êtres humains et la société »[5]. Selon Longo, au lieu de subdiviser les évolutions en « biologique » et « technologique », il serait préférable de parler d'une seule évolution « biotechnologique », au centre de laquelle se trouve le « symbiote » (ou *simbionte*)[6], c'est-à-dire un hybride biotechnologique, né de la composition « homme-plus-environnement »[7]. Longo fait une comparaison entre le rythme rapide de l'évolution technologique et la lenteur de l'évolution humaine. Sur cet écart, il s'exprime ainsi : « La vitesse toujours plus élevée de l'innovation technologique accentue le déséquilibre et pousse à *déléguer* aux machines un nombre croissant d'actions, de fonctions et même de décisions. L'activité cognitive du symbiote est profondément affectée par cette délégation »[8]. Tout cela a, selon Longo, de graves répercussions sur le corps physique de l'individu : « Le corps humain a toujours été amplifié par des instruments et des appareils qui ont étendu et multiplié ses possibilités d'interaction avec le monde, dans un sens aussi bien cognitif qu'opérationnel »[9]. En partant de cette supposition, il serait arbitraire de dire que le corps est enfermé dans ses limites « topologiques », représentées par la peau. L'invention d'outils permet non seulement de suppléer à une capacité compromise ou perdue, mais aussi de donner lieu à une véritable hybridation : « Comme l'homme fait la technologie, la technologie fait l'homme »[10].

L'utilisation des outils implique donc une série de mécanismes de *feedback* qui entraînent des modifications chez la personne même de l'utilisateur. Si cela se produit dans la sphère matérielle, avec des outils techniques qui nous permettent d'être plus précis et d'agir plus efficacement sur notre environnement, cela se vérifie également dans la sphère cognitive, où nous sommes constamment orientés vers certains choix plutôt que d'autres par ces mêmes outils technologiques auxquels nous déléguons nos actions. Toutefois, les concepts exprimés dans la pensée de Longo sont loin d'être exempts de problématiques. En effet,

5 G. LONGO, *Uomo e tecnologia : una simbiosi problematica*, Edizioni Università di Trieste, Trieste 2006, p. 5 (traduction de l'auteur).

6 LONGO (n. 5), p. 48. Sur le symbiote technologique de LONGO, voir aussi : SINI (n. 1), p. 21 ss.

7 LONGO (n. 5), p. 9. L'hybride technologique, ou *homo technologicus*, serait un hybride de biologie et de technologie en constante évolution. L'*homo sapiens* ayant toujours été contaminé par la technologie, il a toujours été un *homo technologicus*.

8 *Idem*, p. 6 (traduction de l'auteur).

9 SINI (n. 1), p. 24 (traduction de l'auteur) ; LONGO (n. 1), p. 58 (traduction de l'auteur).

10 SINI (n. 1), p. 24.

les notions de « nature » et de « naturel » sont contradictoires, à moins qu'elles ne soient elles-mêmes produites par la technologie[11].

Le progrès technologique est irréversible et s'accompagne de nombreux avantages, mais aussi d'inconvénients. Avec l'avènement de l'alphabet écrit, n'avons-nous pas perdu l'usage assidu de la mémoire[12] ? Également la vidéo-écriture, dit Longo, a modifié de manière irréversible le style de l'écriture et a considérablement affaibli notre capacité à tracer les mots à la main[13]. Tant Giuseppe Longo que Carlo Sini estiment que les transformations technologiques entraînent la redéfinition de concepts culturels traditionnels tels que la liberté, la démocratie, l'intelligence, la réalité, l'histoire, le temps et la mémoire[14], ainsi que la liberté de conscience, le caractère raisonnable, la volonté et, par conséquent, la responsabilité[15].

2. *Le paradigme robot-maître et la relation entre le pouvoir, l'intelligence et le contrôle*

Le terme *Robòt* dérive du tchèque *Robota* (travail pénible), mais présente également un lien avec les termes russes *Rabotat* (travailler) et *Rabot* (travail)[16]. La racine commune est « rob », qui en slave signifie « esclave ».

Selon Peter Asaro, la relation homme-machine peut être comparée à la relation maître-esclave, qui, à son avis, est pleine de contradictions intrinsèques[17]. Asaro cite l'auteur Norbert Wiener pour décrire le paradoxe de désirer et en même temps craindre l'intelligence et sa domination. Un système est fiable lorsqu'il remplit sa fonction. Toutefois, plus il est intelligent, plus il est susceptible de produire des conséquences inattendues et indésirables. Wiener écrit :

11 *Idem*, p. 23.

12 *Idem*, p. 35.

13 LONGO (n. 5), p. 26.

14 *Ibid.*

15 Conférence International Society of Public Law (ICON-S), *Le nuove tecnologie e il futuro del diritto pubblico*, Florence, 22-23 novembre 2019, accessible au lien suivant : www.icons-italia.it/firenze-2019/ (consulté le 22.08.2022).

16 Le terme *robòt* dérive de la dramaturgie de science-fiction, notamment d'un texte créé en 1920 par Kalel Čapek, intitulé R.U.R. (*Rossum's Universal Robots*) : K. CAPEK, *R.U.R. Rossum's Universal Robots*, Francesco Bevivino Éditeur, Milan 2015.

17 P. ASARO, « Roberto Cordeschi on Cybernetics and Autonomous Weapons », *Paradigmi*, 03/2015, p. 83-107. Du même auteur et sur le même sujet, voir P. ASARO, « Robots and Responsibility from a Legal Perspective », *Proceedings of the IEEE*, 2007, p. 20-24 ; P. ASARO, *The Liability Problem for Autonomous Artificial Agents*, AAAI Spring Symposia, 2016, p. 190-194 ; P. ASARO, « Algorithms of Violence : Critical Social Perspectives on Autonomous Weapons », *Social Research : An International Quarterly*, vol. 86, n° 2, 2019, p. 537- 555.

« Nous voulons qu'un esclave soit intelligent, capable de nous aider à accomplir nos tâches. Cependant, nous voulons aussi qu'il soit soumis. La soumission totale et l'intelligence totale ne vont pas de pair. Combien de fois, dans l'Antiquité, le philosophe grec intelligent, serviteur d'un esclavagiste romain moins intelligent, a-t-il dû déterminer les actions de son maître plutôt que d'obéir à ses souhaits ! De même, à mesure que les machines deviennent de plus en plus efficaces et fonctionnent à un niveau psychologique plus élevé, la catastrophe [...] de la domination des machines devient de plus en plus proche »[18]. Cependant, contrairement à un esclave en chair et en os dont l'esprit est ouvert à une conscience générique, l'IA n'a aucune intelligence sociale et ne se concentre que sur des tâches spécifiques. Une autre comparaison proposée par Asaro est celle d'un apprenti magicien qui jette un sort pour sécher un sol mouillé, mais qui se trompe et risque de se noyer. « L'apprenti », écrit Asaro, « est capable de déchaîner des forces surnaturelles avec un sortilège, mais il est incapable de les maîtriser quand elles sortent de leur trajectoire »[19].

De ces exemples, nous pouvons déduire la facilité avec laquelle une machine peut causer des dommages irréparables par son comportement imprévisible. Alors, quel est le but d'une machine ? En effet, nous ne pouvons pas dire que le but de la machine est, par exemple, de tuer des ennemis ou d'abattre des avions, car la machine ne sait pas ce qu'est un « ennemi »[20].

Dans le cas où le niveau d'automatisation est tel qu'il n'y a pas d'humains présents pour imposer à la machine certaines exigences, la seule personne humaine qui se trouve en communication avec elle, bien qu'absente au moment où elle agit, est le constructeur/programmeur. Les systèmes d'intelligence artificielle, puisque leur libre arbitre « moral » ne peut être prouvé, peuvent être considérés comme nuisibles, mais pas coupables. Cela donne lieu à un conflit entre le concept de personnalité et la responsabilité pénale.

B. Études sur la morale d'une intelligence artificielle

Atteindre l'IA, c'est comme un « voyage sur la lune »[21]. Les chercheurs gravissent des montagnes toujours plus hautes, convaincus qu'ils atteindront un jour cette destination, mais le résultat est autre : ce n'est pas en marchant que nous atteindrons la lune ; un changement de méthode s'impose.

18 N. WIENER, « Some Moral and Technical Consequences of Automation », *Science*, vol. 131, 6 mai 1960, p. 1355-1358 (traduction de l'auteur) ; ASARO (n. 17), *Roberto Cordeschi*, p. 93.

19 ASARO (n. 17), *Roberto Cordeschi*, p. 92 ss (traduction de l'auteur).

20 ASARO (n. 17), *Algorithms of Violence*, p. 542 ss.

21 S. HÉNIN, *AI : Intelligenza Artificiale tra incubo e sogno*, Hoepli, Milan 2019, p. 214 ss.

La lune de la métaphore est ce que nous appelons « intelligence artificielle générale » (*artificial general intelligence*, AGI), c'est-à-dire un système capable de faire face à toute tâche cognitive humaine et de s'adapter au contexte dans lequel il est placé sans avoir besoin d'être préprogrammé de manière spécifique[22].

À cette fin, l'IA utilise des systèmes d'apprentissage automatique et apprend de sa propre expérience. Dans un sens large, il semble que l'objectif de l'IA soit de reproduire des processus cognitifs similaires à ceux des humains, notamment par l'apprentissage empirique. Certains algorithmes sont même capables d'exécuter de meilleures performances que les êtres humains pour certaines tâches. En effet, on craint que 60 % des emplois intellectuels et plus de 50 % des emplois manuels ne soient remplacés par les nouvelles technologies[23].

Selon la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires, adoptée en 2018 par la Commission européenne, l'intelligence artificielle pourrait être définie comme un « [e]nsemble de sciences, théories et techniques dont le but est de reproduire par une machine des capacités cognitives d'un être humain »[24]. Cependant, l'intelligence artificielle, entendue comme science, théorie ou technique, ne peut être séparée de son « système », c'est-à-dire de sa structure physique[25]. Toujours dans cette même logique, la Commission européenne a qualifié de « système d'intelligence artificielle » tous les « systèmes qui font preuve d'un comportement intelligent en analysant leur environnement et en prenant des mesures – avec un certain degré d'autonomie – pour atteindre des objectifs spécifiques »[26].

Kaplan, un des pionniers de l'intelligence artificielle, a précisé ces systèmes et les environnements dans lesquels ils évoluent. La subdivision de l'intelligence artificielle en différents secteurs est importante, car elle permet de regrouper les domaines d'application de l'IA au sein de quatre macro-domaines de recherche et développement que sont : la robotique, la vision artificielle, la

22 *Ibid.*

23 M. RASETTI, *Il dato è tratto,* Asimmetrie, vol. 14, n° 27, 10/2019, p. 4 ss.

24 COMMISSION EUROPÉENNE, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, Annexe III, Glossaire, Strasbourg, 3-4 décembre 2018.

25 Selon la théorie de l'esprit incarné, la pensée se développerait sur la base des expériences corporelles vécues par l'agent. Ce concept d'origine kantienne est l'objet d'études de philosophes et de cognitivistes depuis la fin du XX^e siècle. Voir notamment F. VARELA/E. THOMPSON/E. ROSCH, *The embodied mind*, MIT Press, Cambridge 1991.

26 COMMISSION EUROPÉENNE, *Communication de la Commission au Parlement Européen, au Conseil Européen, au Conseil, au Comité Économique et Social Européen et au Comité des Régions, L'intelligence artificielle pour l'Europe*, Bruxelles, 25 avril 2018, COM (2018) 237 final.

reconnaissance vocale et le traitement automatique du langage naturel[27]. Chacun de ces domaines a recours aux systèmes les plus modernes de *machine learning*, de techniques d'apprentissage automatique grâce auxquelles l'IA est capable d'apprendre de sa propre expérience, et de *deep learning*, une méthodologie qui permet un apprentissage automatique en imitant les connexions neuronales du cerveau humain. Ainsi le législateur ne peut ignorer la nécessité de s'appuyer sur les experts en la matière pour développer un travail pluridisciplinaire sur ce sujet.

En effet, nous ne pouvons pas imaginer qu'une telle structure articulée soit exempte de risques et de problèmes, notamment sur le plan juridique. La caractéristique intrinsèque de l'IA de ne pas rendre ses processus décisionnels lisibles et « transparents » complique l'analyse des risques et des avantages de cette nouvelle technologie. Les décisions et les actions sont prises par des algorithmes à boîte noire (*black box algorithms*) qui, par définition, en raison de leur opacité, rendent le résultat imprévisible[28].

Cependant, ce n'est pas l'opacité du système qui doit susciter de la crainte, mais plutôt sa neutralité apparente et fallacieuse. En fait, le risque est d'accorder trop de confiance au système électronique en raison de son aptitude à obtenir des résultats beaucoup plus facilement et rapidement que l'esprit humain. Cette confiance est dangereuse, car le système n'est jamais exempt d'erreurs[29]. C'est ce que nous appelons « *fallacia dell'automazione* » (l'automatisation trompeuse), c'est-à-dire la confiance irrationnelle et inconsciente que l'être humain a tendance à accorder à la technologie, en croyant que son fonctionnement est objectif[30]. L'intelligence artificielle ne peut pas remplacer l'activité

27 J. KAPLAN, *Intelligenza Artificiale. Guida al futuro prossimo*, LUISS University Press, Rome 2017, p. 21 ss.

28 A. LAVORGNA/G. SUFFIA, « La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale », *Giustizia penale e nuove tecnologie* 2/2021, p. 88 ss.

29 PARLEMENT EUROPÉEN, Résolution sur l'intelligence artificielle en droit pénal et son utilisation par les autorités policières et judiciaires dans les affaires pénales, 6 octobre 2021, 2020/2016 (INI).

30 G. UBERTIS, « Intelligenza artificiale, giustizia penale, controllo umano significativo », *in* CENTRO NAZIONALE DI PREVENZIONE E DIFESA SOCIALE (édit.), *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Giuffrè Francis Lefebvre, Milan 2021, p. 9 ss ; P. COMOGLIO, « Prefazione », *in* J. NIEVA-FENNOLL (édit.), *Intelligenza artificiale e processo*, Giappichelli, Turin 2018, p. IX. Le concept de « fallacia dell'automazione » ne prend pas du tout de distance, mais finit par confirmer ce que Rouvroy (et Perri) ont affirmé au sujet de la « gouvernabilité algorithmique ». À ce sujet, voir A. ROUVROY/B. STIEGLER, « Il regime di verità digitale. Dalla governamentalità algoritmica a un nuovo Stato di diritto », *La Deleuziana*, n° 3, 2016, p. 6-29 ; P. PERRI, *Sorveglianza elettronica, diritti fondamentali ed evoluzione tecnologica*, Giuffrè Francis Lefebvre, Milan 2020, p. 31.

humaine, car la possibilité d'agir sur la réalité implique une capacité de lecture des circonstances qui est une prérogative du raisonnement humain et qui ne peut jamais être déléguée.

En fait, le problème réside dans le risque d'être piégé dans des dilemmes éthiques qui peuvent provoquer une paralysie morale de la machine (1), ou la pousser vers un jugement éthique obligatoire sur la base de données et de variables préétablies sans qu'elle ait une connaissance réelle et flexible du contexte dans lequel elle se trouve (2).

1. *Les trois lois de la robotique d'Isaac Asimov et la paralysie morale d'une machine*

En 1950, Isaac Asimov avait déjà prévu dans ses histoires de science-fiction des affrontements juridiques entre les lois humaines et l'imprévisibilité de la technologie. Il avait conçu des mécanismes juridiques créatifs pour surmonter l'absence de conciliation entre l'homme et la machine. Parmi ces mécanismes, les « trois lois de la robotique », dans le contexte d'une hypothétique année 2058, sont particulièrement connues :

> « 1. Un robot ne peut porter atteinte à un être humain ni, restant passif, permettre à ce qu'un être humain soit exposé au danger.
>
> 2. Un robot doit obéir aux ordres que lui donne un être humain, sauf si de tels ordres entrent en conflit avec la première loi.
>
> 3. Un robot doit protéger son existence aussi longtemps que cette protection n'entre pas en conflit avec la première ou la deuxième loi. »[31]

Malgré l'apparente simplicité de ces lois, si elles étaient mises en pratique en termes juridiques, cela soulèverait de nombreux problèmes. Premièrement, comment les « dommages » seraient-ils identifiables et quantifiables ? Si, par exemple, un robot est témoin d'une tentative de meurtre de la part d'un homme contre un autre homme, doit-il porter atteinte à l'auteur du crime ou omettre d'intervenir en laissant le crime se dérouler ?

En outre, il faut tenir compte du fait que les ordres donnés par les humains aux robots peuvent ne pas être très précis et tomber dans des contradictions linguistiques ou cognitives. Dans une telle éventualité, un robot risquerait de se retrouver coincé entre l'une et l'autre interprétation de la commande, ne parvenant pas à traduire correctement le langage naturel, avec ses inévitables nuances, figures rhétoriques et jeux de langage en fonction du contexte, dans le langage avec

31 I. ASIMOV, *I, Robot*, Doubleday, Garden City/New York 1950 (traduction de l'auteur).

lequel il est programmé. Nous pourrions dire que, dans ce cas, seule l'ingéniosité humaine, en tant que moteur de recherche de la rationalité, pourrait faire sortir un sujet d'une position de paralysie morale, c'est-à-dire d'une logique binaire, en imaginant une troisième voie de résolution du problème.

De plus, en ce qui concerne la première loi de la robotique d'Asimov, si au lieu de « robots », on utilisait le terme sans doute plus générique et moins évocateur de « machines », on pourrait reconsidérer la situation actuelle sous un jour résolument différent. En effet, il existe des milliers de personnes qui développent des dépendances et des troubles psychologiques de différents types et degrés à la suite d'une utilisation excessive des réseaux sociaux. Ce n'est certainement pas un hasard, mais parce que leurs algorithmes sont spécifiquement conçus pour exploiter à leur avantage certains mécanismes psychologiques qui se sont constitués au cours de milliers d'années d'évolution grâce à des formes de socialisation directe (et non médiatisée)[32].

À noter qu'au-delà des questions juridiques se posent les questions écologiques. Le coût de l'énergie est considérable à l'heure de la crise énergétique mondiale, avec les salles remplies de serveurs consommant sans interruption pour rendre le service disponible à l'utilisateur.

Les trois lois de la robotique illustrent la concrétisation et l'actualité du problème dans la rencontre difficile entre moralité et système pénal par le paradigme du *trolley problem.*

2. *Le trolley problem et l'inévitabilité du jugement éthique*

Le *trolley problem* est un intéressant paradigme de matrice morale dont nous trouvons une analyse approfondie dans l'essai *The Social Dilemma of Autonomous Vehicle*[33] publié dans la revue *Science* en juin 2016. Cette étude met en lumière certaines questions éthiques concernant l'introduction sur le marché des véhicules à conduite autonome.

La question est la suivante : en présence d'un danger imminent pour la vie du conducteur, des passagers de la voiture ou des passants, comment la voiture doit-elle se comporter ? Ou plutôt, qui devrait-elle choisir de sauver en premier

32 Pour en savoir plus sur la responsabilité des *providers* en cas de diffusion de contenus illicites sur les *social networks*, voir B. PANATTONI, « Gli effetti dell'automazione sui modelli di responsabilità : il caso delle piattaforme online », *Diritto Penale Contemporaneo*, 02/2019, p. 33-58.

33 J. BONNEFON/A. SHARIFF/I. RAHWAN, « The Social Dilemma of Autonomous Vehicles », *Science*, vol. 352, n° 6293, 06/2016, p. 1573-1576. Voir ég. A. CAPPELLINI, « Profili penalistici delle self-driving cars », *Diritto Penale Contemporaneo*, 02/2019, p. 333.

dans un choix forcé par des événements imprévisibles ? La voiture pourrait décider de sauver le conducteur en heurtant un ou plusieurs passants ou de sauver les passants en sacrifiant son propre passager[34]. Les réponses au *trolley problem* ont fait l'objet d'une évaluation statistique afin d'identifier les meilleurs choix éthiques dans le cadre de l'utilisation de voitures à conduite autonome[35]. Le plus grand nombre de personnes testées étaient prêtes à sauver la vie de plusieurs personnes au détriment d'une seule ou à sauver un enfant plutôt qu'une personne âgée[36]. Cependant, si un proche se trouve dans la voiture et risque d'être sacrifié avec le conducteur, l'éthique s'inverse, au détriment du plus grand nombre ou de l'enfant[37].

Nous pouvons lire clairement dans ces résultats comment l'homme en tant qu'être social ne peut être enfermé dans un principe d'utilité standardisé. La morale est fondée sur un ensemble de circonstances contextuelles et de variables individuelles, jamais exemptes de la possibilité d'une erreur humaine. Le fait d'imposer des principes moralisants à une machine comporte le risque de porter atteinte au principe de la dignité humaine en procédant à une évaluation de type utilitariste fondée sur le paramètre du « moindre mal » : deux vies ne valent pas plus qu'une[38].

L'avenir réserve un lien encore plus étroit entre l'homme et la machine. L'intelligence humaine et l'intelligence artificielle maintiendront inévitablement l'écart donné par l'expérience, car l'homme ne peut pas égaler la machine dans ses capacités et la machine ne peut pas simuler l'apprentissage culturel de l'esprit humain.

III. La responsabilité pénale d'une intelligence artificielle

Comme déjà mentionné ci-dessus, l'IA utilise des *black box algorithms* qui rendent le processus de décision opaque, même pour le programmeur[39]. Cela soulève de sérieux doutes quant au caractère personnel de la responsabilité pénale. Ce principe est expressément prévu à l'art. 27 par. 1 de la

34 CAPPELLINI (n. 33), p. 333.

35 *Ibid.*

36 F. COSTANTINI/P. L. MONTESSORO, « Il problema della sicurezza tra informatica e diritto : una prospettiva emergente dalle "Smart Cars" », *Informatica e diritto*, XLII, vol. XXV, 2016, n° 1, p. 99 ss.

37 *Ibid.*

38 I. COCA-VILA, « Self-driving Cars in Dilemmatic Situations : An Approach Based on the Theory of Justification in Criminal Law », *Criminal Law and Philosophy*, vol. 12, 2018, p. 59-82 ; repris par CAPPELLINI (n. 33), p. 333.

39 M. DONCIEUX, « Beyond Black-Box Optimization : A Review of Selective Pressures for Evolutionary Robotics », *Evolutionary Intelligence*, 2014, p. 71 ss.

Constitution italienne. Il est également inscrit, en droit français, à l'art. 121-1 du Code pénal, sous réserve des principes de nécessité et de présomption d'innocence contenus aux art. 8 et 9 de la Déclaration des droits de l'homme et du citoyen de 1789[40].

De nombreux facteurs influencent l'évaluation de la responsabilité. En effet, il faut tenir compte du niveau de délégation des fonctions par l'homme à la machine, du degré d'autonomie de la machine, du degré de la faute ou du dol, ainsi que de la structure organisationnelle hiérarchique entre la machine et les personnes humaines impliquées dans sa programmation, construction, distribution et utilisation.

Afin d'aborder correctement la question de la responsabilité pénale, il est nécessaire d'étudier le rôle joué par l'IA dans la commission de l'infraction, en évaluant la possibilité d'attribuer une responsabilité directe à l'agent humain (A) ou à l'agent artificiel (B).

A. Le rôle de l'IA dans la commission de l'infraction et la responsabilité indirecte de l'homme

En ce qui concerne la responsabilité pénale, l'IA pourrait être définie dans la structure de l'infraction soit comme objet, entendu comme un bien juridique protégé, soit comme instrument avec lequel l'infraction est commise, soit comme sujet actif et passif[41].

La qualification du robot en tant qu'objet de protection pénale ne pose pas de problèmes particuliers. Dans le cas où il s'agit d'un bien précieux ou destiné à un usage social, l'IA pourra être traitée comme un bien patrimonial[42]. Parfois, les systèmes d'IA sont si sophistiqués qu'ils pourraient même être considérés

40 Principe reconnu par le système juridique français seulement après son introduction dans le Code pénal français de 1994. La jurisprudence française tend à sacrifier ce principe en se montrant parfois favorable à l'imputation de plusieurs faits concomitants à plusieurs personnes, même si le lien de causalité entre le fait dommageable et le comportement de chaque personne n'est pas reconnu. Y. MAYAUD, « Les systèmes pénaux à l'épreuve du crime organisé », *Revue internationale de droit pénal*, 1997, p. 794 ss ; repris par S. BERNARDI, « La disciplina prevista da nuovo codice penale francese in tema di criminalità organizzata », *in* G. FORNARI (édit.), *Strategie di contrasto alla criminalità organizzata nella prospettiva di diritto comparato*, CEDAM, Padoue 2002, p. 38-41.

41 S. RIONDATO, « Robot : talune implicazioni di diritto penale », *in* P. MORO/C. SARRA (édit.), *Tecnodiritto, Temi e problemi di informatica e robotica giuridica*, Franco Angeli, Milan 2017, p. 85 ss ; F. BASILE, « Intelligenza artificiale e diritto penale : quattro possibili percorsi di indagine », *Diritto penale e uomo*, 09/2019, p. 24 ss.

42 RIONDATO (n. 41), p. 85 ss.

comme des « formes de vie non humaines », auquel cas une protection équivalente à celle des animaux pourrait être adoptée[43].

Il en va différemment lorsque l'IA est qualifiée d'instrument pour la commission d'une infraction. Dans ce cas, la responsabilité doit incomber au programmeur, au fabricant, au distributeur ou à l'utilisateur[44]. Parmi ces personnes, le rôle le plus complexe est celui du programmeur. Il serait raisonnable de le considérer comme responsable dans les limites de l'autonomie décisionnelle de la machine, qui devient de plus en plus large grâce à l'expérience qu'elle fait dans l'environnement extérieur. Il faudra en tout cas vérifier la présence de l'élément matériel, à savoir la commission de l'acte, ainsi que de l'élément psychologique, c'est-à-dire l'intentionnalité ou la non-intentionnalité de l'agent humain.

La dernière hypothèse concerne l'attribution au robot d'une subjectivité juridique, c'est-à-dire d'une personnalité de droit moral, nouvelle en droit pénal. Dans ce cas, nous pourrions envisager une reconnaissance de la responsabilité pénale directement auprès de la machine, qualifiant cette dernière comme agent (auteur) de l'infraction. Toutefois, cela soulèverait des doutes évidents quant à la finalité d'une sanction imposée de cette manière. En fait, que gagnerait la machine à être punie ? Est-ce que le fait de parler d'un droit pénal de l'homme aurait encore un sens, ou devrait-on commencer à concevoir un droit des machines ? Dans le cas où c'est le robot lui-même qui subit un préjudice, il peut être considéré comme le sujet passif de l'infraction et non plus comme le bien juridique protégé. Prenons, par exemple, le cas d'un robot aux caractéristiques humanoïdes utilisé à des fins sexuelles. Ce dernier pourra également être programmé, au moyen de certains algorithmes, avec la capacité de discerner et éventuellement de rejeter un acte sexuel. Pourrions-nous, dans ce cas, le considérer comme une victime si son consentement n'est pas respecté[45] ?

La qualification de l'IA en tant qu'instrument pour la commission de l'infraction nécessite une distinction préalable entre le moyen et l'action dans la théorie générale du crime[46]. L'action doit être attribuée à l'homme aussi bien dans

43 *Ibid.*

44 B. M. MAGRO, « A.I. : la responsabilità penale per la progettazione, la costruzione e l'uso dei robot », *Il Quotidiano Giuridico*, 12.06.2018, p. 1-5.

45 F. Basile (n. 41) p. 32. Pour une mise à jour par le même auteur, voir : F. BASILE, « Intelligenza artificiale e diritto penale : qualche aggiornamento e qualche nuova riflessione », *in* F. BASILE/M. CATERINI/S. ROMANO (édit.), *Il sistema penale ai confini delle hard sciences. Percorsi epistemologici tra neuroscienze e intelligenza artificiale*, Pacini Giuridica, Pise 2021.

46 T. Delegu, « Lo "strumento" nella teoria generale del reato », *Rivista italiana di diritto e procedura penale*, 1974, p. 275 ss ; repris par RIONDATO (n. 41), p. 85 ss.

un sens objectif, car c'est lui qui crée et planifie l'IA, que dans un sens subjectif, car c'est toujours lui qui en a conscience et volonté[47]. D'autre part, l'instrument existe en dehors de l'homme et indépendamment de lui, car il peut être totalement robotisé et être utilisé pour la réalisation d'actions criminelles. Pour ces raisons, l'instrument doit être considéré comme instrumental à l'action criminelle et l'action criminelle doit être imputable à l'être humain[48].

En suivant cette logique, l'IA comprise comme instrument peut constituer un facteur aggravant ou atténuant lors de l'attribution de la peine[49]. En outre, les systèmes d'IA, en tant qu'instruments, pourraient également faire l'objet d'une confiscation et d'une saisie en application de mesures de prévention ou de sécurité ou en exécution d'une condamnation[50]. La responsabilité pénale de l'agent humain pourra être reconnue si une responsabilité intentionnelle ou non intentionnelle lui est imputable (1), en tenant compte du fait que l'IA ne peut pas être impliquée – sauf de par sa fonction instrumentale – puisqu'elle n'a pas d'intelligence et de volonté et, donc, pas de capacité de véhiculer sa propre intentionnalité (2).

1. Responsabilité intentionnelle et non intentionnelle de l'agent humain

L'attribution d'une responsabilité intentionnelle (ou *dolosa*) à un acteur humain ne pose pas de problème particulier. Si le programmeur ou l'utilisateur configure la machine pour qu'elle réalise délibérément une action criminelle, le fait sera directement imputable à l'agent humain, puisque l'événement qui se produit est la réalisation d'un risque spécifique inhérent à l'action initiale[51]. Par ailleurs, il semblerait qu'une tentative d'infraction soit envisageable

47 Riondato (n. 41), p. 85 ss.

48 *Ibid.* D'après l'art. 133 du Code pénal italien : « Le juge doit tenir compte de la gravité de l'infraction déduite de la nature, du genre, des *moyens*, de l'objet, du moment, du lieu et de toute autre modalité de l'action » (traduction de l'auteur). Pour établir la présence d'une faute pénale d'imprudence, le juge français doit estimer que « l'auteur des faits n'a pas accompli les diligences normales compte tenu, le cas échéant, de la nature de ses missions ou de ses fonctions, de ses compétences ainsi que du pouvoir et des *moyens* dont il disposait » en application de l'art. 21-3, al. 3 du Code pénal français.

49 *Ibid.*

50 *Ibid.*

51 Pour le droit italien, voir G. MARINUCCI/E. DOLCINI, *Manuale di Diritto penale. Parte generale*, 5e éd., Giuffrè Francis Lefebvre, Milan 2015, p. 354.

si l'événement voulu par l'agent humain ne se produit pas en raison d'une déviation comportementale imprévisible de la machine[52]. Plus complexe, cependant, est l'attribution de la responsabilité non intentionnelle (ou *colposa*) à l'utilisateur et au programmateur, puisqu'elle dépend du lien de causalité entre l'événement concret et la conduite de l'agent coupable, dont les développements doivent être prévisibles[53].

Malgré la difficile prévisibilité du comportement de la machine, si on veut admettre une responsabilité coupable de l'utilisateur ou du programmeur, cela pourrait résulter du manquement à une obligation de prévention[54]. En ce sens, une extension des mesures de précaution concernant l'utilisation et le déploiement des systèmes d'IA pourrait être envisagée, même en sachant que cela pourrait entraîner un ralentissement ou un arrêt total de la production de tels systèmes[55]. Dans le cas contraire, si la non-prévisibilité est telle qu'aucun lien de causalité ne peut être établi, aucune responsabilité de l'opérateur humain ne peut être retenue[56].

Il conviendrait d'établir différents niveaux de prudence en fonction du degré d'autonomie et de l'utilisation effective de la machine[57]. La Commission européenne est intervenue, sur ce point, avec une Proposition de Règlement visant à classer les systèmes d'IA sur la base du risque, allant jusqu'à interdire l'utilisation de certains de ces systèmes. Parmi ces derniers figurent les systèmes qui réalisent des formes de surveillance intrusive (comme l'identification biométrique à distance) ou qui sont susceptibles de causer des dommages physiques ou psychologiques en manipulant le comportement des personnes et en

52 A. CAPPELLINI, « Machina delinquere non potest ? Brevi appunti su intelligenza artificiale e responsabilità penale », *DisCrimen*, 03/2019, p. 505.

53 Pour le droit italien, voir MARINUCCI/DOLCINI (n. 51), p. 354.

54 Conformément à l'art. 121-3 al. 3 du Code pénal français : « Il y a également délit, lorsque la loi le prévoit, en cas de faute d'imprudence, de négligence ou de manquement à une obligation de prudence ou de sécurité prévue par la loi ou le règlement, s'il est établi que l'auteur des faits n'a pas accompli les diligences normales compte tenu, le cas échéant, de la nature de ses missions ou de ses fonctions, de ses compétences ainsi que du pouvoir et des moyens dont il disposait », et à l'art. 43 al. 1 du Code pénal italien : « L'infraction est *colposa* ou *contro l'intenzione* lorsque l'événement, même s'il est prévisible, n'est pas voulu par l'auteur et survient à la suite de *negligenza* ou d'*imprudenza* ou d'*imperizia*, ou à la suite du manquement aux lois, règlements, ordres ou disciplines » (traduction de l'auteur).

55 MAGRO (n. 44), p. 1-5.

56 *Ibid.*

57 *Ibid.*

contournant leur libre arbitre (*social scoring*, c'est-à-dire des systèmes de notation sociale)[58].

Si, dans un avenir lointain, un système d'IA acquiert un degré d'autonomie tel qu'il ne peut plus être soumis au contrôle de l'homme, les dommages causés par son comportement pourraient également être considérés comme un événement exceptionnel et imprévisible et donner lieu à une cause d'exclusion de la responsabilité pénale pour cas fortuit ou force majeure[59].

2. *L'intelligence et la volonté d'une intelligence artificielle*

La responsabilité pénale s'adresse à des êtres libres, dotés d'intelligence et de volonté[60]. L'intelligence est la capacité de discerner (*capacità d'intendere*), c'est-à-dire la capacité de l'agent à comprendre la signification et les conséquences sociales de ses actes, tandis que la volonté (*capacità di volere*) est la capacité de s'autodéterminer librement[61]. Ce n'est qu'en présence de ces conditions, qui doivent exister au moment de l'acte et en relation avec le fait individuel, que la personne peut être réprimandée pour l'acte commis[62].

Les agents humains peuvent être considérés comme capables de *comprendere* et de *volere*, à moins qu'une cause d'exclusion ne soit présente. Inversement, les agents artificiels, même s'ils sont reconnus comme ayant une subjectivité juridique dans le futur, ne seront jamais capables de s'autodéterminer librement ou d'exercer un libre arbitre « moral »[63].

Les robots pourraient être considérés comme nuisibles, mais pas coupables, et leur comportement pourrait générer un accident, pas un crime[64]. En effet, seul un agent « moral », donc humain, peut être rééduqué, puisque la rééducation

58 COMMISSION EUROPÉENNE, Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernent l'intelligence artificielle et modifiant certaines actes législatives de l'Union, Bruxelles, 21 avril 2021, COM (2021) 206 final. Toujours à propos des risques, COMMISSION EUROPÉENNE, Livre Blanc sur l'intelligence artificielle. Une approche européenne axée sur l'excellence et la confiance, COM(2020) 65 final, 19 février 2020.

59 MAGRO (n. 44), p. 1-5.

60 E. DREYER, *Droit pénal général*, 6e éd., LexisNexis, Paris 2022, p. 679.

61 *Idem*, p. 680 et p. 691 s. Voir aussi MARINUCCI/DOLCINI (n. 51), p. 383.

62 MARINUCCI/DOLCINI (n. 51), p. 383.

63 ASARO (n. 17), *Robots and Responsibility*, p. 20-24.

64 E. DREYER, « Intelligence artificielle et droit pénal », *in* A. BENSAMOUN/G. LOISEAU (édit.), *Droit de l'intelligence artificielle*, LGDJ, Paris 2019, p. 215.

implique le développement et la correction d'un caractère moral[65]. En outre, il n'y aurait aucun sens de punir, ou plutôt de « rééduquer », quelqu'un qui n'est pas responsable de l'acte commis, ni de punir une machine, qui ne gagnerait rien qu'une re-programmation ou un arrêt définitif[66]. La répression n'a pas de sens non plus, car seuls les agents « moraux » sont capables de s'identifier à d'autres agents « moraux » et de reconnaître la similitude entre leurs propres actions et celles d'un autre individu jugé coupable et puni par le système[67].

B. Le processus de transformation de l'IA en un sujet de l'infraction

Le droit pénal est un droit principalement axé sur la personne humaine[68]. Toutes les personnes ne sont pas « humaines ». En effet, la plupart des pays d'Europe continentale reconnaissent la personnalité juridique des sociétés et leur attribuent une responsabilité pénale, indépendante de celle éventuellement prévue pour les personnes physiques agissant en leur nom[69]. La personne « humaine » est reconnue et protégée par de multiples chartes internationales, qui protègent les droits humains et les valeurs fondamentales. Le droit pénal joue un double rôle à l'égard de l'individu, d'une part en le protégeant lorsqu'il est victime d'une infraction, et d'autre part en ne manquant pas de le protéger lorsqu'il subit, en tant que suspect ou accusé d'une infraction, une violation illégale et arbitraire de ses droits fondamentaux[70]. Non seulement les sociétés, mais aussi les animaux ont été reconnus par le système juridique comme des êtres « non humains » et protégés en tant que victimes de crimes. Cependant, ce sont toujours les humains qui décident de ce qui est bien et de ce qui est mal, en restant les détenteurs des intérêts et des valeurs protégés par le système de justice pénale[71].

65 Aux termes de l'art. 27 par. 3 de la Constitution italienne : « Les peines ne peuvent consister en des traitements contraires au sens de l'humanité et doivent tendre à la rééducation du condamné » (traduction de l'auteur).

66 Conférence International Society of Public Law (ICON-S), intervention de Paolo Moro, Aux frontières de la subjectivité. Les figures de responsabilité des machines intelligentes, dans le panel « AI : disciplinata o spregiudicata ?. A.I., diritto e responsabilità », présidé par Andrea Pin, Florence, 22-23 novembre 2019, accessible au lien suivant : www.icons-italia.it/firenze-2019/ (consulté le 05.08.2022).

67 CAPPELLINI (n. 51), p. 512 ss ; ASARO (n. 17), *Robots and Responsibility*, p. 20-24.

68 ASARO (n. 17), *Algorithms of Violence*, p. 550 ss.

69 MARINUCCI/DOLCINI (n. 50), p. 760.

70 S. Riondato, « Robotica e diritto penale (robot, ibridi, chimere, "animali tecnologici") », *in* D. PROVOLO/S. RIONDATO/F. YENISEY (édit.), *Genetics, Robotics, Law, Punishment*, Padova University Press, Padoue 2014, p. 602.

71 *Ibid.*

L'autonomie croissante des systèmes d'IA nous amène à réfléchir sur la possibilité d'attribuer une responsabilité pénale directement à la machine (1). Toutefois, le processus logique qui conduit à cette conclusion diffère sensiblement de celui qui permet de reconnaître la responsabilité pénale d'autres organismes non humains, tels que les personnes morales (2).

1. *Les modèles de responsabilité de Hallevy*

Le principal défenseur de la reconnaissance de la subjectivité d'un agent artificiel est Gabriel Hallevy, qui propose trois modèles de responsabilité : la *perpetration-by-another*, la *natural-probable-consequence* et la *direct liability*[72].

Le premier de ces modèles, la *perpetration-by-another*, considère les systèmes intelligents comme des agents innocents, qui ne peuvent être reconnus comme possédant des caractéristiques proprement humaines[73]. Hallevy compare cette forme d'IA à un enfant ou à une personne en état d'infirmité qui, pour cette raison, ne peut être tenue pénalement responsable de l'acte qu'il a commis. Le robot est donc considéré comme un outil et l'auteur de l'infraction est à rechercher dans la personne qui a orchestré l'infraction.

Le deuxième modèle de responsabilité, la *natural-probable-consequence*, présuppose l'implication des programmeurs et des utilisateurs dans le mécanisme d'imputation et n'achève pas encore complètement le procès de reconnaissance de la responsabilité directe de la machine. Le contexte est celui d'un robot programmé pour effectuer certaines tâches, mais qui, à l'insu des agents humains, commet un crime. Ce modèle de responsabilité, dit Hallevy, « est fondé sur la capacité des programmeurs et des utilisateurs à prévoir la commission potentielle d'infractions »[74]. Contrairement au modèle de la *perpetration-by-another*, l'agent humain ici n'a pas de *mens rea*, c'est-à-dire l'intention de commettre l'infraction : l'infraction est une conséquence naturelle et probable du comportement de cette personne[75].

En revanche, le troisième modèle de responsabilité, la *direct liability*, propose l'attribution directe de la responsabilité à la machine si les exigences de l'élément matériel et de l'élément psychologique sont remplies. Selon Hallevy, l'*actus reus* pourrait être attribué au robot intelligent s'il est conçu pour avoir

72 G. HALLEVY, « "I, Robot – I, Criminal" – When Science Fiction Becomes Reality : Legal Liability of AI Robots Committing Criminal Offenses », *Syracuse Science and Technology Law Reporter*, vol. 22, 2010, p. 1 ss.

73 *Idem*, p. 9.

74 *Idem*, p. 14 (traduction de l'auteur).

75 *Ibid.*

une certaine mobilité des membres et un contrôle mécanique de son corps. Toutefois, sans une structure physique classiquement comprise des « robots à forme humaine », l'*actus reus* ne serait pas facilement discernable. La présence de l'exigence de la *mens rea* serait alors justifiée par le fait que la machine a pu être programmée pour avoir un objectif spécifique[76]. Néanmoins, la *mens rea* est un élément subjectif humain qui ne pourrait être remplacé par le but préprogrammé d'une machine.

2. *Les objections à la direct liability de Hallevy*

Toutefois, le modèle de responsabilité de la *direct liability*, développé par Hallevy, comporte de nombreux éléments critiquables. Riondato, notamment, affirme qu'un tel modèle ne pourrait en aucun cas « contredire l'idée fondamentale selon laquelle même une responsabilité pénale directe des robots serait au service d'objectifs humains (comme le contrôle social) et fondée sur le sens de la justice humaine »[77]. D'autre part, Cappellini, réadaptant le vieux principe *societas delinquere non potest* à la condition des sujets robotiques, affirme que le fait de parler de responsabilité morale et juridique de la machine n'aurait pas de sens, car « elle ne pourrait jamais raisonnablement être réprimandée pour un fait qu'elle a causé, car, contrairement à l'homme, elle n'est pas *libre* mais *déterminée*. C'est pourquoi le manque de liberté d'agir se reflète dans une carence inéluctable de *culpabilité* »[78].

Le principe *societas delinquere non potest* est désormais dépassé, tant en droit italien qu'en droit français. La responsabilité pénale de la personne morale est actuellement reconnue de manière indépendante – et sans exclusion – de celle de la personne physique. En France, cette responsabilité est reconnue par ricochet, sur la base de la condition préalable de la commission de l'infraction par les organes ou les représentants pour le compte de la personne morale. En Italie, cependant, il faut que soit présent le critère minimal de la *colpa d'organizzazione* (faute d'organisation) et que l'infraction soit commise *nell'interesse o a vantaggio* (dans l'intérêt ou à l'avantage) de sujets en position dominante ou de leurs subordonnés.

Le problème, qui porte sur les deux fronts de la responsabilité pénale de la personne morale et de la responsabilité directe de l'IA, trouve son origine dans le concept philosophique de l'existence : « Une personne juridique "existe-t-elle" vraiment, dans une certaine mesure, dans le monde réel ? Ou bien n'est-ce qu'une fiction juridique, une métaphore linguistique résumant un ensemble

76 HALLEVY (n. 72), p. 19-22.

77 RIONDATO (n. 70), p. 603 (traduction de l'auteur).

78 CAPPELLINI (n. 51), p. 502 (traduction de l'auteur) et p. 511 ss.

de règles spéciales applicables à certains domaines des rapports sociaux ? »[79]. Contrairement au robot, la personne morale « n'a pas de réalité naturelle indépendante de ses membres »[80] : elle n'existe qu'en tant que personne morale légalement constituée. L'IA se comporte comme un animal entraîné par l'homme ; par opposition, la personne morale, se comporte comme une marionnette contrôlée par son marionnettiste[81]. Dans le cas de la personne morale, le destinataire de la peine est de toute façon la personne « humaine ». Qui est le destinataire de la peine dans le cas de l'IA ? La réponse est encore une fois à trouver dans l'explication du concept d'existence. Un système d'IA peut avoir un corps physique, toutefois, l'IA revêt rarement une forme anthropomorphique et consiste souvent en la simple installation et programmation de logiciels[82]. Il semblerait donc difficile de concevoir une imputation directe du système d'IA, puisque, même si ce dernier dispose d'un « corps » presque-réel, sa punition ne serait pas justifiée par les finalités, préventives ou répressives, de la peine et laisserait ouvert un véritable vide de responsabilité pour des actes – parfois très graves – commis par des êtres humains au moyen de machines.

D'autres questions se posent en raison de l'arborescence des entreprises impliquées dans la planification, la construction, la diffusion et l'utilisation de l'instrument. Un degré élevé d'autonomie décisionnelle de la machine entraînerait l'imprévisibilité du résultat attendu. Cela rendrait difficile la distinction des rôles et l'identification de celui qui, dans les faits, aurait pu empêcher le dommage de se produire. Une meilleure compréhension du processus de délégation et une intensification des mesures de précaution pourraient, plutôt que de limiter l'usage des systèmes d'IA, permettre une utilisation consciente des risques et des grandes difficultés existant en termes de dommages causés et d'éventuelles obligations de réparation.

IV. Conclusion

Le pionnier de l'intelligence artificielle, Alan Turing, a introduit son « jeu d'imitation » avec ces mots : « Nous ne voulons pas pénaliser une machine pour son incapacité à briller dans les concours de beauté ni pénaliser un homme pour avoir perdu dans une course contre un avion »[83]. Il est clair que l'intelligence humaine et l'intelligence artificielle accomplissent des tâches

79 *Idem*, p. 513 (traduction de l'auteur).
80 *Idem*, p. 515 (traduction de l'auteur).
81 *Ibid.*
82 *Ibid.*
83 A. M. TURING, « Computing Machinery and Intelligence », *Mind*, vol. 49, 1950, p. 435 (traduction de l'auteur).

différentes et ne sont donc comparables ni sur le plan éthique ni sur le plan juridique. Bien que les systèmes d'IA soient des outils dont l'utilité est incontestable, leur utilisation dans la société soulève de nombreuses questions juridiques, qui rendent complexe l'identification de limites et de protections, dans le respect des principes fondamentaux du système juridique. L'éthique et le droit vont de concert, constituant un nouveau profil juridique autour d'un sujet-objet qui s'infiltre à grande vitesse dans tous les domaines de la vie quotidienne, amenant avec lui non seulement innovation mais également désorientation. Si, d'une part, l'IA est capable de résoudre des problèmes que l'homme, sans instrumentation, ne pourrait jamais réaliser, d'autre part, elle est également utilisée pour la gestion de tout ce qui est « *dull, dirty and dangerous* », c'est-à-dire « stupide, sale et dangereux »[84]. Cela est particulièrement vrai dans le contexte de la guerre, où l'utilisation de systèmes automatisés éloigne l'homme du risque de se – rendre sur les lieux de conflit et de subir lui-même des dommages physiques, mais cela peut également se vérifier dans d'autres domaines, comme le trafic de drogue à travers l'utilisation de drones ou de sous-marins autonomes[85].

Si l'on tient compte des coûts de ces outils, on comprend combien l'accès aux systèmes d'IA est souvent réservé aux grandes entreprises et organisations. La détermination de la responsabilité pénale pour des actes commis par le biais de systèmes d'IA est rendue encore plus complexe par la structure de responsabilité à plusieurs niveaux impliquant les personnes physiques et morales qui participent au déploiement, à la construction, à la programmation et à l'utilisation de ces systèmes. Cependant, les conséquences sont également coûteuses en termes de responsabilité morale, puisqu'en remplaçant nos actions par celles mises en œuvre par des systèmes d'IA, nous permettons une déresponsabilisation progressive, déclenchant inévitablement un report de la responsabilité de l'agent humain sur la machine[86].

Dans ce contexte, l'importance de prévoir des mécanismes de responsabilité juridique pour le comportement d'une IA paraît évidente. La société moderne fait de plus en plus confiance aux systèmes d'IA sans bien connaître les méthodes que ces derniers utilisent pour prendre des décisions. En raison de l'opacité des processus décisionnels, il s'avère fondamental d'établir un mécanisme de contrôle et de responsabilité garantissant la protection des individus et protégeant les droits fondamentaux.

84 ISTITUTO DI RICERCHE INTERNAZIONALI (IRIAD), *Droni militari : proliferazione o controllo ?*, 04/2017, p. 8, accessible au lien suivant : www.ecchr.eu/fileadmin/Gutachten/Rapporto_Droni_militari_proliferazione_o_controllo.pdf (consulté le 28.08.2022).
85 *Ibid.*
86 BASILE (n. 41), p. 28 ss.

L'influence des innovations technologiques sur le droit de la responsabilité civile

L'intelligence artificielle : l'occasion d'unifier le droit de la responsabilité civile

ALICE FROCHAUX
Doctorante et assistante diplômée en droit des obligations |
Faculté de droit, des sciences criminelles et de l'administration publique |
Université de Lausanne

Table des matières

I. Introduction

L'intelligence artificielle (ci-après : IA) est perçue aujourd'hui comme la nouvelle technologie qui va s'immiscer peu à peu dans tous les secteurs de la société[1]. Cette technologie soulève certaines questions en matière de responsabilité civile. L'histoire de la responsabilité civile ayant été rythmée par les innovations technologiques, cette contribution propose d'analyser la manière dont ces innovations ont influencé le droit de la responsabilité civile, le système auquel elles ont abouti et comment l'IA peut s'insérer dans ce système.

II. Un système de responsabilité civile centré sur le fait d'une chose

A. L'évolution d'un système centré sur le fait d'une personne à un système centré sur le fait d'une chose

1. Un système de responsabilité civile centré sur la faute

Le Code fédéral des obligations de 1881 faisait de la faute l'élément central de la responsabilité en consacrant une clause générale de responsabilité pour faute (art. 50 aCO), qui est l'ancêtre de notre art. 41 CO[2].

À côté de cette clause générale, il y avait trois chefs de responsabilité spéciaux, repris dans le Code des obligations. Il s'agissait de la responsabilité de l'employeur (art. 62 aCO, art. 55 CO) et de celle du détenteur d'un animal (art. 65 aCO, art. 56 CO) ; ceux-ci étaient tenus pour responsables du préjudice causé par un employé respectivement par un animal sauf preuve de la diligence requise, et de la responsabilité du propriétaire d'ouvrage qui répondait en cas de défaut d'un ouvrage (art. 67 aCO, art. 58 CO)[3]. Ces chefs de responsabilité spéciaux étaient appréhendés comme des variantes de la responsabilité subjective. Selon la doctrine et la jurisprudence, le reproche de manque de diligence requis aux art. 62 aCO et 65 aCO équivalait à une présomption selon laquelle

1 GROUPE DE TRAVAIL INTERDÉPARTEMENTAL « INTELLIGENCE ARTIFICIELLE », *Rapport au Conseil fédéral, Défis de l'intelligence artificielle*, du 13 décembre 2019, p. 17, 22 et 58.

2 Loi fédérale complétant le Code civil suisse (Livre cinquième : Droit des obligations) (CO ; RS 220). Voir P. TERCIER, « Cent ans de responsabilité civile en droit suisse », *in* P. HANS/E. STARK/P. TERCIER (édit.), *Le centenaire du code des obligations, Mélanges*, Fribourg 1982, p. 203, p. 206.

3 TERCIER (n. 2), p. 206.

le comportement du responsable était fautif[4] et le défaut de l'ouvrage, au sens de l'art. 67 aCO, devait être le résultat d'un comportement fautif[5].

2. La révolution industrielle

La révolution industrielle du XIX[e] siècle a conduit au développement de nouvelles technologies (par ex. machine à vapeur, moteur à explosion, électricité) qui ont fait apparaître des risques que la responsabilité subjective était incapable de saisir, puisque la majorité des accidents intervenaient indépendamment de toute faute[6]. Cela a rendu nécessaire l'élaboration d'un système de responsabilité civile qui était mieux adapté à la réparation des préjudices liés aux technologies. Cette adaptation s'est faite de trois manières : par l'objectivation de la notion de faute, le développement des responsabilités objectives simples et l'adoption de chefs de responsabilité pour risque[7].

La faute consacrée à l'art. 50 aCO a été appréhendée, dès l'origine, de manière objective comme un manquement à la diligence due par l'auteur[8], tout en conservant un aspect subjectif dans l'exigence de la capacité de discernement[9].

4 ATF 29 II 485, consid. 3 ; ATF 26 II 103 ; ATF 24 II 128 ; J. CHAMOREL, *La responsabilité de l'employeur pour le fait de ses employés en matière extracontractuelle, Art. 55 CO*, Lausanne 1925, p. 21 ; H. NATER, *Die Haftpflicht des Geschäftsherrn gemäss OR 55 angesichts der wirtschaftlich-technischen Entwicklung*, Glarus 1970, p. 8 ; K. OFTINGER/E. STARK, *Schweizerisches Haftpflichtrecht, Besonderer Teil*, Band II/1, Zurich 1987, § 20 N 2 ; P. TERCIER, « Quelques considérations sur les fondements de la responsabilité civile », *RDS* 1976 I, p. 1, p. 9 ; P. WIDMER/P. WESSNER, *Révision et unification du droit de la responsabilité civile*, Rapport explicatif, Berne 2000, p. 5.

5 B. MÉAN, *La responsabilité du propriétaire de bâtiment ou de tout autre ouvrage, Étude de jurisprudence fédérale*, Lausanne 1904, p. 25 et 67.

6 G. ETIER, *Du risque à la faute : évolution de la responsabilité civile pour le risque du droit romain au droit commun*, Genève/Zurich/Bâle 2006, p. 46 ; H. LANDOLT, « Kurze Geschichte des Schadenausgleichsrechts », *in* S. WEBER/S. FUHRER (édit.), *Retouchen oder Reformen ? Die hängigen Gesetzrevisionen im Bereich Haftung und Versicherung auf dem Prüfstand*, Zurich/Bâle/Genève 2004, p. 67, p. 73 ; K. OFTINGER, « Der soziale Gedanke im Schadenersatzrecht und in der Haftpflichtversicherung », *RSJ* 1943, p. 545, p. 548.

7 V. BRULHART, « Responsabilité pour risque et assurance de la responsabilité : ce qu'elles se doivent l'une à l'autre », *in* F. WERRO/P. PICHONNAZ (édit.), *Les responsabilités fondées sur le risque, Colloque du droit de la responsabilité civile 2017*, Berne 2018, p. 139, p. 142 ; ETIER (n. 6), p. 37 ; WIDMER/WESSNER (n. 4), p. 134.

8 ATF 11 I 56, consid. 6 ; W. FELLMANN/A. KOTTMANN, *Schweizerisches Haftpflichtrecht, Band I : Allgemeineiner Teil sowie Haftung aus Verschulden und Persönlichkeitsverletzung, gewöhnliche Kausalhaftungen des OR, ZGB und PrHG*, Berne 2012, N 538 ; M. JAUN, *Haftung für Sorgfaltspflichtverletzung, Von der Willenschuld zum Schutz legitimer Integritätserwartungen*, Berne 2007, p. 21.

9 FELLMANN/KOTTMANN (n. 8), N 571 ; P. PICHONNAZ/F. WERRO, « La responsabilité fondée sur le risque : un état des lieux et quelques perspectives d'avenir », *in* F. WERRO/

Pour pallier l'apparition des nouveaux risques, on a apprécié de manière toujours plus sévère les devoirs de diligence sous-tendant la notion de faute[10]. Le paroxysme de l'objectivation de la faute a été atteint avec la consécration du principe du *Gefahrensatz*, selon lequel celui qui crée un état de fait dangereux sans prendre les mesures de précaution nécessaires pour empêcher la survenance d'un préjudice est fautif[11].

Les chefs de responsabilité objective simple ont, quant à eux, été créés en écartant toute référence à la faute. La responsabilité de l'employeur (art. 55 CO), celle du détenteur d'un animal (art. 56 CO) et celle du propriétaire d'ouvrage (art. 58 CO) reposent ainsi sur un manque objectif de diligence[12]. Cette solution permet de tenir responsable une personne sans égard à sa capacité de discernement et à son implication personnelle[13].

Enfin, confronté à l'émergence de certains risques extrêmement importants, le législateur a également adopté des chefs de responsabilité objective aggravée où l'obligation de réparer le préjudice résulte exclusivement de la réalisation d'un risque indépendamment de tout comportement humain[14].

La responsabilité civile automobile illustre bien cette évolution. À l'origine, la responsabilité liée à un véhicule automobile relevait du droit commun, soit de l'art. 41 CO ou de l'art. 55 CO. Pour que l'art. 41 CO s'applique, il fallait pouvoir imputer une faute au conducteur du véhicule automobile[15]. Or, cette solution s'est vite avérée insuffisante pour appréhender le nouveau danger créé par

P. PICHONNAZ (édit.), *Les responsabilités fondées sur le risque, Colloque du droit de la responsabilité civile 2017*, Berne 2018, p. 1, p. 8 ; H. REY/I. WILDHABER, *Ausservertragliches Haftpflichtrecht*, 5e éd., Zurich/Bâle/Genève 2018, N 968 ; I. SCHWENZER/CH. FOUNTOULAKIS, *Schweizerisches Obligationenrecht, Allgemeiner Teil*, 8e éd., Berne 2020, N 22.03.

10 ETIER (n. 6), p. 36 ; WIDMER/WESSNER (n. 4), p. 134.

11 ATF 79 II 66, consid. 2 ; ATF 66 II 114, consid. 1 ; ETIER (n. 6), p. 39 s. ; WIDMER/WESSNER (n. 4), p. 118 et p. 134.

12 ATF 34 II 266, consid. 5 (art. 62 aCO) ; ATF 39 II 536, consid. 2 (art. 65a CO) ; C. BURCKHARDT, « Die Revision des schweizerischen Obligationenrechts in Hinsicht auf das Schadenersatzrecht », *RDS* 1903, p. 469, p. 520, p. 544 et p. 559 ; ETIER (n. 6), p. 42; J. METZGER, *La responsabilité du détenteur d'animaux*, Lausanne 1955, p. 90 ; OFTINGER/STARK (n. 4), § 19 N 1 ; TERCIER (n. 4), p. 9 ; WIDMER/WESSNER (n. 4), p. 5.

13 PICHONNAZ/WERRO (n. 9), p. 8 ; P. PICHONNAZ, « La responsabilité de l'intelligence artificielle : un régime de responsabilité objective simple comme solution nuancée », *in* E. M. BESLER/P. PICHONNAZ/H. STÖCKLI (édit.), *Le droit sans frontières, Mélanges pour Franz Werro*, Berne 2022, p. 513, p. 519.

14 FELLMANN/KOTTMANN, (n. 8), N 29 ; REY/WILDHABER (n. 9), N 99 ; F. WERRO, *La responsabilité civile*, 3e éd., Berne 2017, N 32.

15 ATF 51 II 73, consid 1 ; ATF 38 II 487, consid. 2 ; M. GRAF, *Das zivilrechtliche Verschulden des Automobilisten : dogmatisch Bedeutung, Verschuldensbegriff, Sorgfaltspflichten*, Zurich 1945, p. 101 ; A. TANNER, *Die Haftung des Motorfahrzeughalters : ein Beitrag zur Frage der Kausalhaftung*, Berne 1936, p. 17.

les véhicules automobiles[16]. La jurisprudence a alors appliqué le principe du *Gefahrensatz* en considérant le conducteur d'une voiture comme fautif du seul fait d'avoir créé un état de fait dangereux en introduisant un véhicule automobile dans la circulation[17]. La jurisprudence a également développé, à partir de l'art. 55 CO, quasiment une responsabilité pour risque de l'employeur pour les accidents causés par ses chauffeurs[18]. En effet, l'appréciation de la diligence de l'employeur faisant appel à des chauffeurs était si sévère qu'il devenait pratiquement impossible d'apporter la preuve libératoire de l'art. 55 CO. La jurisprudence a ainsi montré la voie au législateur qui a fini par adopter une responsabilité pour risque, que l'on retrouve aujourd'hui à l'art. 58 LCR[19].

Une évolution similaire a eu lieu dans le cadre de la responsabilité du fait des produits. Avec l'industrialisation, les produits mis sur le marché (par ex. appareils électriques, machines motorisées, produits chimiques, médicaments) incarnèrent une nouvelle source de danger pour les consommateurs[20]. L'art. 55 CO a servi d'alibi pour appréhender ce nouveau risque. Dans un arrêt de 1984 consacré à un ouvrier qui avait été grièvement blessé suite à la rupture de l'anneau de suspension d'une dalle en béton préfabriquée, le Tribunal fédéral a objectivé à l'extrême le devoir de diligence de l'employeur, fabricant du produit (art. 55 CO)[21]. Pour prouver le respect de son devoir de diligence, l'employeur doit démontrer la mise en place d'un contrôle final des produits ou, si celui-ci est impossible, un mode de construction qui exclut, avec un haut degré de vraisemblance, les erreurs de fabrication[22]. La libération du producteur est ainsi devenue pratiquement impossible, car la survenance d'un dommage constitue la preuve de l'insuffisance du contrôle final et donc de la violation du

16 FF 1930 II, p. 893 ; GRAF (n. 15), p. 102 s. ; R. GREC, *La situation juridique du détenteur de véhicule automobile en cas de collision de responsabilités*, Lausanne 1969, p. 13 s.

17 GREC (n. 16), p. 15 ; TANNER (n. 15), p. 19.

18 ATF 58 II 29, JdT 1932 I 359.

19 Loi fédérale sur la circulation routière du 19 décembre 1958 (LCR ; RS 741.01). Voir OFTINGER/STARK (n. 4), § 20 N 2 ; WIDMER/WESSNER (n. 4), p. 5.

20 F. GILLARD, « Vers l'unification du droit de la responsabilité », *RDS* 1976 II, p. 193, p. 300 ; REY/WILDHABER (n. 9), N 1407 ; WERRO (n. 14), N 578.

21 ATF 110 II 456, JdT 1985 I 378; ETIER (n. 6), p. 43 ; A.-C. HAHN, « L'ébauche d'un droit européen de la responsabilité civile. Quelques réflexions comparatistes sur les fondements en matière de services et de produits », *in* F. WERRO (édit.), *L'européanisation du droit privé. Vers un Code civil européen ?*, Fribourg 1998, p. 397 ; S. MARCHAND, « Les fondamentaux de la responsabilité du fait des produits. Exposé introductif, sources, for et droit applicable », *in* CH. CHAPPUIS/B. WINIGER (édit.), *La responsabilité du fait des produits, Journée du droit de la responsabilité civile 2016*, Genève/Zurich/Bâle 2018, p. 11 ; WERRO (n. 14), N 578 ; P. WESSNER, « Quelques propos erratiques sur des questions liées à la responsabilité du fait des produits défectueux », *in* CH. CHAPPUIS/B. WINIGER (édit.), *Responsabilités objectives, Journée du droit de la responsabilité civile 2002*, Genève/Zurich/Bâle 2003, p. 61, p. 64.

22 ATF 110 II 456, consid. 2 et 3, JdT 1985 I 378.

devoir de diligence. Cette solution équivaut presque à une responsabilité pour risque du fait du produit[23]. Le législateur suisse a finalement adopté, en 1993, la Loi fédérale sur la responsabilité du fait des produits[24] (LRFP) qui consacre un chef de responsabilité basé sur le défaut du produit (art. 1 LRFP), avec la possibilité pour le producteur d'invoquer certaines causes d'exonération (art. 5 LRFP).

3. *Le système de responsabilité civile centré sur le fait d'une chose*

Depuis la révolution industrielle du XIX^e siècle, le législateur a adopté, pour les technologies qui présentent un risque spécifique, des chefs de responsabilité spéciaux. Cette solution correspond à l'étape finale de l'objectivation du droit de la responsabilité civile. Ces chefs de responsabilité (art. 27 LIE, art. 33 LITC, art. 27 LExpl, art. 39 LRaP, art. 59a LPE, art. 59a^bis LPE, art. 30 LGG, art. 14 LOA, art. 58 LCR, art. 15 LTro, art. 64 LA, art. 40b LCdF, art. 30 LNI, art. 20 LICa, LRFP) se réfèrent tous à une chose ou à un ensemble de choses[25]. Ainsi, dans le domaine des technologies, le système est peu à peu passé d'une responsabilité centrée sur le fait d'une personne (art. 41 CO, art. 55 CO) à une responsabilité centrée sur le fait d'une chose.

23 ETIER (n. 6), p. 43 ; FELLMANN/KOTTMANN (n. 8), N 814 ; H.-J. HESS, *Produktehaftpflichtgesetz (PrHG)*, 3^e éd., Berne 2016, art. 1 N 22 ; WIDMER/WESSNER (n. 4), p. 15.

24 Loi fédérale sur la responsabilité du fait des produits du 18 juin 1993 (LRFP ; RS 221.112.944).

25 Loi fédérale concernant les installations électriques à faible et à fort courant du 24 juin 1902 (LIE ; RS 734.0) ; Loi fédérale sur les installations de transport par conduites de combustibles ou de carburants liquides ou gazeux du 4 octobre 1963 (LITC ; RS 746.1) ; Loi fédérale sur les substances explosibles du 25 mars 1977 (LExpl ; RS 941.41) ; Loi fédérale sur la radioprotection du 22 mars 1991 (LRaP ; RS 814.50) ; Loi fédérale sur la protection de l'environnement du 7 octobre 1983 (LPE ; RS 814.01) ; Loi fédérale sur l'application du génie génétique au domaine non humain du 21 mars 2003 (LGG ; RS 814.91) ; Loi fédérale sur les ouvrages d'accumulation du 1^er octobre 2010 (LOA ; RS 721.101) ; Loi fédérale sur les entreprises de trolleybus du 29 mars 1950 (LTro, RS 744.21) ; Loi fédérale sur l'aviation du 21 décembre 1948 (LA ; RS 748.0) ; Loi fédérale sur les chemins de fer du 20 décembre 1957 (LCdF ; RS 742.101) ; Loi fédérale sur la navigation intérieure du 3 octobre 1975 (LNI ; RS 747.201) ; Loi fédérale sur les installations à câbles transportant des personnes du 23 juin 2006 (LICa ; RS 743.01). Par ailleurs, la Suisse dispose également d'une Loi fédérale sur la responsabilité civile en matière nucléaire du 13 juin 2008 (LRCN ; RS 732.44), mais cette loi reprend le système prévu par la Convention du 29 juillet 1960 sur la responsabilité civile dans le domaine de l'énergie nucléaire. La responsabilité en matière nucléaire suit un régime international et sort du cadre de cet article.

L'évolution s'est faite au fur et à mesure de chaque nouveau progrès technique, avec comme résultat une foisonnance de chefs de responsabilité éparpillés dans diverses lois, ce qui engendre deux inconvénients majeurs.

D'une part, il y a un manque d'homogénéité dû à l'adoption de règles spécifiques pour chaque risque particulier. Les chefs de responsabilité prévoient des solutions différentes à des problèmes *a priori* identiques[26]. Par exemple, certains chefs de responsabilité prévoient que seuls les préjudices corporels et matériels (par ex. art. 27 LIE, art. 33 LITC, art. 58 LCR) sont réparables, tandis que d'autres restent silencieux quant aux préjudices réparables (par ex. art. 27 LExpl, art. 59a LPE, art. 59a^bis LPE, art. 30 LGG). On trouve également divers termes pour désigner la personne du responsable : le détenteur (par ex. art. 58 LCR, art. 59a LPE), l'exploitant (par ex. art. 27 LIE, art. 27 LExpl), l'entreprise (par ex. art. 33 LITC, art. 30a LNI), le producteur (LRFP), etc. Il y a donc une nécessité d'uniformiser le système.

D'autre part, on peut également regretter un manque d'égalité. Selon l'approche actuelle, seul le législateur peut imposer un chef de responsabilité objective en adoptant une disposition spéciale. Ce type de responsabilité est considéré comme une exception strictement délimitée à la clause générale de responsabilité pour faute, de telle sorte que la doctrine majoritaire et la jurisprudence refusent de l'étendre par analogie à d'autres hypothèses que celles prévues par les lois spéciales[27]. Partant, des situations présentant un risque comparable sont traitées de manière différente selon que le législateur s'en est occupé ou non[28].

26 ETIER (n. 6), p. 48 ; K. OFTINGER, *Schweizerisches Haftpflichtrecht, Erster Band : Allgemeiner Teil*, Band I, Zurich 1975, p. 7 ; P. WIDMER, « Le visage actuel de la responsabilité civile en droit suisse », *in* O. GUILLOD (édit.), *Développements récents du droit de la responsabilité civile en droit suisse*, Zurich 1991, p. 7, p. 18.

27 W. FELLMANN, *Schweizerisches Haftpflichtrecht, Band II : Haftung nach der gewöhnlichen Kausalhaftung des StSG und den Gefährdungshaftungen des SVG, des Transportrechts (TrG, EBG, BG Anschlussgleise, BSG und SebG) sowie des LFG*, Berne 2013, N 146 et N 150 ; SH. GRÜNIG, *Die Haftung nach Humanforschungsgesetz, Zugleich eine Untersuchung zum Recht der Gefährdungshaftung*, Zurich/Bâle/Genève 2020, N 418 ; REY/WILDHABER (n. 9), N 96 ; WERRO (n. 14), N 34 ; WIDMER/WESSNER (n. 4), p. 137 ; *contra* : H. HONSELL/B. ISENRING/M. KESSLER, *Schweizerisches Haftpflichtrecht*, 5^e éd., Zurich/Bâle/Genève, § 1 N 22, selon lesquels l'analogie devrait être possible.

28 E. BÜYÜKSAGIS, « De l'opportunité de préciser la portée d'une éventuelle clause générale de responsabilité pour risque », *REAS* 2016, p. 2, p. 2 ; ETIER (n. 6), p. 47 ; B. SCHÖNENBERGER, « Generalklausel für die Gefährdungshaftung – ein sinnvolles Reformorhaben ? », *in* Th. SUTTER-SOMM/F. HAFNER/G. SCHMID/K. SEELMANN (édit.), *Risiko und Recht, Festgabe zum schweizerischen Juristentag 2004*, Berne/Bâle/Genève/Munich 2004, p. 171, p. 177 ; S. WEBER, « OR 2020 – Das neue Deliktsrecht – Doch noch eine Revision des Haftpflichtrechts – erste Eindrücke zu den Vorschlägen im Entwurf OR 2020 », *REAS* 2013, p. 357, p. 357 ; WIDMER (n. 26), p. 18.

Ce système induit également un manque de flexibilité face aux nouvelles technologies, puisqu'il est impossible pour le législateur de suivre toutes les nouvelles situations à risque entraînées par l'évolution de la technique et d'adapter continuellement la législation en la matière[29]. Ce problème ouvre la question d'une clause générale de responsabilité pour les choses dangereuses, c'est-à-dire d'une clause de responsabilité qui serait suffisamment large et souple pour permettre au juge de traiter de manière égale des situations similaires à celles qui, dans le droit en vigueur, font déjà l'objet d'un chef de responsabilité spécial[30].

La nécessité d'une révision du système a été relevée depuis longtemps, mais n'a jamais abouti. La Société suisse des juristes avait constaté le manque de cohérence du système en 1967 déjà et avait alors adopté une résolution où elle exprimait le souhait que soit unifié le droit de la responsabilité civile[31]. Cette résolution avait donné lieu à l'élaboration du projet WIDMER/WESSNER qui proposait une révision totale du droit de la responsabilité civile. Ce projet, qui prévoyait notamment une clause générale de responsabilité pour les activités risquées[32], a finalement été abandonné en 2009[33]. L'idée d'une clause générale de responsabilité pour risque a à nouveau été proposée dans le projet CO 2020[34]. Ce projet n'a jamais abouti, car le Conseil fédéral a jugé que la nécessité d'une révision totale de la partie générale du Code des obligations n'était pas avérée[35].

Malgré l'échec de ces projets, l'unification et la simplification du domaine de la responsabilité civile restent une nécessité pour la prévisibilité et la sécurité du droit. Le bouleversement engendré par l'IA est l'occasion de rouvrir la question d'une unification de la responsabilité civile et de l'introduction d'une clause générale de responsabilité pour risque. Si l'examen complet de la possibilité d'unifier le système de la responsabilité civile et d'introduire une clause

29 SCHÖNENBERGER (n. 28), p. 177 ; F. WERRO/Ch. MÜLLER, « CO 2020 – Le nouveau droit de la responsabilité civile – Introduction », *REAS* 2013, p. 352, p. 353 ; WIDMER (n. 26), p. 18.

30 WIDMER/WESSNER (n. 4), p. 138.

31 SOCIÉTÉ SUISSE DES JURISTES, « Procès-verbal de la 101e assemblée annuelle de la société suisse des juristes, des 22, 23 et 24 septembre 1967 », *RDS* 1967 II, p. 645, p. 819 ; P. TERCIER, « La réforme du droit de la responsabilité civile : Chronique d'une mort annoncée », *in* CH. CHAPPUIS/B. FOËX/L. THÉVENOZ (édit.), *Le législateur et le droit privé, Colloque en l'honneur du Professeur Gilles Petitpierre*, Genève/Zurich/Bâle 2006, p. 25, p. 28 ; WIDMER/WESSNER (n. 4), p. 8.

32 WIDMER/WESSNER (n. 4), p. 134 ss.

33 www.bj.admin.ch/ejpd/fr/home/actualite/news/2009/2009-01-21.html.

34 W. FELLMANN/Ch. MÜLLER/F. WERRO, « Art. 60 CO 2020 », *in* C. HUGUENIN/R. HILTY (édit.), *Code des obligations suisse 2020, Projet relatif à une nouvelle partie générale*, Zurich 2013, N 1 ss.

35 CONSEIL FÉDÉRAL, *Modernisation de la partie générale du Code des obligations, Rapport en réponse aux Postulats 13.3217 Bischof et 13.3226 Caroni*, du 31 janvier 2018, p. 14.

générale de responsabilité pour risque dépasse le cadre de cet article, on peut néanmoins essayer de dégager une certaine systématique dans le système actuel et d'analyser comment l'intelligence artificielle peut s'y insérer.

B. Principes fondamentaux de la responsabilité centrée sur le fait d'une chose

La responsabilité centrée sur le fait d'une chose fonde l'obligation de réparer le préjudice sur une caractéristique particulière d'une chose. On peut distinguer la responsabilité du fait du produit, où il est question du défaut de la chose, et les autres chefs de responsabilité, où il est question du risque de la chose.

1. La responsabilité fondée sur le défaut

La responsabilité du fait du produit est réglée dans la LRFP. Le défaut y est défini comme un manque de sécurité du produit pour les personnes et les choses compte tenu de l'usage auquel il est destiné (art. 4 al. 1 lit. b LRFP)[36].

Le défaut doit exister au moment de la mise en circulation du produit (art. 4 al. 1 lit. c et 5 al. 1 lit. b LRFP), soit au moment où le producteur a abandonné volontairement la maîtrise de fait sur le produit[37]. L'existence du défaut entraine la présomption que celui-ci existait au moment de la mise en circulation, à charge pour le responsable d'apporter la preuve du contraire (art. 5 al. 1 lit. b LRFP). Partant, la LRFP ne vise pas les défauts survenus postérieurement à la mise sur le marché du produit, notamment le défaut de surveillance qui correspond à une violation par le producteur de son obligation de surveiller certains produits mis sur le marché (art. 8 LSPro)[38].

36 ATF 137 III 226, consid. 3.2 ; R. BÜHLER, « Definition des Produktfehlers im Produktehaftpflichtgesetz (PrHG) », *PJA* 1993, p. 1425, p. 1435 ; W. FELLMANN/G. VON BÜREN-VON MOOS, *Grundriss der Produktehaftpflicht*, Berne 1993, N 183 ; FELLMANN/KOTTMANN (n. 8), N 1140 ; HESS (n. 23), art. 4 N 18 ; E. HOLLIGER-HAGMANN, « Art. 2 PrHG, Art. 4 PrHG, Art. 5 PrHG », *in* W. FISCHER/T. LUTERBACHER (édit.), *Haftpflichtkommentar, Kommentar zu den schweizerischen Haftpflichtbestimmungen*, Zurich/Saint-Gall 2016, art. 4 PrHG N 7 ; WERRO (n. 14), N 631.

37 A. BORSARI, *Schadensabwälzung nach dem schweizerischen Produkthaftpflichtgesetz (PrHG)*, Zurich 1998, p. 169 ; FELLMANN/KOTTMANN (n. 8), N 1181 ; HESS (n. 23), art. 5 N 3 ; HOLLIGER-HAGMANN (n. 36), art. 5 PrHG N 16 ; REY/WILDHABER (n. 9), N 1462 ; WERRO (n. 14), N 690.

38 Loi fédérale sur la sécurité des produits du 12 juin 2009 (LSPro ; RS 930.11). FELLMANN/VON BÜREN-VON MOOS (n. 36), N 345 ; REY/WILDHABER (n. 9), N 1473 ; WERRO (n. 14), N 711.

La LRFP exclut également les risques de développement, à savoir les « risques imprévisibles, non identifiables lors de la mise en circulation du produit compte tenu de l'état des connaissances scientifiques et techniques »[39] (art. 5 al. 1 lit. e LRFP). C'est au producteur de démontrer que, dans un cas d'espèce, le manque de sécurité se rapporte à un risque de développement.

Il appartient au lésé d'apporter la preuve du défaut (art. 8 CC)[40]. Ce dernier ne doit pas démontrer la cause du défaut, mais seulement le manque de sécurité[41]. Lorsqu'une preuve stricte n'est pas possible, le degré de preuve doit se limiter à la vraisemblance prépondérante[42]. À cet égard, le Tribunal fédéral a précisé que la démonstration selon laquelle l'utilisation du produit a joué un rôle dans le déroulement des faits qui ont entraîné le préjudice constitue un indice significatif de l'existence d'un défaut en vertu de l'adage *res ipsa loquitur*[43].

Puisque le défaut est défini comme un manque de sécurité pour les personnes et les choses, les préjudices réparables sont les dommages corporels et matériels (art. 1 LRFP) et le tort moral (art. 11 LRFP et art. 47 et 49 CO). L'art. 1 al. 1 lit. b LRFP prévoit une restriction quant à la réparation du dommage matériel : seul le dommage causé à une chose d'un type qui la destine habituellement à l'usage ou à la consommation privés (critère objectif) et qui a été principalement destinée à des fins privées (critère subjectif) est réparable. La réparation des dommages causés aux biens utilisés à des fins professionnelles ou commerciales est ainsi exclue[44]. En outre, l'art. 6 al. 1 LRFP prévoit que le lésé doit supporter le dommage matériel jusqu'à concurrence de 900 francs.

L'art. 1 LRFP désigne le producteur comme sujet de la responsabilité, car il s'agit de la personne à l'origine du défaut. Le producteur a la maîtrise sur la conception du produit et peut influencer les caractéristiques de sécurité de

39 ATF 137 III 226, consid. 4.1.

40 Code civil suisse du 10 décembre 1907 (CC ; RS 210). Voir ATF 137 III 226, consid. 3.2 ; ATF 133 III 81, consid. 4.2.2 ; BORSARI (n. 37), p. 196 ; FELLMANN/KOTTMANN (n. 8), N 1174 ; HESS (n. 23), art. 1 N 122 ; HOLLIGER-HAGMANN (n. 36), art. 4 PrHG N 13 ; REY/WILDHABER (n. 9), N 1446 ; WERRO (n. 14), N 652 ; WESSNER (n. 21), p. 68.

41 ATF 133 III 81, consid. 4.1 ; FELLMANN/KOTTMANN (n. 8), N 1175 ; HOLLIGER-HAGMANN (n. 36), art. 4 PrHG N 15 ; REY/WILDHABER (n. 9), N 1446 ; WERRO (n. 14), N 652.

42 ATF 137 III 226, consid. 3.2 ; ATF 133 III 81, consid. 4.2.2 ; FELLMANN/KOTTMANN (n. 8), N 1177 ; HESS (n. 23), art. 1 N 127 ; REY/WILDHABER (n. 9), N 1446 ; WERRO (n. 14), N 652.

43 ATF 133 III 81, consid. 3.3 ; HESS (n. 23), art. 1 N 133 ; WESSNER (n. 21), p. 68 ; A. MORIN, « Intelligence artificielle et transports, Rapport suisse », *in* O. GOUT (édit.) Responsabilité civile et intelligence artificielle, Recueil des travaux du Groupe de Recherche Européen sur la Responsabilité civile et l'Assurance (GRERCA), Bruxelles 2022, N 64.

44 BORSARI (n. 37), p. 97 ; FELLMANN/KOTTMANN (n. 8), N 1114 ; WERRO (n. 14), N 585.

celui-ci[45]. La LRFP désigne comme producteur toutes les personnes qui participent au processus de production (art. 2 LRFP)[46], à savoir le fabricant du produit fini, d'une matière première ou d'une partie composante (art. 2 al. 1 lit. a LRFP). Il assimile également au producteur « toute personne qui se présente comme producteur en apposant sur le produit son nom, sa marque ou un autre signe distinctif » (art. 2 al. 1 lit. b LRFP) et l'importateur (art. 2 al. 1 lit. c LRFP). Le fournisseur peut également répondre à titre subsidiaire (art. 2 al. 2 et 3 LRFP). Afin d'assurer une juste répartition des risques de la production moderne, la LRFP permet au producteur de se libérer s'il démontre l'existence de certaines circonstances qui l'exonèrent, comme les risques de développement (art. 5 LRFP)[47].

2. *La responsabilité fondée sur le risque*

Parmi les différents chefs de responsabilité pour risque, il est possible de dégager quelques principes généraux, notamment en ce qui concerne le motif spécial justifiant la responsabilité, le préjudice réparable et les critères de rattachement à la personne responsable.

Le motif spécial justifiant la responsabilité consiste dans le risque que présente une chose en elle-même[48]. La chose, en soi parfaite, doit générer un risque important pour les personnes et les choses[49], procédant de la fréquence élevée et/ou de la gravité particulière des préjudices qu'elle peut causer[50]. En outre, ce risque de préjudice doit être inévitable en ce sens que même le comportement le plus diligent ne peut l'écarter avec une garantie suffisante[51]. On trouve ce risque dans les installations qui traitent de substances dangereuses, par exemple les installations électriques (art. 27 LIE) ou les installations où sont fabriqués, entreposés ou utilisés des matières explosives ou des engins pyrotechniques (art. 27 LExpl). Le risque spécifique provient alors des caractéristiques dangereuses des substances traitées dans ces installations, telles l'inflammabilité, la

45 HOLLIGER-HAGMANN (n. 36), art. 2 PrHG N 10.

46 Considérant 4 de la Directive 85/374/CE.

47 Considérant 7 de la Directive 85/374/CE.

48 BÜYÜKSAGIS (n. 28), p. 2 ; SCHÖNENBERGER (n. 28), p. 181 ss ; WIDMER/WESSNER (n. 4), p. 140 s.

49 FELLMANN/KOTTMANN (n. 8), N 29 ; HONSELL/ISENRING/KESSLER (n. 27), § 1 N 19 ; M. KELLER/S. GABI/K. GABI, *Haftpflichtrecht*, 3e éd., Bâle 2012, p. 5 s. ; K. OFTINGER/E. STARK, *Schweizerisches Haftpflichtrecht, Besonderer Teil*, Band II/2, Zurich 1989, § 24 N 6 ; REY/WILDHABER (n. 9), N 99 ; WERRO (n. 9), N 32.

50 FELLMANN/KOTTMANN (n. 8), N 38 ; KELLER/GABI/GABI (n. 49), p. 6 ; OFTINGER/STARK (n. 49), § 24 note 5 ; REY/WILDHABER (n. 9), N 103 ; WERRO (n. 14), N 33.

51 REY/WILDHABER (n. 9), N 99 ; WERRO (n. 14), N 32.

toxicité, l'explosivité, etc. Les moyens de transport, comme les véhicules automobiles (art. 58 LCR), les aéronefs (art. 64 LA) ou les chemins de fer (art. 40b LCdF) présentent également un risque spécifique provenant essentiellement de la locomotion, c'est-à-dire du déplacement rapide de masses relativement importantes à l'aide de forces qui se développent dans l'engin lui-même[52]. Enfin, les organismes pathogènes (art. 59a^{bis} LPE) et les organismes génétiquement modifiés (art. 30 LGG) sont également soumis à une responsabilité pour risque au vu de leur capacité à induire une maladie (art. 5 al. 5^{quater} LPE), respectivement de leurs propriétés nouvelles issues de la modification génétique, de l'instabilité du matériel génétique qui est susceptible de subir des changements ultérieurs et du risque de transmission du matériel génétique modifié à d'autres organismes (art. 30 al. 7 LGG)[53].

Le fait générateur de responsabilité correspond à la réalisation du risque. Par exemple, le risque des installations traitant de substances dangereuses se matérialise lorsque les caractéristiques spécifiques des substances traitées sont à l'origine d'un préjudice. Ainsi, le risque d'une ligne électrique à haute tension (art. 27 LIE) se concrétise lorsque le conducteur d'un camion se fait électrocuter parce que son véhicule se trouve en contact avec ladite ligne[54]. En revanche, tel n'est pas le cas si une personne chute en escaladant le poteau d'une ligne à haute tension. De la même façon, le risque des moyens de transport se réalise lorsque la locomotion provoque un préjudice, typiquement en cas de collision entre un moyen de transport et une personne ou une chose[55]. Il n'y a par contre pas de réalisation du risque lorsque le passager d'un véhicule se blesse en se coinçant les doigts dans une portière[56].

Comme exposé ci-dessus, certains chefs de responsabilité pour risque prévoient que seuls les dommages corporels et matériels sont réparables. Cela découle du

52 ATF 85 II 516, consid. 3a ; ATF 81 II 558, consid. 1 ; ATF 72 II 217, consid. 2b ; FF 2007, p. 4251 ; H. DESCHENAUX/P. TERCIER, *La responsabilité civile*, 2^{e} éd., Berne 1982, § 15 N 102 ; FELLMANN (n. 27), N 1387.

53 FF 2000, p. 2305. W. FELLMANN, *Schweizerisches Haftpflichtrecht, Band III : Haftung nach den Gefärdungshaftungen des JSG, HFG, USG, GTG, EleG, RLG, SprstG, StAG und KHG 2008*, Berne 2015, N 1032 ; CH. HEDIGER, *Die Haftungsbestimmungen des Gentechnikgesetzes (Art. 30-34 GTG), Beurteilung und Vergleich mit der Haftungsregelung des deutschen Gentechnikgeseztes*, Zurich/Bâle/Genève 2009, p. 111 s.

54 ATF 95 II 411.

55 FF 2007, p. 4260 ; R. BREHM, *La responsabilité civile automobile*, 2^{e} éd., Berne 2010, N 178 ; FELLMANN (n. 27), N 366 et N 1247 ; GREC (n. 16), p. 36 ; B. KLETT/M. DE MEURON, « La responsabilité des chemins de fer », *in* F. WERRO/P. PICHONNAZ (édit.), *Les responsabilités fondées sur le risque, Colloque du droit de la responsabilité civile 2017*, Berne 2018, p. 41, p. 58 ; R. KÖNIG, *Die Gefährdungshaftung nach Eisenbahngesetz : Analyse und Kritik der neuen Haftungsregeln*, Zurich/Bâle/Genève 2012, N 65 ; WERRO (n. 14), N 964.

56 ATF 107 II 269, consid. 1a et 2, JdT 1981 I 446.

motif justifiant la responsabilité objective aggravée, à savoir un risque important pour les personnes et les choses. Le tort moral est également réparable[57].

Les différents chefs de responsabilité utilisent divers termes pour désigner le sujet de responsabilité. En réalité, ces notions expriment généralement toutes le même lien subjectif. Conformément au principe de l'intérêt ou de l'utilité, il s'agit de la personne qui tire avantage de l'exploitation de la chose et qui a le pouvoir de disposition direct sur celle-ci[58]. En effet, si on tolère l'utilisation de choses spécifiquement dangereuses, le risque en émanant doit toutefois être à la charge de celui qui en profite[59]. En outre, l'exploitant doit en principe conclure une assurance responsabilité civile obligatoire, ce qui a pour effet de protéger le lésé contre le risque d'insolvabilité de ce dernier (voir art. 35 LITC, art. 63 LCR, art. 70 LA, art. 31 LNI, art. 21 LICa)[60].

III. L'intelligence artificielle

A. Le risque de l'intelligence artificielle

1. Le risque propre à l'intelligence artificielle

L'intelligence artificielle est un terme générique qui englobe les technologies capables d'agir de manière autonome en simulant, à l'aide de méthodes statistiques, certaines facultés propres à l'intelligence humaine dans le but d'atteindre un certain objectif (art. 3 lit. a de la Proposition de Règlement sur l'intelligence artificielle [abrégé ci-après « P-Règ »])[61]. Ces technologies font

57 BORSARI (n. 37), p. 211 s. ; FELLMANN/KOTTMANN (n. 8), N 1124 ; HESS (n. 23), art. 1 N 65 ss ; REY/WILDHABER (n. 9), N 1417 ; SCHWENZER/FOUNTOULAKIS (n. 9), N 53.40 ; *contra* : FELLMANN/VON BÜREN-VON MOOS (n. 36), N 108 ; HONSELL/ISENRING/KESSLER (n. 27), § 21 N 52 ; A. KELLER, *Haftpflicht im Privatrecht*, Band I, 6^{e} éd., Berne 2002, p. 362 ; WERRO (n. 14), N 600.

58 ATF 131 III 61, consid. 2.3 ; ATF 129 III 102, consid. 2.2, JdT 2003 I 500 ; HONSELL/ISENRING/KESSLER (n. 27), § 1 N 22 ; WIDMER/WESSNER (n. 4), p. 147.

59 ATF 131 III 61, consid. 2.3 ; ATF 129 III 102, consid. 2.2, JdT 2003 I 500 ; HONSELL/ISENRING/KESSLER (n. 27), § 1 N 22 ; PICHONNAZ/WERRO (n. 9), p. 8 s. ; WERRO (n. 14), N 34.

60 MORIN (n. 43), N 72 ; WERRO (n. 14), N 985.

61 Un régime de responsabilité civile pour l'intelligence artificielle, Résolution du Parlement européen du 20 octobre 2020 concernant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle (2020/2014(INL)). M. F. LOHMANN, « Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz », *REAS* 2021, p. 111, p. 111 ; I. WILDHABER, « Eine Einführung in die aus-servertragliche Haftung für Künsltiche Intelligenz (KI) », *in* W. FELLMANN (édit.), *Haftpflichtprozess 2021, Haftung für Künstliche Intelligenz, (teil-)automatisierte Fahr-*

appel à des algorithmes destinés à analyser d'immenses quantités de données, à reconnaître parmi celles-ci des tendances et des schémas pour en construire des modèles à l'origine de prédictions servant de base à leurs décisions[62]. L'une des méthodes centrales permettant le développement de logiciels d'IA est l'utilisation d'algorithmes dits de *machine learning* (autoapprentissage), qui peuvent apprendre par eux-mêmes à partir de données ou d'expériences[63].

Le risque principal de l'IA provient de l'impossibilité à prédire ses comportements. Cela découle déjà de son autonomie qui lui confère la faculté de fonctionner dans diverses situations qui ne sont pas entièrement prédéfinies[64]. A cela s'ajoute le fait que le processus de prise de décision résultant de l'autoapprentissage est quasiment impossible à saisir pour les humains (effet boîte noire)[65].

L'IA présente en outre certaines faiblesses. Déjà, la technologie de l'IA est foncièrement complexe vu la logique interne des algorithmes et la multiplicité des composants (capteurs, senseurs, logiciels, fonction de connectivité, etc.)[66],

zeuge, Drohnen und Software, Entwicklungen im Dientsleistungs-, Privatversicherungs-, Prozess-, Staatshaftungs- und Haftpflichtrecht, Zurich/Bâle/Genève 2021, p. 15, p. 16.

62 GROUPE DE TRAVAIL INTERDÉPARTEMENTAL « INTELLIGENCE ARTIFICIELLE » (n. 1), p. 7 ; LOHMANN (n. 61), p. 111.

63 D. LINARDATOS, « Künstliche Intelligenz und Verantwortung », *ZIP* 11/2019, p. 504, p. 504 s. ; PICHONNAZ (n. 13), p. 516.

64 COMMISSION EUROPÉENNE, *Rapport sur les conséquences de l'intelligence artificielle, de l'internet des objets et de la robotique sur la sécurité et la personnalité*, du 19 février 2020, COM(2020)64 final, p. 7 s. ; ELI, *Response, European Commission's Public Consultation on Civil Liability, Adapting Liability Rules to the Digital Age and Artificial Intelligence, Publication European Law Institute 2022*, p. 25 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging Digitaltechnologies, publication de la Commission européenne*, 2019, p. 33 ; W. FELLMANN, « Haftpflichtrecht im Zeichen der Digitalisierung », *REAS* 2021, p. 105, p. 108 ; LOHMANN, (n. 61), p. 112 ; PICHONNAZ (n. 13), p. 518 ; G. WAGNER, « Verantwortlichkeit im Zeichen digitaler Techniken », *VersS* 2020, p. 717, p. 724 ; CH. WENDEHORST, « Strict Liability for AI and other emerging Technologies », *JETL* 2020/2, p. 150, p. 151 ; WILDHABER (n. 61), p. 22 s.

65 COMMISSION EUROPÉENNE (n. 64), p. 2 et p. 10 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 33 ; GROUPE DE TRAVAIL INTERDÉPARTEMENTAL « INTELLIGENCE ARTIFICIELLE » (n. 1), p. 24 ; M. BUITEN/A. DE STRELL/M. PEITZ, *EU Liability Rules for The Age of Artificial Intelligence, Report*, Centre on Regulation in Europe, Mars 2021, p. 27 ; ELI (n. 64), p. 25 ; FELLMANN (n. 64), p. 108 ; LOHMANN (n. 61), p. 117 ; PICHONNAZ (n. 13), p. 518 ; WENDEHORST (n. 64), p. 152 ; WILDHABER (n. 61), p. 20.

66 COMMISSION EUROPÉENNE (n. 64), p. 13 ; ELI (n. 64), p. 24 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 32 s. ; Y. BENHAMOU/J. FERLAND, « Artificial Intelligence and damages : Accessing Liability and Calculating Damages », *in* G. D'AGOSTINO/A. GAON/C. PIOVESAN (édit.), *Leading Legal Disruption : Artificial*

ce qui engendre un risque important d'erreurs dans la production d'un système d'IA et l'impossibilité d'identifier, en cas de dysfonctionnement, quelle est la source de l'erreur et à qui elle est imputable[67]. Ensuite, le développement et le fonctionnement de l'IA sont dépendants des données (données d'apprentissage, données collectées, etc.), de telle manière que des données insuffisantes, erronées ou biaisées compromettent fortement la qualité des résultats obtenus[68]. De surcroît, l'IA obtient souvent ces données grâce à l'interconnexion avec des sources de données ou avec d'autres systèmes d'IA. Un logiciel d'IA doit donc pouvoir s'adapter aux mises à jour ou aux mises à niveau prévues pour les systèmes avec lesquels il interagit[69], ce qui exige une collaboration entre les producteurs et les utilisateurs pour que le produit puisse évoluer et fonctionner correctement[70]. L'interconnexion engendre en sus un risque de cyberattaque[71].

2. *Une responsabilité pour risque ou pour défaut*

Certains auteurs considèrent que le risque spécifique de l'IA en fait une chose spécifiquement dangereuse[72]. On constate toutefois assez vite qu'on ne peut pas qualifier de manière générale tous les systèmes d'IA comme tel, par exemple un système d'IA destiné à planifier des rendez-vous n'engendre pas un danger important[73]. C'est pour cette raison que le Parlement européen proposait de soumettre uniquement les systèmes d'IA à haut risque à une responsabilité objective (art. 4 P-Règ.).

Partant, seuls les systèmes d'IA présentant un risque similaire à celui à la base des responsabilités pour risque peuvent constituer des choses spécifiquement

Intelligence and a Toolkit for Lawyers and the Law, Montréal 2021, p. 165, p. 170 ; FELLMANN (n. 64), p. 109 ; WENDEHORST (n. 64), p. 152 s. ; WILDHABER (n. 61), p. 19.

67 BUITEN/DE STREEL/PEITZ (n. 65), p. 26 ; ELI (n. 64), p. 24 s. ; FELLMANN (n. 64), p. 109 ; WENDEHORST (n. 64), p. 152.

68 COMMISSION EUROPÉENNE, *Livre blanc, intelligence artificielle, Une approche européenne axée sur l'excellence et la confiance*, du 19 février 2020, COM (2020)65 final, p. 14 ; COMMISSION EUROPÉENNE (n. 64), p. 10 ; ELI (n. 64), p. 26 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 33 ; FELLMANN (n. 64), p. 106.

69 EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 33.

70 ELI (n. 64), p. 26.

71 BUITEN/DE STREEL/PEITZ (n. 65), p. 26 ; ELI (n. 64), p. 26 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 33 s. ; WENDEHORST (n. 64), p. 153 ; WILDHABER (n. 61), p. 20.

72 E. BÜYÜKSAGIS, « Responsabilité pour les systèmes d'intelligence artificielle », *REAS* 2021, p. 12, p. 20 ss ; WENDEHORST (n. 64), p. 160.

73 BUITEN/DE STREEL/PEITZ (n. 65), p. 60 ; ELI (n. 64), p. 28 ; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 40 ; PICHONNAZ (n. 13), p. 520 ; WILDHABER (n. 61), p. 59.

dangereuses au sens du droit de la responsabilité civile[74]. Pour cela, les systèmes d'IA doivent exposer les personnes et/ou les choses à un risque particulier. Cette solution se rapproche fortement de la notion de haut risque consacrée à l'art. 3 lit. c P-Règ, ce type de risque étant défini comme « un risque important, dans un système d'IA opérant de manière autonome, de causer un préjudice ou un dommage à une ou plusieurs personnes d'une manière aléatoire et qui va au-delà de ce à quoi l'on peut raisonnablement s'attendre ; l'importance de ce risque dépend de l'interaction entre la gravité de l'éventuel préjudice ou dommage, le degré d'autonomie de décision, la probabilité que le risque se réalise, la manière dont le système d'IA est utilisé et le contexte d'utilisation ». Toutefois, le Parlement européen ne se limitait pas à la protection de l'intégrité corporelle et matérielle des personnes, mais visait également la protection de leur patrimoine en tant que tel (art. 2 al. 1 P-Règ)[75].

Deux critères, qui s'entremêlent, permettent de déterminer si un système d'IA présente un risque important pour les personnes et les choses : le caractère imprédictible du système d'IA et l'environnement dans lequel il est introduit. Ces deux critères rappellent les concepts de degré d'autonomie et du type d'utilisation (manière et contexte) auxquels se réfère la P-Règ pour définir les systèmes d'IA à haut risque.

Le caractère imprédictible dépend de la capacité d'apprentissage et de prise de décision du système[76]. En effet, la possibilité pour un logiciel d'apprendre par lui-même et d'évoluer au fur et à mesure de son utilisation réduit corollairement le caractère prévisible de ses résultats. De même, plus le système d'IA est en mesure d'agir par lui-même et de prendre divers types de décisions, moins il est possible d'anticiper ses comportements. La structure de l'environnement influe nécessairement sur le degré d'autonomie attendu d'un système d'IA. L'autonomie doit être élevée lorsque le système doit fonctionner dans un environnement non structuré, dans lequel il est tenu de répondre à de multiples besoins[77]. En effet, dans un tel cas, le système d'IA doit disposer d'une grande capacité d'apprentissage et de prise de décision, car il peut être exposé à des situations qui diffèrent fortement de celles auxquelles il a été entrainé. En outre, le risque de préjudice est accru dans les environnements où le système d'IA

74 EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 40.

75 BÜYÜKSAGIS (n. 72), p. 22 s. Dans la proposition de Directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), du 28 septembre 2022, COM/2022/496 final, la proposition d'instaurer un régime de responsabilité objective pour certains systèmes d'IA n'a pas été retenue (p. 17).

76 BÜYÜKSAGIS (n. 72), p. 11 ; WILDHABER (n. 61), p. 153 s.

77 M. F. LOHMANN, « Roboter als Wundertüten – eine zivilrechtliche Haftungsanalyse », *PJA* 2017, p. 152, p. 154 ; WILDHABER (n. 61), p. 59 s.

évolue au côté de personnes, notamment de personnes non-utilisatrices[78]. Les systèmes d'IA qui présentent un risque important sont typiquement les voitures et les drones autonomes ou les robots utilisés dans le cadre médical.

B. L'intelligence artificielle comme produit

Selon l'opinion dominante, le logiciel incorporant l'IA s'assimile à un produit au sens de la LRFP[79]. Toutefois, il se distingue des produits traditionnels par le fait qu'il est impossible pour le fabricant de prédéterminer entièrement son fonctionnement au moment de la mise en circulation[80]. Il convient donc d'analyser comment cette caractéristique est appréhendée par la LRFP.

Un système d'IA est défectueux lorsqu'il ne présente pas la sécurité suffisante pour l'usage auquel il est destiné (art. 4 LRFP)[81]. Le système d'IA doit présenter une sécurité quasi absolue lorsqu'il est susceptible, de par son but, d'occasionner d'importants dommages, par exemple une voiture autonome ou un robot chirurgical[82]. Peu importe à cet égard que le préjudice se rapporte à sa capacité d'apprentissage[83]. Le fait qu'une caractéristique d'un produit soit typique et inévitable n'exclut pas la possibilité de retenir un défaut[84].

Le défaut doit exister au moment de la mise en circulation du produit. La mise en circulation de produits numériques soulève quelques questions, car contrairement aux produits traditionnels, le producteur conserve la possibilité d'influencer l'état du produit *via* des mises à jour ou des mises à niveau[85]. On

78 EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 40 ; LOHMANN (n. 77), p. 154 ; WILDHABER (n. 61), p. 59 s.

79 FF 2021 3026, p. 40 ; ELI (n. 64), p. 14 ; FELLMANN (n. 64), p. 107 ; S. HÄNSENBERGER, *Die zivilrechtliche Haftung für autonome Drohnen unter Einbezug von Zulassungs- und Betriebsvorschriften*, Saint-Gall 2018, p. 120 s. ; LOHMANN (n. 61), p. 115 ; MORIN (n. 43), N 19 ; WILDHABER (n. 61), p. 39 s.

80 LOHMANN (n. 61), p. 114.

81 Voir *supra* II.B.I.

82 N. BRAUN BINDER/T. BURRI/M. F. LOHMANN/M. SIMMLER/F. THOUVENIN/K. N. VOKINGER, « Künstliche Intelligenz : Handlungsbedarf im Schweizer Recht », *Jusletter* du 28 juin 2021, N 44 ; BUITEN/DE STREEL/PEITZ (n. 65), p 53.

83 BRAUN BINDER et *al.* (n. 82), N 44 ; LOHMANN (n. 61), p. 116 ; M. F. LOHMANN/M. MÜLLER-CHEN, « Selbstlernenden Fahrzeuge – eine Haftungsanalyse », *RSDA* 2017, p. 48, p. 55 ; WILDHABER (n. 61), p. 45 ; *contra :* HÄNSENBERGER (n. 79), p. 123 ss ; C. WIDMER LÜCHINGER, « Apps, Algorithmen und Roboter in der Medizin : Haftungsrechtliche Herausforderungen », *REAS* 2019, p. 3, p. 11.

84 BRAUN BINDER et *al.* (n. 82), N 44 ; LOHMANN (n. 61), p. 116 ; WILDHABER (n. 61), p. 45.

85 EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES (n. 64), p. 42 ; LOHMANN (n. 61), p. 118 ; WILDHABER (n. 61), p. 42 ; MORIN (n. 43), N 60. Voir ég. ELI, *Draft of Revised*

retient ainsi que la mise sur le marché d'un produit numérique n'intervient pas en une seule fois, mais se répète à chaque mise à jour ou mise à niveau[86].

Le fait que le logiciel d'IA apprenne un certain comportement préjudiciable après sa mise en circulation ne permet pas d'en conclure que le défaut est survenu postérieurement à sa mise en circulation[87]. En effet, il revient au producteur de choisir la meilleure technologie possible de *machine learning*, de fixer certaines limites à la capacité d'apprentissage et décisionnelle, de tester son produit et de donner les instructions et les avertissements nécessaires à son utilisateur[88]. En revanche, l'apprentissage entrepris par l'utilisateur qui s'écarte totalement des instructions fournies par le fabricant peut constituer une utilisation abusive, excluant la notion de défaut. Tel est par exemple le cas si l'utilisateur entraîne volontairement le système d'IA afin qu'il commette un préjudice[89]. Le producteur ne peut pas non plus invoquer le risque de développement pour la capacité d'apprentissage, car le risque de décisions imprévisibles et préjudiciables des logiciels d'IA est connu[90].

L'interconnexion soulève également quelques interrogations. Tout d'abord, on peut se demander si l'on peut retenir un défaut dans la mesure où le dysfonctionnement de l'IA est dû aux données incorrectes reçues d'autres systèmes. Lorsque le système d'IA interagit avec d'autres systèmes, le producteur est tenu de s'assurer de la fiabilité et de l'exactitude des données traitées par son produit[91]. Dès lors, les erreurs dans les données de sources externes constituent un défaut, à moins que toutes les mesures proportionnées permettant de garantir l'exactitude des données (par ex. certificats établis par des organismes dignes de confiance) aient été prises[92]. Ensuite, les systèmes d'IA qui interagissent avec d'autres systèmes doivent régulièrement être *updatés* ou *upgradés* de manière à pouvoir garantir la meilleure opérabilité entre objets connectés[93]. Le dysfonctionnement d'un système d'IA dû à l'omission d'une telle amélioration

Product Liability Directive, Draft Legislative Proposal of the European Law Institute, 2022, p. 21.

86 LOHMANN (n. 61), p. 116.

87 S. HÄNSENBERGER, « Die Haftung für Produkte mit lernfähigen Algorithmen », *Jusletter* du 26 novembre 2018, N 25.

88 BRAUN BINDER *et al.* (n. 82), N 45 ; LOHMANN (n. 61), p. 116 s. ; D. ROSENTHAL, « Autonome Informatiksysteme : Wie steht es mit der Haftung ? », *in* A. KÜNDING/D. BÜTSCHI (édit.), *Die Verselbständigung des Computers*, TA-SWISS 51/2008, p. 131, p. 135 ; WILDHABER (n. 61), p. 42 et p. 46.

89 LOHMANN (n. 61), p. 118 ; WILDHABER (n. 61), p. 45.

90 LOHMANN (n. 61), p. 119 ; MORIN (n. 43), N 62.

91 FF 2021 3026, p. 42.

92 FF 2021 3026, p. 65 ; voir ég. PICHONNAZ (n. 13), p. 523.

93 Voir *supra* III.A.1.

se heurte à l'absence de défaut de surveillance[94]. Afin de tenir compte des spécificités des systèmes d'IA, la LRFP devrait introduire une obligation de suivi et sanctionner la violation de cette dernière[95].

Enfin, la difficulté de la preuve du défaut pour le lésé (art. 8 CC) peut susciter des inquiétudes au vu de la complexité et de l'opacité du fonctionnement de l'IA[96]. Toutefois, la jurisprudence du Tribunal fédéral répond de manière adéquate à ce problème. D'une part, le lésé n'a pas à démontrer la cause du défaut, il suffit d'établir que le système d'IA ne présente pas le niveau de sécurité attendu[97]. D'autre part, en vertu de l'adage *res ipsa loquitur,* on retiendra en principe un défaut lorsque le système d'IA a entrainé un préjudice dans le cadre d'une utilisation conforme[98]. Partant, dans un tel cas, ce sera au producteur d'amener la contre-preuve, en démontrant, par exemple, le respect de toutes les normes et prescriptions techniques ou l'utilisation abusive de la part de l'utilisateur[99].

Cette solution s'avère particulièrement avantageuse pour le cas où le logiciel est interconnecté. Lorsqu'un objet connecté reçoit les données nécessaires à son fonctionnement d'autres objets connectés, il semble extrêmement difficile de remonter à la chose défectueuse à l'origine du préjudice. Par exemple, si un système d'arrosage intelligent pour jardin cause une inondation, cela peut être dû non seulement au défaut du logiciel du système d'arrosage, mais aussi au caractère erroné des données sur l'humidité communiquées par un autre objet connecté[100]. Du moment que les erreurs dans les données de sources externes sont imputables au producteur du système ayant pris une décision à l'origine du préjudice[101], le lésé peut se contenter d'établir que l'utilisation d'une chose connectée a causé un préjudice, quelle que soit la provenance des données traitées. Ce sera alors là aussi au producteur d'apporter une contre-preuve en démontrant qu'il a observé toutes les exigences permettant de garantir l'exactitude des données.

La LRFP permet de prendre relativement bien en compte la spécificité de l'IA, sous réserve de l'introduction d'un devoir de surveillance du producteur dont la violation constituerait un défaut au sens de l'art. 4 LRFP. Vu l'allègement du fardeau de la preuve, elle instaure presque une responsabilité pour risque de l'IA. Toutefois, la protection de la LRFP exclut le dommage causé à une chose

94 LOHMANN (n. 61), p. 118 ; WILDHABER (n. 61), p. 42.

95 Voir LOHMANN (n. 61), p. 118 ; WILDHABER (n. 61), p. 42.

96 Voir *supra* III.A.1.

97 Voir *supra* II.B.1 ; WILDHABER (n. 61), p. 50.

98 Voir *supra* II.B.1 ; WILDHABER (n. 61), p. 50.

99 MORIN (n. 43), N 64 ss.

100 COMMISSION EUROPÉENNE (n. 64), p. 16 ; WENDEHORST (n. 64), p. 152.

101 Voir *supra* III.B.

destinée à un usage professionnel ou commercial. Or, celui qui utilise un système d'IA à des fins professionnelles ou commerciales a le même besoin de protection qu'un consommateur, le premier n'ayant pas plus de contrôle sur les risques présentés par un système d'IA que le second[102]. La sécurité du produit se trouve en effet dans les mains du producteur et non de l'utilisateur, peu importe qu'il s'agisse d'un consommateur ou d'un professionnel. Ce problème est néanmoins relativisé par le fait que le professionnel lésé peut agir contre le producteur sur la base de l'art. 55 CO et peut ainsi profiter de la jurisprudence du Tribunal fédéral consacrant une responsabilité extrêmement sévère du producteur[103].

C. L'intelligence artificielle comme chose spécifiquement dangereuse

On a vu qu'un chef de responsabilité pour risque ne se justifiait que pour certains systèmes d'IA. Parmi ceux-ci, certains sont intégrés à des choses soumises à une responsabilité pour risque, comme les voitures autonomes ou les drones autonomes (art. 58 LCR, art. 64 LA).

Les chefs de responsabilité pour risque sont technologiquement neutres, en ce sens que le risque spécifique est indépendant du recours ou non à l'IA. Ils s'appliquent donc pleinement aux éventuelles installations ou moyens de transport dotés d'un système d'IA[104]. La légitimité de cette solution est mise en doute par certains auteurs arguant que l'exploitation d'une chose par un système d'IA serait plus sûre, car l'IA n'est pas soumise à des faiblesses typiquement humaines, telles que l'ivresse, la fatigue, l'inattention ou encore le manque de réflexe[105]. Néanmoins, l'IA présente ses propres risques[106] et peut entraîner des

102 BORSARI (n. 37), p. 98 ; FELLMANN/KOTTMANN (n. 8), N 1114 ; FELLMANN/VON BÜREN-VON MOOS (n. 36), N 121 ; HESS (n. 23), art. 1 N 82 ; WERRO (n. 14), N 588. Voir ég. ELI (n. 85), p. 13 ss.

103 Voir *supra* II.A.2 ; MORIN (n. 43), N 68.

104 Rapport du Conseil fédéral en réponse au postulat Leutenegger Oberholzer 14.4169 « Automobilité », Conduite automatisée – Conséquences et effets sur la politique des transports, du 21 décembre 2016, p. 26 ; FELLMANN (n. 64), p. 109 ; LOHMANN (n. 61), p. 120 ; LOHMANN/MÜLLER-CHEN (n. 83), p. 51 ; MORIN (n. 43), N 35.

105 LOHMANN/MÜLLER-CHEN (n. 82), p. 53 ; S. MÉTILLE/N. GUYOT, « Le moment venu de reconnaître un statut juridique aux robots », *Plaidoyer* 03/2015, p. 26, p. 28 ; Th. PROBST, « Die Benutzung (teil-)autonomer Motorfahrzeuge im Strassenverkehr aus haftpflichtrechtlicher Sicht », *in* Th. PROBST/F. WERRO (édit.), *StrassenverkehrsTagung 21.-22. Juni 2016*, Berne 2016, p. 1, p. 30 et p. 40 ; G. WAGNER, « Roboter als Haftungssubjekte ?, Konturen eines Haftungsrechts für autonome Systeme », *in* F. FAUST/H.-B. SCHÄFER (édit.), *Zivilrechtliche und rechtökonomische Probleme des Internet und der künstlichen Intelligenz*, Tübingen, p. 1, p. 17.

106 Voir *supra* III.A.1.

accidents qu'un humain aurait aisément évités, comme une collision avec un camion blanc en plein jour sur une route rectiligne suite à une confusion entre le ciel et la couleur blanche du camion[107]. En outre, en cas de défaillance du système d'IA, le risque de préjudices importants est plus élevé[108]. Il paraît donc justifié de soumettre les choses spécifiquement dangereuses ayant recours à l'IA à la responsabilité pour risque. Par conséquent, l'exploitant d'une chose spécifiquement dangereuse pilotée par un logiciel d'IA est responsable dès qu'un dommage corporel, un dommage matériel ou un tort moral est causé par la réalisation du risque spécifique de ladite chose[109].

En droit positif suisse, la responsabilité pour risque est plus intéressante pour le lésé que la responsabilité du fait des produits. La responsabilité pour risque est technologiquement neutre, de telle sorte qu'en matière d'IA, la responsabilité de l'exploitant est engagée dès que le risque spécifique se réalise, quelle que soit la source de l'incident, notamment le rôle joué par le système d'IA[110]. Puisque le risque se rapporte à une chose en soi parfaite, la responsabilité pour risque évite au lésé de s'interroger sur un éventuel défaut, ce qui épargne également toutes les questions délicates soulevées par l'art. 5 LRFP. Enfin, l'assurance responsabilité civile obligatoire garantit au lésé son indemnisation indépendamment du fait que la chose soit ou non équipée d'un système d'IA[111].

Une clause générale de responsabilité pour risque, technologiquement neutre, permettrait de soumettre les exploitants de choses spécifiquement dangereuses à un tel régime de responsabilité objective aggravée en l'absence d'une loi spéciale[112]. Le risque spécifique correspondrait au risque élevé qu'une chose porte atteinte à l'intégrité corporelle ou matérielle d'une personne. La responsabilité de l'exploitant serait engagée dès que le risque spécifique se réaliserait et provoquerait un dommage corporel, un dommage matériel ou un tort moral. Elle s'accompagnerait, en outre, d'une assurance responsabilité civile obligatoire.

107 WAGNER (n. 105), p. 17.

108 PICHONNAZ (n. 13), p. 521.

109 FF 2021 3026 p. 39 ; FELLMANN (n. 64), p. 109 ; M. F. LOHMANN, *Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts*, Baden-Baden 2016, p. 239 ; LOHMANN/MÜLLER-CHEN (n. 83), p. 51 et p. 53 ; MORIN (n. 43), N 35 ; F. WERRO/V. PERRITAZ, « Les véhicules connectés : un changement de paradigme pour la responsabilité civile ? », *in* F. WERRO/Th. PROBST, *Journées du droit de la circulation routière 23-24 juin 2016*, Berne 2016, p. 1, p. 5.

110 WENDEHORST (n. 64), p. 160.

111 FF 2021 3026, p. 39.

112 PICHONNAZ/WERRO (n. 9), p. 2.

D. La rencontre de responsabilités

Vu les avantages de la responsabilité pour risque, la personne lésée par un système d'IA a intérêt à ouvrir action contre l'exploitant si un chef de responsabilité pour risque s'applique. L'exploitant (ou son assurance) peut recourir contre le producteur responsable envers le lésé selon la LRFP, conformément à l'art. 51 CO ou à une règle plus spéciale (par ex. art. 60 LCR), avec comme problème que celui-ci ne sera pas forcément en mesure de lui verser une indemnité étant donné que la LRFP ne prévoit pas d'assurance responsabilité civile obligatoire[113]. Il serait souhaitable d'obliger les producteurs des systèmes d'IA à contracter une assurance responsabilité civile obligatoire[114].

IV. Conclusion

L'intelligence artificielle s'illustre comme la nouvelle technologie à laquelle doit faire face le droit de la responsabilité civile. La véritable question est de savoir comment appréhender le risque propre à cette technologie, c'est-à-dire la capacité d'agir de manière autonome et imprévisible. Pour donner une réponse à cette question qui s'intègre de manière cohérente dans le système actuel, il vaut la peine d'observer comment le législateur a réagi aux innovations technologiques précédentes.

Pour chaque technologie présentant un risque spécifique, le législateur a adopté un chef de responsabilité spécial avec comme résultat une multitude de responsabilités liées aux diverses technologies. Néanmoins, si l'on s'attarde sur cet imbroglio de chefs de responsabilité, on peut en dégager une certaine systématique. Les différents chefs de responsabilité se réfèrent tous au fait d'une chose, soit à une caractéristique d'une chose. Il s'agit d'un défaut, lorsque la chose présente un danger pour les personnes et les choses parce qu'elle ne correspond pas à ce qu'elle doit être (LRFP) et d'un risque, lorsque la chose, supposée parfaite, présente en elle-même un danger important pour les personnes et les choses. La responsabilité incombe à la personne qui dispose en quelque sorte de la maîtrise sur le fait de la chose, à savoir le producteur pour le défaut et l'exploitant pour le risque.

Les systèmes d'IA sont soumis à la LRFP. Cette loi répond relativement bien aux inquiétudes que soulève l'IA, sous réserve de l'instauration d'un devoir de surveillance à la charge du producteur dont la violation constituerait un défaut,

[113] MORIN (n. 43), N 73.

[114] MORIN (n. 43), N 78. Cette solution s'avère notamment intéressante lorsqu'on songe au fait que nombre de développeurs de systèmes d'IA sont des *start-up* qui ne présentent pas nécessairement de certitudes relatives à leurs capacités financières.

de la prise en compte des dommages causés aux choses destinées à un usage professionnel ou commercial et de l'introduction d'une assurance responsabilité civile obligatoire. L'autonomie et l'opacité de l'IA ne profitent pas aux producteurs, comme on pourrait le craindre. En effet, l'impossibilité à identifier la source d'un dysfonctionnement de l'IA ne péjore pas la situation du lésé, lequel peut se contenter d'établir que le système d'IA a provoqué un préjudice dans le cadre d'une utilisation conforme. C'est au final au producteur d'apporter la preuve du contraire et donc d'assumer les inconvénients de l'effet boîte noire et de l'interconnexion de son système.

Certains systèmes d'IA présentent un danger justifiant un chef de responsabilité pour risque. Ceux qui commandent une chose spécifiquement dangereuse, comme les voitures autonomes, sont déjà soumis à une responsabilité objective aggravée. Cette solution présente plusieurs avantages pour le lésé, notamment la simplicité du fait générateur qui est la réalisation du risque spécifique et l'assurance responsabilité civile obligatoire. La responsabilité pour risque étant technologiquement neutre, une clause générale de responsabilité pour les choses spécifiquement dangereuses permettrait de soumettre à un tel régime les systèmes d'IA présentant un risque caractérisé sans qu'il ne soit nécessaire d'adopter un chef de responsabilité spécifique à l'IA.

La preuve par la technologie

Étude comparée en droit de la responsabilité civile et en droit des mineurs

MATTHIEU TOURNIGAND
Doctorant à l'Université Paris II Panthéon-Assas | Assistant de justice à la Chambre des mineurs de la Cour d'appel de Paris

Table des matières

I. Introduction

« Le juge peut, au cours des opérations de vérification, à l'audience ou en tout autre lieu, se faire assister d'un technicien [...] et toute personne dont l'audition paraît utile à la manifestation de la vérité » ; l'art. 181 du Code de procédure civile français, bien qu'il soit peu mobilisé en pratique, illustre l'emploi des technologies par le droit de la preuve. Le technicien, c'est-à-dire la personne faisant usage de techniques, peut être mobilisé pour dégager « la » vérité. Laquelle ? Tout simplement celle dont a besoin le juge pour trancher le litige, et que l'on nommera la vérité juridique[1,2].

1 Par opposition à la vérité scientifique.

2 L'usage du terme de vérité peut paraître inadapté pour le cadre du droit, puisque la réalité est reconstituée au moyen des règles de preuve. Il s'agirait donc d'une construction plus que d'une vérité. Toutefois, il nous semble qu'une fois établie, cette construction est relativement constante du point de vue de l'ordre juridique. Elle devient alors une

Théoriquement, toute réponse à un problème juridique suppose une certaine compréhension du réel. La structure de la règle de droit est généralement décrite de la manière suivante : un effet est attaché à une condition, ou présupposé[3]. Le déclenchement de cet effet est donc subordonné à une appréciation de la réalité.

Dans ce processus d'accès à la connaissance, les juristes peuvent être tentés de s'appuyer sur le développement des techniques scientifiques, espérant ainsi affiner leur compréhension du réel. Nous emploierons ici le terme de « technique » pour désigner l'« ensemble des procédés méthodiques, fondés sur des connaissances scientifiques »[4], soit ce que désigne le terme anglais de « technology »[5].

La technologie, au sens français du terme, « la science qui concerne [les techniques] »[6], doit être appréhendée par l'art juridique en tant qu'outil. Il importe beaucoup que le droit soit lui-même capable de tenir un discours raisonné sur les techniques scientifiques qu'il peut éventuellement utiliser dans la recherche d'une solution, car cet emploi soulève deux séries de questions.

Ces questions peuvent être d'ordre éthique. C'est par exemple le cas lors de la réalisation de tests ADN visant l'établissement de liens de parenté dans un contexte migratoire[7] ou privé[8]. C'est également le cas en matière de divorce, en ce que l'utilisation de preuves obtenues par des procédés techniques scientifiques

vérité ; dont on admet qu'elle décrit justement les faits jusqu'à ce qu'une décision d'appel ou de la Cour de cassation ne la modifie. En outre la distinction est souvent admise en doctrine ; elle présente donc l'intérêt d'être comprise par toutes et tous. Voir par exemple : J. HAUSER, « La preuve biologique, reine des preuves ? », *RTD civ.* 1993, p. 811 ; J.-S. BORGHETTI, « Vaccination contre l'hépatite B et sclérose en plaques : incertitudes scientifiques et divergences de jurisprudence », *JCP* 2011, p. 79 ; J.-L. GILLET, « La croisée des savoirs – Le juge face à des vérités croisées : Vérité scientifique, vérité juridique, vérité judiciaire », *Les cahiers de la justice*, 2018, p. 317.

3 Voir par exemple : B. ANCEL, « Qualification », *Répertoire de droit international*, Dalloz, 1998, n° 85 s.

4 *Dictionnaire Le Robert*, « Technique », Le Robert, édition en ligne, consultée le 13.10.2022, https://dictionnaire.lerobert.com/definition/technique.

5 *Ibid.*

6 MAUSS, « Les techniques et la technologie », *Revue du MAUSS*, 2004/1, n° 23, p. 434, réédition d'un article paru en 1948.

7 En matière migratoire les expertises furent permises dans le cadre du regroupement familial par la loi n° 2007-1631 du 20 novembre 2007 (art. L. 111-6 du Code de l'entrée et du séjour des étrangers et du droit d'asile) ; la disposition fut finalement abrogée par l'ordonnance n° 2020-1733 du 16 décembre 2020, à compter du 1er mai 2021.

8 Un problème éthique se posa par exemple au sujet de la réalisation de tests ADN sur des personnes décédées. La loi n° 2004-800 du 6 août 2004 (art. 16-11 al. 2 *in fine* du Code civil français) a finalement interdit la réalisation de ces tests si la personne n'y avait pas consenti de son vivant.

– des enregistrements audios, vidéos, etc. – est soumise aux art. 259-1 et 259-2 du Code civil français, imposant qu'elles aient été obtenues loyalement[9].

Ces questions peuvent ensuite concerner la mesure dans laquelle les techniques scientifiques peuvent permettre de prouver l'existence de faits ; c'est-à-dire la mesure dans laquelle ces techniques permettent concrètement de répondre au problème juridique.

C'est ce dernier aspect qui constituera le cadre de notre étude. Ce qui nous intéressera donc précisément dans la suite de ces développements, c'est l'utilisation par le droit des techniques scientifiques, en tant que procédés d'accession à une vérité scientifique. Cet aspect de la relation entre la technologie et le droit s'est développé de manière remarquable depuis l'époque moderne.

L'utilisation par le droit de la preuve des techniques scientifiques intervient dans deux hypothèses. Elle intervient d'abord lorsque les faits litigieux se rapportent à un domaine techniquement complexe. C'est par exemple le cas en matière de cybercriminalité, puisque la répression de ce type d'infractions suppose nécessairement l'usage de techniques informatiques. Ainsi l'inforensique, ou *computer forensics*, consiste à utiliser des « techniques d'investigation numérique »[10] afin de prouver la commission d'infractions sur les réseaux. C'est également le cas en matière de preuve de la faute d'un des époux dans une procédure de divorce, notamment à propos de la preuve de l'infidélité de l'époux résultant d'une correspondance électronique par SMS. L'imputabilité des messages à une personne suppose en effet que l'opérateur vérifie la provenance de ces messages[11].

L'utilisation par le droit de la preuve de nouvelles techniques peut aussi intervenir pour améliorer la qualité de la preuve administrée dans des matières qui n'ont pas en elles-mêmes connu de complexification technologique. La matière pénale constitue à nouveau un bon exemple de ce phénomène. Ainsi les analyses balistiques sont aujourd'hui monnaie courante en droit pénal, et permettent de comparer les traces laissées par une arme sur une munition[12]. Le droit civil n'est pas en reste, notamment en matière de preuve d'état civil. Le passeport biométrique, devenu la norme depuis l'entrée en vigueur, le 28 août 2006, du Règlement du 13 décembre 2004[13], permet la « vérification de l'identité du

9 Voir sur ce point : E. ALBOU, « Preuves électroniques et divorce », *AJ fam.*, 2014, p. 483 s.

10 Ch. FERAHL-SCHUHL, *Cyberdroit*, 8e éd., Dalloz, 2020, p. 191.

11 ALBOU (n. 9).

12 M. SCHWENDER, « Police et technique scientifique », *Répertoire de droit pénal et de procédure pénale*, Dalloz, 2016, n° 70 s.

13 Règlement n° 2252/2004 du 13 décembre 2004 établissant des normes pour les éléments de sécurité et les éléments biométriques intégrés dans les passeports et les documents de voyage délivrés par les États membres (JOUE L 385, 29.12.2004, p. 1).

titulaire »[14] grâce à des données telles que la photographie numérisée, l'iris ou l'empreinte digitale. Les nouvelles techniques liées à la blockchain sont également prometteuses en matière de preuve[15].

L'importance des techniques dans la preuve d'un fait ou d'un acte juridique s'est donc étendue à de nombreux champs du droit. Ces deux hypothèses d'utilisation de techniques scientifiques peuvent cependant devenir problématiques.

En effet la relation entre les techniques scientifiques et le droit de la preuve devient difficile lorsque la vérité à laquelle les techniques permettent d'accéder ne correspond pas à la vérité en jeu dans le présupposé de la règle de droit. Autrement dit, une dualité s'instaure entre la réponse demandée par le droit et la réponse apportée par la technologie ; cette dernière n'est pas assez précise pour le droit. Deux exemples topiques permettent d'illustrer cette dualité et les différentes solutions auxquelles elle donne lieu. Nous nous limiterons à ces deux contentieux, car ils sont fondés sur des dynamiques substantielles clairement identifiées, et qu'il nous semble précieux de restituer une analyse précise du droit positif en la matière[16].

Le premier exemple nous vient du droit français de la responsabilité civile, et concerne l'appréciation du lien de causalité entre le fait générateur et le dommage en matière de responsabilité du fait des produits défectueux. Plus précisément, il sera ici question des dommages liés à l'ingestion du diéthylstilbestrol (DES) et à l'inoculation du vaccin contre l'hépatite B.

Le second exemple est tiré du droit relatif à la détermination de la minorité dans le cadre de la prise en charge des mineurs non accompagnés, ou des mineurs isolés étrangers[17], en droit des personnes français.

Les *hiatus* entre vérité scientifique et vérité juridique se posent à des étapes différentes du raisonnement et reçoivent des réponses juridiques différentes. L'étude conjointe des deux contentieux nous permet donc d'affiner notre analyse.

14 Art. 4 § 3, b du Règlement précité.

15 V. MAGNIER, « Enjeux de la blockchain en matière de propriété intellectuelle et articulation avec les principes généraux de la preuve », *Dalloz IP/IT*, 2019, p. 76.

16 Le format de cette contribution ne permet pas de procéder à une analyse aussi précise de l'ensemble du droit, ce qui justifie l'absence de démonstration générale.

17 Ce terme est parfois préféré, à juste titre nous semble-t-il : voir K. PARROT, « Les mineur·es isolé·es à la frontière entre infra-droit et non-droit », *in* E. GALLANT (édit.), *Quelle protection pour les mineurs accompagnés ? Actes du colloque du 21 juin 2018*, IRJS éditions, 2019, p. 62. Toutefois, le terme de MNA est majoritairement employé par l'Aide sociale à l'enfance, les juges, et la plupart des auteurs ; c'est pourquoi il en sera fait usage ici.

L'étude comparée de ces deux domaines peut surprendre, tant les enjeux liés à ces matières diffèrent. Toutefois, du point de vue de la relation entre la technologie et le droit, la situation se pose dans des termes comparables[18]. On remarquera d'ailleurs que les techniques en jeu visent toutes la personne humaine. L'examen des différences entre les contentieux ciblés pourrait permettre d'apporter un éclairage supplémentaire à la problématique posée par le droit des mineurs non accompagnés, dont la solution a déjà fait l'objet de nombreuses critiques[19].

C'est donc bien de technologie et de droit dont il sera question dans cette contribution ; il s'agit de l'étude du discours (le *logos*) sur les techniques en droit français.

Quels sont les discours portés sur les techniques en droit français de la responsabilité civile et en droit des mineurs non accompagnés ? Comment est adapté ce discours en cas de *hiatus* entre vérité scientifique et vérité juridique ?

La technologie est devenue un élément central du droit de la preuve, contribuant grandement à la recherche de la vérité scientifique (II). L'instabilité de cet élément suppose toutefois que le droit encadre de manière appropriée cette aide (III). Une brève conclusion nous permettra de proposer une généralisation des observations dégagées (IV).

II. Illustrations du rôle de la technologie en droit de la preuve

Partons d'abord d'un constat commun et bien identifié en doctrine. Les deux contentieux étudiés s'appuient *a priori*, l'on n'oserait dire naturellement, sur la recherche d'une vérité objective correspondant à une vérité scientifique (A). Dans cette quête, les techniques scientifiques fournissent une aide circonstanciée (B).

18 Voir plus précisément *infra*.

19 Voir par exemple : C. BRUGGIAMOSCA, « La présomption de minorité et l'accès à un recours effectif et suspensif », *AJ fam.* 2020, p. 148 ; F. JAULT-SESEKE, « La constitutionnalité du recours aux tests osseux », *Rev. crit. DIP* 2019, p. 972 ; A.-M. LEROYER, « L'âge des mineurs non accompagnés : quand la détermination de l'âge ne présente pas toujours les garanties suffisantes », *RTD Civ.* 2020, p. 71 ; L. AÏT AHMED/E. GALLANT/H. MEUR (édit.), *Quelle protection pour les mineurs non accompagnés ?, Actes du colloque du 21 juin 2018*, IRJS Editions, 2019 ; S. CORNELOUP, « La jurisprudence de la Cour de cassation relative aux mineurs étrangers à la lumière de la jurisprudence du Conseil constitutionnel », *Titre VII*, Dossier n° 6, avril 2021.

A. La recherche d'une vérité scientifique

Toutes les règles applicables à nos objets d'étude exigent l'examen d'une condition d'une nature qui semble pouvoir relever de la biologie. La recherche de la vérité juridique peut se traduire en des termes scientifiques. Il n'est donc pas incohérent que cette réponse soit recherchée en faisant appel à des techniques scientifiques.

En droit de la responsabilité civile, de nombreux droits nationaux retiennent le même modèle, en exigeant la démonstration d'un dommage, d'un fait générateur et d'un lien de causalité[20]. L'indemnisation des victimes du DES est soumise à ce régime de droit commun en raison de l'ancienneté des faits générateurs. Le droit positif issu de la Directive européenne du 25 juillet 1985 en matière de responsabilité du fait des produits défectueux reprend ces conditions au sein de l'art. 1245-8 du Code civil français : « Le demandeur doit prouver le dommage, le défaut et le lien de causalité entre le défaut et le dommage ». Les dommages faisant suite au vaccin contre l'hépatite B sont, eux, soumis à ces nouvelles dispositions. Dans ces affaires l'enjeu pour le juge est donc de réussir à déterminer si, oui ou non, le produit a causé un dommage. Plus précisément, il s'agit de répondre à la question suivante : le produit de santé, mis sur le marché par la personne morale dont la responsabilité est recherchée, a-t-il causé le dommage spécifiquement subi par la prétendue victime ? La doctrine, et la Cour de cassation[21] distinguent deux membres dans cette question : celui de l'imputabilité du dommage au défaut du produit, d'une part, et celui du lien de causalité entre le produit émis par l'éventuel responsable et le dommage, d'autre part. Or la réponse à cette question de la causalité est, au moins pour partie, adossée à celle de la vérité scientifique. Christophe Radé a pu écrire que « si la causalité juridique ne peut bien entendu pas se confondre avec la causalité scientifique, elle ne peut pas non plus s'en affranchir totalement »[22].

Il semble que le lien entre vérités juridique et scientifique ne soit pas totalement rompu à propos des dommages causés par des produits de santé. La médecine fait partie des domaines dans lesquels l'évolution technologique a participé à la complexification de la matière. Or la complexification – d'un point de vue technique – de la matière dans laquelle s'inscrit le fait générateur impose

20 Ainsi l'art. 1240 du Code civil français est assez proche de l'art. 41 du Code des obligations suisse, puisqu'ils disposent respectivement que « Tout fait quelconque de l'homme qui cause à autrui un dommage oblige celui par la faute duquel il est arrivé à le réparer » et que « Celui qui cause, d'une manière illicite, un dommage à autrui, soit intentionnellement, soit par négligence ou imprudence, est tenu de le réparer ».

21 Civ. 1re 27 juin 2018, nº 17-17.469, P. I. ; *RCA* 2018, Comm. 253, obs. L. Bloch ; *RTD civ.* 2018, p. 925, obs. P. Jourdain.

22 C. RADE, « Causalité juridique et causalité scientifique : de la distinction à la dialectique », *D.* 2012, p. 112.

nécessairement l'usage de techniques scientifiques dans la recherche d'une solution juridique. Si le fait générateur du dommage est issu d'un domaine techniquement complexe, il devient indispensable que l'on s'appuie sur des techniques qui permettent de démêler cette complexité. Les juges s'appuient alors sur diverses expertises et études sur le sujet[23].

Cet appui des techniques scientifiques existe également en droit de la minorité. L'enjeu de la preuve de la minorité est d'assurer aux mineurs non accompagnés une protection suffisante[24] par l'établissement d'une mesure d'assistance éducative[25]. Le mineur est défini par l'art. 388 du Code civil comme « l'individu de l'un ou l'autre sexe qui n'a point encore l'âge de dix-huit ans accomplis ». Retenant une conception chronologique de l'âge, le droit s'appuie sur un critère temporel scientifiquement quantifiable. Aussi, ne faut-il pas s'étonner que l'alinéa suivant autorise le recours à la recherche de l'âge physiologique par des techniques médicales : les « examens radiologiques osseux ».

Le lien qu'entretiennent ces contentieux avec les techniques scientifiques rejoint le constat établi par la théorie générale du droit. Ainsi Xavier Lagarde, après avoir énoncé que le fondement du droit est sa légitimation, relève qu'un des procédés de légitimation courants consiste à faire implicitement appel à l'idée de vérité[26]. Le Doyen Cornu justifie simplement ce lien par l'existence d'un axiome en faveur de la vérité, « *favor veritatis* »[27], la vérité étant « l'or du Droit »[28]. Certains auteurs ont d'ailleurs relevé, en droit comparé, la différence entre la conception française de la vérité et la conception de cette vérité en *common law*[29]. En *common law*, et plus particulièrement au Québec, les experts – choisis par les parties – seraient davantage considérés comme des témoins.

23 Par exemple en matière de DES : CA Versailles 21 décembre 2006, nº 05/06698, TGI Nanterre 10 avril 2014, nº 12/12349.

24 Protection exigée en droit interne (Cons. const. 21 mars 2019, nº 2018-768 QPC, *AJDA* 2019, p. 1448, note Escach-Dubourg ; *D.* 2019, p. 742, note Parinet ; *D.* 2019, p. 709, obs. Fulchiron ; *RDSS* 2019, p. 453, note Caire ; *Rev. crit. DIP* 2019, p. 972, note Jault-Seseke ; *AJ fam.* 2019, p. 222, obs. Bouix ; *Dr. fam.* 2019, nº 107, note Fulchiron), ainsi qu'en droit international (Civ. 1re 14 juin 2005, nº 04-16.942 admettant l'application directe de l'art. 3-1 de la Convention internationale des droits de l'enfant du 26 janvier 1990).

25 Pour un récapitulatif de la procédure de prise en charge des mineurs non accompagnés, voir : A. GUITTON, « Droit des mineurs isolés étrangers tout au long de leur parcours – Tableau récapitulatif », *AJ fam.* 2019, p. 504.

26 Vérité étant ici entendue au sens de vérité scientifique : X. LAGARDE, *Réflexion critique sur le droit de la preuve*, 1994, LGDJ, spéc. p. 17 s.

27 G. CORNU, « Rapport de synthèse », *in La vérité et le droit, journées canadiennes de l'Association Henri Capitant*, Economica, 1987, p. 4.

28 CORNU (n. 27), p. 11.

29 L. KHOURY/E. VERGES, « Le traitement judiciaire de la preuve scientifique : une modélisation des attitudes du juge face à la connaissance scientifique en droit de la responsabilité civile », *Les Cahiers de droit*, 2017, nº 58 vol. 3, p. 517-548.

Plusieurs récits se présentent alors au juge, sur lesquels il s'appuie pour construire son raisonnement. Le droit français supposerait l'existence d'une vérité générale que le juge est chargé de découvrir par le recours à un expert qu'il choisit lui-même[30].

Cela ne signifie pas pour autant que les juristes français ignorent que la science ne peut pas tout prouver. Il est connu que le droit se développe « à partir du réel et de ce que l'on en connaît ou peut en connaître »[31] et que l'ignorance scientifique est prise en compte par le droit. Pour ce qui nous concerne, l'aide des techniques scientifiques doit donc être comprise dans sa stricte mesure.

B. L'insuffisance technique

La difficulté vient de ce que les techniques scientifiques ne permettent pas de trancher le présupposé de la règle de droit ; elles nous disent quelque chose d'une réalité qui n'est pas exactement ce que l'on veut en savoir. Cette difficulté est majorée, en droit français, par l'exigence *a priori* de certitudes et l'impossibilité d'admettre qu'une vérité juridique soit déterminée par probabilités. Ce point de vue n'est pas unanime en droit comparé. Ainsi, le Code civil québécois retient une solution différente puisque son art. 2804 dispose que « la preuve qui rend l'existence d'un fait plus probable que son inexistence est suffisante, à moins que la loi n'exige une preuve plus convaincante ».

Pour mieux cerner le droit français de la preuve, il nous paraît nécessaire de préciser le décalage entre la vérité établie scientifiquement et la vérité juridique dans les domaines étudiés ; et donc l'insuffisance de la science pour répondre aux problèmes juridiques.

Pour les contentieux de responsabilité civile étudiés, l'incertitude n'émerge pas au même stade. Concernant le DES, la difficulté porte sur la causalité, à savoir le rôle de la molécule produite par le laboratoire sur le dommage subi par la victime. Dans ces affaires l'imputabilité n'est plus à démontrer[32] ; elle est constamment admise en jurisprudence[33]. L'indétermination porte donc sur la causalité du dommage par la molécule produite par le laboratoire attrait en justice.

30 KHOURY/VERGÈS (n. 29), p. 522 s.

31 C. LABRUSSE-RIOU, « Rapport français », *in La vérité et le droit, journées canadiennes de l'Association Henri Capitant*, Economica, 1987, p. 104.

32 Une des premières études démontrant le lieu entre l'adénocarcinome et le DES date de 1971 : A. HERBST/H. ULFELDER/D. POSKANZER, « Adenorcarcinoma of the vagina – Association of Maternal Stilbestrol Therapy with Tumor Appearance in Young Women », *The New England Journal of Medicine*, 1971, p. 878 s.

33 Les éléments scientifiques pris en compte sont parfaitement détaillés dans l'arrêt : CA Versailles 14 avril 2016, n° 16/00296. Pour l'admission de cette solution par la Haute juridiction judiciaire française, voir : Civ. 1re 7 mars 2006, n° 04-16.179 et n° 04-16.180, *D.* 2006, p. 812, obs. I. Gallmeister ; *RTD civ.* 2006, p. 565, obs. P. Jourdain ; *RTD com.* 2006, p. 906, obs. B. Bouloc ; *RCA* 2006, p. 164, obs. C. Radé ; Civ. 1re 24 septembre 2009, n° 08-10.081 et n° 08-16.305 ; Civ. 1re 19 juin 2019, n° 18-10.380.

34 Voir par ex. Civ. 1re 24 septembre 2009, n° 08-16.305.

35 Civ. 1re 24 septembre 2009, n° 08-16.305.

36 Car, dans un tel cas, toute technologie dont le fonctionnement trop complexe échapperait à notre compréhension scientifique échapperait au droit de la responsabilité civile.

37 PH. BRUN, « La responsabilité du fait des produits défectueux : l'exemple de la vaccination de l'hépatite B », *RLDC* 2009, p. 29 et s.

38 H. PILLAYRE, « Les victimes confrontées à l'incertitude scientifique et à sa traduction juridique : le cas du vaccin contre l'hépatite B », *Droit et société*, vol. 86, 2014, p. 38 s.

39 Civ. 1re 5 avril 2005, n° 02-11.947 ; CE (Conseil d'État) 9 mars 2007, req. n° 267635 ; Civ. 1re 22 mai 2008, n° 05-20.317 et n° 06-10.967 ; Civ. 1re 10 juillet 2013, n° 12-21.314.

40 Civ. 1re 23 septembre 2003, n° 01-13.063 ; CA Versailles 16 mars 2007, n° 05/09525 ; Civ. 1re 22 janvier 2009, n° 07-16.449 ; Civ. 1re 29 mai 2013, n° 12-20.903 ; Civ. 1re, 20 décembre 2017, n° 16-11.267 ; Civ. 1re, 4 juillet 2019, n° 18-16.809.

41 Par exemple, pour la radiographie du poignet, à l'atlas de Greulich et Pyle ; pour la radiographie dentaire, au test de Demirjian.

Deux sociétés, UCB Pharma et Novartis, se partageaient le marché et les molécules produites étaient identiques. Il est donc scientifiquement impossible de savoir quel est précisément le laboratoire à l'origine de la molécule ingérée[34]. Le problème de droit reste sans réponse scientifique. Partant, la première position de la Cour de cassation consista à refuser d'indemniser la victime qui n'était pas en mesure de prouver qu'elle avait été exposée concurremment aux deux produits[35]. Cela revient à énoncer que s'il n'est pas possible que l'assertion soit vérifiée, alors l'assertion est fausse ; la causalité juridique, calquée sur la causalité scientifique, ne peut la dépasser en même temps qu'elle ne peut s'en satisfaire[36].

Pour les dommages liés au vaccin contre l'hépatite B, les conditions sont identiques. Toutefois, c'est ici la preuve de l'imputabilité du dommage au défaut qui est manquante[37]. La question de savoir si le vaccin contre l'hépatite B peut être à l'origine de scléroses en plaques fait l'objet d'une controverse importante en médecine, notamment au sujet de la méthodologie déployée dans certaines études[38]. Cela se répercute sur le plan juridique puisque l'imputabilité est admise dans certaines décisions[39], et refusée dans d'autres[40]. À nouveau, comme pour le DES, les premières décisions tendaient à refuser d'admettre l'imputabilité juridique en l'absence de la démonstration de l'imputabilité scientifique.

Enfin, pour la détermination de la minorité, la situation est comparable en ce que les techniques scientifiques ne permettent pas de fournir une réponse suffisamment précise pour établir l'âge de l'intéressé. Les examens radiologiques prévus par l'art. 388 al. 2 du Code civil sont généralement de trois sortes ; ils peuvent concerner la dentition, la clavicule et le poignet gauche. Ils permettent, par comparaison avec des standards préétablis[41], de fournir une approximation

de l'âge de l'intéressé. Or il se trouve que ces mesures sont particulièrement imprécises[42]. De nombreuses expertises relèvent que ces standards n'ont pas été élaborés à partir des populations statistiquement visées par ces tests[43]. Il faut ici relever la grande disparité qui existe entre les différents comptes rendus. Certains médecins précisent davantage que d'autres la « marge d'erreur »[44] dont l'indication est obligatoire en vertu de l'art. 388 al. 3 du Code civil depuis une loi du 14 mars 2016. Ainsi, au sein du ressort de la Cour d'appel de Paris, certains experts vont jusqu'à indiquer les références des études relatives aux tests osseux, tandis que d'autres se contentent d'indiquer l'âge physiologique médian et la compatibilité avec l'âge allégué.

En outre, il est prouvé que l'écart-type de l'estimation augmente avec l'âge. Ainsi, pour la radiographie du poignet, il est possible d'obtenir des écarts-types allant jusqu'à quatre ans lorsque la personne a plus de 16 ans[45]. Or, en pratique, le contentieux lié à la détermination de l'âge concerne des personnes dont l'âge allégué est supérieur ou égal à 15 ans. Il en résulte que, assez souvent, les unités médico-judiciaires concluent que l'âge physiologique de l'intéressé est compatible avec l'âge allégué, ou avec un âge inférieur à 18 ans[46]. En résumé, il est impossible, dans de nombreux cas, de se prononcer sur la majorité d'une personne[47].

Les techniques, aboutissant à une vérité scientifique brute, sont donc insatisfaisantes. Le droit doit ainsi se doter d'une technologie au sens propre du terme[48], c'est-à-dire adapter son discours sur les techniques. Il lui revient ainsi de se faire « l'orfèvre de la vérité »[49].

42 Pour un état des lieux technique détaillé : A. GUITTON, *Les examens radiologiques d'âge osseux et l'évaluation de la minorité – Notes d'observation*, InfoMIE, 2019, p. 10 s., www.infomie.net/IMG/pdf/infomie_note_emao_vf.pdf.

43 Les études suivantes sont parfois citées dans les expertises physiologiques rencontrées : B. BÜKEN/A. A. SAFAK/B. YAZICI/E. BÜKEN/A. S. MAYDA, « Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children ? », *Forensic Science International*, 2007, p. 146-153 ; A. ZHUANG/J. W. SAYRE/L. VACHON/B. J. LIU/H. K. HUANG, « Racial differences in growth patterns of children assessed on the basis of bone age », *Radiology*, 2009, vol. 50, n° 1.

44 Qui, d'un point de vue scientifique, est insuffisante : voir rapport InfoMIE (n. 42).

45 Rapport InfoMIE (n. 42), p. 16.

46 Voir par exemple : Civ. 1re 12 janvier 2022, n° 20-17.343.

47 *Ibid.*

48 PH. LE TOURNEAU, *Contrats du numérique. Informatiques et électroniques*, 12e éd., Dalloz, 2022, p. 70.

49 CORNU (n. 27), p. 11.

III. La gestion de l'incertitude technologique

En tant qu'orfèvre de la vérité, le juriste dispose d'un certain nombre d'outils ; les plus connus étant les présomptions. Les droits étudiés s'en sont donc emparés (A). Toutefois, il apparaît que l'usage de présomptions n'est pas miraculeux, et qu'il doit lui-même être réfléchi, surtout en droit des mineurs non accompagnés (B).

A. Le recours à des présomptions

Le recours à des présomptions vise à permettre d'adapter le problème de droit pour pouvoir se satisfaire de la vérité dégagée par la technologie. Bien que ces présomptions interviennent sous différentes formes, leur objectif commun est de s'éloigner de la vérité scientifique pour se rapprocher d'objectifs substantiels (1). Il convient donc d'examiner les objectifs substantiels mis en avant dans les droits étudiés (2).

1. Le rapprochement d'objectifs substantiels

Alors que la recherche de la vérité est souvent présentée comme le penchant naturel du droit de la preuve[50], les présomptions amènent à tempérer ce constat. L'insuffisance technologique rend caduc le problème de droit. En effet, dans ce cas, il n'est plus possible de lui donner une réponse dont le fardeau ne pèse pas systématiquement sur le demandeur. Ce demandeur étant la plupart du temps une partie « faible », le droit de la preuve ne peut se satisfaire de ce résultat, au terme duquel l'application abstraite de règles de preuve prive pratiquement la partie protégée par les objectifs juridiques substantiels ; « justesse n'est pas justice »[51]. L'incertitude scientifique pousse le droit de la preuve à privilégier « une situation incertaine en elle-même, qui est néanmoins préférée à tous les autres possibles »[52]. Cela se traduit, virtuellement, soit par une modification du problème de droit, lui permettant de recevoir une réponse technologique substantiellement satisfaisante, soit par une modification des réponses acceptables au problème de droit.

Il faut ici souligner qu'il ne s'agit pas de la seule fonction des présomptions, qui peuvent être employées afin de tenir pour vrai scientifiquement ce qui n'a

50 Voir *supra* II.A.

51 M.-L. MATHIEU, *Logique et raisonnement juridique*, PUF, rééd. 2015, p. 3.

52 R. LIBCHABER, « Les présomptions, entre fonction probatoire et rôle substantiel », *Droit & Philosophie*, 2019, p. 11.

pas été prouvé techniquement. Dans ces cas, les présomptions visent à traduire le « sentiment de vérité »[53] éprouvé à l'égard de la situation. Elles illustrent alors l'adage *praesumptio ex eo quod plerumque fit*. Il ne nous semble pas que cela soit complètement le cas, ni en matière de droit de la responsabilité du fait des produits de santé, ni en matière de preuve de la minorité.

2. *Des présomptions polymorphiques*

Au contraire, des objectifs substantiels affleurent dans ces présomptions. L'analyse des décisions montre qu'elles peuvent intervenir selon des modalités variées.

En matière de réparation des dommages liés au DES, une présomption de causalité fut créée par l'arrêt du 24 septembre 2009[54]. La Cour de cassation y a ainsi énoncé « qu'il appartenait à chacun des laboratoires de prouver que son produit n'était pas à l'origine du dommage »[55]. Il s'agit donc d'une présomption de droit[56]. Il nous semble que, dans ce cas, le problème de droit n'est plus « la production du laboratoire X est-elle à l'origine du dommage ? », mais plutôt « le laboratoire X a-t-il participé à la production de la molécule à l'origine du dommage ? »[57]. Cela correspond à la définition doctrinale de la présomption de droit, qui est réputée présumer « un acte ou un fait en tenant un autre acte ou fait pour certain »[58]. La modification de ce problème de droit vise à s'adapter au savoir technique mobilisé par la victime au soutien de sa prétention, et permettre son indemnisation. La question de la contribution effective à la dette et des éventuels recours entre les codébiteurs solidaires n'intervient qu'après cette indemnisation[59]. L'objectif substantiel est assez clair ici. Il s'agit

53 *Ibid.*

54 Civ. 1re 24 septembre 2009, n° 08-16.305.

55 La présomption sera explicitée dans l'arrêt Civ. 1re 19 juin 2019.

56 Le juge étant autorisé à créer de telles présomptions bien que le Code civil semble réserver cette prérogative au seul législateur : E. VERGES/G. VIAL/O. LECLERC, *Droit de la preuve*, PUF, 2015, n° 244.

57 Comp. C. QUEZEL-AMBRUNAZ, « La fiction de la causalité alternative », *D.* 2010, p. 11 : selon l'auteur, retenir la « responsabilité de chacun alors même qu'il est établi que certains ne sont pour rien dans la genèse du préjudice » fait échec à la qualification de présomption. Or, justement, il n'était pas établi que l'un des deux laboratoires n'ait joué aucun rôle dans la production du médicament et il est incertain que seul l'un des deux médicaments ait été utilisé.

58 J. GHESTIN/H. BARBIER, *Introduction générale, Traité de droit civil*, t. 2, dir. J. Ghestin, LGDJ, 5e éd., 2020, p. 96.

59 La répartition du poids de la dette n'a pas encore été tranchée, certaines décisions ayant divisé le montant de la dette en deux (CA Paris 26 octobre 2012, n° 10/18297), d'autres ayant opté pour l'application du critère de *market share liability* (CA Versailles 14 avril 2016, n° 16/00296).

de favoriser l'indemnisation des victimes, mouvement identifié de longue date[60], dans un contexte de collectivisation du risque technique[61].

L'objectif substantiel est identique en matière de vaccin contre l'hépatite B. Plutôt que de rejeter systématiquement l'imputabilité de maladies démyélinisantes au vaccin, la jurisprudence fait usage de présomptions de fait. Se substituant au discours scientifique[62], elle permet aux juges du fond de s'appuyer sur certains indices pour admettre l'imputabilité du dommage au vaccin. Ainsi, la Cour de cassation a admis que l'apparition des symptômes peu de temps après la dernière injection était de nature à justifier l'élaboration d'une présomption[63], de même qu'une prise en compte de l'histoire familiale, de l'origine ethnique et du nombre d'injections[64]. Cette présomption est donc une présomption de fait, et non de droit[65]. Selon nous, il ne s'agit pas de la modification de la question de droit, mais de la multiplication des procédés de preuve admissibles au-delà de la technologie scientifique. En d'autres termes, le problème de droit reste : « le défaut du vaccin peut-il causer ce dommage ? », mais il ne sera plus indispensable que la réponse à cette question soit valide d'un point de vue scientifique. Les juges disposent alors d'une souplesse accrue dans la preuve de l'imputabilité, ce qui permettra d'indemniser de nouvelles victimes. Selon Alessia Farano le juge devient alors un « apprenti sorcier »[66], s'éloignant de la vérité scientifique pour pouvoir conclure sur la *quaestio juris* de manière plus satisfaisante[67]. Ce procédé est certes moins protecteur qu'une présomption de droit, mais vise, dans une certaine mesure, à préserver l'existence d'un lien entre vérité scientifique et vérité juridique. Ce faisant, il permet également de préserver les producteurs du vaccin[68].

Ainsi, tant à propos des dommages liés au DES qu'à propos de la vaccination contre l'hépatite B, une certaine distance a été trouvée entre les techniques scientifiques et la solution juridique ; et que « c'est tantôt la science tantôt les sentiments qui triomphent en matière de causalité »[69].

60 L. CADIET, « Sur les faits et les méfaits de l'idéologie de la réparation », *in Le juge entre deux millénaires. Mélanges offerts à P. Drai*, Dalloz, 2000, p. 495.

61 G. VINEY, *Introduction à la responsabilité*, 3e éd., LGDJ, 2008, p. 136.

62 KHOURY/VERGES (n. 29), spéc. p. 543 s.

63 Civ. 1re 25 novembre 2010, n° 09-16.556.

64 Civ. 1re 10 juillet 2013, n° 12.21-314.

65 Concernant la Directive 85/374/CEE du Conseil du 25 juillet 1985 en matière de responsabilité du fait des produits défectueux (OJ L 210, 7 août 1985, p. 29), la CJUE admet ces présomptions, à l'inverse des présomptions de droit : CJUE 21 juin 2017, *N. W e.a. c/Sanofi Pasteur*, aff. C-621/15.

66 A. FARANO, « L'évaluation de la preuve scientifique », *Droit & Philosophie*, n° 11, 2019, p. 33-45.

67 *Ibid.*

68 Voir CJUE 21 juin 2017, *N. W e.a. c/Sanofi Pasteur*, aff. C-621/15, spéc. § 44 s.

69 P. BRUN, « Causalité juridique et causalité scientifique », *RLDC*, 2007, n° 40, p. 16.

En droit des mineurs non accompagnés, une présomption résulte également de l'art. 388 al. 3 du Code civil. Il s'agit donc d'une présomption légale, selon laquelle lorsqu'un doute subsiste, à la suite d'un test osseux, sur l'âge de l'intéressé, ce dernier doit être considéré comme mineur. D'un point de vue théorique, il s'agit d'une présomption de fait, puisque les conclusions médicales « ne peuvent à elles seules permettre de déterminer si l'intéressé est mineur »[70]. Le juge prend en compte les autres éléments du dossier, comme l'évaluation sociale ou l'examen des documents d'identité par la Police aux frontières[71]. Aussi, la présomption de l'art. 388 al. 3 du Code civil semble n'être qu'un rappel de l'art. 1382 du Code civil.

D'autres États européens ont adopté une présomption identique[72]. En Italie, par exemple, l'art. 19^bis § 8 du Décret n° 142 du 18 août 2015[73] dispose que « *[i]f, after the social and medical assessment, doubts about the person's minor age still persist, minor age shall be presumed* »[74]. L'objectif substantiel est en apparence clairement désigné ici : compenser l'incertitude des tests osseux par la prise en charge de personnes pour lesquelles une hésitation subsiste[75]. Cette incertitude justifierait par ailleurs que cette présomption de minorité ne joue pas si aucun test osseux n'a eu lieu[76].

Exposée de manière neutre, cette présomption paraît répondre de manière satisfaisante à l'insuffisante précision technique des tests osseux. Cela n'est cependant pas certain, et il nous semble que le droit de la preuve de la minorité est largement insuffisant et gagnerait à s'inspirer des techniques développées en droit de la responsabilité civile.

B. Les insuffisances de la présomption de minorité en droit français

La présomption de minorité faisant suite aux tests osseux est défectueuse pour plusieurs raisons. Cette défectuosité est sans doute liée à un

70 Art. 388 al. 3 Code civil.
71 Civ. 1re 22 mai 2019, n° 18-22.738 ; Civ. 1re 20 septembre 2019, n° 19-15.262 ; Civ. 1re 12 janvier 2022, n° 20-17.343.
72 À l'inverse d'autres États qui interdisent purement ces tests, comme le Royaume-Uni, ou le soumettent à des conditions restrictives, comme la Norvège.
73 Issu de la loi n° 47 du 07 avril 2017.
74 Traduction issue de l'arrêt CourEDH 21 juillet 2022, n° 5797/17, *Darboe et Camara c. Italie*, § 47.
75 Cons. const. 21 mars 2019, n° 2018-768 QPC, *AJDA* 2019, p. 1448, § 11.
76 A. GOUTTENOIRE, « La présomption de minorité cantonnée aux tests osseux », *JCP* 2019, p. 1088-1890.

dévoiement du droit de la protection de l'enfance français, polarisé, dans cette question, autour de principes qui lui sont étrangers.

La présomption de minorité faisant suite aux tests osseux est d'abord insuffisante en raison de sa place dans le « circuit probatoire » de la minorité. Elle constitue en effet un cercle vicieux[77], car la réalisation de tests osseux suppose deux conditions selon l'art. 388 al. 2 du Code civil : l'absence de documents d'identité valables et l'invraisemblance de l'âge allégué. En d'autres termes, il est donc nécessaire qu'il existe un doute sur l'âge de la personne. Ce doute, s'il subsiste à la suite du test, donne lieu à une présomption de minorité. Or il est très rare qu'il ne subsiste aucun doute sur la minorité de l'intéressé à la suite du test osseux. Il est même très fréquent que le test osseux conclue qu'il est possible que la personne soit mineure. Face à cette situation délicate, les juges disposent alors d'une option : faire jouer systématiquement la présomption de minorité, ce qui revient à déformer l'esprit de l'art. 388 al. 3 du Code civil, ou s'appuyer sur les autres éléments du dossier pour vérifier la conformité des résultats avec les autres éléments de preuve à disposition. Cette deuxième possibilité est, légitimement, appliquée par les juges. On en revient donc à l'état antérieur au test osseux.

Pour estimer l'existence d'un doute, les juges, mués en « apprentis sorciers »[78], utilisent alors des éléments de preuve contestables et aggravant l'insuffisance de la présomption de minorité. Ainsi, les documents d'identité, alors qu'ils ont nécessairement été reconnus non valables, peuvent redevenir un élément déterminant si l'âge qu'ils mentionnent n'est pas compris parmi les âges « possibles » retenus par les tests osseux[79]. D'autres éléments, plus triviaux, sont également pris en compte. Les évaluations sociales réalisées sont à nouveau examinées. Or ces évaluations sociales s'appuient sur des critères très évanescents et peu techniques. Ainsi, la cohérence du parcours migratoire est examinée, de même que l'apparence physique de l'intéressé, ou son « degré de maturité »[80] ; et ce, alors que la plupart des personnes évaluées ne parlent pas ou peu le français.

Qu'est-ce à dire ? Il nous semble que la présomption de fait ne vise plus l'objectif substantiel de protection des mineurs, mais consiste à s'assurer qu'il n'existe aucune fraude. La preuve ne vise plus alors la minorité mais la fraude. En l'état, on pourrait même s'avancer à dire qu'il existe une présomption de droit, à l'issue du test osseux, établissant la fraude de l'intéressé et donc sa majorité. La gestion de l'incertitude technologique est retournée contre les personnes supposées substantiellement protégées. Cela illustre le dévoiement des

77 F. JAULT-SESEKE, « La constitutionalité du recours aux tests osseux », *Rev. crit. DIP* 2019, p. 972.
78 Expression tirée de FARANO (n. 66).
79 Civ. 1re 12 janvier 2022, n° 20-17.343.
80 Voir par exemple : Civ. 1re 3 décembre 2020, n° 20-19.942.

principes directeurs de ce contentieux maintes fois souligné en doctrine[81], et lié à l'intrusion des considérations de politique publique en droit des personnes[82]. La légitimité des décisions n'est plus fondée sur l'idée de vérité, mais, pour employer l'idée développée par Xavier Lagarde, s'appuie sur « l'imputation à une partie d'une attitude répréhensible »[83].

En outre, il n'est pas certain que le cantonnement de la présomption de minorité à la réalisation de tests osseux soit conforme aux décisions du Comité des Nations Unies[84]. Il est également très probable que la décision récente de la Cour européenne des droits de l'homme[85] contraigne la France à modifier sa position[86].

Quelles sont alors les voies d'amélioration ? Deux voies nous semblent possibles du point de vue de la preuve. La voie la moins ambitieuse consiste à s'inspirer du contentieux lié au vaccin contre l'hépatite B et la pratique judiciaire des présomptions de fait. On a observé que les juridictions utilisent, en soutien de la présomption qui doit être établie, des critères très divers et parfois extra scientifiques, mais dont le but est de favoriser un objectif substantiel. En droit des mineurs non accompagnés, les juges en font également usage, mais cette utilisation vise la plupart du temps à appuyer la majorité de l'intéressé. Pour compenser l'importance des tests osseux, il serait profitable que la minorité puisse se rapporter non plus à une vérité biologique mais une vérité sociologique ; l'âge n'étant pas une notion univoque[87]. Ainsi, d'autres éléments pourraient entrer dans le faisceau d'indices de cette présomption de fait, comme par exemple la vulnérabilité de la personne[88]. Cela permettrait de restaurer l'objectif substantiel de protection des mineurs.

81 Voir note 7.

82 G. SALAME, *Le devenir de la famille en droit international privé (une perspective postmoderne)*, Presses Universitaires d'Aix-Marseille, 2006, p. 147 s.

83 LAGARDE (n. 26), p. 409.

84 CORNELOUP (n. 19).

85 CourEDH 21 juillet 2022, n° 5797/17, *Darboe et Camara c. Italie*.

86 La Cour de cassation avait en effet admis la compatibilité de la solution avec l'art. 3 de la Convention européenne des droits de l'Homme : Civ. 1re 21 novembre 2019, n° 19-15.890.

87 Voir, pour une approche sociologique de cette problématique : A. LENDARO, « Mineur jusqu'aux os ? La juge des enfants et l'âge du jeune étranger au prisme des tests osseux », *Ethnologie française*, vol. 50 n° 2, 2020, spéc. p. 381.

88 Voir par exemple la contribution de S. CORNELOUP au colloque de la Cour de cassation, accessible en ligne, intitulé « La preuve de l'état des personnes : questions d'actualité », 17.03.2022, www.courdecassation.fr/agenda-evenementiel/la-preuve-de-letat-des-personnes-questions-dactualite (consulté le 31.8.2022).

La seconde voie, plus ambitieuse, est la création d'une présomption de droit consistant à estimer que toute personne se prétendant mineure doit être considérée comme telle[89]. Il semble que la dernière décision de la Cour européenne des droits l'homme oblige les États parties à prendre en charge ces personnes jusqu'à l'existence d'une décision judiciaire. En l'espèce, le requérant avait été placé en centre pour majeurs à son arrivée en Italie et subissait un test osseux un mois plus tard, avant de finalement saisir le juge. Au regard de ces éléments, la Cour européenne des droits l'homme décide que « *while the national authorities' assessment of the age of an individual might be a necessary step in the event of doubt as to his or her minority, the principle of presumption implies that sufficient procedural guarantees must accompany the relevant procedure* »[90]. Elle conclut ainsi à la violation de l'art. 8 CEDH. Il semblerait donc que la présomption de minorité doive précéder les tests osseux[91]. Cette présomption rapprocherait la solution de celle adoptée en matière de DES, en estimant que lorsqu'il n'est pas démontré que la personne est majeure, alors elle doit être considérée comme mineure. L'objectif de protection des mineurs serait alors restauré.

Il faut espérer que le droit des mineurs non accompagnés soit en mesure de s'adapter à ces exigences technologiques en s'interrogeant davantage sur les objectifs substantiels qu'il poursuit, à l'image du droit de la responsabilité civile.

IV. Conclusion générale

L'usage de techniques scientifiques en droit de la preuve paraît s'imposer dans une large mesure. Les avancées techniques contemporaines offrent la possibilité d'une nouvelle compréhension des faits, présumée plus précise, et il est tout naturel que le droit en fasse usage : « la tentation est grande pour le juriste d'indexer son jugement sur celui de la science »[92].

89 Cette présomption existe déjà, mais elle prend fin à la notification par l'Aide sociale à l'enfance au mineur-accompagné du refus de sa prise en charge, soit quelques jours après l'évaluation sociale ; alors que la police aux frontières n'a pas encore analysé les documents d'identité et que les tests osseux n'ont pas été réalisés.

90 CourEDH 21 juillet 2022, n° 5797/17, *Darboe et Camara c. Italie*, § 154.

91 En droit français, les mineurs allégués sont présumés mineurs, et mis à l'abri, jusqu'au refus de leur prise en charge par l'Aide sociale à l'enfance, qui intervient environ une dizaine de jours après leur arrivée dans le service. À compter de ce refus de prise en charge, ils sont livrés à eux-mêmes jusqu'à l'éventuelle décision du juge des enfants.

92 BRUN (n. 69).

Au terme de cette étude, il nous semble cependant que le progrès représenté par l'usage des nouvelles techniques scientifiques en droit de la preuve doit être circonstancié. Ce progrès n'est pas sans limites, et ne nous permet pas d'apporter une preuve certaine de n'importe quel fait. L'étude comparée réalisée ici permet d'affirmer que l'insuffisance technique est plus ou moins bien cernée selon les domaines du droit et que l'incertitude technologique est importante.

Ces disparités dans la compréhension des techniques scientifiques se répercutent alors sur le plan des règles de preuve entourant les techniques scientifiques. Les présomptions, instruments utilisés pour répondre aux difficultés en matière de preuve, sont inégalement employées en droit de la responsabilité des produits de santé et en droit des mineurs non accompagnés. Cette inégalité montre que l'incertitude technique peut être instrumentalisée. Elle peut être ignorée, comme lorsque les tests osseux sont pris pour un instrument permettant de déterminer la minorité ou la majorité d'une jeune personne. Même lorsqu'elle est prise en compte et corrigée par une présomption, il est possible que cette présomption ne soit pas à la mesure de l'incertitude et des enjeux de la matière.

Il apparaît donc que l'usage de techniques scientifiques ne permet pas toujours d'échapper aux problématiques que posent les matières juridiques. Il contribue plutôt au renouvellement des tensions qui leur sont inhérentes. Il faut alors prendre garde à ce qu'une position *a priori* favorable à l'utilisation des techniques scientifiques n'aboutisse à émousser une vision critique du droit de la preuve et à affecter les objectifs substantiels défendus par nos systèmes juridiques.

Le rôle du droit pour contrer la discrimination algorithmique dans le recrutement automatisé

FABIAN LÜTZ
Doctorant en droit européen | Pôle numérique | Faculté de droit, des sciences criminelles et de l'administration publique | Université de Lausanne

Table des matières

Le droit suisse permet-il de réprimer les *deepfakes* ?

QUENTIN JACQUEMIN
Doctorant et assistant diplômé en protection des données et droit pénal informatique | Faculté de droit, des sciences criminelles et de l'administration publique | Université de Lausanne

Table des matières

I. Introduction

Si les premiers cas constatés d'utilisation des *deepfakes* à des fins criminelles peuvent remonter jusqu'en 2012[1], la majorité des sources s'accordent toutefois pour dire que c'est en 2017 que ce phénomène a véritablement pris une dimension internationale. À cette période, un utilisateur du média social Reddit nommé « *deepfakes* » a créé un canal de discussion sur lequel il a commencé à poster des vidéos à caractère pornographique qui superposaient des visages de célébrités féminines sur ceux des actrices pornographiques du film original[2]. Ces vidéos ont connu un fort succès[3] et ont mené au développement de nombreux *subreddits*[4] postant des vidéos similaires[5]. Il a fallu attendre le 7 février 2018 pour que Reddit adopte de nouvelles règles communautaires et décide de fermer ces canaux[6].

1 En effet, selon une étude réalisée sur la base d'environ 1 % des données publiques de Twitter, 155 tweets mentionnent « *fake porn* » en 2012, et 3595 incluent spécifiquement les termes « *deep nudes* » et « *deep fake porn* » entre 2012 et 2018. S. MADDOCKS, « A Deepfake Porn Plot Intended to Silence Me : Exploring Continuities between Pornographic and Political Deep Fakes », *Porn Studies* 2020, 7:4, p. 415 ss, p. 417.

2 K. KOBRIGER *et al.*, « Out of Our Depth With Deep Fakes : How the Law Fails Victims Of Deep Fake Nonconsensual Pornography », *Rich. J. L. & Tech* 2021, 28:2, p. 204 ss, p. 207.

3 En un mois, plus de 15 000 personnes se sont abonnées à ce canal. R. SPIVAK, « "Deepfakes" : The Newest Way to Commit One of the Oldest Crimes », *Georgetown Law Technology Review* 2019, 3:2, p. 339 ss, p. 345.

4 Il s'agit de canaux de discussion dédiés à certains thèmes spécifiques.

5 K. DHRUVA, « Deepfakes, Online Platforms, end a Novel Proposal For Transparency, Collaboration And Education », *Rich. J. L. & Tech* 2021, 27:3, p. 1 ss, p. 6.

6 À la fermeture du canal *r/deepfakes*, il n'y avait pas moins de 90 000 abonnés. K. DHRUVA (n. 5) p. 7.

Depuis 2017, l'utilisation de cette technologie n'a pourtant pas décru. Suivant une étude de l'Université de Londres (UCL), les *deepfakes* constituent une des menaces principales pour notre société[7]. Du même avis, une étude d'avril 2022 menée par *l'Innovation Lab* d'Europol, regroupant environ 80 experts juridiques, est arrivée à la conclusion que les *deepfakes* représentent le problème technologique qui impactera le plus leur travail jusqu'en 2030[8]. Cela va de pair avec certaines études qui estiment que 90 % du contenu généré en ligne en 2026 sera synthétique, en ce sens qu'il aura été créé ou modifié par de l'intelligence artificielle[9].

Contrairement à ce qui prévaut aux États-Unis et en Europe, ce phénomène n'a – à notre connaissance – jamais fait l'objet d'études juridiques en Suisse. Cette contribution a pour but d'expliquer les *deepfakes* (II) et de voir quelles solutions existent actuellement (III). Parmi celles-ci, il s'agira de déterminer si le droit suisse, spécialement sous l'angle pénal, est suffisamment adapté pour pallier ce phénomène (III.B). En guise de conclusion, nous attirerons l'attention sur d'autres problèmes que soulève l'utilisation de cette technologie à des fins illicites (IV).

II. Qu'est-ce qu'un *deepfake* ?

A. D'un point de vue général

Le phénomène doit son nom à l'utilisateur de Reddit « *deepfakes* » qui créa ce canal Reddit de vidéos pornographiques en 2017[10]. En soi, le terme « *deepfake* » ou « *deep fake* » est un amalgame lexical des mots anglais « *deep learning* » (un sous-ensemble de l'intelligence artificielle) et « *fake* » (un faux)[11]. À l'origine, cette terminologie ne visait que le fait de superposer des visages

7 www.ucl.ac.uk/news/2020/aug/deepfakes-ranked-most-serious-ai-crime-threat (consulté le 11.08.2022).

8 EUROPOL, *Facing Reality ? Law Enforcement and the Challenge of Deepfakes, An Observatory Report from the Europol Innovation Lab*, Publications Office of the European Union, Luxembourg 2022, p. 4.

9 EUROPOL (n. 8), p. 5.

10 J. VINCENT, « Why We Need a Better Definition of "Deepfake" », *The Verge* 2018, www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news (consulté le 11.08.2022) ; M. BODI, « The First Amendment Implications of Regulating Political Deepfakes », *Rutgers Computer & Technology Law Journal* 2021, vol. 47, p. 143 ss, p. 146.

11 J. VINCENT (n. 10). En français, on trouve parfois les termes « vidéos truquées » ou « vidéos hypertruquées » pour parler de *deepfake*. C. LANGLAIS-FONTAINE, « Démêler le vrai du faux : étude de la capacité du droit actuel à lutter contre les *deepfakes* », *La Revue des droits de l'homme* 18/2020, p. 1 ss, p. 1.

dans des vidéos pornographiques en utilisant des outils fondés sur l'intelligence artificielle au sens large[12].

Il faut toutefois admettre que ce terme a désormais pris une signification plus large. Même s'il n'existe pas de définition unanime, il ressort de la littérature certaines caractéristiques essentielles que nous pouvons résumer dans la définition suivante : un *deepfake* est une manipulation de contenus audiovisuels, fondée sur des procédés utilisant du *deep learning*, dont le résultat aboutit à une synthèse des composants originaux (*synthetic media*). D'une manière générale, le terme de « médias synthétiques » est utilisé pour désigner le résultat de façon neutre et générale[13], alors que l'utilisation du mot « *deepfake* » est associée aux usages criminels ou abusifs de cette technologie[14]. Pour certains, ce qui différencie un *deepfake* de tout autre média synthétique, c'est la volonté de son auteur de tromper le public[15].

Il découle de cette définition que la manipulation peut viser tant des images, des vidéos (changer le visage ou l'apparence d'une personne par exemple), des audios (modifier uniquement la voix d'une personne), ou une combinaison de ces éléments (changer l'apparence et la voix d'une personne figurant dans une vidéo)[16]. L'exemple le plus représentatif et certainement le plus connu pour illustrer cette technologie reste la vidéo du cinéaste Jordan Peele dans la peau de Barack Obama visant à sensibiliser la population américaine face aux dangers des *deepfakes*[17].

B. D'un point de vue technique

Cette contribution juridique ne prétend aucunement décrire précisément le fonctionnement technique des *deepfakes*, mais il s'agit ici de fournir les informations principales pour des personnes non averties en la matière. En outre, il s'agit d'aborder le *deep learning* (1), et plus spécialement le *Generative Adversial Network* et le *Facial Renconstruction Autoencoder Neural Network* (2), dans la mesure où ce sous-ensemble de l'intelligence artificielle

12 J. VINCENT (n. 10) ; B. CHESNEY/D. CITRON, « A Looming Challenge for Privacy, Democracy and National Security », *California Law Review* 2019, vol. 107, p. 1753 ss, p. 1757.

13 Twitter par exemple dispose d'une politique en matière de médias synthétiques et manipulés, https://help.twitter.com/fr/rules-and-policies/manipulated-media (consulté le 11.08.2022).

14 EUROPOL (n. 8), p. 5.

15 J. VINCENT (n. 10).

16 B. CHESNEY/D. CITRON (n. 12), p. 1757 ; R. SPIVAK (n. 3), p. 351 ss ; J. VINCENT (n. 10). Cf. également *infra* II.B.3 pour les formes que peuvent prendre les *deepfakes*.

17 https://youtu.be/cQ54GDm1eL0 (consulté le 11.08.2022).

constitue la pierre angulaire de ce phénomène d'un point de vue technique. Enfin, nous énoncerons les formes les plus courantes de *deepfakes* fondées sur ces technologies (3).

1. Deep learning

L'intelligence artificielle (IA) désigne des systèmes qui visent à modéliser les capacités cognitives des humains pour effectuer des tâches et qui peuvent s'améliorer de manière itérative[18]. L'IA englobe plusieurs sous-ensembles dont l'apprentissage par la machine (*machine learning*), qui lui-même comprend la sous-catégorie de l'apprentissage profond (*deep learning)*[19]. Le *machine learning* se réfère à l'étude des systèmes informatiques qui apprennent et s'adaptent automatiquement par l'expérience, sans être explicitement programmés[20].

Le *deep learning* est une sous-catégorie de *machine learning* qui se caractérise par le fait qu'il permet, de par sa structure, de résoudre des problèmes particulièrement complexes et d'identifier des modèles dans des données non structurées telles que des images, du son, de la vidéo ou du texte[21]. Contrairement au *machine learning*, le *deep learning* n'a pas besoin de données structurées en amont, peut traiter une plus grande quantité de données, et requière globalement beaucoup moins d'intervention humaine[22].

Les algorithmes de *deep learning* sont organisés sous la forme de réseaux neuronaux, qui comprennent une couche constituée des données d'entrée (*input layer*), un certain nombre de couches cachées (*hidden layers*) et une couche

18 M. SCHREYER/A. GIERBL *et al.*, « Artificial Intelligence Enabled Audit Sampling », *EF* 4/22, p. 106 ss, p. 108 ; www.oracle.com/ch-fr/artificial-intelligence/what-is-ai/ (consulté le 26.08.2022).

19 M. SHRUTI, *Discover the Differences Between AI vs. Machine Learning vs. Deep Learning*, 13.07.2022, www.simplilearn.com/tutorials/artificial-intelligence-tutorial/ai-vs-machine-learning-vs-deep-learning (consulté le 26.08.2022).

20 COURSERA, *Deep Learning vs. Machine Learning : Beginner's Guide*, 11.08.2022, www.coursera.org/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide (consulté le 26.08.2022).

21 P. GILLIERON, « Intelligence artificielle : la titularité des données », *RSDA* 2021, p. 435 ss, p. 437.

22 E. KAVLAKOGLU, *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks : What's the Difference ?*, 27.05.2020, www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks (consulté le 26.08.2022).

avec les données de résultat (*output layer*)[23]. L'apprentissage est qualifié de « profond » en référence au nombre de couches du réseau[24].

Chaque couche est composée d'un grand nombre de nœuds (*nodes*) qui sont des unités de calcul à l'intérieur du réseau neuronal[25]. Leur nom et leur structure sont d'ailleurs inspirés du cerveau humain, imitant la façon dont les neurones biologiques se signalent les uns aux autres[26]. L'ensemble de ces couches forme un modèle de *machine learning* capable d'apprendre et de s'améliorer en se fondant sur les données d'entrées[27].

Lorsqu'il s'agit de créer des logiciels de *deepfakes*, le développeur va rentrer une grande quantité de données, par exemple des images du visage d'une personne A. et d'autres données qui ne sont pas des images du visage de A. L'algorithme s'entraîne alors à découvrir des motifs et à extraire les traits (*features*) qui permettent de déterminer quelles images appartiennent au visage de A[28].

Pour être plus spécifique, les *deepfakes* se fondent sur deux technologies principales : les *Generative Adversial Networks* (*GAN*) et le *Facial Reconstruction Autoencoder Neural Network* (*FRANN*)[29]. Toutes deux utilisent des réseaux neuronaux, ce qui signifie que les *deepfakes* reposent en réalité sur un ensemble de réseaux neuronaux.

2. *GAN et FRANN*

Le FRANN est au cœur des *deepfakes*. Cette technologie permet d'entraîner un algorithme sur deux sets de données (par exemple des images de la personne d'origine, d'une part, et des images de la personne qu'on aimerait superposer à celles de la personne d'origine, d'autre part). Le but du réseau neuronal est d'identifier les traits élémentaires caractéristiques des sets de données (par exemple les éléments propres au visage de A et ceux propres au visage de B) et, à partir de ces traits élémentaires, de pouvoir reconstruire l'image

23 P. GILLIERON (n. 21), p. 437.

24 Dès lors que le réseau a plus de deux couches, il peut être qualifié de profond (IBM CLOUD EDUCATION, *Neural Networks*, 17.08.2020, www.ibm.com/cloud/learn/neural-networks#toc-what-are-n-2oQ5Vepe, consulté le 26.08.2022).

25 K. KOBRIGER *et al.* (n. 2), p. 211.

26 IBM CLOUD EDUCATION (n. 24).

27 K. KOBRIGER *et al.* (n. 2), p. 211.

28 EUROPOL (n. 8), p. 7 ; K. KOBRIGER *et al.* (n. 2), p. 211.

29 J. CLAVE, *Deepfakes with the First Order Model Method*, 27.09.2020, https://colab.research.google.com/github/JaumeClave/deepfakes_first_order_model/blob/master/first_order_model_deepfakes.ipynb#scrollTo=dAJ-w-MaDpDM (consulté le 28.08.2022) ; K. KOBRIGER *et al.* (n. 2), p. 213.

d'origine[30]. De manière simplifiée, le *deepfake* est créé par le fait d'utiliser les éléments caractéristiques du visage de A pour reconstruire le visage de B[31].

Le GAN[32] permet quant à lui de créer des *deepfakes* plus réalistes. Il met en œuvre deux réseaux neuronaux, l'un appelé le discriminant et l'autre le générateur[33]. Le but du générateur est de créer des contenus (par exemple des visages) synthétiques proches des données de départ, de sorte à tromper le discriminant[34]. Le discriminant doit pour sa part déterminer si le contenu qu'il analyse est issu des données d'entraînement de départ (le visage de A et de B) ou s'il s'agit de contenus créés par le générateur[35]. À la fin des itérations, le discriminant est incapable de déterminer avec plus de 50 % de certitude si le contenu qu'il analyse est réel ou synthétique, autrement dit s'il s'agit d'un *deepfake* ou non[36].

Plus il y a de données disponibles et variées, meilleur sera le résultat de l'algorithme[37]. Par exemple, un algorithme entraîné uniquement avec des données d'hommes blancs avec des cheveux noirs ne sera pas efficace pour créer des *deepfakes* de femmes asiatiques avec des cheveux blonds[38]. En revanche, plus on aura de photos et de vidéos différentes de la personne qu'on souhaite superposer sur une autre, plus le *deepfake* sera convaincant.

3. *Formes*

La combinaison de ces deux technologies permet ainsi d'échanger le visage d'une personne avec celui d'une autre sur une vidéo (*face swap*), de remplacer la voix de quelqu'un par celle d'une autre personne (*deepfake audio*)[39], de changer des caractéristiques d'une personne dans une vidéo par

30 J. CLAVE (n. 28) ; D. NELSON, *What is an Autoencoder ?*, 20.09.2020, www.unite.ai/what-is-an-autoencoder/ (consulté le 28.08.2022) ; H. BANDYOPADHYAY, *Autoencoder in Deep Learning : Tutorial & Use Cases [2022]*, 19.07.2022, www.v7labs.com/blog/autoencoders-guide (consulté le 28.08.2022).

31 K. KOBRIGER *et al.* (n. 2), p. 213.

32 Créé en 2014 par IAN GOODFELLOW, un chercheur de Google à l'Université de Montréal qui a publié le résultat de ses recherches pour la première fois dans : I. GOODFELLOW *et al.*, *Generative Adversial Networks (GANs)*, 10.07.2014, https://arxiv.org/abs/1406.2661 (consulté le 28.08.2022).

33 B. CHESNEY/D. CITRON (n. 12), p. 1760 ; EUROPOL (n. 8), p. 8 ; R. SPIVAK (n. 3), p. 343.

34 M. BODI (n. 10), p. 145 ; R. SPIVAK (n. 3), p. 343.

35 R. SPIVAK (n. 3), p. 343, lequel cite I. GOODFELLOW *et al.* (n. 31).

36 https://developers.google.com/machine-learning/gan/training (consulté le 28.08.2022).

37 EUROPOL (n. 8), p. 9.

38 EUROPOL (n. 8), p. 9.

39 J. VINCENT (n. 10).

I. Introduction

Sans que le grand public en ait nécessairement conscience, la grande majorité des entreprises multinationales utilisent des outils d'intelligence artificielle (IA) pour le recrutement de leurs employé(e)s[1]. Cela change la donne concernant la discrimination dans le marché du travail, y compris en ce qui concerne les différences de traitement non justifiées et fondées sur le genre. Dès lors, une analyse des possibles risques de discrimination algorithmique dans le recrutement automatisé s'impose à la lumière des caractéristiques et du fonctionnement des algorithmes.

Des algorithmes sont utilisés dans plusieurs phases liées au marché du travail, de la publication des avis de vacance de poste, pour le recrutement au sens large (tri et analyse automatique des CV et lettres de motivation, gestion automatique et suivi des candidatures, invitation à un entretien ou rejet de la candidature, entretien par vidéo[2], etc.), l'évaluation du travail, la fixation des rémunérations et des bonus, jusqu'à la fin du contrat, ainsi que les bénéfices (indemnités de chômage, formations, etc.) accordés aux chercheurs d'emploi[3]. Si la présente contribution se concentre sur le recrutement automatisé *stricto sensu* et la discrimination sur la base du sexe, les conclusions pourront en principe être reprises *mutatis mutandis* pour les autres discriminations algorithmiques qui apparaissent dans le contexte du marché du travail.

1 Voir T. KNOBLAUCH/C. HUSTEDT, *Der maschinelle Weg zum passenden Personal*, Bertelsmann Stiftung, Stiftung neue Verantwortung, 2019, p. 10-13. Voir en général sur l'IA et l'apprentissage automatique, A. FRÜH/D. HAUX, *Foundations of Artificial Intelligence and Machine Learning*, Berlin, Weizenbaum Institute, 2022, spéc. p. 4 s., ainsi que Y. LE CUN, *Quand la machine apprend : la révolution des neurones artificiels et de l'apprentissage profond*, Paris, Odile Jacob, 2019.

2 A. FORMAN/N. GLASSER/S. JONES/M. AIBEL, « Companies Using Video Interviews Beware : New Obligations for Positions Based in Illinois », *The Computer & Internet Lawyer*, vol. 37(2), 2020, p. 1 ; pour un aperçu des différends applications, E. T. ALBERT, « AI in talent acquisition: a review of AI-applications used in recruitment and selection », *Strategic HR Review*, vol. 18 (5), 2019, p. 215-222, spéc. p. 215 s., https://doi.org/10.1108/SHR-04-2019-0024 (consulté le 5.8.2022).

3 Voir Annexe A de cette contribution – Les phases d'une possible discrimination.

De multiples logiciels de recrutement automatisé sont désormais disponibles afin de faciliter le recrutement non seulement pour les entreprises, mais également pour le public ; des entreprises comme Google offrent aux candidats des logiciels d'IA afin de s'entraîner pour des entretiens de recrutement[4]. Ce marché des produits de recrutement et son fonctionnement ne peuvent être exposés de manière exhaustive ici[5]. Afin d'en éclairer les enjeux, l'auteur se limitera à offrir un aperçu des principaux logiciels qui facilitent, avec l'aide de l'IA, le recrutement pour les employeurs. Ainsi, le tri et le classement de CV sont souvent automatisés par les entreprises. On peut distinguer les *Applicant Tracking System* (ATS)[6] – des logiciels qui facilitent et regroupent le processus de sélection des candidats – et les *Recruiting Management Systems* (RMS)[7], qui intègrent toutes les fonctionnalités d'un ATS mais incluent également la gestion des relations avec les candidats, facilitant et organisant les différentes phases du recrutement pour les employeurs. Les ATS sont utilisés par 99 % des entreprises Fortune 500 et fonctionnent de plus en plus avec de l'IA :

> *« An ATS that incorporates AI can process complex data sets, helping identify the skills that differentiate top performers* [...]. *By isolating the factors that determine success, employers and recruiters can create more targeted job postings, screen and assess skills on a more granular level, and help address weaknesses in the hiring process. Refined algorithms will provide a more nuanced, detailed portrait of an ideal candidate, and create a more personalized, frictionless experience for job seekers. »*[8]

In fine, il suffit pour l'analyse de constater que malgré l'utilisation largement répandue de ces outils, plusieurs chercheurs ont questionné leur valeur ajoutée[9], attirant l'attention sur de multiples problèmes de biais et de discriminations. Ces préoccupations rendent nécessaire de réfléchir et d'approfondir la question

4 GOOGLE, Interview WarmUp, https://grow.google/certificates/interview-warmup/ (consulté le 5.8.2022).

5 Pour une présentation plus exhaustive, voir A. KÖCHLING/M. C. WEHNER, « Discriminated by an Algorithm : A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development », *Business Research*, vol. 13, 2020, p. 795-848, https://doi.org/10.1007/s40685-020-00134-w.

6 ORACLE, *What is an Applicant Tracking System ?*, www.oracle.com/human-capital-management/recruiting/what-is-applicant-tracking-system/ (consulté le 9.8.2022).

7 Des exemples sont : *Zoho recruit* (www.zoho.com/recruit/), *Greenhouse* (www.greenhouse.io), *Freshteam* (www.freshworks.com/hrms/), *Jobvite* (www.jobvite.com) (consultés le 5.8.2022).

8 ORACLE (n. 6).

9 N. TIPPINS/F. OSWALD/S. MCPHAIL, « Scientific, Legal, and Ethical Concerns about AI-Based Personnel Selection Tools : A Call to Action », *Personnel Assessment and Decisions*, vol. 7(2), 2021, p. 1.

de la meilleure manière pour éventuellement réguler ces logiciels. Ainsi, par exemple, le fait de chercher le ou la candidat(e) « idéal(e) » ouvre la porte aux biais et stéréotypes, car l'algorithme va rechercher le profil idéal à travers les objectifs définis dans l'algorithme (par l'humain), son paramétrage ainsi que les données fournies. Une récente étude a mis en lumière le fait que ces logiciels de recrutement automatique mettent souvent à l'écart de bons profils et négligent la diversité parmi les candidats, et ceci en raison du paramétrage des algorithmes[10]. Le fonctionnement de ces logiciels peut mener à créer des inconvénients surtout pour les femmes, par le fait que les algorithmes ne cherchent pas seulement des aspects positifs dans les CV mais essaient d'analyser les éventuels « trous » dans les CV. Très souvent, ces « trous » dans la carrière résultent d'une grossesse (congé maternité souvent suivi par un congé parental) ou d'une situation familiale nécessitant d'assumer le rôle d'aidant[11]. Dans le cadre de notre analyse, il sera important d'observer le fonctionnement des mécanismes de régulation qui seront mis en place prochainement, tels que la loi de la ville de New York sur les algorithmes de recrutement qui est entrée en vigueur en janvier 2023 (voir *infra* III. D.).

La présente contribution est consacrée à la discrimination algorithmique et traite plus particulièrement de la discrimination sur la base du sexe causée par des algorithmes de recrutement. Afin de poser le cadre d'analyse, les problèmes liés à la discrimination algorithmique seront tout d'abord exposés, notamment en spécifiant le principe de discrimination, en définissant la discrimination algorithmique et en énonçant des stratégies et des solutions réglementaires pour éliminer ou réduire de telles discriminations (II.). Sur cette base, des exemples concrets de régulation des algorithmes émanant de la Commission européenne et de la ville de New York seront examinés et comparés (III.). En guise de conclusion, les multiples approches suivies pour la régulation des algorithmes de recrutement feront l'objet d'une analyse critique, avec un accent sur les perspectives du rôle du droit dans la sauvegarde des droits humains à l'âge des algorithmes (IV.).

10 Pour un apercu, voir J. FULLER/M. RAMAN/E. SAGE-GAVIN/K. HINES, « Hidden Workers : Untapped Talent », *Harvard Business School Project on Managing the future of Work and Accenture* 2021, www.hbs.edu/managing-the-future-of-work/Documents/research/hiddenworkers09032021.pdf (consulté le 9.8.2022).

11 *Ibid.*, p. 22 ; R. KARAYAN, « Les effets pervers des logiciels de recrutement, une réalité », 2021, L'usine digitale, www.usine-digitale.fr/article/etude-les-effets-pervers-des-logiciels-de-recrutement-une-realite.N1141652 (consulté le 9.8.2022).

II. La discrimination algorithmique : problèmes et solutions

Après avoir détaillé le principe de non-discrimination et défini la discrimination algorithmique (A.), cette section expose les problèmes liés à la discrimination algorithmique (B.) et évalue plusieurs solutions suggérées dans la doctrine (C.).

A. Le principe de non-discrimination et la discrimination algorithmique

Partout en Europe, le droit consacre le principe de non-discrimination, qui trouve ses origines et est ancré dans les droits humains ainsi que dans les droits fondamentaux nationaux[12]. Parmi les caractéristiques protégées par le droit se trouvent notamment l'origine ethnique, la race et le sexe. La discrimination fondée sur le sexe/le genre servira d'exemple pour la présente analyse. Par conséquent, la section suivante analysera d'abord l'interdiction de discrimination fondée sur le genre en général (1.), pour ensuite définir plus spécifiquement la discrimination algorithmique dans le cadre de l'égalité des genres (2.).

1. Le principe de non-discrimination

En droit de l'Union européenne, le principe de non-discrimination sur la base du sexe dans le domaine du marché du travail découle des art. 14 (1)(a) et 2 (1)(a), (b) de la Directive 2006/54/CE[13]. Une discrimination classique (sans intervention d'une technologie d'IA) peut se présenter sous forme d'une dichotomie, car la discrimination peut survenir d'une manière directe ou indirecte. La discrimination directe est définie en droit européen comme une « situation dans laquelle une personne est traitée de manière moins favorable en raison de son sexe qu'une autre ne l'est, ne l'a été ou ne le serait dans une situation comparable »[14].

12 S. FREDMAN/B. GOLDBLATT, « Gender Equality and Human Rights : Background Paper for UN Women's Progress of the World's Women Report », *UN Women Discussion Paper Series* 1, 2015, p. 4.

13 Directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte), JO L 204 du 26.7.2006, p. 23-36 (« Dir. 2006/54 »).

14 Art. 2 (a) Dir. 2006/54. Voir aussi en général C. TOBLER, *Indirect Discrimination : a Case Study into the Development of the Legal Concept of Indirect Discrimination under EC Law*, Bruxelles, Intersentia, 2005.

La discrimination indirecte constitue, quant à elle, une « situation dans laquelle une disposition, un critère ou une pratique apparemment neutre désavantagerait particulièrement des personnes d'un sexe par rapport à des personnes de l'autre sexe, à moins quecette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour parvenir à ce but soient appropriés et nécessaires »[15].

Le principe d'interdiction de discrimination est posé en droit suisse à l'art. 8[16] de la Constitution fédérale et concrétisé à l'art. 3 de la Loi fédérale sur l'égalité entre femmes et hommes (LEg)[17]. En droit français, une discrimination est prohibée par les art. 225-1 à 225-4 du Code pénal : « toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe [...] ».

En droit américain, on parle plutôt de *disparate impact*, mais la *anti-discrimination law* américaine poursuit un objectif identique au droit européen[18].

Bien qu'en théorie, ces définitions de la discrimination soient suffisamment larges afin d'inclure toute discrimination causée par des algorithmes, ce phénomène mérite une propre définition permettant de mieux saisir toute sa dimension et spécificité.

2. *La discrimination algorithmique*

À la lumière de ces définitions, une discrimination algorithmique peut être définie comme le traitement différent non justifié sur la base d'une caractéristique protégée par la loi dans le cadre d'une décision automatisée impliquant un algorithme. Comme l'utilisation de l'IA a souvent pour but de créer une décision individuelle et, de ce fait, d'opérer une sorte de discrimination technique, la question d'une discrimination algorithmique est généralement

15 Art. 2 (b) Dir. 2006/54. Voir aussi S. MOREAU, « What is Discrimination ? », *Philosophy & Public Affairs*, 2010, p. 143-179.

16 « Nul ne doit subir de discrimination du fait notamment de son origine, de sa race, de son sexe », art. 8 (2) Cst., et « L'homme et la femme sont égaux en droit. La loi pourvoit à l'égalité de droit et de fait, en particulier dans les domaines de la famille, de la formation et du travail », art. 8(3) Cst.

17 Loi fédérale sur l'égalité entre femmes et hommes du 24 mars 1995 (LEg ; RS 151.1). « Il est interdit de discriminer les travailleurs à raison du sexe, soit directement, soit indirectement [...] », art. 3(1) LEg.

18 M. SCHERER/A. KING/M. MRKONICH, « Applying Old Rules to New Tools : Employment Discrimination Law in the Age of Algorithms », *South Carolina Law Review*, vol. 71, 2019, p. 449-522, spéc. p. 459 ; MOREAU (n. 15), p. 143.

soulevée en tant que question juridique de fond[19]. L'élément clé est donc l'utilisation d'un algorithme dans le processus décisionnel, par exemple un algorithme de recrutement, qui prend ou influence largement la décision au lieu d'un humain[20]. C'est cette définition qui permettra de tracer le champ d'application d'une éventuelle réglementation, y compris les conditions applicables.

Le concept de discrimination algorithmique ayant ainsi été délimité, la section suivante énonce les problèmes liés à la discrimination algorithmique, notamment les biais et le bruit (*noise*).

B. Les problèmes

L'algorithme surmontant les erreurs décisionnelles de l'humain relève encore de la science-fiction. La littérature générale et juridique évoque habituellement le phénomène des biais comme principal défaut des algorithmes[21]. À ce titre, les sciences comportementales nous ont enseigné que les décisions des humains peuvent être entachées de biais et de bruit (*noise*)[22]. Les algorithmes seraient-ils en mesure de surmonter les biais qui grèvent leurs programmateurs[23] ? Dans la mesure où ils fonctionnent sur la base des données qui les nourrissent, les algorithmes sont tributaires du caractère discriminatoire et des biais de ces données, ainsi que de l'absence de certaines données. Dans le domaine des algorithmes[24], la littérature distingue les biais cognitifs et les biais

19 DÉPARTEMENT FÉDÉRAL DES AFFAIRES ÉTRANGÈRES (DFAE), *Intelligence artificielle et réglementation internationale, Rapport à l'attention du Conseil fédéral*, Berne, 2022, p. 8 s.

20 Pour une telle définition, voir par exemple la loi de New York City, Int 1894-2020 : A Local Law to amend the administrative code of the city of New York, in relation to automated employment decision tools (subchapter 25), § 20-870.

21 A. KÖCHLING/M. WERNER, « Discriminated by an Algorithm : a Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development », *Business Research*, vol. 13, 2020, p. 795-848, https://doi.org/10.1007/s40685-020-00134-w (consulté le 31.8.2022) ; J. A. PRASSL, « What if Your Boss Was an Algorithm ? The Rise of Artificial Intelligence at Work », *Comparative Labor Law & Policy Journal*, vol. 41(1), 2019, p. 123 ; J. A. PRASSL, « Regulating Algorithms at Work : Lessons for a "European Approach to Artificial Intelligence" », *European Labour Law Journal*, vol. 13(1), 2022, p. 30-50.

22 D. KAHNEMAN, *Thinking, Fast and Slow*, Macmillan 2011 ; A. TVERSKY/D. KAHNEMAN, « Judgment under Uncertainty : Heuristics and Biases », *Science*, 185 (4157), 1974, p. 1124-1131 ; D. KAHNEMAN/O. SIBONY/C. R. SUNSTEIN, *Noise : a Flaw in Human Judgment*, New York City, Little Brown, 2021.

23 B. A. WILLIAMS/C. F. BROOKS/Y. SHMARGAD, « How Algorithms Discriminate Based on Data they Lack : Challenges, Solutions, and Policy Implications », *Journal of Information Policy*, vol. 8(1), 2018, p. 78-115.

24 Voir, par exemple, pour dix différents types de biais, TILBURG INSTITUTE FOR LAW, TECHNOLOGY, AND SOCIETY, « Handbook on Non-Discriminating Algorithms », *Summary*

statistiques, ces derniers étant souvent subdivisés en biais de données, biais de variables omises, biais de sélection et biais d'endogénéité[25]. Le premier biais de données, souvent appelé *garbage in, garbage out*, évoque le fait que le bon fonctionnement et l'entraînement d'un algorithme dépendent largement des données qui lui sont fournies. Le deuxième biais de variables omises se produit notamment lorsque certaines qualités – notamment des qualités requises sur le marché du travail – ne peuvent pas être quantifiées et rendues opérables pour un algorithme. Le troisième biais de sélection s'observe lorsque les données ou les caractéristiques d'un groupe de personnes analysées se distinguent de l'ensemble des personnes de la catégorie. Le quatrième biais d'endogénéité s'observe lorsque les algorithmes se basent sur les données du passé afin de prédire le futur sans pour autant comprendre que des événements passés ne se reproduiront pas nécessairement dans le futur de la même manière.

L'origine du biais peut également se trouver dans le design de l'algorithme[26]. Les biais peuvent en effet avoir leur origine soit dans le design soit dans les données. On parle souvent de *machine bias*[27] pour décrire le phénomène des algorithmes biaisés. Dans de tels cas, la discrimination algorithmique est en partie due à des problèmes d'informations, soit il n'y en a pas ou pas assez, soit il y en a trop ou les données ne sont pas correctes, soit il y a des informations qui causent un traitement différent pouvant engendrer une discrimination.

Afin de limiter les décisions biaisées et potentiellement discriminatoires, certains auteurs ainsi que des propositions législatives préconisent d'inclure l'expertise humaine dans le processus décisionnel[28].

Research Report 2021, p. 5, www.tilburguniversity.edu/about/schools/law/departments/tilt/research/handbook (consulté le 6.9.2022).

25 P. BERTAIL/D. BOUNIE/S. CLÉMENÇON/P. WAELBROECK, « Algorithmes : bias, discrimination et équité », 2019, p. 10-12, www.telecom-paris.fr/algorithmes-biais-discrimination-et-equite (consulté le 6.9.2022).

26 Pour des questions de genre et le design des architectures des choix sous l'angle des sciences comportementales, voir I. BOHNET, *What Works*, Harvard University Press, Boston, 2016.

27 H. FRY, *Hello World : How to be Human in the Age of the Machine*, New York City, Random House, 2018, p. 76.

28 Voir en général F. PASQUALE, *New Laws of Robotics : Defending Human Expertise in the Age of AI*, Cambridge, Belknap Press, 2020 ; il en va de même pour la proposition de la Commission européenne, spéc. art. 14.

C. Les solutions

Dans la doctrine, un courant optimiste[29] s'oppose à un courant pessimiste[30], alors que des voix plus nuancées émergent, complétant graduellement le champ de discussion (1.). Tous ces courants proposent des solutions différentes, parfois similaires ou complémentaires, afin d'éviter ou de corriger les décisions algorithmiques discriminatoires (2.).

1. L'état des lieux de la doctrine

Le courant optimiste[31] regroupe principalement des représentants des sciences informatiques ou économiques ainsi que du monde entrepreneurial actifs dans la création ou l'utilisation des algorithmes. Ses représentants plaident contre une régulation (trop stricte) des algorithmes, soulignant leurs effets bénéfiques[32].

Une approche plus prudente et pessimiste envers les algorithmes rassemble notamment les pouvoirs publics, les régulateurs, les parlements et les organisations internationales, qui plaident majoritairement en faveur d'une régulation et qui ont déjà proposé des projets dans ce sens[33].

29 J. KLEINBERG/H. LAKKARAJU/J. LESKOVEC/J. LUDWIG/S. MULLAINATHAN, « Human Decisions and Machine Predictions », *The Quarterly Journal of Economics*, 133(1), 2018, p. 237-293 ; J. KLEINBERG/J. LUDWIG/S. MULLAINATHAN/C. R. SUNSTEIN, « Discrimination in the Age of Algorithms », *Journal of Legal Analysis*, vol. 10, 2018, p. 113-174, https://doi.org/10.1093/jla/laz001 ; J. KLEINBERG/J. LUDWIG/S. MULLAINATHAN/C. R. SUNSTEIN, « Algorithms as Discrimination Detectors », *Proceedings of the National Academy of Sciences*, vol. 117(48), 2020, p. 30096-30100.

30 Loi de New York City, Int 1894-2020 (n. 20) ; COMMISSION EUROPÉENNE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle)* (COM/2021/206 final) ; CONSEIL DE L'EUROPE, *Recommandation CM/Rec (2020)1 du Comité des Ministres aux États Membres sur les impacts des systèmes algorithmiques sur les droits de l'homme*, adoptée par le Conseil des Ministres le 8 avril 2020, https://search.coe.int/cm/pages/result_details.aspx?ObjectId=09000016809e1124 (consulté le 31.8.2022).

31 J. KLEINBERG/H. LAKKARAJU/J. LESKOVEC/J. LUDWIG/S. MULLAINATHAN (n. 29), p. 237-293 ; J. KLEINBERG/J. LUDWIG/S. MULLAINATHAN/C. R. SUNSTEIN (n. 29), p. 113-174 ;

32 J. KLEINBERG/J. LUDWIG/S. MULLAINATHAN/C. R. SUNSTEIN (n. 29), p. 30096-30100. Notamment les entreprises du numérique (voir *infra* n. 50) et l'OCDE (voir *infra* n. 35)

33 semblent favoriser un droit souple.
Voir COMMISSION EUROPÉENNE (n. 30) ; CONSEIL DE L'EUROPE (n. 30). Voir aussi DFAE (n. 19), p. 27 : « La Suisse est bien placée pour participer à la conception du cadre réglementaire international de l'IA ». P. HACKER/E. WIEDEMANN/M. ZEHLIKE, « Towards a Flexible Framework for Algorithmic Fairness », 2020, arXiv preprint arXiv: 2010.07848 ; S. WACHTER/B. MITTELSTADT/C. RUSSELL, « Why Fairness Cannot Be

Enfin, une approche plus nuancée peut être identifiée dans la doctrine[34] et parmi quelques organismes internationaux qui, tout en reconnaissant la nécessité de régulation, préfèrent souvent une forme plus douce de régulation à des règles juridiques contraignantes[35].

2. *Les stratégies pour limiter la discrimination algorithmique*

Les différents courants doctrinaux présentés dans la présente section évoquent plusieurs stratégies concrètes afin de lutter contre une possible discrimination algorithmique[36].

a) La neutralisation des caractéristiques protégées

Une possibilité évidente, afin que l'algorithme ne puisse plus discriminer, consisterait à éliminer des données, les informations relatives aux caractéristiques protégées, telles que le sexe d'un candidat. Toutefois, bien que les données soient ensuite dépourvues d'informations explicites sur le sexe, l'analyse des données par l'algorithme, aidée par les corrélations entre différentes informations, pourrait révéler le sexe qui a été justement dissimulé ou neutralisé dans les informations. Ce phénomène est aussi connu sous le terme de *proxies*[37], des informations qui permettent aux algorithmes d'approximer une caractéristique protégée sans pour autant qu'elle soit explicitement mentionnée

Automated : Bridging the Gap between EU Non-Discrimination Law and AI », *Computer Law & Security Review*, vol. 41, 2021, p. 105567 ; D. JOOS/K. MEDING, « Anforderungen bei der Einführung und Entwicklung von KI zur Gewährleistung von Fairness und Diskriminierungsfreiheit », *Datenschutz Datensicherheit*, vol. 46, 2022, p. 376-380, https://doi.org/10.1007/s11623-022-1623-6.

34 P. BERTAIL/D. BOUNIE/S. CLÉMENÇON/P. WAELBROECK (n. 25), p. 10-12.

35 Voir, à titre d'exemple, OCDE, Recommendation du Conseil sur l'intelligence artificielle du 22 mai 2019 (OECD/LEGAL/0449), https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449 (consulté le 31.8.2022).

36 Pour une classification différente de solutions, voir P. BERTAIL/D. BOUNIE/S. CLÉMENÇON/P. WAELBROECK (n. 25), p. 13-16.

37 Pour plus d'informations sur la discrimination par proxies ou par corrélation, voir notamment F. LÜTZ, « Discrimination by Correlation. Towards Eliminating Algorithmic Biases and Achieving Gender Equality », *in* S. QUADFLIEG/K. NEUBURG/S. NESTLER (édit.), *(Dis)Obedience in Digital Societies – Perspectives on the Power of Algorithms and Data*, Bielefeld, Transcript Verlag, 2022, p. 250-293, spéc. p. 258 s.

dans les données[38]. Plusieurs études ont démontré qu'une telle stratégie de neutralisation de certaines données ne fonctionne pas ou en tous cas pas très bien[39].

b) L'utilisation des informations sensibles

Dans la mesure où l'élimination des données sensibles telles que le sexe ne permet pas de lutter contre toute discrimination, est-ce qu'il serait possible de réduire les discriminations en utilisant précisément ces informations sensibles ? L'origine de cette stratégie se trouve dans le droit européen de la protection des données, et plus particulièrement dans le RGPD qui prohibe une certaine utilisation des données sensibles ainsi que dans le Règlement IA qui autorise l'utilisation de ces données sensibles afin de mener des audits des systèmes d'IA[40]. Sans une telle autorisation, un *bias audit* visant par exemple à détecter des risques de discrimination sur la base du sexe serait quasiment impossible[41]. Donc, même si l'existence des informations sensibles constitue souvent la porte d'entrée pour des discriminations, ces mêmes informations pourraient être utiles afin de tester des systèmes et les mettre en conformité avec les règles de non-discrimination.

Les informations sur le genre pourraient en outre permettre des actions positives afin de promouvoir le sexe sous-représenté. Plus précisément, à travers la *Rooney rule*, des algorithmes pourraient par exemple promouvoir automatiquement les femmes dans la prochaine étape du processus de sélection lors d'un

38 M. VEALE/R. BINNS, « Fairer Machine Learning in the Real World : Mitigating Discrimination without Collecting Sensitive Data », *Big Data & Society*, vol. 4(2), 2017, p. 1 et 4.

39 Voir, par exemple, UNESCO, *Recommendation on Ethics of Artificial Intelligence*, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000381137 ; EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR JUSTICE AND CONSUMERS, *Algorithmic Discrimination in Europe : Challenges and Opportunities for Gender Equality and Non-Discrimination Law*, Brussels, Publications Office, 2021, https://data.europa.eu/doi/10.2838/544956 ; C. ORWAT, *Diskriminierungsrisiken durch Verwendung von Algorithmen : eine Studie, erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes*, Berlin, Nomos, 2019.

40 Voir M. BEKKUM/F. BORGESIUS, « Using Sensitive Data to Prevent Discrimination by AI : Does the GDPR Need a New Exception ? », 10.48550/arXiv.2206.03262, 2022. Sur le lien entre protection des données et discrimination, voir aussi, P. HACKER, « Teaching Fairness to Artificial Intelligence : Existing and Novel Strategies against Algorithmic Discrimination under EU Law », *Common Market Law Review*, vol. 55(4), 2018, p. 1143-1185.

41 M. BEKKUM/F. BORGESIUS (n. 40), p. 11.

recrutement[42]. Cette démarche est possible uniquement si les données contiennent des informations sur le sexe.

c) La modification des données avant leur utilisation par les algorithmes

Une proposition attractive est suggérée par COFONE, lequel part du postulat que le problème se situe principalement au niveau de l'information : « *instead of ineffectively blocking or passively allowing attributes in training data, we should modify them. We should use existing pre-processing techniques to alter the data that is fed to algorithms to prevent disparate impact outcomes* »[43].

Cette stratégie vise en quelque sorte à surmonter les inconvénients des stratégies précédentes (a) et (b), sans utiliser ni supprimer les données en question, mais en les modifiant. Plus précisément, COFONE suggère d'utiliser des techniques de modification des données afin de diminuer le risque d'une discrimination. La question de savoir comment identifier ces données en premier lieu reste cependant sans réponse lorsqu'on doit se confronter avec les problématiques évoquées ci-dessus (données sensibles et protection des données). À cela s'ajoute qu'au vu des coûts induits par une telle modification pour les entreprises, il est peu probable que celles-ci procéderont à une telle modification des données sans règles contraignantes. De ce fait, seules les entreprises poursuivant une réelle politique d'égalité entre hommes et femmes procéderont à une telle démarche coûteuse et gourmande en temps.

Comme d'autres auteurs l'ont relevé, les algorithmes de recrutement discriminant notamment les femmes, la modification des données pourrait s'inscrire dans le but d'atteindre l'objectif d'égalité entre hommes et femmes, mais une telle modification permettrait aussi d'atteindre une égalité au sens plus large[44]. BORGESIUS admet que, en dehors des aspects techniques, une modification des

42 E. CELIS/C. HAYS/A. MEHROTRA/N. VISHNOI, « The Effect of the Rooney Rule on Implicit Bias in the Long Term », *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, p. 678-689, https://dl.acm.org/doi/abs/10.1145/3442188.3445930 (consulté le 31.8.2022).

43 I. N. COFONE, « Algorithmic Discrimination Is an Information Problem », *Hastings Law Journal*, vol. 70, 2018, p. 1389.

44 A. KELLY-LYTH, « Challenging Biased Hiring Algorithms », *Oxford Journal of Legal Studies*, vol. 41(4), 2021, p. 899-928, spéc. p. 900.

règles juridiques existantes s'impose ; il relève cependant que des règles spécifiques en matière de discrimination algorithmique ne résoudraient pas tous les problèmes[45].

Bien qu'attractive à première vue, une telle solution s'avère difficile à mettre en œuvre en pratique. Elle peut néanmoins se révéler prometteuse si elle est adoptée par les entreprises spontanément ou en raison du fait qu'elle est obligatoire de par la loi.

d) Les principes généraux du droit et la flexibilité du droit

Au lieu de suggérer des méthodes nouvelles afin de réduire toute possibilité de discrimination par les algorithmes, certains auteurs proposent l'utilisation des outils du droit existants. En droit de l'Union européenne, les outils classiques du droit de la non-discrimination, telle la Directive 2006/54/CE consacrant le principe de non-discrimination sur la base du sexe dans le domaine du marché du travail, pourraient bien évidemment être invoqués dans le cadre d'une plainte ou d'une action en justice par un citoyen de l'Union européenne estimant avoir subi une discrimination par un algorithme. Les limites du droit créé avant l'émergence de la technologie IA/algorithmes ont déjà été exposées dans la présente contribution (voir *supra* II. B.), notamment la difficulté de prouver une discrimination algorithmique. Ces questions liées à la détection et à la charge de la preuve conduisent à mener des réflexions sur d'éventuelles précisions législatives ou sur l'adoption d'une approche régulatrice encadrant certains problèmes liés aux biais et discriminations des algorithmes, afin de faciliter la mise en œuvre des règles classiques du droit de la non-discrimination.

Bien que les outils existants, comme les principes généraux du droit ou le droit classique de la non-discrimination, puissent être utilisés afin de combattre les discriminations algorithmiques, il est néanmoins clair qu'en l'absence de jurisprudence en la matière, l'insécurité juridique persistera[46].

e) La théorie dite artificial immutability

D'autres auteurs encore suggèrent une nouvelle théorie dite *artificial immutability*, qui tente d'élargir le champ d'application des règles classiques

45 J. F. ZUIDERVEEN BORGESIUS, « Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence », *The International Journal of Human Rights*, vol. 24(10), 2020, p. 1572-1593 ; J. GERARDS/J. F. BORGESIUS, « Protected Grounds and the System of Non-Discrimination Law in the Context of Algorithmic Decision-Making and Artificial Intelligence », *Colorado Technology Law Journal*, vol. 20, 2022, p. 1.

46 M. SCHERER/A. G. KING/M. J. MRKONICH, (n. 18), p. 449.

du droit de la non-discrimination[47]. Cette approche considère que les algorithmes créent des catégories ou groupes de personnes (*algorithmic groups*) – par exemple la famille monoparentale ou le détenteur d'un chien – qui subissent une discrimination mais ne sont pas protégées par le droit comme c'est notamment le cas pour la discrimination basée sur le sexe ou la race[48]. D'où la nécessité selon WACHTER de protéger ces nouvelles catégories algorithmiques.

Bien que la réalité confirme ce phénomène dérangeant, la question se pose de savoir si la formation par l'algorithme de catégories aléatoires et non prévisibles en constante mutation se prête à être inscrite dans une règle de droit, comme c'est le cas pour l'interdiction de la discrimination fondée sur le sexe. Afin d'être reconnue et protégée par la loi, une catégorie doit en effet être dotée d'un certain intérêt à être protégée dans la société, en plus d'une certaine stabilité, cohérence et prévisibilité.

f) Conclusion intermédiaire

Il est clair que les stratégies présentées de manière non exhaustive ci-dessus présentent des avantages et des inconvénients. Néanmoins, elles sont toutes en mesure de contribuer à diminuer des discriminations algorithmiques, que ce soit dans un cadre non contraignant, si elles sont spontanément appliquées par les concepteurs et utilisateurs des algorithmes, ou dans le cadre d'une loi. Idéalement, ces stratégies seront combinées et intégrées dans un cadre juridique contraignant afin de garantir au maximum le respect du principe de non-discrimination pour le plus grand nombre de situations. Les modalités de la mise en œuvre d'une telle régulation seront discutées à travers deux exemples concrets issus de l'Union européenne et de la ville de New York (voir *infra* III. B. et C.).

III. Comment réguler le recrutement automatisé ?

Après avoir abordé des questions générales relatives aux possibilités de régulation en matière d'algorithmes (A.), l'analyse portera sur la proposition de l'Union européenne concernant l'utilisation de l'intelligence artificielle dans le cadre du recrutement automatisé (B.), d'une part, ainsi que la proposition de la ville de New York sur les algorithmes de recrutement (C.), d'autre part. Une comparaison entre ces deux propositions permettra de conclure cette section relative à la question de la régulation (D.).

47 S. WACHTER, « The Theory of Artificial Immutability : Protecting Algorithmic Groups Under Anti-Discrimination Law », arXiv preprint arXiv:2205.01166, 2022.

48 Voir S. WACHTER (n. 47).

A. Questions préliminaires et générales

Après quelques réflexions générales sur l'architecture et les modalités de la régulation des algorithmes (1.), cette sous-section traitera de deux moyens spécifiques permettant de contrôler les algorithmes pour éviter les problèmes de biais et discriminations (2.).

1. Réflexions générales sur la régulation des algorithmes

La doctrine a déjà suggéré plusieurs modèles de régulation[49], variant du simple laissez-faire déléguant aux entreprises et au marché le soin de s'occuper du problème[50], au droit souple (*soft law*)[51], à l'utilisation de contrats pour garantir l'*algorithmic accountability*[52], à la définition de standards par des organismes comme l'*International Standard Setting Organisation* (ISO)[53], ou à l'adoption d'une législation spécifique[54]. Les modèles de régulation doivent être adaptés à l'objet de la régulation, dans le cas d'espèce le recrutement automatisé, tout en sachant qu'aussi bien des règles générales que des règles spécifiques concernant le recrutement automatisé sont susceptibles d'atteindre les objectifs poursuivis.

49 Pour une optique globale, voir Y. MENECEUR, *L'intelligence artificielle en procès : plaidoyer pour une réglementation internationale et européenne*, Bruxelles, Bruylant, 2020 ; pour plus de détails voir F. LÜTZ (n. 37).

50 Les entreprises activent dans l'IA (par exemple Google, Microsoft, Meta) ont souvent développé des standards, *best practices* ou principes afin de minimiser les biais ou ces discriminations, mais ceux-ci lient seulement l'entreprise elle-même. Voir, par exemple, https://ai.google/principles ; www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6 ; https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/ (consultés le 5.8.2022).

51 En général sur le *soft law*, voir A. FLÜCKIGER, « (Re-) faire la loi : Traité de légistique à l'ère du droit souple », Berne, éditions Stämpfli, 2019 ; voir aussi l'exemple de l'OCDE (n. 35).

52 C. COGLIANESE/E. LAMPMANN, « Contracting for Algorithmic Accountability », *Administrative Law Review Accord*, vol. 6, 2021, p. 175-199, spéc. p. 184 ; M. E. KAMINSKI, « Understanding Transparency in Algorithmic Accountability », *in* W. BARFIELD (édit.), *Cambridge Handbook of the Law of Algorithms*, Cambridge University Press, 2020, p. 121-138.

53 L'Organisation internationale de normalisation (ISO), par exemple, développe des standards qui peuvent être intégré lors de l'élaboration des algorithmes par les développeurs, mais aussi pris en compte par les États dans leur processus législatif, ISO/IEC JTC 1/SC 42, Artificial intelligence, www.iso.org/committee/6794475.html ; ISO/IEC 23053:2022, Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML), www.iso.org/fr/standard/74438.html (consultés le 5.8.2022).

54 Pour une loi qui est entrée en vigueur en janvier 2023, voir Loi de New York City, Int 1894-2020 (n. 20). Voir *infra* III. D.

2. *Assurer le bon fonctionnement des algorithmes : entre analyses d'impact et bias audit*

De nombreux débats sont menés pour déterminer le meilleur moyen de vérifier le bon fonctionnement des algorithmes, notamment en ce qui concerne les biais et les discriminations. Ces propositions portent notamment sur les analyses d'impact (*impact assessment*)[55] (a.), les *bias audits*[56] (b.) et les contrôles internes opérés par l'entreprise créatrice de l'IA (c.).

a) Les analyses d'impact pour les algorithmes

L'objectif des analyses d'impact est d'identifier les éventuelles conséquences d'une action (de régulation) proposée[57]. Connu notamment dans les domaines de la protection des données[58], des droits humains[59], ainsi que dans l'éthique et le social, ce moyen d'évaluation se positionne de plus en plus pour vérifier les impacts des algorithmes. Ainsi, les entreprises pourraient procéder à une analyse de leurs algorithmes pour détecter des risques de biais, stéréotypes ou discriminations, soit à leur propre initiative soit en vertu d'une loi rendant obligatoire une telle analyse comme la loi de New York pour les algo-

55 Voir le Rapport du EUROPEAN LAW INSTITUTE, *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*, 2022 ; A. MANTELERO « AI and Big Data : A Blueprint for a Human Rights, Social and Ethical Impact Assessment », *Computer Law & Security Review*, vol. 34(4), 2018, p. 754-772 ; J. YAM/J. A. SKORBURG, « From Human Resources to Human Rights : Impact Assessments for Hiring Algorithms », *Ethics and Information Technology*, vol. 23(4), 2021, p. 611-623.

56 S. BROWN/J. DAVIDOVIC/A. HASAN, « The Algorithm Audit : Scoring the Algorithms that Score us », *Big Data & Society*, vol. 8(1), 2021, p. 1-8 ; B. VECCHIONE/K. LEVY/S. BAROCAS, *Algorithmic Auditing and Social Justice : Lessons from the History of Audit Studies*, 2021, p. 1-9 ; P. T. KIM, « Auditing algorithms for discrimination », *University of Pensylvania Law Review Online*, vol. 166, 2017, p. 189 ; D. METAXA *et al.*, « Auditing algorithms : Understanding Algorithmic Systems from the Outside in », *Foundations and Trends® in Human–Computer Interaction*, vol. 14(4), 2021, p. 272-344 ; E. KAZIM/A. S. KOSHIYAMA/A. HILLIARD/R. POLLE, « Systematizing Audit in Algorithmic Recruitment », *Journal of Intelligence*, vol. 9(3), 2021, p. 46.

57 J. YAM/J. A. SKORBURG, « From Human Resources to Human Rights : Impact Assessments for Hiring Algorithms », *Ethics and Information Technology*, vol. 23(4), 2021, p. 611-623, spéc. p. 617.

58 Voir, par exemple, l'analyse d'impact relative à la protection des données prévue à l'art. 35 du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (Règlement général sur la protection des données, « RGPD », JO L 119 du 4.5.2016, p. 1).

59 A. MANTELERO (n. 55).

rithmes de recrutements (voir *infra* C.). Les *impact assessments* sont fréquemment mentionnés dans les études et exigés dans quelques propositions de réglementation[60].

b) Les bias audits

Les *bias audits* peuvent être décrits comme étant un moyen de vérification plus spécifique, car ils se concentrent sur les biais et les discriminations uniquement. De tels *audits* peuvent être envisagés dans le droit souple (*soft law*) et dans les propositions de réglementation.

Au vu de la complexité et du manque de transparence du processus de décision des algorithmes et afin d'assurer la qualité du recrutement, les *bias audits* sont un moyen d'identifier des problèmes de biais et discriminations en amont de l'utilisation du système par les entreprises. Notamment en vue de l'importance du recrutement automatisé pour les individus, selon l'avis de l'auteur, un tel contrôle *ex ante* s'impose afin de vérifier que l'IA ne produise pas des résultats biaisés ou discriminatoires[61].

Les services de *bias audits* sont offerts par des entreprises et permettent de vérifier la capacité des algorithmes à éviter des biais et des discriminations[62]. Toutefois, ces services soulèvent un certain scepticisme, au vu du jeune stade de développement de la compétence en matière de *AI auditing*. La Commission européenne, par exemple, a préféré opter pour des contrôles internes effectués par les fournisseurs d'IA[63].

c) Les contrôles internes (ex-ante conformity ou self assessment)

Une troisième option afin de pouvoir s'assurer du bon fonctionnement des algorithmes consiste à faire effectuer des contrôles internes par les entreprises qui fournissent l'IA[64]. Un contrôle interne signifie que les experts de

60 Par exemple, dans les propositions de la COMMISSION EUROPÉENNE (n. 30), p. 14, du CONSEIL DE L'EUROPE (n. 30), point 3.3 (« *Regular testing, evaluation, reporting and auditing against state-of-the-art standards related to completeness, relevance, privacy, data protection, other human rights, unjustified discriminatory impacts* »), de l'OCDE (n. 35), point 2.3 b) (« *assessment mechanisms* »).

61 Pour plus des détails et des idées pour un cadre systématique d'analyse et d'audit des algorithmes de recrutement, voir E. KAZIM/A. S. KOSHIYAMA/A. HILLIARD/R. POLLE (n. 56), p. 46.

62 Par exemple, O'Neil Risk Consulting & Algorithmic Auditing, https://orcaarisk.com (consulté le 31.8.2022).

63 Voir COMMISSION EUROPÉENNE (n. 30), p. 16.

64 *Ibid.*

l'entreprise ayant développé l'algorithme vérifient sans l'intervention d'un tiers le bon fonctionnement de l'IA et le détaillent dans un rapport. Les experts pourront être aidés dans leur démarche par des logiciels d'IA qui facilitent la recherche d'éventuels biais ou l'identification des résultats discriminatoires. Cette technique de régulation est connue dans d'autres domaines, comme par exemple pour les produits pharmaceutiques où les entreprises elles-mêmes procèdent à la vérification de l'efficacité et la sécurité des médicaments pendant les essais cliniques ; ce contrôle interne est ensuite validé par les autorités compétentes avant la mise sur le marché.

Même si une telle approche, qui met la responsabilité de la qualité de l'IA sur l'entreprise qui la développe, pourrait fonctionner en principe, il est primordial que toute régulation de l'IA envisage un contrôle adéquat du processus de contrôle interne notamment par des ressources humains et techniques.

d) Conclusion intermédiaire

Un processus de contrôle interne des algorithmes avant leur mise sur le marché a clairement l'avantage d'impliquer les compétences nécessaires dans le processus, car l'entreprise qui a développé l'IA est chargée elle-même d'assurer le bon fonctionnement du contrôle. Néanmoins, ce processus de régulation pose des problèmes en termes d'impartialité et de crédibilité si les *self-assessments* ne sont pas vérifiés par une autorité de régulation. Les *bias audits* quant à eux ont l'avantage de présenter davantage d'impartialité, car une entité externe à l'entreprise a la charge de l'audit. Pour l'analyse d'impact, son efficacité dépend de la personne qui est en charge de tester les impacts des algorithmes.

Même si ces solutions présentent de nombreux avantages, reste la question de savoir à qui il incombe de juger si les auditeurs (privés) effectuant ces *bias audits* ou l'analyse d'impact effectuent leur travail correctement afin de sauvegarder les droits individuels et réduire les biais et les discriminations dans le processus de recrutement[65]. Finalement, les biais et les discriminations ne seront évités ou, au moins, réduits vraisemblablement que lorsqu'il y aura des règles légales impératives.

65 S. COSTANZA-CHOCK/I. D. RAJI/J. BUOLAMWINI, « Who Audits the Auditors ? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem », *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, p. 1571-1583, https://doi.org/10.1145/3531146.3533213 (consulté le 31.8.2022).

B. La proposition de l'Union européenne

Bien que de nature horizontale, la Proposition de Règlement européen sur l'intelligence artificielle (IA)[66], qui peut être considérée comme une réglementation sur les produits sans pour autant se baser sur une approche avec des droits individuels, mentionne le recrutement automatisé en tant qu'exemple pour les systèmes d'IA à haut risque :

> « Tout au long du processus de recrutement et lors de l'évaluation, de la promotion ou du maintien des personnes dans des relations professionnelles contractuelles, les systèmes d'IA peuvent perpétuer des schémas historiques de discrimination, par exemple à l'égard des femmes [...]. »[67]

Concernant les questions d'emploi, de gestion de la main-d'œuvre et d'accès à l'emploi indépendant, la Proposition de Règlement prévoit d'être applicable dans les domaines suivants qui sont considérés comme des systèmes d'IA « à haut risque » :

> « (a) les systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, notamment pour la diffusion des offres d'emploi, la présélection ou le filtrage des candidatures, et l'évaluation des candidats au cours d'entretiens ou d'épreuves ;
>
> (b) l'IA destinée à être utilisée pour la prise de décisions de promotion et de licenciement dans le cadre de relations professionnelles contractuelles, pour l'attribution des tâches et pour le suivi et l'évaluation des performances et du comportement de personnes dans le cadre de telles relations. »[68]

L'art. 6 par. 2 de la Proposition de Règlement est le centre névralgique des systèmes d'IA à haut risque et ouvre la voie aux conditions spéciales à respecter, notamment en ce qui concerne un système de gestion des risques (art. 9), les données et la gouvernance des données (art. 10), la documentation technique (art. 11), l'enregistrement (art. 12), la transparence et la fourniture d'informations aux utilisateurs (art. 13).

La question se pose également de la mesure dans laquelle l'art. 5, qui établit entre autres la prohibition du *social scoring* (art. 5 par. 1(c), pour les autorités

66 COMMISSION EUROPÉENNE (n. 30). Pour plus de détails sur la proposition de Règlement sur l'intelligence artificielle, voir F. LÜTZ, « Gender Equality and Artificial Intelligence in Europe. Addressing Direct and Indirect Impacts of Algorithms on Gender-Based Discrimination », *ERA Forum* 2022, p. 33-52.

67 Considé. 36.

68 COMMISSION EUROPÉENNE (n. 30), Annexes 1-9, l'art. 4 (a) et (b) de l'Annexe III qui spécifie les systèmes d'IA à haut risque visé par l'art. 6 par. 2.

publiques), pourrait trouver une utilité dans le recrutement automatisé, si des algorithmes de recrutement pour des postes auprès de l'État utilisent des données qui ne sont pas strictement nécessaires pour le profilage des candidats. Si un algorithme de recrutement se rapproche du *social scoring*, on pourrait se demander s'il s'agit d'un système d'IA à haut risque, en principe autorisé uniquement sous conditions en vertu de l'art. 6, ou bien s'il s'agit plutôt d'une pratique d'IA prohibée par l'art. 5.

Enfin, l'art. 14 pose le principe d'un contrôle humain afin de faire mieux respecter les droits fondamentaux. Compte tenu du fait que l'architecture de la Proposition de Règlement suggère qu'une substitution de l'humain par les algorithmes est en principe envisageable et autorisée, ce qui implique que le recrutement n'est plus une prérogative humaine, un contrôle renforcé par l'humain est impératif afin d'assurer un recrutement automatisé en conformité avec le droit de la non-discrimination et les droits humains.

Sans faire l'objet d'une régulation spécifique afin d'éviter les discriminations, le recrutement automatisé sera ainsi soumis à une régulation et pourra de cette manière faire objet d'un certain contrôle juridique à travers les différentes conditions mentionnées. Une telle régulation pourrait faciliter la détection de biais et de discriminations et une éventuelle protection des droits devant les administrations ou les tribunaux, notamment en ce qui concerne la collecte des preuves.

Actuellement négocié au Conseil de l'Union européenne et au Parlement Européen, la version finale du Règlement sur l'IA qui sera adoptée pourrait évoluer sans pour autant changer cette approche générale et horizontale[69].

C. La proposition de la ville de New York

La ville de New York a adopté en décembre 2021 une loi qui adresse spécifiquement la problématique du recrutement automatisé afin d'éviter des biais et des résultats discriminatoires[70]. Cette loi, qui entrera en vigueur en janvier 2023, définit un système de recrutement automatisé (*automated employment decision tool*) comme : « *any computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence,*

69 Pour le *statu quo* du progrès législatif, voir OBSERVATOIRE LÉGISLATIF DU PARLEMENT EUROPÉEN, *Législation sur l'intelligence artificielle* (2021/0106(COD)), https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en (consulté le 31.8.2022).

70 Loi de New York City, Int 1894-2020 (n. 20).

that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions that impact natural persons »[71].

L'élément clé de la régulation est l'obligation d'effectuer un *bias audit* avant la première utilisation, ce qui nécessite : « *an impartial evaluation by an independent auditor. Such bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool's disparate impact on persons* [...] »[72].

Une décision en matière d'emploi inclut toute sélection pour des candidats à l'embauche ou en vue d'une promotion dans la ville de New York. Plus précisément, la loi s'applique aux entreprises localisées dans la ville de New York ainsi qu'aux résidents de la ville[73].

Le principe général posé par la loi est qu'il est interdit pour les employeurs ou les agences d'emploi d'utiliser un tel système de recrutement automatisé, sauf si un *bias audit* a été effectué au moins une année avant l'utilisation du logiciel et qu'un résumé des résultats du *bias audit* le plus récent a été rendu public sur un site web avant l'utilisation[74].

Concernant les conditions d'information et de publicité, il y a une obligation d'informer sur l'utilisation d'un système de recrutement automatisé, les qualifications et caractéristiques utilisées par le système, ainsi que les informations et types de données récoltés[75].

Des pénalités sont imposées en cas de non-respect de cette loi, avec un maximum de USD 500 pour une première violation, et entre USD 500 et USD 1 500 pour chaque violation consécutive[76]. Il faut noter que chaque jour où un système de recrutement automatisé est utilisé compte comme une violation[77]. L'omission d'informer le candidat ou la candidate de l'utilisation d'un système de recrutement automatisé constitue également une violation[78].

71 Loi de New York City, Int 1894-2020 (n. 20), § 20-870.

72 *Ibid.*

73 Loi de New York City, Int 1894-2020 (n. 20), § 20-871 b.

74 Loi de New York City, Int 1894-2020 (n. 20), § 20-871 a (1).

75 Loi de New York City, Int 1894-2020 (n. 20), § 20-871 b (1), (2), (3). En Suisse, à partir de 2023, l'art. 23 de la nouvelle Loi fédérale sur la protection des données prévoira un devoir d'informer en cas de décision individuelle automatisée. Pour ne pas avoir une décision automatisée, l'employeur devrait garder un contrôle sur la décision et pouvoir s'éloigner de la décision.

76 Loi de New York City, Int 1894-2020 (n. 20), § 20-872 a.

77 Loi de New York City, Int 1894-2020 (n. 20), § 20-872 b.

78 Loi de New York City, Int 1894-2020 (n. 20), § 20-872 c.

D. Comparaison entre les propositions de l'Union européenne et de la ville de New York

D'un côté, on peut voir que la proposition de l'Union européenne prend la forme d'une régulation générale et horizontale qui, même si elle n'adresse pas directement la question de la discrimination algorithmique dans le domaine du marché du travail, semble avoir la problématique en tête. De l'autre, la proposition de la ville de New York fait le choix d'établir une loi qui s'adresse uniquement aux algorithmes de recrutement.

Il est trop tôt pour savoir si la régulation proposée par l'Union européenne va atteindre les buts souhaités, notamment à travers les différents critères de transparence et de responsabilité, ou si une régulation plus spécifique sera nécessaire en supplément. En ce qui concerne la loi de la ville de New York, même si l'approche suivie semble la mieux adaptée pour limiter les discriminations fondées sur des caractéristiques protégées, la question se pose de savoir si le moyen choisi pour y parvenir (les *bias audits*) est le plus efficace. Comme élaboré plus haut, le contrôle des biais s'effectuerait notamment par des entreprises privées actives dans les *bias audits* et non pas par des autorités publiques, ce qui pourrait engendrer un certain risque de manque d'objectivité. Ce problème a été déjà constaté et rendu mondialement connu dans la crise financière pour les agences de notation, en lien avec l'octroi d'un rating moyennant paiement par les entreprises. Même si un contrôle effectué par des entreprises privées n'est pas problématique en tant que tel – un tel contrôle est, par exemple, pratiqué avec succès dans le secteur des médicaments –, il est souhaitable qu'un contrôle ou une vérification des entreprises établissant ces *bias audits* ou des audits eux-mêmes soit effectué ou au moins vérifié par une autorité publique.

Au vu des enjeux majeurs pour les droits fondamentaux, une implication plus forte de la puissance publique s'impose. Dans cette optique, il manque aussi dans la proposition de l'Union européenne un système explicite permettant aux individus ou aux groupes de protection des consommateurs de faire valeur leurs droits, comme c'est le cas dans le système mis en place par le RGPD[79]. Cette lacune ne peut être compréhensible que si on comprend la Proposition de Règlement sur l'IA comme un outil de régulation qui soit nécessitera des compléments législatif dans le futur, soit trouvera des compléments dans le droit existant – par exemple dans le droit de la non-discrimination –, afin que les individus disposent des droits nécessaires pour déposer une plainte ou saisir la justice.

79 Voir art. 77, 78 et 79 du RGPD (n. 58).

IV. Conclusion

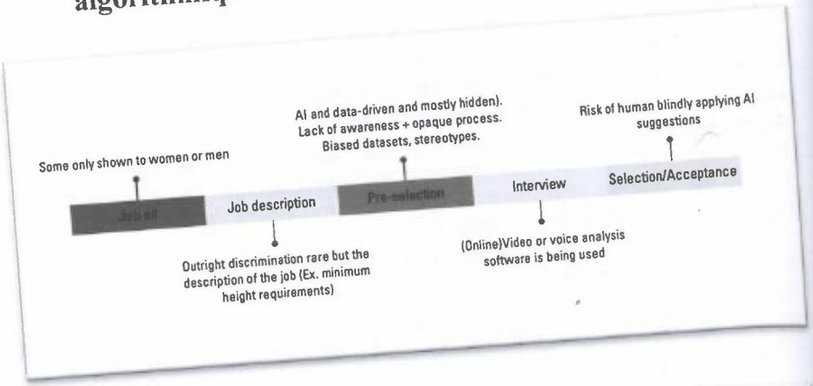
Considérant l'importance du marché du travail dans la vie de tous les citoyens, une analyse approfondie des risques de discrimination basée sur le sexe (ainsi que d'autres caractéristiques protégées par la loi) est nécessaire en cas d'utilisation d'un processus algorithmique lors du recrutement ou dans le cadre de l'évaluation des performances du travail effectué, notamment en vue d'une promotion et d'un licenciement. Une régulation s'impose dans les cas où le principe de non-discrimination ne peut être garanti que par le biais du droit. Le trio information, régulation et *effective enforcement* du droit de la non-discrimination s'impose. Sans information sur l'utilisation d'un algorithme, les victimes ne peuvent pas faire valoir leurs droits. Sans régulation, le risque de biais et de discriminations est plus élevé et les outils existants sont moins performants. Sans application effective des règles juridiques applicables en matière de discrimination algorithmique par l'administration et les juges, les discriminations algorithmiques ne vont pas cesser.

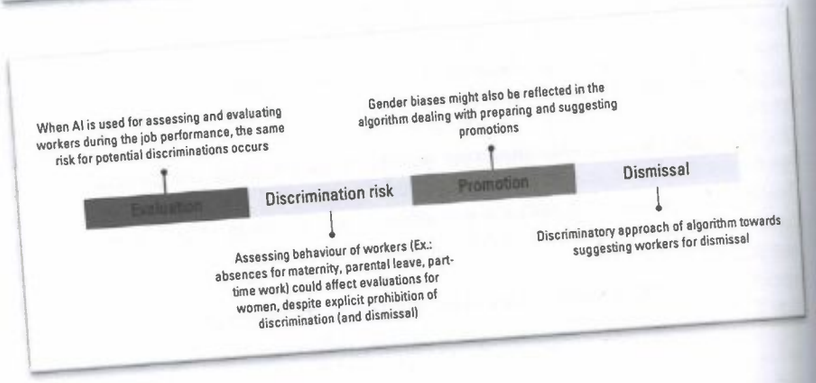
Le rôle du droit est primordial pour organiser notre société en général et poser des limites juridiques spécifiques si des nouvelles technologies telles que des algorithmes mettent en danger des droits humains, comme l'égalité des chances entre hommes et femmes, notamment dans le domaine du marché du travail. Finalement, vu le mode opératoire global des algorithmes, une régulation qui se contente d'être nationale ou régionale peinerait à suffisamment empêcher les biais et les discriminations algorithmiques. C'est la raison pour laquelle, une approche mondiale de régulation en matière d'IA s'impose, position qui est soutenue par la Suisse dans le récent rapport du DFAE sur l'IA et la réglementation internationale[80]. Une telle régulation mondiale pourra être élaborée par l'Organisation Internationale du Travail (OIT) qui regroupe, par sa nature *tripartite*, les représentants des États, des employeurs et des employées[81].

80 DFAE (n. 19), p. 27.

81 À ce stade, l'OIT n'a pas encore développé des standards ou conventions sur la question de la discrimination algorithmique des systèmes de recrutement et la réflexion sur l'IA et la discrimination dans le monde du travail reste rare. Voir par exemple les réflexions des collaborateurs du département de recherche de l'OIT, E. EKKEHARDT/R. MEROLA/D. SAMAAN, « Economics of Artificial Intelligence : Implications for the Future of Work », *IZA Journal of Labor Policy* 9, vol. 1, 2019.

V. Annexe : Les phases d'une possible discrimination algorithmique





Le *Healthy Smart Nudging* : quels enjeux juridiques ?

Les technologies cognitives comme instruments de contrainte étatique douce pour promouvoir la santé publique

AUDE GUILLOT
Doctorante dans le cadre du projet FNS Eccellenza* | Institut de droit de la santé | Faculté de droit | Université de Neuchâtel

Table des matières

* « The increasing weight of regulation : the role(s) of law as a public health tool in the prevention state » (n° 181125).

I. Introduction

L'augmentation mondiale de la prévalence des maladies non transmissibles (MNT)[1], aujourd'hui cause première de mortalité[2], se trouve au centre des préoccupations politiques internationales[3] et nationales[4] et invite ces dernières à renforcer leurs stratégies de lutte dans ce domaine. Les États se dotent ainsi de nouveaux outils, issus des domaines des sciences comportementales et du marketing social[5], les *nudges*[6] (« coups de pouce »). S'intégrant à notre quotidien, parfois même à notre insu[7], ces derniers viennent influencer sans contraintes nos choix, fondant par là même un « environnement comportemental

1 Les principales maladies non transmissibles définies par l'Organisation Mondiale de la Santé (OMS) sont l'infarctus du myocarde, les accidents vasculaires cérébraux, les cancers, le diabète et les affections respiratoires chroniques. Pour plus d'informations, voir le site de l'OMS, disponible sous : www.who.int/fr/health-topics/cardiovascular-diseases/noncommunicable-diseases#tab=tab_1 (consulté le 20.08.2022).

2 OMS, *Global Status Report on Noncommunicable Diseases 2014*, 2014, disponible sous : https://apps.who.int/iris/handle/10665/148114 (consulté le 20.08.2022). Pour la Suisse : *Nationale Strategie Herz- und Gefässkran-kheiten, Hirnschlag und Diabetes, 2017-2024*, CardioVasc Suisse, Berne 2016, disponible sous : www.sgedssed.ch/fileadmin/user_upload/1_ueber_uns/Nationale_Strategie_Herz-_und_Gefaesskrankheiten__Hirnschlag_und_Diabetes_2017-2024.pdf (consulté le 20.08.2022). Voir ég. OBSAN Swiss Health Observatory, *Gesundheit in der Schweiz – Fokus chronische Erkrankungen, Nationaler Gesundheitsbericht 2015*, Neuchâtel 2015, disponible sous : www.obsan.admin.ch/en/node/3112 (consulté le 20.08.2022).

3 Voir à cet égard le point 3.4 des « Objectifs de développement durable des Nations unies et le Plan d'action mondial pour la lutte contre les maladies non transmissibles 2013-2020 » de l'OMS. Cf. site Internet de l'OMS : *OMS. Maladies non transmissibles, 1er juin 2018*, disponible sous : www.who.int/fr/news-room/fact-sheets/detail/noncommunicable-diseases (consulté le 11.04.2022).

4 Voir par exemple, concernant la Suisse, les politiques de santé définies par la *Stratégie MNT 2017-2024* et la *Santé 2030*, réalisées par l'Office fédéral de la Santé publique (OFSP). Voir ég. OFSP, *Plan de mesures 2021-2024 de la Stratégie nationale Prévention des maladies non transmissibles (stratégie MNT) 2017-2024*, août 2020, disponible sous : www.bag.admin.ch/dam/bag/fr/dokumente/nat-gesundheitsstrategien/ncd-strategie/ncd-massnahmenplan-2021-2024.pdf.download.pdf/NCD_Massnahmenplan%202021-2024_FR.pdf (consulté le 11.04.2022). Voir aussi OFSP, *Santé 2030,* disponible sous : www.bag.admin.ch/bag/fr/home/strategie-und-politik/gesundheit-2030.html (consulté le 11.04.2022).

5 Voir à cet égard, V. BEZENÇON, *L'application des approches comportementales à l'action publique, Analyse internationale et pistes de réflexion pour la promotion de la santé en Suisse*, Document de travail n° 60, Promotion Santé Suisse, Berne/Lausanne, 2021.

6 R. H. THALER/C. R. SUNSTEIN, *Nudge : The Gentle Power of Choice Architecture*, New Haven, 2008, p. 6.

7 A. FLÜCKIGER, « Gouverner par des "coups de pouce" (*nudges*) : instrumentaliser nos biais cognitifs au lieu de légiférer ? », *Les Cahiers du droit*, vol. 59, n° 1, 2018, p. 203-207 et p. 218.

incitateur »[8] (*Choice Architecture*)[9]. Ils s'envisagent comme « tout aspect de l'architecture du choix qui modifie le comportement des personnes de manière prévisible sans interdire aucune option ni modifier de manière significative leurs incitations économiques »[10]. Spécifiquement appliqués au domaine de la santé publique, les *healthy nudges*[11], en tant que techniques d'incitation douce[12], influent sur nos décisions pour la promotion d'un mode de vie sain[13]. L'objectif recherché est de rendre le choix de la santé le plus aisé, voire même, celui défini par défaut[14].

Si les *nudges* matériels ne sont plus à présenter tant les stickers de distanciation sociale nous ont marqués pendant la pandémie du Covid-19, grâce aux outils informatiques, à l'intelligence artificielle (IA) et aux objets connectés, ils s'immiscent aujourd'hui dans tous les aspects de nos vies. Encore assez méconnu, le « *smart nudging* » se conçoit comme toute utilisation des technologies cognitives (domaines informatiques imitant les fonctions du cerveau humain)[15] visant à influencer le comportement des individus, sans limiter les options qui leur sont proposées[16]. La portée des sciences comportementales se trouve remise en question, notamment quant à la compréhension conventionnelle des facteurs influençant la capacité d'agir, dont l'impact sur les personnes lors de leur mise en œuvre par les technologies intelligentes reste mésestimé[17]. Ainsi, l'évolution et l'utilisation croissante du *smart nudging*, notamment dans le domaine de la santé, suscitent beaucoup d'interrogations quant à la sphère de protection de nos droits fondamentaux, mettant par là même une pression sur le droit qui se doit d'appréhender ces avancées[18].

8 *Idem*, p. 199.

9 THALER/SUNSTEIN (n. 6), p. 6.

10 *Ibidem*. Voir ég. C. MELE *et al.*, « Smart Nudging : How Cognitive Technologies enable Choice Architectures for Value Co-creation », *Journal of Business Research*, vol. 129, 2021, p. 951 s. et réf. cit.

11 T. MARTEAU *et al.*, « Judging Nudging : Can Nudging Improve Population Health ? », *The British Medical Journal (BMJ) Clinical Research*, vol. 342, 2011, p. 263-265. Voir ég. A. GUILLOT/M. LÉVY, « Les Healthy nudges : quel potentiel comme outil de santé publique ? », *Revue médicale suisse*, vol. 18, 2022, p. 1398-1401.

12 Voir à cet égard, le communiqué de presse de l'UniGE du 17 janvier 2022, disponible sous : www.unige.ch/communication/communiques/2022/inciter-au-lieu-de-contraindre-le-nudge-prouve-son-efficacite (consulté le 29.08.2022).

13 L. GOSTIN, « Public Health Law in a New Century, Part. 1 », *The Journal of American Medical Association (JAMA)*, vol. 283, n° 21, 2000, p. 2837.

14 *Ibidem*.

15 Datafranca, Wikipédia, *L'encyclopédie libre*, disponible sous : https://datafranca.org/wiki/Technologies_cognitives (consulté le 25.08.2022).

16 MELE *et al.* (n. 10), p. 949.

17 C. MELE/T. RUSSO-SPENA, *Innovation in Sociomaterial Practices : The case of IoE in the healthcare ecosystem*, Handbook of service science, 2019, p. 517-544.

18 MELE *et al.* (n. 10), p. 949.

Cette contribution souhaitant analyser les problématiques juridiques, sous l'angle du droit suisse, que peut emporter l'utilisation du *healthy smart nudging* dans la lutte contre les maladies chroniques (III), il apparaît nécessaire, dans un premier temps, de comprendre comment ces nouvelles technologies modifient un environnement contextuel déterminé (II). Nous aborderons ainsi succinctement les notions de l'économie comportementale, du *healthy nudge*, du *smart nudging* et de l'informatique cognitive. Une description des types d'architecture du choix créés par ces *nudges* digitaux nous permettra d'envisager la manière dont ils personnalisent en temps réel nos environnements pour influer sur nos décisions personnelles et la manière dont ils pourraient contribuer à la réalisation d'une santé publique de précision[19].

II. Modulation des écosystèmes de services dans un objectif de promotion de la santé

« [L']architecture n'est rien d'autre qu'un façonnage permanent de l'immersion »[20].

Empruntée au domaine de l'art, la définition de Peter Sloterdijk illustre la manière dont la création et la modification des environnements qui nous entourent sont possibles. S'il convient, dans un premier temps, d'étudier les nouvelles formes d'interactions homme-machine (II.A), nous analyserons ensuite plus particulièrement le *smart nudging* et les architectures numériques qui peuvent découler de leurs différentes combinaisons (II.B).

A. Du *nudge* aux technologies cognitives

Mettant en avant les biais cognitifs existants chez tous les êtres humains, Richard H. Thaler et Cass R. Sunstein proposaient en 2008, la théorie du « coup de pouce » afin d'améliorer la compréhension du comportement décisionnel. L'éclairage de l'économie comportementale a permis la mise en exergue d'une rationalité limitée, influencée par nos biais cognitifs[21] et notre environnement

19 G. MARKS SULTAN *et al.*, « Santé personnalisée : définition, caractéristiques et perspectives pour le futur », *Revue médicale suisse*, vol. 17, 2021, p. 654-657.

20 P. SLOTERDIJK, *L'architecture comme art de l'immersion*, trad. réalisée par O. MANNONI, *Le Visiteur n° 26*, Paris, Société française des architectes et Infolio, 2021, publié dans le Journal Grand Continent le 10 août 2022.

21 FLÜCKIGER (n. 7), p. 211.

sociétal[22]. L'objectif est d'inciter les personnes au passage à l'action, abstraction faite de toute contrainte, force de loi ou restriction du choix[23].

Dès 1999, Lawrence Lessig[24] met en avant la possibilité et les différentes modalités permettant de recourir à une architecture d'Internet/du web[25]. En 2000, il a rappelé que de nombreux facteurs régulent les comportements ; la loi n'étant qu'une possibilité parmi d'autres[26]. Pour ce qui est du cyberespace, il convient de considérer l'« architecture » comme étant un régulateur. La particularité de cette structure réside dans le fait que ces « contraintes sont appliquées par le pouvoir physique d'un contexte ou d'un environnement »[27].

Concevoir Internet comme une architecture numérique, *i.e.* une plateforme constituée d'un ensemble de codes[28] définissant la manière dont les utilisateurs vont interagir avec le cyberespace, renforce l'idée de Lawrence Lessig selon laquelle cet environnement est quelque chose de socialement construit[29]. Additionnées à cette structure, les technologies cognitives (dont le *smart nudging*) peuvent faciliter la réglementation des gouvernements[30]. L'architecte peut définir quelles améliorations sont possibles au sein d'une architecture définie[31]. Pensée en termes concurrentiels, cette faculté peut moduler non seulement l'efficacité mais également engendrer des discriminations[32], puisque affectant le contenu autorisé[33]. À l'échelle politique, les luttes de pouvoir peuvent se jouer « sur cette scène définie par l'architecture de l'espace »[34] offerte aux individus. Se posent ainsi les questions de savoir quels acteurs détiennent cette compétence et dans quelle mesure, ainsi que, en santé publique, la légitimité de l'État à intervenir et dans quelles limites[35].

22 MELE *et al.* (n. 10), p. 951 et réf. cit.

23 R. H. THALER/C. R. SUNSTEIN, *Nudge : Improving Decisions about Health, Wealth, and Happiness*, New Heaven 2008, traduit de l'anglais par M.-F. PAVILLET, *Nudge, Comment inspirer la bonne décision*, Paris, 2010, p. 5. Voir ég. FLÜCKIGER (n. 7), p. 199-227.

24 L. LESSIG, *Codes and other Laws of Cyberspace*, New York, Basic Books, 1999, p. 86 s.

25 FLÜCKIGER (n. 7), p. 206.

26 L. LESSIG, « Architecting for Control, Draft 1.0 », 2000, p. 2, disponible sous : https://cyber.harvard.edu/works/lessig/camkey.pdf (consulté le 09.08.2022).

27 *Idem*, p. 3 (traduction de l'autrice).

28 *Idem*, p. 5.

29 *Ibidem*.

30 *Idem*, p. 7.

31 *Idem*, p. 9 s.

32 Concernant les préjugés fondés sur le genre et la discrimination algorithmique dans le contexte de l'intelligence artificielle, voir F. LÜTZ, « Gender Equality and Artificial Intelligence in Europe, Addressing Direct and Indirect Impacts of Algorithms on Gender-Based Discrimination », *ERA Forum*, vol. 23, 2022, p. 33-52.

33 LESSIG (n. 26), p. 10.

34 *Idem*, p. 13 (traduction de l'autrice).

35 *Ibidem*.

B. *Smart nudging* et architectures du choix

De nos jours, la majorité des choix et décisions est réalisée avec le support des appareils numériques ou connectés. La combinaison de différents *nudges* digitaux (B.1) crée une architecture du choix ou encore un environnement de choix numérique (*Digital choice environments*), *i.e.* des interfaces utilisateurs ou écrans affichés impliquant une prise de décision de la part des personnes connectées (B.2)[36].

1. Nudges digitaux et co-création de valeur

La réponse des personnes aux différents mécanismes d'action utilisés par le *digital nudging* est essentielle pour la conception des nouveaux sites numériques. À ce jour, des chercheurs proposent l'utilisation d'un modèle de processus de conception des *nudges* numériques qui appréhende la manière dont les systèmes d'information eux-mêmes peuvent personnaliser ces derniers pour correspondre aux caractéristiques uniques des utilisateurs connectés. Parallèlement, se développent également les appareils intelligents ou encore les applications aptes à faire des incitations douces à certains comportements. L'optimisation de la portée de ces « coups de pouce » est fondée sur le suivi et l'analyse des données en temps réel et sur la personnalisation de l'interface utilisateur (p.ex. notifications, de popup[37], messages dans la barre d'état, etc.)[38].

Les logiciels intelligents (*Intelligent Software Agents (ISAs)*) peuvent mettre à profit cette technologie du *smart nudging* et font correspondre options et préférences de l'utilisateur ciblé, à différents moments opportuns[39]. Les mécanismes d'action privilégiés sont l'usage d'éléments de conception des interfaces utilisateur pour guider le choix et pour influencer les entrées des différents utilisateurs en ligne[40]. Trois éléments coordonnés sont déterminants : technologies cognitives, architecture du choix et co-création de valeur (*value co-creation*)[41]. Concernant cette dernière et par analogie aux analyses présentées par Stephen L. Vargo et Robert F. Lusch[42] ainsi que par Christian Grönroos[43], on retient, dans ce contexte, qu'une co-création de valeur est une notion considérant un écosystème de services[44] au sein duquel les ressources sont considérées comme utiles à la personne[45] ; l'expression possible de son avis par cette dernière sur un site Internet donné est une source de valeur et d'autosatisfaction[46]. Ainsi, par leur navigation, les personnes sont créatrices de valeur[47].

L'intégration du *smart nudging* aux différentes technologies cognitives existantes permet de faciliter le passage à l'action et l'augmentation de la co-création de valeur[48]. Quatre types d'utilisation de *nudges* ont été principalement identifiés, selon leur incorporation à[49] :

- des objets connectés (*smart wearables)* : dispositifs technologiques portés comme accessoires, stockant les données et le contexte des acteurs en temps réel ;
- des outils intelligents : manipulables mais non portables, alimentés par l'IA et doté d'apprentissage automatique ;
- des agents conversationnels (ou robots sociaux) : différents types d'interactions sociales et de cognition, pouvant réagir à un environnement social ; et
- des plateformes intelligentes : exploitant l'apprentissage automatique dans un but d'amélioration de la fiabilité, des performances, de la sécurité du traitement et de l'analyse des données, ainsi que de l'interconnexion entre les acteurs.

Les plateformes intelligentes ont également la possibilité de s'intégrer directement aux applications mobiles ou à Internet. Leur utilisation permet d'accroître les objectifs visés, de manière indépendante à la localisation de l'utilisateur ou de l'appareil usité[50].

36 MELE *et al.* (n. 10), p. 952 et réf. cit.

37 Les fenêtres popup apparaissant sur la page Internet des utilisateurs effectuant des recherches sur des sites privés sont déjà bien connues. On peut citer en exemple, des popup informant du nombre de consommateurs intéressés par le même produit, ou encore du nombre de produits restants, etc. Les mécanismes utilisés sont alors l'urgence et l'effet de rareté qui invitent la personne à commander rapidement.

38 MELE *et al.* (n. 10), p. 952 s. et réf. cit.

39 *Idem,* p. 953 et réf. cit.

40 M. WEINMANN *et al.*, « Digital Nudging », *Business and Information Systems Engineering*, vol. 58, nº 6, 2016, p. 433-436.

41 MELE *et al.* (n. 10), p. 954 et réf. cit.

42 S. L. VARGO/R. F. LUSCH, « Evolving to a New Dominant Logic for Marketing », *Journal of marketing*, vol. 68, nº 1, 2004, p. 1-17.

43 C. GRÖNROOS, « Value Co-creation in Service Logic : A Critical Analysis, Marketing Theory », vol. 11, nº 3, 2011, p. 279-301.

44 MELE *et al.* (n. 10), p. 950 et réf. cit.

45 GRÖNROOS (n. 43), p. 279-301, spécifiquement p. 290 s. (traduction de l'autrice).

46 P.-N. SCHWAB, *The Antecedents and Consequences of Politeness in a Complaint Handling Setting*, thèse (Université libre de Bruxelles, Solvay Brussels School of Economics and Management), 2015, p. 9.

47 MELE *et al.* (n. 10), p. 950 s. Voir ég. l'analyse de SCHWAB (n. 46), p. 36 s.

48 MELE *et al.* (n. 10), p. 954 et p. 958.

49 *Idem,* p. 954.

50 MELE *et al.* (n. 10), p. 954.

2. *Modélisation des architectures du choix numériques*

Combinés, les *nudges* digitaux permettent la conception de différents types d'architectures du choix, telles que[51] :

- L'Architecture du choix élargissant l'accessibilité aux ressources : accessibilité accrue aux bases de données en temps réel (données non disponibles autrement ; p.ex. informations, base de données massives, relations et interactions multiples) qui facilite la prise de décision[52]. Les applications pratiques sont nombreuses et s'illustrent notamment dans les vêtements intelligents, les robots sociaux[53], les applications comme *Nutrino*[54], ou encore le *Fibtbit Coach Ace*[55,56]. Ces métadonnées documentent et reconstruisent de manière continue la vie quotidienne des acteurs afin de fournir des recommandations adéquates pour l'avenir[57].
- L'Architecture du choix d'engagement étendu : stimulant la cognition, les *nudges* vont suivre en temps réel les habitudes des personnes qui se sentent engagées dans le but d'améliorer leurs objectifs et leur santé. Intégrés à notre quotidien, on trouve ainsi un outil d'IA tel que la *Fourchette HAPIfork* influençant les habitudes alimentaires[58], les robots intelligents tels que *ElliQ* ou *Mabu* peuvent aider les patients à suivre leur régime de médication face aux défis des maladies chroniques (la structure des conversations se fonde alors sur les modèles psychocomportementaux qui facilitent le changement de comportement)[59].
- L'Architecture du choix renforçant l'action humaine en favorisant l'automatisme et l'autonomisation : élaborer « des données sur le comportement d'un acteur pour l'informer sur ses actions, ses intérêts et ses choix au fil du temps »[60]. Pour un exemple, le *CaféWell Concierge* utilise l'IA (langage naturel adjoint à une intelligence spatio-temporelle) afin d'encourager les patients à effectuer des choix à tout moment de manière autonome[61].

51 *Ibidem.*

52 *Ibidem.*

53 Ces robots sociaux augmentent la capacité d'action des utilisateurs en fournissant des contenus dynamiques proposés en fonction de leurs choix et comportements récurrents.

54 Cette application encourage les futures mères à faire les bons choix pour leur santé via notamment des conseils nutritionnels adaptés en temps réel, personnalisés et contextuels.

55 Dispositif portatif qui analyse les activités quotidiennes des enfants et leurs comportements, notamment les minutes actives et le temps de sommeil. Il propose des activités personnalisées selon les objectifs prédéfinis à atteindre.

56 MELE *et al.* (n. 10), p. 954 s.

57 *Idem,* p. 954 s. et réf. cit.

58 *HAPIfork* est un appareil intelligent ayant la capacité d'inciter à la modification du comportement individuel en émettant des voyants lumineux et en réalisant des vibrations douces afin d'avertir les utilisateurs qui mangeraient trop vite. Une action influençant la pondération peut être envisagée avec les tendances diététiques. Les données récoltées portent sur les repas de l'utilisateur ainsi que sur les minutes/intervalles par « portion de fourchette ».

59 MELE *et al.* (n. 10), p. 955 s.

III. *Healthy smart nudging* : enjeux juridiques

Aborder la question du *healthy smart nudging* dans le domaine de la santé publique suppose l'étude de l'usage des technologies cognitives par les autorités publiques. Quelle légitimité aurait un État à recourir à de telles architectures numériques/instruments de contrainte douce ? Quels acteurs pourraient se prévaloir de la compétence d'œuvrer en tant qu'architecte[62] ? Quelle autorité aurait la compétence de pouvoir trancher entre les intérêts divergents ? Quel type de responsabilité envisager[63] ?

À l'instar de toute activité étatique, cette approche fondée sur un écosystème de services multisystémiques caractérise une nouvelle forme de gouvernance en santé et ne doit servir que des objectifs d'intérêt public[64], dont l'un d'entre eux s'identifie dans la lutte contre les MNT[65]. À ce jour, le rapport de l'unité Surveillance, suivi et notification du Département des MNT de l'OMS[66] interpelle quant au recul des mesures politiques de lutte prises contre ces pathologies chroniques. Plusieurs campagnes ont montré leur efficacité mais il convient désormais de renforcer celle-ci, par adjonction d'autres outils novateurs ou encore par le développement de la coopération entre les différents secteurs.

60 *Idem,* p. 956 (traduction de l'autrice).

61 *Ibidem.*

62 Le terme d'architecte est envisagé dans cette contribution comme désignant toute personne ou entité qui créé une architecture du choix déterminée. Voir à cet égard THALER/SUNSTEIN (n. 6), p. 6 s.

63 Pour une réflexion similaire applicable au *nudges*, voir FLÜCKIGER (n. 7), p. 211.

64 *Idem,* p. 210.

65 Les MNT, ou pathologies chroniques, affectent aujourd'hui majoritairement la population suisse en termes de mortalité et morbidité et représentaient déjà il y a dix ans, 80 % des coûts de santé. À ce sujet, voir : OFFICE FÉDÉRAL DE LA STATISTIQUE (OFS), *Maladies, Maladies et problèmes de santé (2017)*, disponible sous : www.bfs.admin.ch/bfs/fr/home/statistiques/sante/etat-sante/maladies.html (consulté le 12.08.2022) ; voir ég. OFSP, *Faits et chiffres : Maladies non transmissibles (2013),* disponible sous : www.bag.admin.ch/bag/fr/home/zahlen-und-statistiken/zahlen-fakten-nichtuebertragbare-krankheiten.html (consulté le 12.08.2022). Voir ég. S. WIESER *et al., Die Kosten der nichtübertragbaren Krankheiten in der Schweiz, Schlussbereicht im Auftrag des Bundesamts für Gesundheit (BAG)*, Abteilung Nationale Präventionsprogramme, Berne 2014.

66 OMS, *Suivi des progrès dans la lutte contre les maladies non transmissibles* [*Non communicable diseases progress monitor 2022*], 2022, p. IV, disponible sous : https://apps.who.int/iris/rest/bitstreams/1424793/retrieve (consulté le 25.07.2022).

La santé communautaire revêtant la qualité d'intérêt public[67], elle justifie l'intervention des autorités. Toutefois, l'application des technologies cognitives à ce domaine soulève des défis juridiques d'importance (III.A). Le cadre d'analyse normatif déterminant défini par le droit de la santé publique apporte un éclairage quant aux exigences à respecter lors des différentes utilisations numériques et technologiques (III.B).

A. Utilisation par l'État d'instruments de contrainte douce

Envisager l'utilisation du *smart nudging* par une autorité étatique pose des questions de faisabilité, d'une part, et du respect du cadre légal, d'autre part. Ces *nudges* incorporés aux technologies offrent une grande flexibilité et permettent de promouvoir l'efficacité des politiques de santé publique. Ces actes non obligatoires[68] servent à exécuter de manière plus souple les objectifs définis dans les dispositions législatives[69] ; ils doivent respecter les conditions de mise en œuvre de l'activité publique (A.1) et le principe de la transparence (A.2).

1. *De l'activité étatique*

Si les connaissances informatiques et techniques offrent aujourd'hui de nombreuses possibilités d'actions, pour être implémentées, ces dernières doivent respecter certaines conditions. Ainsi, à teneur de l'art. 5 de la Constitution fédérale de la Confédération suisse (Cst. féd.)[70], outre une base légale,

67 Le Tribunal fédéral a d'ailleurs souligné, dans un arrêt majeur de 1992, que les politiques publiques et mesures associées qui ont un but de lutter contre les maladies fondent un intérêt public à l'amélioration et la protection de la santé des individus. Le Tribunal fédéral a mis en exergue qu'outre les maladies transmissibles, la santé publique englobe également la prévention des pathologies non transmissibles et tous les aspects de promotion de la santé. Si la distinction entre les intérêts privés et publics n'est pas aisée dans ce domaine, une approche globale permet de s'assurer que les intérêts de tous soient préservés (ATF 118 Ia 427, consid. 6 lettre b). Voir ég. A. S. DUPONT *et al.*, *Le doit à la santé, une perspective de droit comparé, Suisse*, EPRS, Service de recherche du Parlement européen, Unité Bibliothèque de droit comparé, mai 2022, p. 59 et réf. cit.

68 Le Tribunal fédéral a explicitement reconnu que l'intervention étatique n'était pas limitée aux décisions, mais peut être exécutée de manière informelle (ATF 128 II 156, 153).

69 A. FLÜCKIGER, « Régulation, dérégulation, autorégulation : l'émergence des actes étatiques non obligatoires », *Société suisse des juristes (SSJ), Rapports et communications*, 2004, vol. 2, § 13-15, p. 173 et réf. cit.

70 Constitution fédérale de la Confédération suisse du 18 avril 1999 (RS 101).

l'existence d'un intérêt public et le respect du principe de proportionnalité sont à considérer. Ainsi, si l'État souhaite instaurer un environnement incitateur propice à influer sur la santé des individus, l'exigence de la légalité se posera, d'une part, eu égard à la norme autorisant ou contraignant les architectes à l'instaurer[71] et, d'autre part, relativement à l'activité étatique elle-même. Concernant la prise en considération du principe de proportionnalité, la flexibilité des technologies cognitives elle-même pourra être prise en compte[72].

De surcroît, si les mesures mises en œuvre dans les politiques publiques, quand bien même qualifiées d'interventions étatiques douces, par leur portée ou leurs effets, portent atteinte à la sphère de protection des droits fondamentaux, elles doivent répondre aux exigences posées par l'art. 36 Cst. féd. À titre d'exemple, l'intégration de systèmes dits « par défaut » sur Internet portera atteinte en premier lieu à la liberté personnelle de l'utilisateur[73]. Il conviendra d'analyser au cas par cas le respect ou non des droits fondamentaux.

Alors que plusieurs études ont démontré le rôle primordial des déterminants de la santé et notamment les effets des habitudes et du milieu de vie[74], le rôle du législateur est fondamental dans la conception des politiques publiques dont l'efficience dépend totalement de leur capacité à atteindre le public visé (cf. notion de santé publique de précision)[75]. La segmentation[76] (*i.e.* le ciblage d'une certaine catégorie de personnes) revêt un rôle cardinal afin d'identifier et d'atteindre les populations qui seraient vulnérables, luttant par là même contre les inégalités et inéquités en santé.

2. *Healthy smart nudging et principe de la transparence*

L'information transmise à toute personne concernée relative à la politique publique, les objectifs visés ainsi que les mécanismes d'action usités permettent de tenir compte des exigences relatives au principe de la transparence

71 FLÜCKIGER (n. 7), p. 223.

72 FLÜCKIGER (n. 69), § 16, p. 174 et réf. cit.

73 Pour un exemple concret, on peut mentionner les systèmes de type « opt-in » ou « opt-out » choisis par les États en matière de dons d'organe. À cet égard, voir M. LEVY MADER, *Le don d'organes entre gratuité et modèles de récompense : quels instruments étatiques face à la pénurie d'organes ?*, Bâle, 2011.

74 Pour des exemples, voir O. GUILLOD/V. STAUFFER, « Le droit de la santé : bilan et perspectives », *in Droit de la santé : fondements et perspectives, Actes de la 10e journée de droit de la santé*, Institut Droit de la Santé, Neuchâtel, vol. 7, Genève, 2004, p. 43-69.

75 MARKS SULTAN *et al.* (n. 19), p. 654-657.

76 BEZENÇON (n. 5), p. 6.

dans lequel s'inscrit toute activité étatique. Ce principe a pour finalité « de contribuer à l'information du public [...] »[77] et ainsi « de garantir la libre formation de l'opinion publique et de favoriser la participation [...] »[78]. Dans la pratique, ce principe se concrétise notamment par « l'information du public par les autorités sur leurs activités »[79] et vise à « promouvoir la confiance des citoyens dans les institutions étatiques et leur fonctionnement (art. 1 de la Loi fédérale sur le principe de la transparence dans l'administration (LTrans)) »[80].

Les économistes ont découvert tôt l'avantage que représentait l'exploitation des failles pouvant impacter notre rationalité dans un objectif d'influer sur nos décisions et nos comportements[81]. L'analyse première de ces mécanismes a eu pour finalité la prédétermination du comportement des individus en les engageant, à leur insu, vers une ou plusieurs conduites en amont[82]. Bien que les technologies cognitives aient pour vocation de *nudger* une personne sur un court instant, sans contrainte et dans un but d'adoption d'un comportement souhaité, il n'en demeure pas moins qu'une entité étatique ne saurait cacher ou manipuler inconsciemment des citoyens[83]. L'atteinte à la liberté personnelle (art. 10 al. 2 Cst. féd.) s'illustrerait alors dans le sentiment trompeur de disposer d'une réelle liberté d'action, alors qu'elle serait faussée par le « paternalisme libertarien »[84]. Le principe de la transparence impose à l'État la transmission d'une information détaillée sur l'objectif visé et les mécanismes d'action usités. Cette exigence peut elle aussi se concrétiser dans les options de mise à

77 Art. 1 al. 1 de la Loi fédérale sur le principe de la transparence dans l'administration du 17 décembre 2004 (LTrans ; RS 152.3).

78 Art. 1 al. 1 de la Loi sur la transparence des activités étatiques (LTAE) du 28 juin 2006, canton de Neuchâtel (RSN 150.50).

79 Art. 1 al. 2 let. b LTAE.

80 ATF 1C_462/2018 du 17 avril 2019, consid. 3.2 (traduction de l'autrice). Voir ég. ATF 142 II 340 consid. 2.2 ; ATF 142 II 324 consid. 3.4 ; ATF 1C_462/2018 du 17 avril 2019 consid. 3.2 ; ATAF 2016/18 consid. 4.1, 2014/24 consid. 3.1. Voir ég. l'avis du Préposé à la protection des données et à la transparence (PPDT) des cantons du Jura et de Neuchâtel n° 2020.3096, *Conditions pour restreindre l'accès à un document officiel (2020.3096)*, du 14 janvier 2020, publié sur le site Internet du PPDT, disponible sous : www.ppdt-june.ch/fr/Activites/Avis/2020/Conditions-pour-restreindre-l-acces-a-un-document-officiel-20203096.html (consulté le 15.08.2022). Pour le Tribunal fédéral, ce principe de la transparence contribue lui aussi à la concrétisation de la liberté d'information de l'art. 16 Cst. féd. (ATF 1C_462/2018 du 17 avril 2019, consid. 3.2).

81 FLÜCKIGER (n. 7), p. 202-204.

82 S. BAGGIO, *Psychologie sociale*, Bruxelles, 2006, p. 28.

83 FLÜCKIGER (n. 7), p. 203-207 et p. 218.

84 *Idem,* p. 205 et réf. cit. Voir ég. R. THALER/C. SUNSTEIN, « Libertarian Paternalism », *American Economic Review*, vol. 93, n° 2, 2003, p. 175.

disposition d'un QR code[85] ou encore d'un lien hypertexte permettant d'accéder à ces différents éléments.

B. *Healthy smart nudging* et droits fondamentaux

Le législateur suisse doit prendre « des mesures afin de protéger la santé »[86], dans le respect du droit en vigueur et, notamment, relativement au respect de la sphère de protection des droits fondamentaux des individus. La synergie entre *Big data* et technologies cognitives soulève plusieurs problématiques à cet égard.

1. Santé publique : qualité de l'information et liberté personnelle

La question de la qualité de l'information transmise ou nécessaire à la mise en œuvre du *nudge* pose celle des caractéristiques que doit revêtir l'information en santé publique[87].

Au travers du prisme des droits de l'homme, les questions relatives à l'autodétermination nécessaire à la réalisation de ses propres choix en matière de santé et donc à l'exercice de sa liberté personnelle, sont liées à la liberté d'opinion et dépendent de l'information accessible[88]. Les normes internationales qui fondent le socle du droit à une liberté d'opinion et du respect de la sphère privée sont les art. 8 et 10 de la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CEDH)[89], ainsi que l'art. 19 du Pacte international relatif aux droits civils et politiques (Pacte ONU II)[90]. L'art. 9 de la Convention des Nations Unies relatives aux droits des personnes handicapées[91] exige éga-

85 Cette technologie a déjà gagné le grand public et permet l'ouverture d'un lien hypertexte sur son propre téléphone portable.

86 Art. 118 al. 1 Cst. féd. À noter à cet égard qu'il existe, en Suisse, une répartition des compétences entre la Confédération et les cantons.

87 À cet égard, nous renvoyons à la thèse de N. JOSET en cours de rédaction, *Devoir d'information en santé publique*, IDS, Université de Neuchâtel.

88 Pour un exemple concernant la liberté personnelle, voir la réflexion du Tribunal fédéral dans l'ATF 133 I 110, consid. 5 s.

89 Convention de sauvegarde des droits de l'homme et des libertés fondamentales du 4 novembre 1950 (CEDH ; RS 0.101).

90 Pacte international relatif aux droits civils et politiques du 16 décembre 1966 (Pacte ONU II ; RS 0.103.2).

91 Convention relative aux droits des personnes handicapées du 13 décembre 2006 (RS 0.109).

lement que les États assurent l'accès à l'information que les autorités ont publiée. La Déclaration universelle des droits de l'homme (DUDH)[92] envisage elle aussi ces droits relatifs à la liberté d'opinion et à la recherche d'informations en son art. 19[93].

Relativement à la thématique des technologies cognitives, un accord intergouvernemental réalisé par 43 pays a été signé, y compris par la Suisse, lors de la 74e Assemblée générale des Nations Unies en septembre 2019. Cet accord, qui aborde plus spécifiquement les questions en lien avec la mise en place des architectures du choix, prend la forme d'un Partenariat international pour l'information et la démocratie[94]. Il prône la diffusion d'une information fiable (art. 25 al. 2), d'une part, et de faire preuve de transparence concernant l'organisation de contenu par algorithmes (notamment quant à la modération des procédures de décisions humaines et techniques), la collecte des données personnelles, ainsi que relativement à tous types d'accords passés avec des gouvernements ou des entités privées (art. 26 al. 3), d'autre part. Ce partenariat se base sur les idées mises en avant par une Commission internationale lors du Forum international sur l'Information et la démocratie en 2018 qui avait relevé plusieurs problématiques. Les conclusions de cette Commission ont été regroupées dans une Déclaration internationale sur l'information et la démocratie[95] qui relève que « [q]uand elles créent les moyens techniques, les normes et les architectures du choix, les entités structurantes – entendues comme les entités contribuant à la structuration de l'espace et de l'information et de la communication – doivent respecter les principes et garanties qui assurent la nature démocratique de cet espace »[96].

Concernant le droit positif suisse, « [l]e droit à l'autodétermination est, d'un point de vue constitutionnel, rattaché à la liberté personnelle garantie par l'art. 10 Cst., avec ses prolongements dans la [...] [CEDH], le Pacte ONU II ou

92 Déclaration universelle des droits de l'homme du 10 décembre 1948 (DUDH).

93 S. GHIELMINI *et al.*, *Droits fondamentaux et droits humains à l'ère numérique*, Berne, 2021, p. 100. Voir ég. art. 8 al. 4 Cst. féd., art. 14 de la Loi fédérale sur l'élimination des inégalités frappant les personnes handicapées du 13 décembre 2002 (LHand ; RS 151.3) et art. 10 de l'Ordonnance sur l'élimination des inégalités frappant les personnes handicapées du 19 novembre 2003 (OHand ; RS 151.31).

94 Partenariat international pour l'information et la démocratie, 74e Assemblée générale des Nations Unies de septembre 2019, disponible sous : https://informationdemocracy.org/fr/principes/ (consulté le 31.08.2022). Voir ég. France, Ministère de l'Europe et des affaires étrangères, www.diplomatie.gouv.fr/fr/politique-etrangere-de-la-france/la-france-et-les-nations-unies/l-alliance-pour-le-multilateralisme/partenariat-information-et-democratie/ (consulté le 31.08.2022).

95 Déclaration internationale sur l'information et la démocratie du 5 novembre 2018, disponible sous : https://informationdemocracy.org/fr/declaration-internationale-information-democratie/ (consulté le 31.08.2022).

96 *Idem*, p. 7.

encore la Convention d'Oviedo »[97]. De manière générale, toute information officielle doit, pour renseigner le public, être faite « en temps utile et de manière détaillée » (art. 180 al. 2 Cst. féd. et art. 10 LOGA)[98, 99]. Le Centre suisse de compétence pour les droits humains (CSDH) est d'avis que « la publication par les pouvoirs publics d'informations importantes sur leurs sites Internet contribue à renforcer la liberté d'information »[100] (art. 16 Cst. féd.). L'exactitude et la qualité de ces données doivent permettre l'autodétermination de la personne. Une information, même simple, n'est pas anodine et peut modifier le comportement du destinataire[101]. Ce trait est d'autant plus marqué en santé publique, domaine dans lequel, l'information est souvent orientée ou incitatrice[102] dans un objectif de poursuite de l'intérêt public visé.

Patrice Meyer-Bisch est d'avis qu'une information appropriée permet non seulement de combattre l'asymétrie entre les acteurs, mais également de lutter contre la désinformation[103]. Cet auteur définit plusieurs critères quant au contenu, à la structure, à l'accès, au sujet, à la réalisation ainsi qu'aux responsabilités. Si plusieurs caractéristiques sont déterminantes, il convient notamment de mentionner l'accent mis sur la cohérence (diversité, croisement et contrôle des sources, etc.), la clarté (pas de complication abusive, pas de lissage de l'information, etc.) et la transparence[104]. Un utilisateur naviguant sur le web pour trouver des renseignements sur le site de la Confédération ne revêt pas la qualité de patient[105], pour autant, la qualité de l'information transmise ne devrait pas être moindre que celle transmise dans le cadre d'un contrat de soins[106].

97 ATF 2C_451/2020 du 9 juin 2021, consid. 6.2.1.

98 Loi sur l'organisation du gouvernement et de l'administration du 21 mars 1997 (LOGA ; RS 171.010).

99 Voir ég. les art. 10a, 11, 34, 40 et 54 LOGA, ainsi que FLÜCKIGER (n. 69), p. 205 et réf. cit.

100 GHIELMINI (n. 93), p. 58.

101 FLÜCKIGER (n. 69), § 73, p. 204 et réf. cit.

102 On distingue l'information orientée ou subjective en santé du *nudge* par l'absence d'adjonction d'outils de marketing social qui inciterait au passage à l'action (*call to action*). Voir ég. FLÜCKIGER (n. 69), p. 205 et réf. cit.

103 P. MEYER-BISCH, « La vie communicationnelle, Une triangulation de droits culturels, condition pour l'effectivité de tous les droits humains », *in* P. MEYER-BISCH/S. GANDOLFI/G. BALLIN (édit.), *L'interdépendance des droits de l'homme au principe de toute gouvernance démocratique, Commentaire de souveraineté et de coopérations*, Genève, 2019, p. 153-158.

104 Sous réserve d'exceptions mentionnées dans les dispositions y relatives.

105 Voir O. GUILLOD, *Droit médical*, Neuchâtel, 2020, chap. 6, p. 229 s. Voir ég. ATF 133 III 121, 123.

106 À cet égard, voir GUILLOD (n. 105), p. 287 s., notamment, relativement à l'information, § 363, p. 301 s. ; voir ég. R. CHRISTINAT, « L'autonomie du soignant comme limite au droit à l'autodétermination du patient », *in* A.-S. DUPONT/O. GUILLOD (édit.), *Réflexions romandes en droit de la santé : Mélanges offerts à la Société suisse des juristes par*

Dans le domaine de la santé et en raison de l'intérêt public poursuivi, l'information doit être donnée « en termes clairs, intelligibles et aussi complets que possible »[107].

La connaissance des citoyens des outils de marketing social[108], bien que notoires pour certains, ne suffit pas à remplacer une information complète[109] et précise sur les objectifs poursuivis par la politique de santé publique et sur les mesures mises en place. Nous sommes d'avis que la possibilité offerte à l'utilisateur d'accéder à cette information, en tout temps, par le biais de la mise à disposition sur le site d'un QR code ou d'un lien hypertexte renvoyant à une autre page, est indispensable. À ce niveau, la dichotomie qui intervient entre les mécanismes d'action des *nudges*, fondés sur les biais cognitifs des individus, par essence inconscients, et les exigences relatives à la mise en œuvre du principe de transparence interpelle et invite à la réflexion[110]. Une question additionnelle subsiste quant à savoir si la qualité de l'information doit être de surcroît personnalisée lorsque, contrairement à une simple navigation, la personne s'identifie via un *login* (espace privé). Enfin, les options de mise à disposition par l'État de logiciels ou d'applications pour téléphone présentent l'avantage, par adhésion volontaire de l'utilisateur, d'autoriser la transmission d'une information complète, avec des éléments interactifs et didactiques.

2. *De la proportionnalité et du degré de contrainte*

Pour toute mesure étatique, doivent être notamment considérés tous les aspects de restriction de l'autonomie, d'atteinte à la sphère privée (art. 13 Cst. féd.), à la liberté personnelle (art. 10 al. 2 Cst. féd.) et autres intérêts juridiquement protégés. Les actes matériels, à l'instar des normes, peuvent porter atteinte à nos droits, à différents degrés, par manipulation de notre environnement comportemental. Ainsi, un choix pour la santé défini par défaut peut être

l'Institut de droit de la santé de l'Université de Neuchâtel à l'occasion de son congrès annuel 2016, Zurich, 2016, p. 1-20.

107 ATF 133 III 121, 129. Voir ég. ATF 105 II 284.

108 Les techniques marketing sont aujourd'hui bien connues. Pour des exemples, on peut citer l'utilisation du principe de la rareté pour vendre (p.ex. il ne reste plus que deux produits), le principe de l'urgence, les stratégies de différenciation, la disposition des denrées alimentaires dans une grande surface, etc.

109 FLÜCKIGER (n. 69), p. 267.

110 Voir à cet égard H. BRUNS/E. KANTOROWICZ-REZNICHENKO, « Can Nudges Be Transparent and Yet Effective ? », *Journal of Economic Psychology*, vol. 65, 2018, p. 41-59.

anodin ou lourd de conséquences (p.ex. être ou non donneur d'organes, consentement présumé)[111]. À quel moment un environnement « incitatif » peut devenir « impératif »[112] ? Comment s'assurer du libre consentement d'une personne à effectuer une action (p.ex. s'enregistrer pour une vaccination, cliquer sur un lien hypertexte, etc.), si cette dernière a, à plusieurs reprises, été *nudgée* ? Quelle similarité avec le *consent washing*[113, 114] ? Une telle configuration n'est-elle pas propice à revêtir, par accumulation de mécanismes incitateurs, un certain degré de contrainte ?

« Être libre et agir ne font qu'un »[115]. Cette vision d'Hanna Arendt peut se voir mise à mal par les nouvelles possibilités offertes avec l'IA. La question des mesures à prendre dans un but de protection et promotion de la santé communautaire[116] doit s'évaluer au regard du principe de proportionnalité (art. 5 al. 2 Cst. féd.). La modularité du *smart nudging* permet, en effet, de solliciter une personne à plusieurs reprises, en temps réel, via divers mécanismes d'action implémentés à tous niveaux structurels et dans notre quotidien. Même s'il apparaît, au premier abord, que les mesures incitatives sont moins intrusives que celles issues de normes contraignantes[117], tant leur possibilité d'accès à la sphère privée que la récurrence possible de ces outils offerte par les mécanismes d'apprentissage de l'algorithme, remettent en question cette apparence.

111 À ce sujet, en Suisse, le principe du consentement présumé a été accepté par votation le 15 mai 2022. Pour plus d'informations, nous renvoyons au site de l'OFSP, disponible sous : www.bag.admin.ch/bag/fr/home/medizin-und-forschung/transplantationsmedizin/rechtsetzungsprojekte-in-der-transplantationsmedizin/indirekter-gegenvorschlag-organspende-initiative.html (consulté le 20.08.2022). Voir ég. LEVY MADER (n. 73).

112 FLÜCKIGER (n. 7), p. 219 s.

113 L. MAUREL, alias Calimaq, *StopCovid, la subordination sociale et les limites du « Consent Washing »*, Site -S.I.Lex-, article du 1er mai 2020, disponible sous : https://scinfolex.com/2020/05/01/stopcovid-la-subordination-sociale-et-les-limites-au-consent-washing/ (consulté le 20.08.2022).

114 Terme apparu dans le contexte de la pandémie du Covid-19 et la mise en place, en France, de l'application *StopCovid*. On entend par *consent washing*, selon la présidente de la Commission nationale de l'informatique et des libertés (CNIL), « un consentement libre et éclairé sujet à caution », *i.e.* « l'instrumentalisation possible du consentement des personnes dans un contexte de crises où des pressions de toutes sortes vont […] s'exercer […] [afin d'inciter] à utiliser ce dispositif » (MAUREL (n. 113)). Voir à cet égard l'avis de la Commission nationale consultative des droits de l'homme (CNCDH), « Un consentement libre et éclairé sujet à caution », *in Avis sur le suivi numérique des personnes*, 28 avril 2020, p. 6-11, disponible sous : www.erebfc.fr/documentation/ressource/Avissuivinum%C3%A9rique+des+personnes_CNCDH_28avril_compressed.pdf?id=216 (consulté le 10.10.2022) ; voir ég. l'avis de la présidente de la CNIL, disponible sous : www.cncdh.fr (consulté le 20.08.2022).

115 H. ARENDT, *La condition de l'homme moderne*, traduit de l'anglais par G. FRADIER, Paris, 1983, p. 43 s.

116 GOSTIN (n. 13).

117 FLÜCKIGER (n. 7), p. 224.

Dans le prisme d'une activité étatique et si l'on se concentre sur les possibilités les plus aisées d'utilisation numérique de ces outils, il est possible d'envisager, au-delà de la simple implémentation de *nudges* sur le site de la Confédération, celle des technologies cognitives et du *smart nudging* (fenêtres popup, adjonction de phrases, etc.). Si ces logiciels d'IA nécessitent un *data lake* (*i.e.* une grande base de données)[118] pour leur fonctionnement, les données personnelles de préférence utilisateurs ou de navigation (notamment la traçabilité, les cookies) sont aptes à fournir des données suffisantes, quand bien même la personne ne se serait pas identifiée. La simplicité des informations collectées lors de nos connexions nous fait souvent sous-estimer l'importance et la valeur de celles-ci. Il est ainsi possible de concevoir une interface personnalisée même en l'absence d'un profilage[119] approfondi. Le scandale de Google relatif à la diffusion par inadvertance de la vidéo « *The Selfish Ledger* » (le « Registre égoïste ») produite en 2016 par Nick Foster[120], met en exergue la dimension collective des « données personnelles »[121]. Chaque usage des appareils connectés produit un enregistrement de nos données, pouvant rendre compte de chacune des

118 Elles sont généralement créées via l'utilisation de cookies ou encore via des éléments de traçabilité.

119 Le profilage se distingue du profil par l'automatisation du traitement des données dans un but d'analyse et de prédiction des comportements (art. 4 al. 4 du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (RGPD)). Voir à cet égard, le site de la CNIL, *Profilage et décision entièrement automatisée*, avis du 29 mai 2018, disponible sous : www.cnil.fr/fr/profilage-et-decision-entierement-automatisee (consulté le 20.08.2022).

120 Vidéo « *The Selfish Ledger* », N. FOSTER, 2016, disponible sous : www.youtube.com/watch?v=LUSZfEBTwRc et sa transcription complète, disponible sous : https://vpnrevie.ws/selfish-ledger-transcript/ (consultés le 28.09.2022). Voir ég. L. MAUREL, alias Calimaq, *Google, les données sociales et la Caverne des Habitus*, Site -S.I.Lex-, article du 28 juin 2018, disponible sous : https://scinfolex.com/2018/06/28/google-les-donnees-sociales-et-la-caverne-des-habitus/ (consulté le 15.08.2022).

121 Ces données personnelles s'inscrivent notamment dans le *social graph* (graphe social). Selon F. PISANI, « le graphe social tel que conçu par BERNERS-LEE est une couche d'abstraction supplémentaire qui permet de représenter les liens entre les gens mesurés au travers de leurs interactions online (et non plus seulement les ordinateurs dans le cas de l'Internet ou les pages dans le cas du web). Le graphe sémantique en ajoute une de plus : celle des métadonnées permettant à des machines de lire les interactions en question » (F. PISANI, « GGG et graphe sémantique », article publié en ligne le 11 décembre 2007, disponible sous : www.francispisani.net/ggg-et-graphe-semantique/ (consulté le 10.10.2022)). Pour un exemple, on peut citer le « réseau des données liées » qui est particulièrement bien exploité par les réseaux sociaux tels Méta ou Twitter. Voir à cet égard, H. GUILLAUD, « Comprendre le graphe social », site Internet actu.net, article du 28 septembre 2007, disponible sous : www.internetactu.net/2007/09/28/comprendre-le-graphe-social/ (consulté le 15.08.2022).

actions réalisées. Tant l'« épigénome lamarckien »[122] de Nick Foster que le concept d'« *habitus* » de Pierre Bourdieu[123] permet à Google (ou tout autre navigateur), selon Lionel Maurel, « d'en déduire les causes qui déterminent les individus à agir dans un sens ou dans un autre »[124]. Cette connaissance fonde un *data lake* suffisant pour moduler l'architecture numérique que l'on souhaite. Il devient ainsi aisé, par addition d'un algorithme d'apprentissage, d'inviter les utilisateurs à adopter un comportement souhaité « au gré de suggestions indirectes mais programmées dans les interfaces »[125]. L'identification de la personne via un *login* permet évidemment d'augmenter la segmentation et la personnalisation donc l'efficience de l'effet escompté, et ce en temps réel.

Ces premières hypothèses emportent la possibilité de *nudger* une personne à plusieurs reprises, lors de sa navigation volontaire et anonyme (ou de manière non identifiable) sur un site officiel, avec différentes modalités d'action[126]. Une réflexion doit être ici portée sur le respect du principe de proportionnalité (art. 5 al. 2 Cst. féd.) eu égard au but visé. Si l'avantage certain que représentent les modulations en temps réel et personnalisables de l'IA, l'autorité publique ne peut en faire un usage excessif. La question de l'efficience des *nudges* sur le long terme a été soulevée à diverses reprises[127], aussi, avec les technologies cognitives, la variation des mécanismes d'action usités présente la qualité certaine de ne pas fatiguer, par habitude ou redondance, la personne *nudgée*. En effet, il convient de relever que la diminution de l'efficacité de ces « coups de pouce » (liés ou non aux technologies intelligentes), lorsqu'ils sont couplés aux

122 Dans sa vidéo, « *The Selfish Ledger* », NICK FOSTER vient détailler les techniques utilisées par Google afin de réaliser le profilage des individus dans un but de modifier leur comportement utilisateur. Pour NICK FOSTER, le « Registre égoïste » se comprend comme une accumulation de connaissances comportementales, enregistrées sur le long terme et concernant la vie d'un individu. Ce registre peut être assimilé à un épigénome numérique (FOSTER (n. 120)). Voir ég. MAUREL (n. 120).

123 P. BOURDIEU, *Esquisse d'une théorie de la pratique, Précédé de « Trois études d'ethnologie kabyle »*, Genève, 1972, p. 282.

124 MAUREL (n. 120).

125 F. TRÉGUER/L. BEAUMAIS, *Gestion techno-policière d'une crise sanitaire*, SciencesPo, Centre de recherches internationales, article du 6 mai 2020, disponible sous : www.sciencespo.fr/ceri/fr/content/gestion-techno-policiere-d-une-crise-sanitaire (consulté le 15.08.2022).

126 Un *nudge* ou l'utilisation du *smart nudging* sur un site officiel à destination d'une personne qui y est soumise lors d'une navigation volontaire et anonyme sur Internet (p.ex. visite du site de la Confédération pour y trouver tout renseignement sanitaire) diffère de toute utilisation des technologies cognitives qui atteindrait cette personne directement dans sa sphère privée (p.ex. mailing, réseaux sociaux, etc.).

127 M. MAIERA *et al.*, « No Evidence for Nudging after Adjusting for Publication Bias », *The Proceedings of the National Academy of Sciences* (PNAS), vol. 119, nº 31, réf. e2200300119, 2022. Voir ég. sur la question de l'efficacité des mesures préventives, A. FLÜCKIGER, « Légiférer sans arbitraire dans l'incertain, Le principe de proportionnalité entre précaution et expérimentation », *Weblaw, LeGes*, vol. 32, 2021, p. 12 s.

dispositions contraignantes qui ont le même objectif sanitaire n'a, à notre connaissance, pas été démontrée[128]. L'évidence pousse cependant à reconnaître qu'une hypersollicitation serait délétère au regard de l'objectif poursuivi et pourrait engendrer, par fatigue, l'effet opposé à celui initialement escompté. Le changement et la nouveauté évitent tout sentiment de lassitude, permettant la prise en compte d'une même information ou encore l'adoption d'un même comportement plus facilement. Nous sommes d'avis qu'il conviendrait de limiter le nombre de sollicitations possibles ou encore de modulations différentes des interfaces numériques par les technologies cognitives, sous peine de créer un « environnement harceleur ».

D'autres possibilités en lien avec les technologies cognitives sont offertes à l'État. L'autorité publique en charge de la santé pourrait collaborer avec des entités privées (p.ex. réseaux sociaux comme Métavers, Instagram, etc., ou des entreprises développant des agents conversationnels, etc.), ou bien mettre elle-même à disposition des logiciels informatiques ou des applications pour téléphone. Concernant une collaboration avec les entreprises privées, tant l'implémentation des outils numériques que l'achat de robots connectés se réalisent sur la base du volontariat et impliquent l'adhésion des personnes aux conditions générales d'utilisation. Leur mise en œuvre rejoint donc les utilisations courantes d'applications du domaine privé. Eu égard à ces différentes options, une des possibilités d'encadrement juridique serait, à tout le moins, le respect des exigences du droit de la consommation (notamment art. 97 Cst. féd. ; CO[129], LIC[130]) pour éviter toute adjonction de clauses abusives[131] (p.ex. géolocalisation, etc.), ainsi que celles du droit de la protection des données (art. 1 s. de la Loi fédérale sur la protection des données (LPD)[132]) (p.ex. traitement des données, durée de conservation, etc.). Se pose la question de savoir dans quelle mesure une entité privée, sous mandat d'une autorité publique, pourrait émettre des sollicitations, modulées aux besoins, afin d'agir sur différents biais cognitifs des utilisateurs ? On distinguera selon que le mandat émane d'une autorité fédérale ou cantonale. Si le mandat émane d'une autorité fédérale, l'application de l'art. 12 al. 1 LPD interdit toute atteinte illicite à la personne et l'al. 2 du même article, toute violation des principes généraux. Concernant ces derniers,

128 FLÜCKIGER (n. 7), p. 224.

129 Loi fédérale complétant le Code civil suisse, Livre cinquième, Droit des obligations du 30 mars 1911 (CO ; RS 220).

130 Loi fédérale sur l'information des consommatrices et des consommateurs du 5 octobre 1990 (LIC ; RS 944.0).

131 Art. 8 de la Loi fédérale contre la concurrence déloyale du 19 décembre 1986 (LCD ; RS 241). Voir ég. la page Internet du Secrétariat à l'économie (SECO), Clauses générales abusives, disponible sous : www.seco.admin.ch/seco/fr/home/Werbe_Geschaefts methoden/Unlauterer_Wettbewerb/Missbrauchliche_Geschaftsbedingungen.html (consulté le 19.08.2022).

132 Loi fédérale sur la protection des données du 19 juin 1992 (LPD ; RS 235.1).

les art. 4 al. 2 et al. 4 LPD posent en règles les principes de la bonne foi, de la proportionnalité et de la reconnaissance de la finalité du traitement des données par la personne concernée. Les applications pour téléphone, les robots sociaux ou autres présentant la particularité de pouvoir atteindre leurs utilisateurs en tout temps et directement dans leur sphère privée (art. 13 Cst. féd.), le respect du principe de proportionnalité est donc à analyser au regard des conditions de l'art. 36 Cst. féd. Pour toute analyse de la légalité du traitement et de l'utilisation des données dans le cadre d'un mandat émis par une autorité cantonale, il convient de se référer au droit cantonal concerné. En outre, les possibilités d'utilisation du *healthy smart nudging* par des logiciels ou applications pour téléphone mis à disposition par l'autorité étatique elle-même doivent également respecter les exigences du droit relatif à la protection des données et de l'art. 36 Cst. féd.

Enfin, une obligation étatique imposée aux citoyens d'implémenter ce genre de logiciels ou d'applications numériques, sans égard au support ou objet connecté servant à sa mise en œuvre, est à exclure et constituerait une mesure disproportionnée (art. 36 Cst. féd.) au regard du but poursuivi et de l'atteinte portée à la sphère de protection de la liberté personnelle (art. 10 al. 2 Cst. féd.).

3. *De la proportionnalité et de la surveillance*

La surveillance des individus est très encadrée[133] : en dehors du contexte pénal, elle s'envisage également dans le contexte civil à des conditions très particulières (p.ex. droit des assurances sociales, mesures d'éloignement, etc.). L'idée d'utiliser les géodonnées[134] pour prévenir les pathologies n'est pas nouvelle. La géomédecine trouve des applications tant dans la lutte contre les MNT[135] que pour celle des épidémies. Une illustration récente est la *heatmap*[136] (carte thermique) pour le suivi de la propagation du virus du Covid-19. Les objets

133 Nous ne pourrons pas développer le cadre normatif applicable dans cette contribution et renvoyons aux ouvrages spécifiques dédiés.

134 Sont considérées comme étant des géodonnées, les informations numériques pouvant être collectées ou non en temps réel et auxquelles une position géographique précise peut être attribuée. Pour des exemples d'utilisation, voir la brochure de l'OFFICE FÉDÉRAL DE LA TOPOGRAPHIE SWISSTOPO, *Géodonnées pour tous*, 2020, disponible sous : www.swisstopo.admin.ch/fr/connaissances-faits/geoinformation.html (consulté le 10.10.2022).

135 À ce sujet, voir A. GUILLOT/M. LEVY, « *Healthy urban planning* et droit administratif, Utiliser les données de la géomédecine et les outils de l'aménagement du territoire pour promouvoir un urbanisme salutogène », *Jusletter* du 31 janvier 2022.

136 Pour plus d'informations à ce sujet, voir Heidi.news.com : www.heidi.news/sante-alimentation/l-initiative-d-un-ex-medecin-cantonal-pour-une-carte-precise-du-corona virus-en-suisse (consulté le 19.08.2022).

connectés offrant cette possibilité, des applications ont également été créées, *SwissCovid*[137] en Suisse, *StopCovid*[138] puis *TousAntiCovid*, en France[139], avec téléchargement sur la base du volontariat. L'obtention du consentement des personnes aux conditions d'utilisation permet, entre autres, le partage de leurs données relatives à la géolocalisation. Tant en France qu'en Suisse, un contrôle du traitement de ces données sensibles a été réalisé. Si dans le premier cas, outre les dispositions normatives, le rendu public du code source a été nécessaire, au sein de la Confédération helvétique les systèmes de traçage ont été encadrés par le biais d'ordonnances[140].

Les technologies cognitives ouvrent plusieurs options dont celle d'une synergie entre géomédecine et *healthy smart nudging*. *Nudger* une personne via un objet connecté, dès son entrée dans une certaine zone géographique ou encore après un temps de présence défini préalablement, est aujourd'hui réalisable. Appliquée au domaine des maladies chroniques, une telle hypothèse pourrait s'avérer disproportionnée eu égard à la durée de la surveillance (cf. conditions de l'art. 36 Cst. féd.). L'atteinte à la sphère privée et à la liberté personnelle ne

137 Pour plus d'informations, nous renvoyons au site Internet de l'OFSP : www.bag.admin.ch/bag/fr/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/swisscovid-app-und-contact-tracing.html (consulté le 10.08.2022).

138 Pour plus d'informations, nous renvoyons au site Internet de la CNIL, notamment à l'avis du 26 mai 2020 : www.cnil.fr/fr/la-cnil-rend-son-avis-sur-les-conditions-de-mise-en-oeuvre-de-lapplication-stopcovid (consulté le 10.08.2022) ; ainsi qu'à la Délibération n° 2020-056 du 25 mai 2020 portant avis sur un projet de décret relatif à l'application mobile *StopCovid* (demande d'avis n° 20008032) : www.cnil.fr/sites/default/files/atoms/files/deliberation-2020-056-25-mai-2020-avis-projet-decret-application-stopcovid.pdf (consulté le 10.08.2022).

139 Pour plus d'informations, nous renvoyons au site Internet de la Commission Européenne : https://ec.europa.eu/commission/presscorner/detail/fr/ip_20_670 (consulté le 10.08.2022). Nous renvoyons également, pour les lignes directrices 4/2020 relatives à l'utilisation des données de localisation et d'outils de recherche de contacts dans le cadre de la pandémie de Covid-19, au site Internet *European Data Protection Board* (edpb) : https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-042020-use-location-data-and-contact-tracing_fr (consulté le 10.08.2022).

140 À cet égard et relativement aux conditions de mise en œuvre, nous renvoyons à l'Ordonnance sur l'arrêt du système de traçage de proximité pour le coronavirus SARS-CoV-2 et du système visant à informer d'une infection possible au coronavirus SARS-CoV-2 lors de manifestations du 30 mars 2022 (RS 818.101.25) ; l'Ordonnance du 24 juin 2020 sur le système de traçage de proximité pour le coronavirus SARS-CoV-2 (OSTP), en vigueur du 25.06.2020 au 01.04.2022 (RS 818.101.25) ; l'Ordonnance du 13 mai 2020 sur l'essai pilote du système suisse de traçage de proximité visant à informer les personnes potentiellement exposées au nouveau coronavirus (Covid-19) (Ordonnance Covid-19 essai pilote traçage de proximité), en vigueur du 14.05.2020 au 25.06.2020 (RS 818.101.25).

serait, en dehors de tout contexte d'urgence sanitaire mondiale, pas acceptée[141]. En outre, l'adhésion du public pourrait se révéler faible tant la réticence à l'acceptation d'une surveillance étatique (même encadrée) existe en général, et ce, quand bien même chaque individu consent sans limites à la transmission de ses données personnelles aux entreprises privées qui utilisent les mêmes technologies, voire des technologies plus intrusives[142].

Plus anecdotique, la question d'utilisation en interne, au sein des différents services de l'administration (p.ex. état civil et fisc), de logiciels d'IA qui permettraient un recoupement des données massif en temps réel et donc diverses possibilités de segmentation de la population. Cette hypothèse est également vite écartée par le non-respect du principe de proportionnalité au regard du but visé (cf. conditions de l'art. 36 Cst. féd.). De surcroît, de tels échanges de données seraient contraires au principe de finalité du traitement des données personnelles de l'art. 4 al. 4 LPD[143]. L'intérêt public défini par l'objectif de santé publique poursuivi[144] ne saurait justifier d'envisager le glissement d'un État sanitaire vers un État policier.

4. *Données personnelles et vie privée : quelle protection ?*

En Suisse, le législateur a élaboré, dans la droite lignée de l'art. 8 CEDH, un cadre normatif visant à protéger nos données, en se fondant, entre autres, sur plusieurs droits fondamentaux comme la liberté personnelle et la sphère privée, qui comprend le droit à l'autodétermination informationnelle[145]. Tant les « données personnelles »[146] que les « données sensibles »[147] font l'objet d'une protection particulière[148] et leur utilisation doit respecter le cadre juridique en vigueur, délimité en grande partie par la LPD et, en ce qui concerne

141 Cf. art. 1 et 3 al. 7 let. a de la Loi fédérale sur les bases légales des ordonnances du Conseil fédéral visant à surmonter l'épidémie de Covid-19 du 25 septembre 2020 (Loi Covid-19 ; RS 818.102).

142 Pour des exemples, nous renvoyons aux conditions générales des applications des *smart wearables* qui posent là encore la problématique du consentement à un suivi en temps réel et au partage des données de santé.

143 Lorsque les personnes confient leurs données personnelles à une entité publique, elles le font pour une finalité déterminée qui doit être reconnaissable pour la personne concernée.

144 Cf. n. 67.

145 Voir ég. à ce sujet, A. FLÜCKIGER/S. DAHMEN, « Jurisprudence actuelle en matière de protection des données », *in* A. EPINEY/D. NÜESCH (édit.), *Big data et droit de la protection des données*, Zurich, 2016, p. 127-141.

146 Art. 3 al. 1 LPD.

147 Art. 3 let. c ch. 2 LPD.

148 Voir à cet égard, F. ÉRARD, *Le secret médical, Étude des obligations de confidentialité des soignants en droit suisse*, Zurich, 2021, p. 385 s.

la recherche sur les personnes, par la Loi fédérale relative à la recherche sur l'être humain (LRH)[149].

À l'ère du numérique et du *Big data*, une prise de conscience est faite de la valeur que représente la détention de données. Si le Commissaire à la protection des données du Conseil de l'Europe[150], Jean-Philippe Walter, s'inquiète de l'avenir de notre sphère privée, Nick Foster avait, pour sa part, dès 2016, mis en avant un nouveau paradigme : ce ne serait pas les individus qui produiraient des données personnelles mais l'inverse[151, 152]. Les entités privées collectant massivement nos données personnelles (sous couvert de notre consentement)[153] et réalisant diverses segmentations, l'utilisation de technologies cognitives variées pour influer sur nos décisions est envisageable. Si les données personnelles ne revêtent pas forcément le caractère de « sensible » au sens de l'art. 3 LPD, leur recoupement est apte à avoir cette qualité. Des auteurs, tels que Frédéric Érard[154], ont déjà suggéré le renforcement de la protection relative à celles-ci lorsqu'elles sont agrégées. Cet auteur a souligné que l'accumulation de données selon le contexte et la temporalité permettait d'obtenir des informations sur l'état de santé des individus qui ont alors un caractère de données sensibles, appelant une protection[155].

Dans le domaine de la santé publique, la disponibilité des données représente un enjeu majeur. Un exemple concret est celui de la géomédecine qui permet la détermination de zones géographiques prioritaires[156] pour la mise en place

149 Loi fédérale relative à la recherche sur l'être humain du 30 septembre 2011 (LRH ; RS 810.30).

150 J.-P. WALTER, « La protection des données à l'ère du numérique, un regard européen », *Le Temps*, 18.08.2022, disponible sous : www.letemps.ch/opinions/protection-donnees-lere-numerique-un-regard-europeen (consulté le 20.08.2022). Pour plus d'informations, voir la Convention 108 et ses Protocoles, site du Conseil de l'Europe, disponible sous : www.coe.int/fr/web/data-protection/convention108-and-protocol#:~:text=Convention%20pour%20la%20protection%20des,de%20la%20protection%20des%20donn%C3%A9es (consulté le 20.08.2022).

151 FOSTER retient que « [l]orsque nous utilisons les technologies contemporaines, un flux d'informations est créé sous la forme de données. Une fois analysé, il décrit nos actions, nos décisions, nos préférences, nos mouvements et nos relations. Cette version codifiée de qui nous sommes devient de plus en plus complexe, évolutive, changeante et déformante en fonction de nos actions » (FOSTER (n. 120)). Voir ég. MAUREL (n. 120).

152 MAUREL (n. 120).

153 Par exemple, *via* les cookies.

154 ÉRARD (n. 148).

155 *Idem*, § 1472, p. 554.

156 GUILLOT/LEVY (n. 135). Voir ég. S. JOOST *et al.*, « Persistent Spatial Clusters of High Body Mass Index in a Swiss Urban Population as Revealed by the 5-Year Geocolaus Longitudinal Study », *BMJ Open Science*, vol. 6, nº 1, réf. e010145, 2016, p. 1. Voir également S. JOOST *et al.*, « De la géomédecine pour une santé publique de précision, et des médecins à la direction de l'urbanisme », *Revue Tracés*, 2018, disponible sous : www.researchgate.net/publication/325119842_De_la_geomedecine_pour_une_sante_publique_de_precision_et_des_medecins_a_la_direction_de_l%27urbanisme (consulté le 10.10.2022) et S. JOOST, *Données médicales et génétiques géoréférencées au lieu de résidence pour un service de santé publique de précision*, Swiss Public Health Conference 2017, Bâle, 2017. Voir ég. J. SIMOS *et al.*, *Healthy Cities, The Theory, Policy, and Practice of Value-Based Urban Planning*, New York, 2017, et D. DE RIDDER, *Geospatial approaches for precision public*, thèse, 2021.

des mesures de prévention et promotion de la santé sur le terrain. Les données permettent donc la création de la santé publique de précision[157], pendant de la médecine personnalisée[158].

S'envisage également la question des collaborations entre le secteur public et le secteur privé. En effet, la segmentation ou la détermination d'un groupe-cible est d'autant plus facile grâce à l'utilisation des *data lake* des réseaux sociaux qui recoupent nos données. Un exemple illustrant parfaitement l'efficience que peuvent avoir ces techniques de ciblage est la campagne de la fondation *Know Your Lemons*[159], qui, dans un but de lutte contre le cancer du sein, a diffusé par l'intermédiaire des réseaux sociaux (*in casu* Facebook) et à destination de femmes d'un certain âge, une image mettant en avant plusieurs citrons dans une boîte à œufs, citrons présentant une altération de leur peau et incitant toute personne présentant les mêmes symptômes à consulter. Si cet exemple démontre l'efficacité d'une collaboration entre entités privées, de manière similaire, les collaborations public-privé permettraient d'augmenter la portée de certaines campagnes de prévention en santé publique au travers des réseaux sociaux. Le mandat public doit ici également prendre en considération, en amont, les questions relatives à la protection des données.

157 MARKS SULTAN *et al.* (n. 19), p. 654-657.

158 ACADEMY OF MEDICAL SCIENCES, *Stratified, personalized or P4 medicine : a new direction for placing the patient at the centre of healthcare and health education*, Summary of a joint FORUM meeting held on 12 May 2015, p. 4 ; A. POKORSKA-BOCCI *et al.*, « Personalized medicine : what's in a name ? », *Personalized Medicine*, vol. 11, nº 2, 2014, p. 203. En Suisse, l'exploitation du potentiel de données généré par les institutions de santé dans un but d'améliorer les différentes actions de prévention et de promotion a été envisagée sur tout le territoire dans un but de recherche. L'ASSM, en collaboration avec le SIB Institut Suisse de Bioinformatique, a créé un système national d'interrogation des données cliniques. Cet échange de données est envisagé pour la création d'une santé personnalisée (*Swiss Personalized Health Network* (SPHN)), disponible sous : https://sphn.ch/fr/home/ ; voir ég. le *Fact-Sheet 2020* du SPHN, disponible sous : https://sphn.ch/wp-content/uploads/2020/11/201112_SPHN_Factsheet_web_DEF.pdf (consultés le 20.08.2022).

159 Fondation *Know Your Lemons*, créée par C. ELLSWORTH BEAUMONT. Voir https://fr.knowyourlemons.org/ (consulté le 10.08.2022).

IV. Conclusion

La santé publique a un rôle déterminant à jouer au sein de nos sociétés pour lutter contre l'augmentation de la prévalence des MNT. Cette urgence sanitaire, couplée à l'impératif de maîtrise des coûts[160], nous oblige à repenser la gouvernance dans le domaine de la santé, notamment quant à ses effets sur le long terme. Ce dernier critère relatif à la temporalité met en exergue l'un des principaux défis à relever dans la lutte contre ces maladies chroniques : l'efficience des politiques publiques définies sur une longue échéance. Cette problématique invite les autorités à mettre en place des mesures variées et à trouver des outils novateurs.

À l'ère du numérique et des objets connectés, l'utilisation des technologies cognitives apparaît indispensable. Outre les possibilités offertes de lutter contre les inégalités et inéquités en santé par l'amélioration de la qualité de l'information et par la transmission de celle-ci à un public cible (p.ex. des populations vulnérables) défini par une segmentation préalable, le *healthy smart nudging* a cet avantage, via une utilisation plus générale, d'influer également sur nos différentes habitudes de vie, concrétisant par là même les objectifs durables.

L'utilisation de ces nouveaux outils de contrainte étatique douce offre diverses options et une grande modularité, s'adaptant aisément au fédéralisme helvétique, puisque pouvant être mise en œuvre à différents échelons. Elle doit cependant s'inscrire dans un cadre normatif défini afin de préserver les intérêts et droits en présence[161]. Le respect de la liberté personnelle, du libre arbitre, de la sphère privée et de la protection des données personnelles est la principale difficulté et, à cet égard, toute mesure envisagée doit, de manière casuistique, être évaluée au regard du principe de proportionnalité. L'adhésion du public à ces nouvelles technologies cognitives est déterminante puisqu'il reste l'acteur et le destinataire principal des bénéfices recherchés. L'implémentation des sciences comportementales aux développements numériques récents permet d'envisager des nouvelles modalités d'application des mesures de prévention et de promotion de la santé, contribuant ainsi à l'élaboration d'une santé publique de précision[162].

160 J. MARTI *et al.*, « Économie comportementale, santé et médecine », *Revue médicale suisse*, vol. 15, 2019, p. 1982-1986. Voir ég. OFSP, *Politique de la santé : stratégie du Conseil fédéral 2020-2030*, disponible sous : www.bag.admin.ch/bag/fr/home/strategie-und-politik/gesundheit-2030/gesundheitspolitische-strategie-2030.html (consulté le 30.08.2022).

161 Cette question est l'objet de la thématique de recherche de ma thèse : A. GUILLOT, *L'architecture des Healthy Nudges : Quels enjeux juridiques ?*, IDS, Université de Neuchâtel.

162 MARKS SULTAN *et al.* (n. 19), p. 654-657.

Le développement du *Quantified Self*

De l'adoption d'un meilleur mode de vie à une nouvelle forme de science citoyenne

DYLAN HOFMANN
Doctorant dans le cadre du projet FNS Eccellenza* | Institut de droit de la santé | Faculté de droit | Université de Neuchâtel

Table des matières

* « The increasing weight of regulation : the role(s) of law as a public health tool in the prevention state » (n° 181125).

I. Introduction

De nos jours, il n'est pas rare de rencontrer une personne, un membre de notre famille, une collègue ou un ami qui porte une montre connectée ou un bracelet mesurant ses performances sportives. Parfois, on tombe sur une publication sur un réseau social d'une connaissance qui a parcouru 10 kilomètres en course à pied et qui partage son exploit. Ces situations sont devenues courantes dans notre vie quotidienne et peut-être même que vous, qui êtes en train de lire ces lignes, portez un de ces appareils d'automesure connectés.

Le développement des appareils d'automesure a contribué à la création d'un mouvement citoyen nommé « *Quantified Self* ». Ce mouvement se fonde sur la pratique de l'automesure et s'est démocratisé avec l'accessibilité des appareils de suivi connectés.

En parallèle, les développeurs d'applications ont rapidement compris que les données collectées par les participants au *Quantified Self* devaient être enregistrées et traitées pour permettre la visualisation des efforts, des progrès, des comportements à modifier ou la conservation d'un historique des activités mesurées. Dans cet élan, nous pouvons constater le développement d'une multitude d'applications dites de « santé et bien-être » qui favorisent la lecture et permettent le traitement des données collectées pour établir des recommandations et des projections de santé.

La présente contribution a pour vocation de définir le *Quantified Self* dans sa forme de mouvement citoyen et d'exposer les risques et enjeux majeurs que la pratique de l'automesure et le partage de données personnelles liées à la santé peuvent avoir pour les individus. Notre propos débute par une brève introduction qui démontre l'omniprésence des appareils d'automesure connectés dans notre quotidien (I.). Le développement se poursuit avec une définition du *Quantified Self* (II.). Il s'agira de procéder à quelques distinctions importantes par rapport à des notions connexes et d'examiner comment ce mouvement s'intègre dans le contexte des maladies transmissibles et des maladies non transmissibles. Cette partie nous permettra également de procéder à un survol des normes applicables aux dispositifs médicaux, qui peuvent éventuellement s'appliquer aux appareils d'automesure et aux applications de santé et bien-être. L'exposé mettra ensuite en évidence les dangers majeurs liés au *Quantified Self* (III.). Il sera notamment question d'évaluer l'impact des acteurs dominants sur le marché des appareils d'automesure connectés et des applications de santé et bien-être, et de souligner les enjeux économiques nés autour des données de santé. Ce point mettra en exergue les risques directs et indirects pour les utilisateurs de ces instruments. Ces observations nous mèneront à une réflexion quant aux perspectives de développement du *Quantified Self* et des enjeux sociétaux et juridiques que cette évolution peut représenter (IV.). En guise de conclusion, il s'agira de rappeler l'importance d'une protection complète des données de santé collectées dans le cadre du *Quantified Self* et de la transparence quant aux algorithmes et pratiques commerciales mises en place dans ce contexte (V.).

II. Le *Quantified Self* : d'un mouvement citoyen vers un comportement institutionnalisé ?

A. Définition du *Quantified Self*

Le *Quantified Self* est un mouvement initié et fondé par deux journalistes américains, Kevin Kelly et Gary Wolf, en 2007. Il s'agit d'une communauté citoyenne et internationale de personnes qui utilisent ou créent des instruments d'automesure et qui partagent un intérêt commun pour la connaissance de soi au travers des chiffres[1].

Ce mouvement connaît d'ailleurs un véritable engouement avec le développement des technologies portables. En effet, ces outils permettent la mesure en temps réels de certains éléments du quotidien (par exemple : le nombre de pas effectué par jour). En plus de ce suivi, ces outils offrent une visualisation simple et accessible des données collectées et des résultats obtenus après une analyse algorithmique[2].

La participation au *Quantified Self* se fonde sur l'idée que les individus peuvent avoir une meilleure connaissance d'eux-mêmes au travers d'un suivi quotidien et qu'ils sont en mesure de prendre des décisions libres et éclairées sur la base de cette automesure. On retrouve cette idéologie dans le credo du mouvement « connais-toi toi-même au travers des chiffres »[3].

---

1 K. CRAWFORD/J. LINGEL/T. KARPPI, « Our Metrics, Ourselves : A Hundred Years of Self-tracking from The Weight Scale to The Wrist Wearable Device », *European Journal of Cultural Studies*, vol. 18(4-5), 2015, p. 480 ; J. D. PRINCE, « The Quantified Self : Operationalizing the Quotidien », *Journal of Electronic Resources in Medical Libraries*, vol. 11(2), 2014, p. 92 s. ; E. CHOE/N. LEE/B. LEE/W. PRATT/J. KIENTZ, *Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data*, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, p. 1143 ; U. MEIDERT/M. SCHEERMESSER/Y. PRIEUR/S. HEGYI/K. STOCKINGER/G. EYYI/M. EVERS-WÖLK/M. JACOBS/B. OERTEL/H. BECKER, « Quantified Self – Schnittstelle zwischen Lifestyle und Medizin », *TA-TaSwiss*, vol. 67, 2018, p. 42-44 ; voir le site Internet du *Quantified Self*, https://quantifiedself.com/about/what-is-quantified-self/ (consulté le 29.08.2022).

2 M. HENDRIKS, *Self-Tracking Data and its Commercial Uses*, www.diggitmagazine.com/articles/self-tracking-data-and-its-commercial-uses (consulté le 29.08.2022) ; CRAWFORD/LINGEL/KARPPI (n. 1), p. 481.

3 Traduction de l'anglais par l'auteur ; site Internet du *Quantified Self* (n. 1).

Au vu de ce qui précède, le *Quantified Self*, peut être défini de la manière suivante : il s'agit d'un mouvement au travers duquel un individu entreprend le suivi de toutes sortes d'informations physiques, environnementales, comportementales et/ou biologiques relatives à son quotidien. Aujourd'hui, ce suivi se fait essentiellement au moyen d'appareils d'automesure connectés, mais il est tout à fait imaginable de le faire à l'aide d'un cahier de suivi… mais l'exercice s'avérera plus périlleux, notamment lors de la phase d'analyse. Une fois les données collectées, l'individu peut adopter ou modifier son comportement dans le sens des recommandations issues de son suivi[4].

Sur la base des constats qui ont été faits dans le cadre de la pratique du *Quantified Self*, des acteurs publics et privés, de milieux plus ou moins proches de la santé, s'intéressent aux développements du mouvement afin d'éventuellement l'encadrer et l'institutionnaliser.

B. Quelques notions connexes et délimitations

Dans le cadre de cette contribution, il est important de distinguer le *Quantified Self* sous sa forme de mouvement citoyen de l'automesure en tant que comportement – parfois institutionnalisé comme nous le verrons un peu plus loin. Cette distinction revêt une importance dans le sens où la participation au mouvement débute avec la volonté d'un individu, sans recommandation ni incitation d'un tiers ou d'une institution. L'automesure, en revanche, est une notion plus large qui décrit le comportement d'un individu qui collecte des données relatives à son mode de vie, sa santé, ou tout autre élément dont il souhaite faire une analyse[5]. Ainsi, l'automesure peut, éventuellement, être issue d'une recommandation ou d'une incitation, voire d'une obligation.

La *eHealth*, ou *cybersanté*, désigne l'emploi des technologies de l'information et de la communication dans la mise en œuvre, le support et l'interconnectivité des processus du système de santé[6]. La *cybersanté* est une notion qui englobe « notamment les services de soins de santé, la surveillance sanitaire, les publications, l'éducation, les connaissances et la recherche dans le domaine de la

4 M. ALMALKI/K. GRAY/F. SANCHEZ, « The Use Of Self-Quantification Systems for Personal Health Information : Big Data Management Activities and Prospects », *Health information Science and Systems* 2015 3(Suppl 1):S1, p. 2 ; M. SWAN, « The Quantified Self : Fundamental Disruption in Big Data Science and Biological Discovery », *Big Data*, vol. 1(2), 2013, p. 86 s.

5 MEIDERT *et al.* (n. 1), p. 42-44.

6 S. ABIDI, « Introduction », *in* S. ABIDI (édit.), *Mobile Health : A Technology Road Map*, Cham 2015, p. 1 s. ; ORGANISATION MONDIALE DE LA SANTÉ, *Rapport du secrétariat*, document A58/21, avril 2005, p. 1.

santé »[7]. Le dossier électronique du patient est un exemple d'instrument central de la *cybersanté*.

La *mHealth*, ou *santé mobile*, désigne l'emploi des technologies portables dans le domaine des soins de santé, de la médication et des programmes de préventions des maladies[8]. En d'autres termes, il s'agit des outils connectés ou connectables à disposition de la *cybersanté*. À titre d'exemple, il est possible d'illustrer la santé mobile au moyen de la connexion entre un smartphone et un glucomètre. Le patient mesure son taux de glucose et transmet le résultat à son médecin au moyen du smartphone et d'une application dédiée. La distinction avec le *Quantified Self* peut, parfois, s'avérer complexe car les technologies du *mHealth* peuvent être reprises par les adeptes du *Quantified Self* dans le cadre de leur suivi quotidien.

La télémédecine est une discipline de la *cybersanté*. Il s'agit de la pratique de la médecine à distance. Il est, dès lors, possible de faire de la *téléconsultation*, *téléexpertise*, *télésurveillance*, *téléassistance* ou de la *régulation médicale à distance*[9]. Ici, la distinction avec le *Quantified Self* est marquée par le fait qu'en télémédecine, il y a une interaction avec un professionnel de santé et que les outils d'automesure connectés ne sont pas essentiels à cette pratique.

C. Le *Quantified Self* et les maladies transmissibles

En guise de remarque liminaire, l'Organisation mondiale de la santé (OMS) définit les maladies transmissibles comme des maladies « causées par les agents pathogènes (bactéries, virus, parasites et champignons) et se [propageant], directement ou non, d'une personne à une autre »[10].

L'essence même du *Quantified Self* est de se connaître au travers des chiffres. Quand bien même nous verrons que ce mouvement s'inscrit bien plus facilement dans le contexte des maladies non transmissibles, il trouve tout de même une

7 ORGANISATION MONDIALE DE LA SANTÉ, *Cybersanté*, www.emro.who.int/fr/health-topics/ehealth/#:~:text=Selon%20l'OMS%2C%20la%20cybersant%C3%A9,%2C%20l'% C3%A9ducation%2C%20les%20connaissances (consulté le 29.08.2022).

8 ABIDI (n. 6), p. 1 s. ; R. SALVETER, « mHealth : que peuvent apporter les applications mobiles à la santé ? », *Spectra*, nº 121, septembre 2018, p. 2.

9 FÉDÉRATION DES MÉDECINS SUISSES, *Télémédecine*, www.fmh.ch/fr/themes/ehealth/telemedecine.cfm (consulté le 29.08.2022).

10 ORGANISATION MONDIALE DE LA SANTÉ, *Maladies infectieuses*, disponible sous : www.emro.who.int/fr/health-topics/infectious-diseases/index.html (consulté le 29.08.2022).

application en matière de maladies transmissibles sous la forme du « *Quantified Flu* »[11].

Le *Quantified Flu* a pour objectif de déterminer si les différents paramètres physiologiques suivis par les appareils portables peuvent aider à prédire le moment ou les raisons pour lesquelles nous tombons malades (par exemple : grippe, Covid-19, etc.). Ce mouvement est plus récent que le *Quantified Self*. Partant, les individus qui y participent ne font que partager des données collectées pendant une maladie passée et contribueront à enrichir la base de données en répertoriant leurs symptômes lorsqu'ils tomberont malades à l'avenir[12].

Dans son état actuel, le *Quantified Flu* permet, au mieux, une analyse rétroactive d'une période de maladie. Cette forme d'automesure reste encore très éloignée des possibilités de prévention et d'amélioration personnelle recherchées par de nombreux adeptes du *Quantified Self*.

D. Le *Quantified Self* et les maladies non transmissibles

Selon l'OMS, les maladies non transmissibles sont également appelées « maladies chroniques [et] tendent à être de longue durée et résultent d'une association de facteurs génétiques, physiologiques, environnementaux et comportementaux »[13]. En outre, « [l]es principaux types de maladies non transmissibles sont les maladies cardiovasculaires (accidents vasculaires cardiaques ou cérébraux), les cancers, les maladies respiratoires chroniques (comme la bronchopneumopathie chronique obstructive ou l'asthme) et le diabète »[14]. Ces maladies sont considérées comme une pandémie silencieuse. À elles seules, elles représentaient plus de 70 % des décès dans le monde selon les données de 2018 mises à disposition par l'OMS[15]. Dans l'élaboration des programmes de prévention des maladies non transmissibles, une priorité est accordée à la connaissance des facteurs de risque et à l'importance de réduire les facteurs directement liés aux comportements[16].

Dans ce contexte, le *Quantified Self* se pose en réel outil de mesure concernant les facteurs comportementaux et, dans une certaine mesure, les facteurs environnementaux. En effet, les outils connectés employés dans le *Quantified Self*

11 Site Internet du *Quantified Flu* : *About the Quantified Flu*, https://quantifiedflu.org/about/ (consulté le 29.08.2022).

12 *Ibid.*

13 ORGANISATION MONDIALE DE LA SANTÉ, *Maladies non transmissibles*, www.who.int/fr/news-room/fact-sheets/detail/noncommunicable-diseases (consulté le 29.08.2022).

14 *Ibid.*

15 *Ibid.*

16 OFFICE FÉDÉRAL DE LA SANTÉ PUBLIQUE, *Stratégie nationale – Prévention des maladies non transmissibles (stratégie MNT) 2017-2024*, Berne 2016, p. 4-7 et p. 16-19.

permettent de mesurer en temps réel et en permanence des données relatives aux habitudes et à l'hygiène de vie des utilisateurs. De plus, ce suivi permet, dans un second temps, de générer des conclusions et des recommandations que les utilisateurs peuvent choisir de suivre ou d'appliquer dans le but d'adopter un mode de vie plus sain et de réduire l'impact négatif des facteurs comportementaux ou environnementaux sur leur santé[17].

D'ailleurs, pour illustrer l'impact du *Quantified Self* sur le développement du marché des applications de santé et bien-être, il convient de regarder quelques statistiques en matière d'applications de santé et bien-être et de mise sur le marché d'appareils portables connectés. Ainsi, on pouvait compter 65 300 applications disponibles sur le *Google Play Store* à la fin du dernier trimestre 2021[18]. On en comptait 54 000 sur l'*App Store* d'Apple à la même période[19]. En matière d'appareils portables connectés, on dénombrait environ 722 millions d'exemplaires sur le marché en 2019 avec une projection à plus d'un milliard d'ici à la fin de l'année 2022[20]. Ces quelques données démontrent la popularité et le développement du secteur de l'automesure au moyen des technologies connectées.

E. Survol du cadre juridique applicable aux appareils connectés et aux applications de santé et bien-être

Les lois topiques en matière d'appareils connectés et d'applications de santé et bien-être[21] sont, essentiellement, la Loi fédérale du 15 décembre 2022 sur les médicaments et les dispositifs médicaux (LPTh)[22], l'Ordonnance du

17 AMALKI/GRAY/SANCHEZ (n. 4), p. 1 s.

18 Selon le site statista.com : *Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 1st quarter 2022*, www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ (consulté le 29.08.2022)

19 Selon le site statista.com : *Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 1st quarter 2022*, www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ (consulté le 29.08.2022).

20 Selon le site statista.com : *Number of connected wearable devices worldwide from 2016 to 2022*, www.statista.com/statistics/487291/global-connected-wearable-devices/ (consulté le 29.08.2022).

21 Littérature complémentaire sur le sujet : G. AEBISCHER, « Les applications mobiles de santé – De véritables dispositifs médicaux ? », *PJA*, 2017, p. 63-72 ; J. DRITTENBASS/I. WILDHABER, « Regulation of Medical Robots in Switzerland : The Example of Robotic Applications in Minimally Invasive Surgery », *Life Science Recht*, 2020, p. 11-19 ; L. CELLIER/S. GHERNAOUTI, « SwissCovid, un dispositif médical ? », *Jusletter* du 22 mars 2021 ; S. LEINS-ZURMÜHLE, « Mobile Applikationen als Medizinprodukte : Qualifikation und Pflichten nach der revidierten Schweizer Medizinprodukte-verordnung », *Life Science Recht*, 2021, p. 137-147.

22 RS 812.21.

1er juillet 2020 sur les dispositifs médicaux (ODim)[23], la Loi fédérale du 30 septembre 2011 relative à la recherche [avec] l'être humain (LRH)[24] et l'Ordonnance du 1er juillet 2020 sur les essais cliniques de dispositifs médicaux (OClinDim)[25]. Les normes du droit suisse ont été harmonisées avec les normes européennes pour éviter une exclusion de la Suisse du marché européen et pour faciliter les processus de reconnaissance. Le législateur suisse a repris, dans la mesure du possible, les dispositions du Règlement du 5 avril 2017 relatif aux dispositifs médicaux (RDM)[26] et du Règlement du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro (RDIV)[27].

Très succinctement, les nouvelles normes européennes ont été adoptées suite à de nombreux incidents impliquant des dispositifs médicaux. Pour l'essentiel, les modifications s'articulent autour de trois axes principaux : le renforcement de la sécurité des dispositifs, l'amélioration de leur traçabilité et la simplification de l'information destinée aux patients et au public[28].

Étant donné que le propos de la présente contribution porte sur les appareils portables connectés et les applications de santé et bien-être dans le cadre du mouvement *Quantified Self*, il paraît important de contextualiser juridiquement ces instruments, notamment sous l'angle de la législation applicable aux dispositifs médicaux. Le propos se limitera à exposer les points essentiels relatifs à la qualification en tant que dispositif médical. De manière très succincte, le processus d'autorisation de mise sur le marché d'un dispositif médical suit les étapes clés suivantes :

1. La qualification du produit en tant que dispositif médical[29] ;
2. La classification du dispositif sur la base de son potentiel de risque[30] ;
3. L'évaluation des exigences relatives à la sécurité et à la performance des dispositifs[31] ;

23 RS 812.213.
24 RS 810.30.
25 RS 810.306.
26 Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) nº 178/2002 et le règlement (CE) nº 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (JO L 117 du 5 mai 2017, p. 1).
27 Règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro et abrogeant la directive 98/79/CE et la décision 2010/227/UE de la Commission (JO L 117 du 5 mai 2017, p. 176).
28 Rapport explicatif de l'OFSP sur la révision totale de l'ordonnance sur les dispositifs médicaux et l'ordonnance sur les essais cliniques de dispositifs médicaux (nouvelle réglementation sur les dispositifs médicaux), mai 2019, p. 7 s.
29 Art. 4 LPTh et art. 3 ODim.
30 Art. 15 ODim.
31 Art. 45 LPTh et art. 6 ODim.

4. L'évaluation de conformité selon le risque (selon la classification)[32] ;
5. La délivrance de la déclaration ou du certificat de conformité (selon la classification)[33] ;
6. La matériovigilance et la matériosurveillance une fois le dispositif médical mis sur le marché[34].

En ce qui concerne la qualification de dispositif médical pour les appareils d'automesure connectés, rappelons que le développement concerne des appareils de type *Apple Watch* ou *FitBit*. Notre propos ne porte pas sur les outils propres à la santé mobile (par exemple : glucomètre connecté). Cette limite posée, pour qu'une montre connectée puisse être qualifiée de dispositif médical, il faut qu'elle remplisse les critères des art. 4 al. 1 let. b LPTh et 3 ODim, à savoir qu'elle doit être destinée « à un usage médical, ou [présentée] comme [telle], dont l'action principale n'est pas obtenue par un médicament ». Compte tenu de cette définition, il est difficile, à l'heure actuelle, de qualifier un appareil portable connecté comme l'*Apple Watch* ou *la FitBit* en tant que dispositif médical. Une nuance doit tout de même être apportée pour l'*Apple Watch* qui est considérée comme un dispositif médical de classe II par la *Food and Drug Administration*[35]. Ce propos pourrait très prochainement devenir obsolète puisque des rumeurs existent autour du prochain modèle d'*Apple Watch* pour lequel *Apple* aurait fait une demande d'autorisation auprès de la *Food and Drug Administration* aux États-Unis. En effet, ce nouveau modèle serait doté d'un glucomètre[36]. Cet ajout engendrerait une fonctionnalité hybride qui pourrait conduire à la qualification dans une classe de dispositif médical plus avancée de l'*Apple Watch*.

Concernant la qualité de dispositif médical des applications de santé et bien-être, l'exercice est un peu plus complexe. En effet, avec la modification des règlements européens relatifs aux dispositifs médicaux, le support du logiciel n'est plus le critère de qualification. Partant, pour qu'un logiciel de santé ou de bien-être soit qualifié de dispositif médical, il faut que quatre critères clés soient remplis[37] :

32 Art. 46 LPTh et art. 21, 23 et 24 ODim.
33 Art. 25, 29 et 46 ODim.
34 Art. 58 LPTh et art. 17, 56, 60, 64, 65, 66, 75 et 76 ODim.
35 L. REYNOUARD, *L'Apple Watch 4 : un dispositif médical ?*, https://apleb.fr/e-sante-et-les-nouvelles-tendances-de-l-intelligence-artificielle/l-apple-watch-4-un-dispositif-medical/ (consulté le 03.10.2022).
36 *Apple Watch glucose monitoring : Myth or Reality ?*, https://digitalhealthcentral.com/2021/04/19/apple-watch-glucose-monitoring/ (consulté le 29.08.2022).
37 SWISSMEDIC, *Aide-mémoire : Logiciels médicaux*, Berne 2021, p. 1 s. ; IDEM, « Quand une application devient dispositif médical », *Spectra*, nº 121, septembre 2018, p. 7 ; COMMISSION EUROPÉENNE, *Is your Software a Medical Device ?* ; GROUPE DE COORDINATION EN MATIÈRE DE DISPOSITIFS MÉDICAUX, *Guidance on Qualification and*

1. Il doit tout d'abord s'agir d'un logiciel ;
2. Son usage doit être fait à des fins médicales ;
3. Il doit profiter à une personne individuelle et non à un ensemble populationnel ;
4. Le logiciel doit permettre un traitement de données qui ne se limite pas au stockage, à l'archivage, à la communication, à la recherche simple, ni à la compression sans perte des données.

La modification de ces critères démontre un changement de paradigme important. En effet, l'ancien système accordait une importance au support sur lequel le logiciel était installé. Sous le régime actuel, le support n'a plus d'importance. C'est la fonctionnalité même du logiciel qui sera examinée pour savoir s'il s'agit ou non d'un dispositif médical. Cela a pour conséquence qu'un grand nombre d'applications de santé et de bien-être présentes sur les magasins virtuels d'*Apple* et de *Google* peuvent être qualifiées de dispositifs médicaux et être soumises au processus d'autorisation prévu par la législation idoine.

III. Aperçu des dangers relatifs au *Quantified Self*

A. La position dominante des acteurs privés

Le domaine des technologies médicales est rapidement devenu un terrain d'investissement et d'innovation important dans lequel des entreprises de la *Big Tech* se sont empressées d'investir. Aujourd'hui, ces mêmes multinationales sont devenues incontournables lorsqu'il s'agit de fournir des outils ou produits informatiques dans le domaine de la santé. Citons, à titre d'exemple, *IBM Watson* qui s'est spécialisé dans le développement de l'intelligence artificielle dans le milieu hospitalier, ou *Google* qui, par l'intermédiaire de sa société *Alphabet* et de cinq sociétés-filles, est actif dans le secteur de la santé, de la recherche et de l'analyse de données[38].

Dans le domaine du *Quantified Self* et de l'automesure dans un sens plus large, les entreprises de la *Big Tech* fournissent un large éventail de prestations. Celles-ci s'étendent de la production d'appareils d'automesure connectés à la mise à disposition des écosystèmes pour le portage des applications de santé et bien-être, tout en passant par la fourniture des serveurs et *clouds* pour le stockage des données collectées et l'offre en matière d'algorithmes pour le traitement

Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR, octobre 2019, p. 10 s.

38 H. CHARDONNIÈRE, « L'offensive des GAFAM et des BigTech dans la santé », *Les Échos Études*, www.lesechos-etudes.fr/blog/actualites-21/loffensive-des-gafam-et-des-bigtech-dans-la-sante-9584 (consulté le 29.08.2022).

et l'analyse de ces mêmes données. Pour illustrer la force de leur position, nous pouvons mentionner qu'*Apple* était le principal acteur du marché de la montre connectée en 2018 avec 45 % de parts de marché pour ce genre d'appareil[39].

Cette posture dominante a permis le développement rapide des technologies médicales. En effet, ces sociétés disposent de moyens financiers importants leur permettant une grande capacité d'investissement et favorisant le développement et la recherche dans ce secteur. En outre, les entreprises de la *Big Tech* ont été fortement sollicitées pendant la pandémie de Covid-19 pour lutter contre la désinformation, étudier la propagation du virus ou accélérer la recherche d'un traitement ou d'un vaccin. Leur puissance économique leur confère un statut de référence et les oblige à d'importantes responsabilités[40].

Cependant, la forte présence des entreprises de la *Big Tech* impacte le droit des consommateurs et leur liberté de choix quant au fournisseur d'appareils d'automesure connectés ou de serveurs pour conserver leurs données, par exemple. En effet, ces sociétés dominent le marché au point de ne pas laisser de réelles alternatives aux utilisateurs. Leur omniprésence est telle que même les gouvernements peinent à travailler avec d'autres prestataires ou à créer leurs propres systèmes autonomes et indépendants[41].

La dernière problématique que soulève le rôle des entreprises de la *Big Tech* dans le domaine des technologies médicales concerne leur modèle d'affaires. En effet, ces sociétés ne s'intéressent pas ou que très peu à la prévention en matière de santé. Ces entreprises se sont intéressées à ce domaine uniquement en raison de la masse de données qu'il génère et du potentiel commercial que les profils individualisés constituent[42].

B. Les enjeux commerciaux grandissants

Comme nous l'avons mentionné précédemment, les données collectées dans le cadre du *Quantified Self* sont des données personnelles sensibles qui concernent la santé du participant. Elles présentent un potentiel d'utilisation important, surtout si leur collecte ou leur conservation sur des serveurs est sou-

39 Selon le site statista.com : *Number of Connected Wearable Devices Worldwide from 2016 to 2022*, www.statista.com/statistics/487291/global-connected-wearable-devices/ (consulté le 29.08.2022).

40 *Le rôle des GAFAM dans la eSanté*, https://ellcie-healthy.com/le-role-des-gafa-dans-la-e-sante/ (consulté le 29.08.2022).

41 M. ROY, « Les GAFAM, nouveaux maîtres de la santé mondiale ? », *Annuaire français de relations internationales*, vol. 20, 2019, p. 245-247.

42 ROY (n. 41), p. 242-245.

mise à des conditions générales autorisant leur utilisation à des fins commerciales. Les données personnelles collectées présentent un intérêt pour de très nombreux acteurs économiques, gouvernementaux, de l'industrie pharmaceutique et même dans le domaine de la cybercriminalité. Une fois traités par ces différents acteurs, le corpus de données pourra servir à l'établissement de campagnes publicitaires ciblées ou de *nudging* sur les réseaux sociaux, par exemple[43]. Pour illustrer cette métamorphose économique des données de santé, on peut citer le rachat de la société *FitBit* par *Google* pour un montant de 2,1 milliards de dollars. *Google* a ainsi mis la main sur l'enregistrement de plus de « 275 000 milliards de pas et plus de 15 milliards d'heures de sommeil [mesurés] depuis 2009 grâce aux 120 millions de capteurs d'activité et montres distribuées [sur le marché] »[44].

LUPTON a constaté qu'au même titre que les échantillons biologiques humains, les assemblages de données numériques liées à la santé sont devenus des marchandises à part entière. En effet, ces données comprennent deux sortes de valeur : une première qui est liée à leur commercialisation et une seconde qui découle de la capitalisation du corps humain. La valeur attribuée aux données numériques de santé est appelée biovaleur. En outre, LUPTON insiste sur le fait que la pratique de l'automesure est génératrice d'un biocapital numérique[45]. Cette valorisation est même perçue par certains sociologues comme une nouvelle forme du capitalisme de surveillance[46].

En outre, le potentiel économique des données collectées dans le cadre du *Quantified Self* se retrouve dans certaines activités criminelles. En effet, les *hackers* ayant réussi à mettre la main sur des données médicales peuvent les revendre sur le marché noir au même titre qu'ils revendraient des données bancaires ou des identifiants professionnels[47].

43 D. LUPTON, « Lively Data, Social Fitness and Biovalue : The Intersections of Health and Fitness Self-tracking and Social Media », *The SAGE Handbook of Social Media*, 2015, p. 596 ss.

44 T. DELOZIER, *Fitbit officiellement racheté par Google*, article publié le 15.01.2021, www.lesnumeriques.com/montre-connectee/fitbit-officiellement-rachete-par-google-n159267.html (consulté le 29.08.2022).

45 LUPTON (n. 43), p. 572 s. ; D. LUPTON, *The Quantified Self : A Sociology of Self-Tracking*, Cambridge 2016, p. 147.

46 LUPTON (n. 43), p. 572 s.

47 J. HAYEK, *Les données de santé, nouvel or noir des cyberpirates*, www.hospitalia.fr/Les-donnees-de-sante-nouvel-or-noir-des-cyberpirates_a2710.html (consulté le 29.08.2022).

C. Les risques directs et indirects pour les utilisateurs

1. Le vol, la perte et les traitements illicites des données

De manière assez évidente, le risque primaire auquel les participants au *Quantified Self* sont exposés concerne les données qu'ils collectent et leur protection au moment du transfert puis de leur conservation sur un *cloud*[48]. En effet, les données de santé sont considérées comme des données sensibles, tant sous l'angle du Règlement européen du 27 avril 2016 sur la protection des données (RGPD)[49] que celui de la Loi fédérale du 19 juin 1992 sur la protection des données (LPD)[50], et nécessitent dès lors une protection particulière.

Le vol, la perte ou le traitement illicite de ce type de données peuvent porter lourdement atteinte aux droits de la personnalité des utilisateurs. En effet, au moyen des données de santé, il est possible d'usurper l'identité d'une personne[51]. Ces données sont parfois intimes et peuvent apparaître comme compromettantes pour leur titulaire. Un individu mal intentionné peut s'en servir pour exercer des pressions et faire du chantage[52]. D'autres données concernent le positionnement GPS de l'utilisateur. Souvent, il s'agit d'une géolocalisation en temps réel qui permet à un tiers non autorisé qui y aurait accès de connaître en permanence la position de l'utilisateur. Il lui est donc possible de le pister et le harceler[53]. De plus, il est complexe d'arrêter la vente des données de santé une fois qu'elles sont mises en vente. En effet, le titulaire ne saura pas exactement quelles sont les données vendues et sous quelles formes elles peuvent être revendues. De plus, contrairement à des données bancaires, pour lesquelles il est possible de bloquer le compte pour empêcher des virements frauduleux par exemple, le titulaire ne pourra pas adresser de demande de blocage pour rendre obsolètes les données de santé volées[54].

48 V. PEUGEOT, « Données de santé : contours d'une controverse », *Alternatives économiques*, vol. 4, nº 80, 2018, p. 36 s.

49 Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (JO L 119 du 4 mai 2016, p. 1).

50 RS 235.1. Définition des données personnelles : art. 3 let. a LPD ; définition des données sensibles : art. 3 let. c LPD.

51 M. BALLANO BARCENA/C. WUEEST/H. LAU, *How Safe is Your Quantified Self ? : security response*, 11 août 2014, p. 20.

52 BALLANO BARCENA/WUEEST/LAU (n. 51), p. 21.

53 BALLANO BARCENA/WUEEST/LAU (n. 51), p. 21 ; M. HOY, « Personal Activity Trackers and the Quantified Self », *Medical Reference Services Quarterly*, vol. 35, nº 1, 2016, p. 97.

54 HAYEK (n. 47).

Le problème majeur avec les appareils d'automesure connectés concerne la multiplication de leurs fonctionnalités. En effet, certains rapports constatent une augmentation du risque de fuite ou de vol des données personnelles avec l'augmentation des fonctionnalités de transfert et de connectivité. Chacune de ces fonctionnalités crée des sortes de brèches dans le flux des données qui permettent des intrusions ou des erreurs de transfert[55].

En cas d'infraction à la Loi sur la protection des données, l'individu atteint dans ses droits peut intenter une action à l'encontre de la personne responsable du traitement, conformément aux art. 15, 34 et 35 LPD.

Afin de faire valoir des prétentions à l'encontre de l'auteur d'un vol ou d'un traitement illicite des données de santé, l'utilisateur dispose d'actions qui relèvent du droit pénal, civil ou de la protection des données. Tout d'abord en matière pénale, nous pouvons mentionner que le Code pénal suisse du 21 décembre 1937 (CP)[56] sanctionne les infractions suivantes[57] :

- la soustraction de données[58] ;
- la soustraction de données personnelles[59] ;
- l'accès indu à un système informatique[60] ;
- la mise à disposition d'informations permettant l'accès indu à un système informatique[61] ;
- la détérioration de données[62] ; et
- l'extorsion et le chantage[63].

En ce qui concerne les prétentions à faire valoir en cas d'atteinte à la personnalité, l'art. 15 LPD renvoie aux art. 28, 28a et 28l du Code civil suisse du 10 décembre 1907 (CC)[64]. Sur la base de ces dispositions, l'utilisateur « peut requérir en particulier que le traitement des données, notamment la communication à des tiers, soit interdit ou que les données soient rectifiées ou détruites »[65]. Mentionnons également l'art. 28b CC qui permet à l'utilisateur de

55 BALLANO BARCENA/WUEEST/LAU (n. 51), p. 16-18.
56 RS 311.0.
57 Pour une analyse détaillée de ces infractions voir S. MÉTILLE/J. AESCHLIMANN, « Infrastructures et données informatiques : quelle protection au regard du code pénal Suisse ? », *Revue pénale suisse*, vol. 132, 2014, p. 283-317.
58 Art. 143 CP.
59 Art. 179^novies CP.
60 Art. 143^bis al. 1 CP.
61 Art. 143^bis al. 2 CP.
62 Art. 144^bis ch. 1 CP.
63 Art. 156 CP.
64 RS 210.
65 Art. 15 LPD *in fine*.

se défendre contre les violences, les menaces et le harcèlement en saisissant un juge et faire valoir ses prétentions à l'encontre de l'auteur du trouble.

Enfin, en cas d'atteinte illicite à sa personnalité, l'utilisateur peut prétendre à une indemnisation de son tort moral sur la base de l'art. 49 du Code des obligations du 30 mars 1911 (CO)[66].

2. *Les atteintes corporelles, morales ou psychiques*

Les utilisateurs d'appareils d'automesure connectés s'exposent également à des risques d'atteintes corporelles, morales ou psychiques. En effet, certains défauts de fabrication peuvent causer des brûlures graves, notamment en raison d'une surchauffe de la batterie d'un appareil d'automesure connecté ou de la matière recouvrant l'appareil[67]. Les utilisateurs peuvent également s'exposer à un risque de blessure si les recommandations de l'appareil qu'ils décident de suivre sont erronées en raison d'une erreur de mesure[68]. Il en va de même lorsque les paramètres de base ne sont pas corrects lors de l'enregistrement d'un nouvel utilisateur[69]. Enfin, la course à l'amélioration et à la performance peut causer une augmentation du stress chez certains utilisateurs et les affecter dans leur santé mentale.

En cas de défaut de fabrication d'un appareil, la Loi fédérale du 18 juin 1993 sur la responsabilité du fait des produits (LRFP)[70] permet d'actionner la responsabilité du fabricant d'un produit défectueux qui a causé un dommage à autrui. Sur la base de l'art. 1 LRFP, un « producteur répond du dommage » causé par un produit défectueux. Le dommage peut consister en la mort d'une personne ou des lésions corporelles[71].

Une autre voie possible pour sanctionner la personne responsable des lésions corporelles est de passer par l'art. 125 CP qui concerne les lésions corporelles par négligence. Mais la preuve de la négligence pourrait s'avérer complexe à

66 RS 220.
67 Exemples de rappels de produits aux États-Unis par la *Consumer Product Safety Commission* : www.cpsc.gov/Recalls/2014/Fitbit-Recalls-Force-Activity-Tracking-Wristband (consulté le 29.08.2022) ; www.cpsc.gov/Recalls/2022/Fitbit-Recalls-Ionic-Smartwatches-Due-to-Burn-Hazard-One-Million-Sold-in-the-U-S (consulté le 29.08.2022).
68 M. RODGERS/G. ALON/V. PAI/R. CONROY, « Wearable Technologies for Active Living and Rehabilitation : Current Research Challenges and Future Opportunities », *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 6, 2019, p. 4.
69 A-C. PERROY, « La m-santé, à l'ère de la e-santé. Promesses, enjeux et responsabilités », *Annales pharmaceutiques françaises*, vol. 74, n° 6, 2016, p. 426 s. ; COMMISSION EUROPÉENNE, *Livre vert sur la santé mobile*, 2014, p. 11 s.
70 RS 221.112.944.
71 Art. 1 al. 1 let. a LRFP.

démontrer en raison de la technicité de certains appareils connectés. Autre point problématique : la preuve de la causalité adéquate en cas de lésions corporelles causées par une recommandation liée à une erreur de mesure ou de calcul. Il faudrait être en mesure de connaître parfaitement le fonctionnement d'un algorithme pour pouvoir démontrer en quoi le calcul qui a été fait a conduit l'utilisateur à adopter un comportement dangereux pour sa santé[72].

3. *La surveillance, le contrôle et le profilage des utilisateurs*

Le *Quantified Self* consiste à collecter des données concernant les habitudes des utilisateurs. En général, l'usage d'appareils d'automesure connectés et d'applications de santé et bien-être implique l'acceptation de conditions générales d'utilisation. Souvent, celles-ci impliquent le droit d'utiliser les données collectées pour améliorer, personnaliser ou développer les services proposés, voire prévoient le droit d'utiliser les données collectées à des fins commerciales. Dans cette configuration, le danger pour les utilisateurs est de voir l'assemblage de leurs données de santé vendu et revendu à des entreprises dans une perspective de ciblage publicitaire. Ce genre de pratique, compte tenu du degré sensible et intime des données, permet un ciblage précis des utilisateurs au point de leur proposer une publicité ultra personnalisée[73].

Une autre facette des dangers liés à la segmentation par les données de santé concerne le risque de discrimination. La segmentation désigne le processus qui permet de diviser une population, un marché, en différents groupes ayant les mêmes besoins et les mêmes envies[74]. En effet, les données collectées dans le cadre du *Quantified Self* permettent d'avoir une vision d'ensemble de l'état de santé et des comportements « à risque » d'une personne[75]. Cela risque de conduire à l'exclusion d'individus de certains programmes d'assurances ou de les soumettre à une liste importante de réserves visant à exclure la prise en charge de certaines prestations. Pour l'instant, cette pratique reste limitée aux assurances privées. On assiste tout de même à une pression des assureurs pour intégrer ces programmes assuranciels dans l'assurance-maladie de base. Or, ce genre d'approche s'avère dangereux pour les fondements de notre système

72 H. HALSE, *Positive and Negative Effects of Exercise*, 2019, www.livestrong.com/article/459374-positive-negative-effects-of-exercise/ (consulté le 29.08.2022).
73 BALLANO BARCENA/WUEEST/LAU (n. 51), p. 20.
74 M. MCDONALD/I. DUNBAR, *Market Segmentation : How to Do It and How to Profit from It*, 4e éd., Chichester 2012, p. 9.
75 BALLANO BARCENA/WUEEST/LAU (n. 51), p. 20 ; PEUGEOT (n. 48), p. 37 s.

d'assurances sociales puisqu'elle contribue à l'érosion du principe de solidarité, pilier fondateur de ce système[76].

4. *La fatigue de l'automesure et l'addiction au sport*

La fatigue de l'automesure correspond à la perte de volonté et de motivation à pratiquer le suivi de ses paramètres de santé et à prendre part aux activités recommandées par une application de santé et bien-être. Cette fatigue aboutit, souvent, à un abandon du suivi et des activités relatives à l'automesure avec une détérioration des bonnes habitudes qui avaient été prises jusque-là[77].

La fatigue de l'automesure apparaît lorsque le participant au *Quantified Self* décide de mesurer trop de paramètres ou de suivre son activité trop intensivement. Il en résulte une démultiplication des données à collecter et de recommandations à suivre. L'utilisateur se voit alors découragé par cette masse d'informations et perd tout intérêt à l'automesure[78].

Il convient de souligner que la multiplication des notifications peut réduire le bien-être et la volonté des utilisateurs de continuer leur activité. En effet, aux États-Unis, un rapport a constaté une baisse de la productivité chez les employés submergés de notifications. Une telle mécanique peut être transposée dans la pratique de l'automesure[79]. L'accumulation d'applications, de données à saisir et de notifications à lire peut créer de l'inconfort et péjorer le bien-être mental des utilisateurs.

À l'inverse, la collecte de données et le besoin d'atteindre un objectif quotidien, encouragés par le battage médiatique autour des bienfaits de l'exercice physique, peuvent devenir obsessionnels[80]. La littérature spécialisée dans le domaine s'accorde tout de même sur le fait que la pratique d'un sport de manière excessive peut engendrer un épuisement et accroître le risque de blessure. Une telle dépendance peut se développer plus rapidement lorsque l'exercice physique constitue une échappatoire à des sentiments désagréables, des situations difficiles ou une obsession pour son poids. Finalement, l'addiction au

76 D. HOFMANN/M. LEVY, « Solidarité et santé publique », *Revue médicale suisse*, vol. 18, nº 790, 2022, p. 1395.
77 J. ETKIN, « The Hidden Cost of Personal Quantification », *Journal of Consumer Research*, 2016, p. 967-984.
78 CHOE *et al.* (n. 1), p. 1147.
79 Voir *Rapport sur l'anatomie du travail d'Asana*, 2022, https://asana.com/fr/resources/anatomy-of-work#form (consulté le 29.08.2022).
80 S. MEKKY, *Wearable Computing and the Hype of Tracking Personal Activity*, Stockholm 2014, p. 2 s.

sport peut conduire à une mauvaise estime de soi[81]. Il est cependant difficile d'estimer à partir de quel point une activité physique peut être considérée comme excessive ou obsessionnelle.

Pour remédier à cela, il est important que les utilisateurs soient informés et conseillés sur la manière de suivre leurs paramètres de santé de manière efficace sur le court, le moyen et le long terme. Cela passe par une sélection attentive des paramètres qu'ils souhaitent suivre ou, si la charge mentale est trop importante, l'abandon du suivi.

IV. Perspectives de développement autour du *Quantified Self*

A. L'adoption d'un standard de qualité pour le *Quantified Self* ?

L'unique certification que l'on peut trouver sur les appareils d'automesure connectés est le marquage « CE » pour la conformité aux normes de fabrication et de sécurité européennes. Cette certification ne concerne que la sécurité générale du produit et ne porte aucunement sur la fiabilité des mesures et des recommandations que ces appareils effectuent[82].

Une nouvelle certification permettrait une meilleure information des consommateurs. On pourrait envisager l'introduction d'une sorte « d'indice de fiabilité » qui attribuerait une note permettant d'informer les utilisateurs sur la précision des mesures de leurs appareils et sur le degré de confiance qu'ils peuvent accorder aux recommandations obtenues après le traitement de leurs données.

À titre d'illustration, nous pouvons mentionner l'instauration de l'indice de réparabilité français. Il a été créé après l'adoption de la Loi n 2022-105 du 10 février 2020 relative à la lutte contre le gaspillage et à l'économie circulaire. Issu du secteur privé, cet indice a pour objectif de sensibiliser les consommateurs sur la possibilité de faire réparer un appareil électronique. L'indice de réparabilité est calculé par les fabricants, ce système reposant sur une déclaration volontaire. Cependant, des grilles de calcul sont mises à disposition afin d'assurer une application et une notation uniforme des produits. Les fabricants déclarent ensuite leur note à la société *Spareka* qui est responsable de la tenue

81 E. THORNTON/S. SCOTT, « Motivation in the Committed Runner : Correlations Between Self-Report Scales and Behaviour », *Health Promotion International*, vol. 10, n° 3, p. 177-184 ; T. MOFFETT, *Positive & Negative Effects of Exercise*, 2011, www.2bstronger.com/article/fitness/positive-negative-effects-of-exercise-36906.html (consulté le 29.08.2022).

82 P. ESPINOZA, « La santé connectée : une réponse aux défis actuels de suivi des patients ? », *Horizon pluriel*, n° 28, janvier 2015, p. 7.

du site web « *indicereparabilite.fr* », qui recense toutes les notes. La notation repose sur les critères suivants[83] :

- disponibilité de la documentation ;
- démontabilité, accès et outils ;
- disponibilité des pièces détachées ;
- prix des pièces détachées ; et
- critère spécifique à la catégorie d'équipements concernée.

L'examen de ces critères aboutit à une note allant de 0 (impossible à réparer) à 10 (aisément réparable) et un code couleur de rouge à vert[84].

Sur la base de cet exemple, il serait aisé d'imaginer la création d'un indice de fiabilité destiné aux appareils d'automesure connectés. Parmi les critères qui pourraient servir à l'évaluation de la fiabilité, nous pouvons mentionner les suivants :

- qualité des paramètres initiaux et de la facilité et diversité des options de paramétrage par l'utilisateur ;
- qualité de la mesure ;
- possibilité de contourner le système de mesure et de « grossir » les chiffres pour maquiller la réalité ;
- niveau de transparence de l'algorithme et du flux de données ;
- diversité dans la personnalisation de l'appareil par rapport à chaque utilisateur ; et
- gestion du flux de données par l'utilisateur.

Le même standard de qualité pourrait être appliqué aux applications de santé et bien-être.

Il resterait encore à déterminer quelle serait l'autorité compétente pour évaluer l'indice de fiabilité et, dans la mesure où l'organisme compétent serait public, de créer une base légale encadrant son activité.

83 *Indice de réparabilité*, www.indicereparabilite.fr/(consulté le 29.08.2022) ; *Choisir un appareil avec un bon indice de réparabilité*, https://longuevieauxobjets.gouv.fr/acheter-durable/indice-de-reparabilite (consulté le 27.09.2022).

84 *Choisir un appareil avec un bon indice de réparabilité*, https://longuevieauxobjets.gouv.fr/acheter-durable/indice-de-reparabilite (consulté le 27.09.2022) ; art. 3 de l'Arrêté du 29 décembre 2020 du ministère de la Transition écologique relatif aux modalités d'affichage, à la signalétique et aux paramètres généraux de calcul de l'indice de réparabilité (JORF n° 0316 du 31.12.2020).

B. L'intégration du *Quantified Self* dans la société actuelle

Le *Quantified Self* et ses outils sont déjà bien implantés dans notre quotidien.

Tout d'abord, mentionnons le cas du « *Quantified Employee* »[85]. Cette notion décrit l'implémentation des appareils d'automesure connectés dans le cadre professionnel. Dans la majorité des cas, l'employeur fournit l'outil de mesure à ses employés. Ce genre de pratique peut se faire dans un contexte vertueux. En effet, l'employeur a l'obligation de protéger la santé de ses employés. L'utilisation d'appareils d'automesure connectés lui permet d'identifier les activités à risque au sein de son entreprise et d'améliorer les conditions de travail en conséquence. Cependant, une limite peut vite être franchie lorsque l'employeur décide de détourner cette fonctionnalité pour savoir à quel moment les employés arrivent sur leur lieu de travail, quelles sont les fréquences de leurs pauses ou quelle est la rapidité d'exécution des tâches qui leur sont attribuées. Un exemple peut être trouvé en Chine, par exemple, où une agence gouvernementale pour l'environnement utilise des bracelets connectés pour savoir si les employés commencent et terminent leur journée de travail conformément aux horaires.

Un autre développement intéressant du *Quantified Self* dans la société actuelle est le programme « *Pas à Pas* + » créé par Unisanté en Suisse[86]. Ce programme est destiné aux personnes qui ne sont pas ou insuffisamment actives ou qui font partie d'un groupe à risque pour les maladies non transmissibles. Les professionnels de la santé peuvent prescrire une participation (gratuite) à ce programme dans lequel un humain joue le rôle motivateur et fait les recommandations nécessaires au participant.

Enfin, le *Quantified Self* a trouvé une résonnance particulière chez les assureurs maladies complémentaires. En Suisse, ces dernières années, nous avons vu naître des offres basées sur l'optimisation de la santé au moyen d'objectifs quantifiés par les appareils d'automesure connectés. En échange d'une réduction de prime ou de bons d'achat, les assurés s'engagent à pratiquer une activité physique ou à atteindre un nombre de pas quotidien[87]. À titre d'exemple, nous

85 C. STEELE, *The Quantified Employee : How Companies Use Tech to Track Workers*, www.pcmag.com/news/the-quantified-employee-how-companies-use-tech-to-track-workers (consulté le 29.08.2022) ; J. BERSIN, *Quantified Self : Meet the Quantified Employee*, www.forbes.com/sites/joshbersin/2014/06/25/quantified-self-meet-the-quantified-employee/?sh=fe76787c5fe4 (consulté le 29.08.2022).

86 *Le projet Pas à Pas +*, www.pas-a-pas.ch/le-projet-pas-a-pas/(consulté le 29.08.2022).

87 À ce sujet, l'auteur travail dans le cadre de la rédaction de sa thèse de doctorat sur l'émergence de ce phénomène d'assurances dites « entreprenantes » et de leur usage dans le cadre de la prévention des maladies non transmissibles.

pouvons citer le programme *MyStep* de la CSS assurance[88] ou le programme *Benevita* de l'assurance Swica[89].

C. Le *Quantified Self* dans la société de demain

En raison de la dimension prise par la pratique de l'automesure et le *Quantified Self*, il convient de se poser la question de son influence sur notre société et sur la manière dont cette pratique pourrait modeler le rapport de l'État face aux individus et le rapport de solidarité entre ces derniers. Une étude en sociologie faite en 2021 par SAMOCHOWIEC et MÜLLER dresse quatre scénarios dystopiques – ayant de faibles chances de se produire dans leur entièreté – qui nous poussent à réfléchir à une intégration juste et adéquate de ces technologies et de l'automesure.

Le premier scénario est celui du « *Big Government* »[90]. Il s'agit d'une configuration extrême du rôle étatique avec un État qui prend une fonction paternaliste sanitaire. Les bases de données sont centralisées et gérées exclusivement par le gouvernement. Les individus se voient dans l'obligation de transmettre leurs données et de pratiquer l'automesure sous peine d'être sanctionnés. Les individus qui adoptent un comportement malsain et dangereux pour leur santé sont également fortement sanctionnés et recadrés pour changer de mode de vie. Une tendance illustrative de ce genre de processus serait le système de « crédit social » appliqués depuis peu en Chine[91].

Le deuxième scénario est celui du « *Big Business* »[92]. Dans cette constellation, le marché des soins de santé est totalement dérégulé. L'État n'intervient plus, mais la solidarité entre les individus est institutionnalisée. Les services de santé s'organisent sous la forme d'un libre marché. Concernant le remboursement des soins de santé, des groupes de personnes seraient organisés sous la forme de *pools* de risques pair-à-pair ou des assurances pair-à-pair. Cette structure finance les frais de chacun et le surplus est restitué aux membres en fin d'année. Ces groupes sont établis sur la base des données collectées. Chaque membre du groupe dispose d'un accès à la totalité des données des autres membres et

88 CSS, *Informations à la clientèle : myStep. Indemnisation pour les pas effectués. La récompense de votre activité physique*, 2018.

89 Présentation du programme sur le site Internet de l'assurance SWICA : www.swica.ch/fr/kampagnen/avantages/benevita#ant2 (consulté le 29.08.2022).

90 J. SAMOCHOWIEC/A. MÜLLER, « La montre connectée nuit-elle à la solidarité ? Scénarios pour un système de santé basé sur les données », *GDI* 2021, p. 18-25.

91 A. LEE, *What is China's social credit system and why is it controversial ?*, www.scmp.com/economy/china-economy/article/3096090/what-chinas-social-credit-system-and-why-it-controversial (consulté le 29.08.2022).

92 SAMOCHOWIEC/MÜLLER (n. 90), p. 27-32.

la communauté peut exclure ou refuser un membre qui ne correspondrait pas à la catégorie de risque. Ce genre de pratique conduirait inévitablement à des discriminations importantes à l'encontre des personnes présentant un risque élevé – que cela soit lié à leur comportement ou non. Une tendance actuelle qui permet d'illustrer ce scénario est le système d'assurance « *pay as you live* »[93]. Ces assurances fondent principalement leurs primes et les exclusions de prestations (réserves d'assurance) sur la base du mode de vie des assurés et des risques qu'ils prennent.

Le troisième scénario est celui du « *Big Self* »[94]. Cette perspective place l'État en facilitateur de modes de vie sains. Il n'existe aucune obligation de collecte ou de transfert de données ni d'obligation de comportement. Cependant, les individus sont fortement encouragés à prendre connaissance des éléments qui péjorent leur santé pour pouvoir prendre des décisions éclairées. Le *nudging* (ou paternalisme libéral)[95] est une tendance qui s'intègre dans les mécanismes mis en place dans ce scénario. Il s'agit d'un moyen « d'influencer les décisions en adaptant l'environnement sans coercition ni incitation financière »[96].

Le dernier scénario est celui du « *Big Community* »[97]. Dans cette configuration, le partage de données est devenu une norme sociale. Les individus contribuent collectivement à l'enrichissement d'une base de données commune en matière de santé. L'inclusion des personnes à risque est tout aussi importante parce que leurs données contribuent à l'amélioration des bases de données et permettent ainsi une meilleure approche en matière de prévention des maladies. Le *Quantified Self*, qui prendrait alors la forme d'une science citoyenne, illustre bien ce changement de norme sociale avec des individus qui collectent leurs propres données de santé et les partagent en temps réel, tout en effectuant leur propre analyse de ces données[98].

Comme mentionné en introduction de ce paragraphe, il y a très peu de chances pour qu'un de ces scénarios se réalise entièrement. En revanche, si nous examinons quelques mesures prises dans le contexte de la pandémie de Covid-19,

93 *Erster Versicherer macht « Pay as you live » im Neugeschäft zur Bedingung*, www.versicherungsbote.de/id/4871660/Versicherer-als-Life-Coach-John-Hancock/ (consulté le 29.08.2022).
94 SAMOCHOWIEC/MÜLLER (n. 90), p. 33-40.
95 Pour plus d'informations sur le *nudging*, voir R. THALER/C. SUNSTEIN, *Nudge : Improving decisions about health, wealth, and happiness*, New Heaven/Londres 2008.
96 SAMOCHOWIEC/MÜLLER (n. 90), p. 39.
97 SAMOCHOWIEC/MÜLLER (n. 90), p. 41-46.
98 Pour un exemple de science citoyenne en santé : N. PEEL, *Citizen Scientists can spot cancer cells like pathologists, so what happens next ?*, https://news.cancerresearchuk.org/2015/10/01/citizen-scientists-can-spot-cancer-cells-like-pathologists-so-what-happens-next/ (consulté le 29.08.2022).

il est possible d'opérer un classement dans chacun des scénarios[99]. Ainsi, l'exigence du « pass sanitaire » dans certains lieux publics relève du « *Big Government* », les campagnes d'information et d'encouragement à la vaccination tombent dans le « *Big Self* », la volonté de ne pas rembourser les soins des personnes non vaccinées relève du « *Big Business* » et le débat autour de la levée des brevets sur les vaccins serait du ressort du « *Big Community* ».

D. Les développements juridiques actuels et à venir susceptibles d'impacter le *Quantified Self*

1. En matière de protection des données

L'adoption du RGPD a permis aux États membres de l'Union européenne de se doter de nouveaux instruments de sanction et d'actualiser le droit de la protection des données. Ce Règlement est d'autant plus intéressant qu'il peut avoir des implications concrètes pour la Suisse. Premièrement, le critère de l'établissement[100] ouvre deux possibilités de rattachement : en cas de traitement de données personnelles relatives à une activité provenant d'une succursale européenne appartenant à une entreprise suisse et en cas de sous-traitance en Suisse pour le compte d'une société européenne. Deuxièmement, le critère du ciblage[101] ajoute deux rattachements supplémentaires lors du traitement de données personnelles de résidents européens par une société suisse dans une perspective d'offre de biens et services au sein de l'Union européenne ou en cas de suivi du comportement de personnes dans l'Union européenne.

Dans la mesure où un responsable du traitement suisse (ou un sous-traitant) entre dans l'un des critères de rattachement mentionné ci-dessus, il sera soumis au RGPD et au respect des obligations prévues dans le Règlement, notamment :

- la mise en place de mécanismes de sécurité suffisants compte tenu du risque encouru par les individus[102] ;
- le respect des garanties en matière de protection des données dès la conception des produits et services[103] ;
- la tenue d'un registre relatif aux activités de traitement de données[104] ;

99 SAMOCHOWIEC/MÜLLER (n. 90), p. 48-50.
100 Art. 3 al. 1 RGPD et considérant 22.
101 Art. 3 al. 2 RGPD et considérants 23 et 24.
102 Art. 24 et 32 RGPD.
103 Art. 25 RGPD.
104 Art. 30 RGPD.

- la réalisation d'une étude d'impact en matière de protection des données dans la mesure où un risque accru existe pour les libertés et les droits des individus concernés[105] ;
- la notification à l'autorité de contrôle et à la personne concernée en cas de violation des données à caractère personnel[106] ; et
- la désignation d'un délégué à la protection des données[107].

Un autre apport du RGPD concerne le traitement « portant sur des catégories particulières de données à caractère personnel »[108]. Cet apport est important dans le contexte du *Quantified Self* car les données collectées concernent la santé, les comportements, la vie sexuelle et potentiellement des informations génétiques des utilisateurs. Cette intégration ajoute un degré de protection important et nécessaire pour cette catégorie de données en interdisant leur traitement par principe (art. 9 par. 1 *in fine* RGPD).

Sous l'impulsion des modifications récentes du droit européen de la protection des données, les autorités suisses ont également entrepris une révision totale de la LPD. L'entrée en vigueur de la nouvelle Loi fédérale sur la protection des données est fixée au 1er septembre 2023 (nLPD)[109].

Il convient de souligner que le législateur fédéral a introduit dans la nLPD les notions de « profilage » et de « profilage à risque »[110] et a renforcé le principe de consentement explicite pour la collecte et le traitement de données personnelles[111]. Les données collectées par les adeptes du *Quantified Self* sont d'une telle diversité qu'elles permettent de dresser des profils biologiques et comportementaux particulièrement représentatifs et fidèles de chaque utilisateur.

Enfin, nous constatons l'inscription du « droit à la remise ou à la transmission des données personnelles »[112]. Ce droit permet à un adepte du *Quantified Self* de demander au responsable du traitement qu'il lui remette les données communiquées. Partant, il pourra se rendre compte de la masse de données qu'il a

105 Art. 35 RGPD.
106 Art. 33 et 34 RGPD.
107 Art. 37 RGPD.
108 Art. 9 RGPD.
109 Message du 15 septembre 2017 concernant la loi fédérale sur la révision totale de la loi fédérale sur la protection des données et sur la modification d'autres lois fédérales, FF 2017 p. 6565 ss ; Loi fédérale (projet) sur la protection des données, FF 2020 p. 7397 ss ; Conseil fédéral, Nouveau droit de la protection des données, FF 2020 1er septembre 2023, 31 août 2022, www.bj.admin.ch/bj/fr/home/aktuell/mm.msg-id-90134.html (consulté le 27.09.2022).
110 Art. 5 let. f et g nLPD.
111 Art. 6 al. 7 nLPD.
112 Art. 28 nLPD.

collectées et communiquées. Ce droit peut servir un but d'éducation et de sensibilisation.

2. *En matière de droit des dispositifs médicaux*

L'Union européenne s'est dotée récemment de deux nouveaux règlements relatifs aux dispositifs médicaux :

- Le Règlement du 5 avril 2017 relatif aux dispositifs médicaux[113] ;
- Le Règlement du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro[114].

L'adoption de ces deux règlements sur le plan européen a engendré la modification des ordonnances d'application dans le domaine des dispositifs médicaux en Suisse. Les nouvelles versions de l'ODim et de l'OClin-Dim sont entrées en vigueur le 26 mai 2021. Comme mentionnée plus haut, cette harmonisation du droit suisse avec le droit européen sert à éviter des entraves au commerce de dispositifs médicaux entre la Suisse et l'Union européenne. Il s'agit également pour les autorités suisses de reprendre les améliorations en matière de sécurité des patients, de transparence en matière d'information et de surveillance du marché des dispositifs médicaux[115].

Si les normes européennes permettent de qualifier certaines applications de santé et bien-être de dispositifs médicaux[116], ces nouvelles réglementations n'apportent pas de changement spécifique pour les appareils d'automesure connectés. En ce qui concerne les normes suisses, il n'y a pas vraiment d'autres commentaires à faire, puisque le droit suisse en matière de dispositifs médicaux reprend le droit de l'Union européenne. Certaines applications de santé et bien-être doivent donc faire l'objet d'une certification pour être mises sur le marché. Cela doit permettre d'assurer une meilleure sécurité du produit. Cependant, les appareils qui effectuent les mesures ne doivent pas faire l'objet d'une telle certification… il existe donc un risque d'erreur de mesure à la source.

V. Conclusion

Le nombre d'appareils d'automesure connectés mis sur le marché, le développement croissant des applications de santé et bien-être et les intérêts

113 Cf. n. 26.
114 Cf. n. 27.
115 Rapport explicatif de l'OFSP (n. 28), p. 8 s.
116 Art. 2 ch. 1 et considérant 19 RDM.

économiques liés aux données de santé font que le *Quantified Self* est en train de connaître son âge d'or. Si ce mouvement peut se targuer de permettre aux individus de mieux connaître leur santé au travers de l'analyse de leurs données personnelles, les données ainsi recueillies attirent également les convoitises des développeurs et des entreprises de la *Big Tech*.

De plus, le développement des bases de données et le fait que les utilisateurs prennent conscience des éléments péjorant leur santé font que le *Quantified Self* et l'automesure présentent un intérêt non négligeable pour la santé publique et l'établissement des programmes de prévention des maladies non transmissibles. Cependant, il n'est pas encore possible d'affirmer totalement ce point en raison du manque d'études sur l'efficacité de cette pratique à long terme. Un autre avantage du *Quantified Self* est que les utilisateurs bénéficient d'un réel *empowerment*. Ils deviennent connaisseurs de leur santé et peuvent agir en connaissance de cause sur la base de l'analyse précise et personnalisée de leurs comportements et habitudes. En quelque sorte, chacun devient proactif et expert de sa propre santé et peut confronter ses mesures à celles effectuées par les professionnels de la santé.

En revanche, les contours de la pratique vertueuse de l'automesure sont délimités par des lignes poreuses qui peuvent être facilement franchies. Le *Quantified Self* présente plusieurs risques dont l'impact est encore difficilement mesurable. Premièrement, il y a un risque d'érosion de la solidarité sociale, car on peut imaginer que les individus adoptant des comportements sains et mesurant leurs données de santé en permanence ne veulent plus couvrir les soins des personnes qui ne font pas suffisamment ou pas d'effort pour améliorer leur santé. Deuxièmement, la protection des données personnelles est mise en danger. En effet, l'automesure engendre la collecte d'une quantité massive de données personnelles liées à la santé qui doivent bénéficier d'une protection accrue. Ces données sont entrées dans des algorithmes qui créent des profils et effectuent des recommandations. Or, les mesures mises en place pour sécuriser les transferts des données et leur conservation ne sont pas suffisantes à l'ère des cyberattaques massives. À ce risque s'ajoute également l'attrait commercial des données personnelles pour les entreprises qui détiennent les données ou qui fabriquent les appareils d'automesure connectés. Troisièmement, l'instauration d'une idéologie du contrôle et de la surveillance individuelle permanente peut mener à un risque de dérives importantes en cas d'institutionnalisation, comme cela a été présenté dans l'hypothèse du scénario du « *Big Government* ».

La pratique de l'automesure est prometteuse dans le sens où elle permet d'agir directement sur les facteurs de risque de nature comportementale. Elle engendre une prise de conscience et propose des améliorations. Malheureusement, il demeure trop d'inconvénients qui touchent directement les droits et libertés des adeptes du *Quantified Self* pour que cette pratique puisse être

recommandée sans avertissement préalable. Une amélioration de la sécurité des appareils d'automesure connectés et un renforcement en matière de protection des données permettraient d'atténuer certains points négatifs.

d'autres traits (par exemple la couleur de cheveux), de transférer les expressions faciales d'une personne sur celles d'une autre personne figurant dans une vidéo (*face re-enactment*), de créer une vidéo de quelqu'un qui parle à partir d'un enregistrement audio et d'une séquence vidéo de son visage (*lip-synching*)[40], voire même d'élaborer des contenus entièrement synthétiques[41].

C. D'un point de vue juridique

Sur le plan juridique, il n'est pas certain que la composante d'intelligence artificielle de la définition proposée[42], spécialement du *deep learning*, importe. Par exemple, est-ce vraiment déterminant que l'auteur d'un *deepfake porn* ait utilisé un logiciel fondé sur du *deep learning* plutôt qu'un autre n'utilisant pas de tels procédés tels qu'*Adobe After Effects* ? Nous pensons que cette composante n'est pas déterminante sur le plan juridique. Partant, il convient de considérer que toute manipulation de contenus audiovisuels aboutissant à une synthèse des composants originaux suffit sur le plan juridique. Cela inclut ce qui est parfois qualifié de « *shallow fake* »[43], à savoir des manipulations souvent de moins bonne qualité, comme des images modifiées par le biais de *Photoshop* ou des vidéos altérées avec *Adobe After Effects*[44].

D. Cas d'application concrets positifs et criminels

Si toute technologie en soi est neutre, il découle de la technologie *deepfake* tant des applications bénéfiques (1), que des criminelles (2).

1. *Cas d'application positive*

L'utilisation la plus connue du *deepfake* à des fins non criminelles reste certainement les filtres sur les réseaux sociaux, tels que ceux permettant de changer de sexe, de visage et d'âge sur *Snapchat*[45], ou ceux de l'application

40 J. VINCENT (n. 10).
41 EUROPOL (n. 8), p. 9 ; J. VINCENT (n. 10) ; K. KOBRIGER *et al.* (n. 2), p. 214. Pour un exemple d'images de personnes générées par GAN, voir : https://thispersondoesnotexist.com/ (consulté le 28.08.2022).
42 Cf. *supra* II.A.
43 Voir par exemple : S. MADDOCKS (n. 1), p. 416 ; M. BODI (n. 10), p. 149.
44 B. CHESNEY/D. CITRON (n. 12), p. 1761.
45 Voir par exemple : https://francoischarron.com/sur-le-web/infos/snapchat-utilise-la-technologie-du-deepfake-pour-sa-nouvelle-fonction/3cHPaYidJi/ (consulté le 30.08.2022).

chinoise *Zao* qui permet de superposer son visage sur des scènes de films ou séries TV[46].

Cette technologie est couramment utilisée dans l'industrie du cinéma, notamment pour permettre de rajeunir, vieillir, voire ressusciter certains acteurs[47]. Elle est également utilisée dans le divertissement audiovisuel pour critiquer la société sous la forme de satires[48], et pour obtenir des sous-titres instantanés en plusieurs langues[49].

Pour rester dans le domaine artistique, on peut également citer l'exposition de l'EPFL « *Deep Fakes : Art and Its Double* » qui interroge la capacité des copies numériques de trésors artistiques universels à provoquer, auprès du public, une réaction émotionnelle durable[50].

Cela étant, d'autres applications se développent notamment dans le domaine médical, par exemple pour générer des scans d'IRM synthétiques présentant une situation anormale (comme une tumeur) dans le but de pouvoir créer une base de données de recherche (par exemple pour lutter contre les tumeurs) sans pour autant utiliser les données de vrais patients[51]. Certains auteurs estiment finalement que cette technologie pourrait être utilisée à des fins pédagogiques, par exemple pour avoir dans un cours d'histoire l'avatar d'Abraham Lincoln qui récite aux étudiants le discours de Gettysburg[52].

46 K. DHRUVA (n. 5), p. 7 ; K. KOBRIGER *et al.* (n. 2), p. 214 ; M. BODI (n. 10), p. 147.
47 Par exemple, lorsque l'actrice Carrie Fisher interprétant la princesse Leia dans *Star Wars* est décédée et qu'il a fallu créer des scènes et des dialogues supplémentaires car dans le film son personnage était encore vivant, voir : B. CHESNEY/D. CITRON, (n. 12), p. 1770 ; https://gizmodo.com/it-took-some-movie-magic-to-complete-carrie-fishers-lei-1821121635 (consulté le 12.08.2022). Voir aussi *MyHeritage* qui a pour but d'animer le visage des défunts sur des photographies (www.bbc.com/news/technology-56210053, consulté le 12.08.2022).
48 Ci-après une vidéo superposant Donald Trump sur une séquence de la série *Breaking Bad* relative au blanchiment d'argent : https://youtu.be/Ho9h0ouemWQ (consulté le 12.08.2022).
49 Par exemple le logiciel *Synthesia* : www.synthesia.io/ (consulté le 12.08.2022) ; https://blog.richardvanhooijdonk.com/en/the-good-the-bad-and-the-future-of-deepfakes/ (consulté le 12.08.2022).
50 https://epfl-pavilions.ch/fr/exhibitions/deep-fakes-art-and-its-double (consulté le 12.08.2022).
51 R. VAN HOOIJDONK, *The Good, the Bad, and the Future of Deepfakes*, 23.04.2021, https://blog.richardvanhooijdonk.com/en/the-good-the-bad-and-the-future-of-deepfakes/ (consulté le 12.08.2022).
52 K. DHRUVA (n. 5), p. 10. Pour d'autres utilisations à des fins pédagogiques, voir aussi : B. CHESNEY/D. CITRON (n. 12), p. 1770.

2. *Cas d'application criminelle*

Certaines études ont montré une tendance des criminels à vouloir être précurseurs en matière de nouvelles technologies afin d'obtenir rapidement les connaissances nécessaires pour utiliser, créer et vendre des outils facilitant la cybercriminalité[53].

Le *deepfake* ne fait pas exception à cette tendance, car les premières utilisations criminelles de cette technologie visaient à créer des logiciels utilisant des procédés d'intelligence artificielle permettant la superposition de visages de célébrités féminines sur ceux d'actrices pornographiques[54]. Ces vidéos pornographiques altérées sont couramment dénommées « *deepfake porn* » ; parmi les victimes notoires figurent notamment Gal Gadot, Daisy Ridley, Jessica Alba, Natalie Dormer, Scarlett Johansson, Taylor Swift, Katty Perry, Cara Delevigne et Emma Watson[55].

L'évolution de la technique fait qu'il est désormais possible de modifier n'importe quel élément de la vidéo originale par le biais d'outils relativement facilement accessibles, si bien que les victimes ne sont plus forcément des personnalités connues[56]. En outre, ces logiciels sont parfois utilisés pour se venger de son ou sa partenaire après une rupture, ce qui constitue une forme de *revenge porn*[57].

Il existe aussi des *deepnudes*, en référence à l'application « *DeepNude* » qui permettait de transformer n'importe quelle photo d'une personne habillée, en sa version dénudée[58].

Ces deux activités criminelles, et plus globalement les *deepfakes* à caractère pornographique, constitueraient plus du 90 % de tous les *deepfakes* en ligne[59]. Si ces *deepfakes* peuvent concerner en théorie tant un homme qu'une femme, la majorité des victimes restent toutefois des femmes[60].

53 Ce que l'on appelle aussi « *CaaS* » pour *Crime as a Service*, voir : EUROPOL (n. 8), p. 10.

54 K. KOBRIGER *et al.* (n. 2), p. 207 ss ; M. BODI (n. 10), p. 146 ; S. MADDOCKS (n. 1), p. 1 ; S. COLE, *We Are Truly Fucked : Everyone Is Making AI-Generated Fake Porn Now*, 24.01.2018, www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley (consulté le 12.08.2022).

55 R. SPIVACK (n. 3), p. 345 s. ; S. COLE (n. 54).

56 Un utilisateur du média social Discord a par exemple fait un *deepfake porn* d'une fille avec laquelle il était au collège en utilisant 380 photos de ses comptes Instagram et Facebook. B. CHESNEY/D. CITRON (n. 12), p. 1773 ; A. DODGE *et al.*, « Using Fake Video Technology To Perpetuate Intimate Partner Abuse », *Domestic Violence Advisory* 2018, p. 6.

57 K. KOBRIGER *et al.* (n. 2), p. 210.

58 M. BODI (n. 10), p. 149 ; K. DHRUVA (n. 5), p. 10 ; K. KOBRIGER *et al.* (n. 2), p. 207.

59 S. MADDOCKS (n. 1), p. 416 ; M. BODI (n. 10), p. 149 ; K. KOBRIGER *et al.* (n. 2), p. 208.

60 K. KOBRIGER *et al.* (n. 2), p. 208 ; B. CHESNEY/D. CITRON (n. 12), p. 1773.

Il faut encore mentionner l'utilisation des *deepfakes* à des fins de désinformation ou de propagande[61]. Même s'ils sont moins nombreux que ceux à caractère sexuel, les conséquences potentielles de telles manipulations peuvent être considérables pour la société[62]. Comme exemple, on peut mentionner la vidéo de Boris Johnson qui encourage son électorat à voter pour son concurrent Jeremy Corbyn et vice-versa[63], et plus récemment la fausse vidéo du président ukrainien qui incite sa population à déposer les armes face à la Russie[64].

Les *deepfakes* peuvent être utilisés dans un grand nombre d'autres circonstances, parmi lesquelles porter atteinte à la réputation d'une personne[65], escroquer de l'argent[66], usurper l'identité d'autrui ou se créer une fausse identité[67], ou encore à des fins d'espionnage[68].

61 Ce que l'on qualifie parfois de « *deepfake politique* » ou de « *deepfake news* ». M. BODI (n. 10), p. 150.

62 B. CHESNEY/D. CITRON (n. 12), p. 1777 ss.

63 www.bbc.com/news/av/technology-50381728 (consulté le 12.08.2022).

64 G. CLULEY, *Deepfake President Zelensky Calls on Ukraine To Surrender, As TV Station Hacked*, 17.03.2022, www.bitdefender.com/blog/hotforsecurity/deepfake-president-zelensky-calls-on-ukraine-to-surrender-as-tv-station-hacked/?utm_source=DiploMail&utm_campaign=f9ca82810a-EMAIL_CAMPAIGN_2022_06_03_06_15&utm_medium=email&utm_term=0_4510155485-f9ca82810a-120712296 (consulté le 12.08.2022).

65 On peut citer à cet égard la vidéo de Nancy Pelosi modifiée pour la faire apparaître saoule, retweetée par l'avocat de Donald Trump : « What is wrong with Nancy Pelosi ? » (www.nytimes.com/2019/05/24/us/politics/pelosi-doctored-video.html (consulté le 12.08.2022) ; M. BODI (n. 10), p. 150), ou encore une vidéo du ministre malaisien de l'économie Azmin Ali engagé dans une activité sexuelle avec un autre homme (ce qui constitue dans ce pays un crime passible de prison, voir : www.thestar.com.my/opinion/columnists/one-mans-meat/2019/06/15/is-it-azmin-or-a-deepfake (consulté le 12.08.2022)).

66 Par exemple à des fins d'hameçonnage (*phishing*), en usurpant la voix du directeur d'une entreprise (www.theverge.com/2019/9/5/20851248/deepfakes-ai-fake-audio-phone-calls-thieves-trick-companies-stealing-money, consulté le 12.08.2022) ou dans des cas d'arnaques aux sentiments (*romance scam*).

67 EUROPOL (n. 8), p. 12.

68 Certains auteurs ont étudié la « *deepfake geography* » qui consiste à altérer artificiellement des images satellites, par exemple pour modifier la configuration de certaines villes ou cacher des infrastructures essentielles. B. ZHAO *et al.*, « Deep Fake Geography ? When Geospatial Data Encounter Artificial Intelligence », *Cartography and Geographic Information Science* 2021, 48:4, p. 338 ss.

III. Comment lutter contre les *deepfakes* ?

A. Solutions techniques

Si d'importants progrès ont été réalisés en matière de détection et de prévention des *deepfakes*, et qu'il y a une participation de plus en plus importante des acteurs privés à contribuer aux recherches sur ce sujet[69], il n'existe pourtant aucune mesure qui, à notre connaissance, permet de prévenir complètement l'utilisation criminelle de cette technologie[70].

En outre, il existe passablement d'études dans lesquelles les chercheurs mettent au point des méthodes permettant de contourner les mécanismes de prévention et de détection établis dans d'autres études[71]. En parallèle, il existe dans ce domaine la même logique que la course aux virus et aux antivirus, en ce sens que les criminels mettent au point des algorithmes de production de *deepfakes* de plus en plus élaborés notamment grâce au GAN, ce qui rend quasiment impossible la tâche de créer un parfait algorithme de détection à ce stade[72]. Ainsi, il convient d'examiner si des solutions autres que techniques permettent de pallier les *deepfakes*.

B. Solutions légales

1. Remarques liminaires

Pour les *deepfakes*, comme pour les autres nouvelles technologies, on peut noter trois tendances législatives pour réglementer les problématiques liées à leur utilisation. La première consiste à adopter des législations visant spécialement à lutter ou réguler les *deepfakes*. C'est l'approche choisie par les

69 Facebook a par exemple mis sur pied le « *Deepfake Detection Challenge* » en partenariat avec le monde académique et privé pour élaborer plus rapidement des logiciels efficaces de détection, https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/ (consulté le 30.08.2022). Voir aussi : K. KOBRIGER *et al.* (n. 2), p. 209.

70 B. CHESNEY/D. CITRON (n. 12), p. 1787 s.

71 K. KOBRIGER *et al.* (n. 2), p. 217 ss et les références citées.

72 K. KOBRIGER *et al.* (n. 2), p. 219.

États de Californie[73], du Texas[74], de la Virginie[75] et au niveau fédéral[76] aux États-Unis. La deuxième met en place un cadre qui, bien que ne traitant pas spécifiquement des *deepfakes*, régule toutefois indirectement le phénomène. Cela concerne par exemple le futur règlement sur l'intelligence artificielle en Europe[77] ou l'adoption de la loi contre la manipulation de l'information en France[78]. Le troisième choix fait par certains pays, dont la Suisse, consiste à ne pas adopter de loi sectorielle ou propre aux *deepfakes*, en partant du principe que les dispositions générales en vigueur suffisent pour réguler ce phénomène.

L'aspect principal de cette section est d'aborder, sous l'angle du droit pénal suisse, comment il est possible d'incriminer la personne qui crée ou utilise des *deepfakes* à des fins criminelles (2). Nous passerons ensuite rapidement sur les moyens civils permettant à la victime d'obtenir la réparation de son préjudice (3) et sur les possibilités de responsabiliser les plateformes numériques (4).

2. La mise en ligne d'un deepfake est-elle en soi punissable en droit pénal ?

Il convient de rappeler qu'il n'existe pas de loi ni d'article réglementant spécifiquement les *deepfakes* en droit suisse, c'est pourquoi nous analyserons les infractions figurant dans la partie spéciale du Code pénal[79].

73 La Californie a adopté la loi AB-602 (disponible sous : https://openstates.org/ca/bills/20192020/AB602/, consulté le 18.08.2022) qui vise uniquement les *deepfakes* utilisés à des fins pornographiques et qui offre des voies de droits spéciales aux victimes. K. DHRUVA (n. 5), p. 29 ; M. BODI (n. 10), p. 156.

74 Le Texas s'est muni d'une loi SB 751 (disponible sous : https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751F.htm, consulté le 18.08.2022) interdisant l'utilisation des *deepfakes* afin de manipuler des élections.

75 L'art. 5, § 18.2-386.2 du Code de Virginie (disponible sous : https://law.lis.virginia.gov/vacode/title18.2/chapter8/section18.2-386.2/, consulté le 18.08.2022) réprime, à l'instar de la Californie, l'utilisation des *deepfakes* à des fins pornographiques. K. DHRUVA (n. 5), p. 29 ; C. LANGLAIS-FONTAINE (n. 11), p. 4.

76 Voir H. R. 6088 (disponible sous : www.congress.gov/bill/116th-congress/house-bill/6088/text, consulté le 18.08.2022) qui interdit, comme le Texas, la manipulation d'élections par le biais de cette technologie.

77 Voir la proposition de règlement du parlement européen et du conseil établissant des règles harmonisées concernant l'intelligence artificielle du 21 avril 2021 (COM/2021/206 final), disponible sous : https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52021PC0206 (consulté le 18.08.2022).

78 Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559 (consulté le 18.08.2022).

79 Code pénal suisse du 21 décembre 1937 (CP ; RS 311).

En outre, la question qui nous intéresse principalement est celle de savoir si la mise en ligne d'un *deepfake* constitue en elle-même une infraction pénale (*per se*). Pour ce faire, nous limiterons l'examen aux *deepfakes* à caractère pornographique (*deepfakes porn* et *deepnudes*) (a) et à des fins de désinformation ou de manipulation électorale (*deepfakes news*) (b), car il s'agit des utilisations criminelles les plus importantes et les plus courantes. Enfin, nous donnerons un bref aperçu d'autres infractions qui peuvent être réalisées par le biais de *deepfakes* (c), autrement dit des situations où l'on ne s'intéresse pas à la punissabilité des *deepfakes* en soi, mais à leur utilisation.

a) Deepfakes porn et *deepnudes*

Il convient d'analyser les *deepfakes* à caractère pornographique sous l'angle des infractions contre l'intégrité sexuelle (aa), des infractions contre l'honneur (bb), des infractions contre le domaine secret ou privé (cc), et des infractions contre le droit d'auteur (dd).

aa) Infractions contre l'intégrité sexuelle

aaa) Importuner autrui avec de la pornographie (art. 197 ch. 2 CP)

Lorsque l'auteur met en ligne un *deepfake porn* ou un *deepnude*, il est intéressant de savoir si l'intégrité sexuelle, principalement de la personne qu'on superpose sur l'acteur pornographique, mais plus généralement aussi de toute personne qui tombe sur un tel contenu, est protégée.

À cet égard, l'art. 197 ch. 2 CP sanctionne quiconque importune autrui avec de la pornographie. Pour que l'infraction soit réalisée, il faut un objet ou une représentation (1), à caractère pornographique (2), sans valeur culturelle ni scientifique (3), une présentation en public ou à une personne sans y avoir été invité (4) et une intention (5)[80]. Les éléments constitutifs de l'infraction (2) et (4) méritent une attention particulière dans le contexte des *deepfakes*.

En effet, il faut que le *deepfake* soit présenté au public, c'est-à-dire à un cercle indéterminé – même restreint – de personnes, « sans y avoir été invité »[81]. Cela exclut par exemple les *deepnudes* envoyés en privé à la victime, notamment à des fins de chantage, ou dans des groupes de personnes déterminées qui veulent

80 A. CAMBI FAVRE-BULLE, *in* A. MACALUSO/L. MOREILLON/N. QUELLOZ (édit.), *CR Code pénal II*, Bâle 2017, art. 197 N 1 ss ; B. CORBOZ, *Les infractions en droit suisse*, Volume I, 3e éd., Berne 2010, art. 197 N 36 ss.

81 A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 36.

consommer de tels contenus. En outre, si le contenu pornographique n'est pas directement perceptible ou que les spectateurs sont avertis en avance de la nature du contenu, alors la punissabilité est exclue[82]. Un site Internet ou un canal de discussion sur Reddit intitulé « *deepfakes porn* », dont l'accès au contenu nécessite un clic ou une action, ne remplit pas cet élément constitutif, car celui qui accède au contenu n'y est pas confronté inopinément. Ainsi, l'élément (4) restreint considérablement le champ d'application de l'art. 197 ch. 2 CP dans le contexte des *deepfakes* à caractère pornographique, puisque ceux-ci se trouvent majoritairement sur des sites Internet.

Ensuite, si le caractère pornographique ne pose pas de problème dans le cas des *deepfakes porn*, il peut toutefois en poser pour les *deepnudes*. En effet, même si la notion de pornographie reste une notion juridique indéterminée[83], toute représentation du corps humain nu, de ses attributs ou de la sexualité en général ne saurait être qualifiée de pornographique[84]. Ce qui est érotique n'est donc pas nécessairement pornographique[85]. La jurisprudence et la doctrine retiennent que la représentation pornographique doit avoir pour but de provoquer une excitation sexuelle de la personne qui y est confrontée (par exemple, par une pose sexuellement évocatrice) et doit insister exagérément sur les parties génitales dans le sens d'une sexualité sans connotation humaine et émotionnelle[86]. Or, dans le cadre d'un *deepnude* où l'on transforme l'image d'une femme habillée en sa version dénudée, il n'y a pas forcément d'accent qui est mis sur les parties génitales, ni de position sexuellement explicite. En l'absence de ces éléments, on ne saurait retenir le caractère pornographique des *deepnudes*. L'élément constitutif de l'infraction (2) exclut ainsi potentiellement beaucoup de *deepnudes* du champ d'application de l'art. 197 ch. 2 CP.

En résumé, cet article n'est susceptible de trouver application que lorsqu'un internaute est confronté inopinément à un *deepfake* à caractère pornographique, par exemple lorsqu'un tel contenu apparaît spontanément en déroulant son fil d'actualité Facebook ou Instagram. S'il s'agit d'un *deepnude*, il faut encore qu'il vise à exciter sexuellement en ayant un focus particulier sur les parties génitales pour être illicite.

82 ATF 128 IV 260, consid. 2.1.

83 ATF 128 IV 260, consid. 2.1.

84 B. ISENRING/M. A. KESSLER, *in* M.-A. NIGGLI/H. WIPRÄCHTIGER (édit.), *Basler Kommentar Strafrecht II*, 4e éd., Bâle 2019, art. 197 N 9 et 14 b ; A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 9.

85 A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 9.

86 ATF 128 IV 260, consid. 2.1 ; M. DEPUIS *et al.* (édit.), *PC Code pénal*, 2e éd., Bâle 2017, art. 197 N 15 s. ; A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 7 ; B. ISENRING/M. A. KESSLER (n. 84), art. 197 N 14.

L'auteur qui met en ligne un *deepfake* illicite peut être punissable selon l'art. 197 ch. 2 CP. En revanche, les internautes qui consomment de tels contenus ne sont pas punissables dans la mesure où la consommation de pornographie douce n'est pas punissable[87].

bbb) Pornographie infantile (art. 197 ch. 4 et 5 CP)

Aux États-Unis, la question des *deepfakes* à caractère de pornographie infantile a été tranchée dans deux arrêts importants. Il faut noter à ce sujet que le gouvernement américain avait voulu étendre la protection accordée par le *Child Pornography Prevention Act* de 1996 (CPPA) contre les contenus pédopornographiques en bannissant

> *« any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture » that « is, or appears to be, of a minor engaging in sexually explicit conduct »*[88].

Toutefois, dans le premier arrêt en la matière, *Ashcroft v. Free Speech Coalition*, la Cour suprême a considéré que le CPPA était contraire à la liberté d'expression accordée par la Constitution américaine, car il était formulé trop largement[89]. La Cour a limité la portée de ce texte en jugeant que le droit américain ne protège pas les *deepfakes porn* infantiles qui ne concernent pas un enfant existant réellement[90].

Dans le second arrêt, *United States v. Hotaling*, une Cour d'appel a retenu que le fait de superposer les visages d'enfants de sexe féminin existant réellement sur ceux d'actrices pornographiques adultes qui se livrent à des actes sexuels n'est pas protégé par la liberté d'expression[91]. Cet arrêt confirme que seule l'utilisation de contenus audiovisuels d'enfants existant réellement est illicite aux États-Unis.

En Suisse, la question n'a pas été tranchée à notre connaissance. Selon nous, il convient d'admettre que tant la situation dans laquelle l'image ou la vidéo d'un enfant existant réellement a été modifiée, de sorte à le dénuder (*deepnude*), ou a été superposée sur l'image d'un acteur pornographique (majeur ou mineur) se livrant à des actes d'ordre sexuel (*deepfake porn*), que celle où on crée des contenus pédopornographiques sans pour autant utiliser l'image d'un enfant

87 B. ISENRING/M. A. KESSLER (n. 84), art. 197 N 19.

88 § 2256(8)(B) du CCPA, disponible sous : www.congress.gov/bill/104th-congress/house-bill/4123/text (consulté le 30.08.2022).

89 *Ashcroft v. Free Speech Coalition*, 535 U.S. 234 (2002).

90 K. KOBRIGER *et al.* (n. 2), p. 224.

91 *United States v. Hotaling*, 634 F.3d 725, 728, 730 (2 d Cir. 2011).

existant réellement constituent des *deepfakes* pédopornographiques (pornographie dure[92]) réprimés à l'art. 197 ch. 4 et 5 CP[93].

Il convient d'émettre une réserve par rapport à la jurisprudence du Tribunal fédéral dans l'hypothèse des *deepnudes* infantiles. D'une manière générale, le Tribunal fédéral admet que des images d'enfants nus peuvent être considérées comme pornographiques, même sans accent particulier sur les parties génitales, notamment lorsque l'enfant est mis en scène dans une position objectivement provocante[94]. Néanmoins, il ressort de la jurisprudence que le caractère pornographique n'est pas réalisé lorsque l'auteur n'a manifestement exercé aucune influence sur l'enfant lors de la prise de vue[95]. Cette précision jurisprudentielle ne saurait être applicable aux *deepnudes* d'enfants, car cela permettrait d'exempter de tout caractère pénal la manipulation *a posteriori* d'une photo où l'enfant habillé posait sans influence des parents quant à la prise de vue.

Le droit pénal sanctionne un certain nombre de comportements associés à la pédopornographie. Ainsi, sont punissables conformément aux al. 4 et 5 de l'art. 197 CP, non seulement les personnes qui créent les *deepfakes* et les *deepnudes* à caractère pédopornographique, mais également celles qui les mettent en ligne, celles qui les hébergent et même celles qui les consomment ou en possèdent.

ccc) Synthèse

Sous l'angle des infractions protégeant l'intégrité sexuelle, la mise en ligne d'un *deepfake* est en soi punissable lorsqu'il porte atteinte au droit de tout individu de ne pas être confronté contre son gré à de la pornographie (art. 197 ch. 2 CP), et lorsqu'il s'agit de représentations ou d'objets pédopornographiques (art. 197 ch. 4 et 5 CP). En revanche, on peut regretter qu'il n'existe

92 On vise les objets ou représentations ayant pour contenu des actes d'ordre sexuel avec des mineurs ; A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 53. Ces actes peuvent être effectifs ou non effectifs, ces derniers incluant les cas sans participation d'acteurs mineurs « réels » ; B. ISENRING/M. A. KESSLER (n. 84), art. 197 N 22d.

93 Du même avis : A. CAMBI FAVRE-BULLE (n. 80), art. 197 N 55 ; B. ISENRING/M.A. KESSLER (n. 84), art. 197 N 22 d, qui considèrent qu'il s'agit de pornographie dure dès lors que les personnes représentées paraissent avoir moins de 18 ans.

94 ATF 6S.345/2004 du 8 mars 2005 ; M. FELBER, « 18. Urteil 6S.345/2004 vom 8. März 2005 », *RSJ* 101/2005, p. 1 ss, p. 273 ; ATF 133 IV 31, consid. 6.1.2.

95 ATF 131 IV 31, consid. 6.1.2. Dans cet arrêt, il s'agissait d'une photo d'une jeune fille nue avec les jambes écartées prise à la plage par le père. Le Tribunal fédéral a estimé que la photo immortalisait un souvenir de vacances et que le père n'avait pas encouragé une telle position, cette dernière étant naturelle.

aucune disposition qui protège l'intégrité sexuelle de la personne dont on usurpe l'image ou d'autres attributs pour créer des *deepfakes* pornographiques.

bb) Infractions contre l'honneur

À défaut d'une protection de son intégrité sexuelle, l'honneur de la personne mise en scène dans le *deepfake* est-il protégé ? Trois infractions rentrent en ligne de compte : la diffamation (art. 173 CP), la calomnie (art. 174 CP) et l'injure (art. 177 CP).

aaa) Injure (art. 177 CP)

L'injure réprime quiconque, qui aura, par la parole, l'écriture, l'image, le geste ou par des voies de fait, attaqué autrui dans son honneur (art. 177 al. 1 CP). Si l'image et le geste constituent des faits attentatoires à l'honneur, force est d'admettre qu'une vidéo est également concernée. Ainsi, tant les *deepnudes* que les *deepfakes porn* rentrent dans le champ d'application de cette infraction. Toutefois, le fait attentatoire à l'honneur doit être adressé directement à la victime et non pas à un tiers[96], car dans ce dernier cas, la diffamation ou la calomnie prime l'injure[97].

L'injure ne concerne donc que les *deepfakes* adressés exclusivement à la victime, par exemple dans le cas d'un *deepnude* envoyé à la victime à des fins de chantage. En revanche, si le *deepfake* est rendu également accessible à des tiers, ce qui est le cas d'un *deepfake porn* figurant sur un site Internet, l'injure ne trouve plus application.

Les concours d'infractions sont possibles lorsque les biens juridiquement protégés sont différents[98], raison pour laquelle l'injure est subsidiaire à la diffamation et à la calomnie qui protègent aussi l'honneur de la victime[99]. En revanche, un concours reste envisageable avec les infractions de pornographie

96 M. DEPUIS *et al.* (n. 86), art. 177 N 17 ; S. TRECHSEL/M.-J. LEHMKUHL, *in* S. TRECHSEL/M. PIETH (édit.), *PK Schweizerisches Strafgesetzbuch*, 4e éd., Bâle 2021, art. 177 N 2 ; B. CORBOZ (n. 80), art. 177 N 20-21.

97 L. RIEBEN/M. MAZOU, *in* A. MACALUSO/L. MOREILLON/N. QUELLOZ (édit.), *CR Code pénal II*, Bâle 2017, art. 177 N 11 ; F. RIKLIN, *in* M.-A. NIGGLI/H. WIPRÄCHTIGER (édit.), *Basler Kommentar Strafrecht II*, 4e éd., Bâle 2019, art. 177 N 35.

98 M. DEPUIS *et al.* (n. 86), art. 177 N 33.

99 Cf. Chapitre 1 du Titre 3 CP « délits contre l'honneur ».

(art. 197 CP)[100], de chantage (art. 156 CP) et de menace ou de contrainte (art. 180 ou 181 CP).

bbb) Calomnie et diffamation (art. 173 et 174 CP)

Pour déterminer si des *deepfakes* pornographiques peuvent être constitutifs d'une calomnie ou d'une diffamation, il s'agit dans les deux cas de savoir s'ils constituent une atteinte à l'honneur (1), communiquée à un tiers (2) avec intention (3)[101]. Par « communication à un tiers », il faut que l'auteur accuse ou jette le soupçon (a) d'une allégation de fait (b), adressée à un tiers (c) qui porte atteinte à une personne reconnaissable (d)[102].

Lorsque l'auteur d'un *deepfake* pornographique le met en ligne, il accuse ou jette volontairement le soupçon sur une victime d'adopter un certain comportement, en l'occurrence de jouer dans un film pornographique ou de poser nue. Cela constitue une allégation de fait – par opposition à un jugement de valeur – car on affirme, par une vidéo ou une image, qu'une personne s'est comportée d'une manière déterminée, que ce soit activement ou passivement, ou qu'elle se trouve dans un état de fait déterminé[103].

De plus, en rendant un *deepfake* accessible en ligne, par exemple sur un site Internet ou sur un réseau social, l'auteur accepte qu'au moins un tiers puisse prendre connaissance de son allégation, ce qui remplit également cet élément[104]. En revanche, si le *deepfake* n'est communiqué qu'à la victime, seule l'injure rentre en ligne de compte, sauf s'il est en sus rendu accessible à des tiers, par exemple si l'auteur transmet d'abord le *deepnude* à la victime pour la contraindre à adopter un certain comportement et que, suite à sa non-exécution, il le publie en ligne sur les réseaux sociaux. Dans ce cas, seule la calomnie sera retenue étant donné le caractère subsidiaire de l'injure[105].

Pour que l'élément constitutif de la communication à un tiers soit réalisé, il faut encore que l'atteinte soit dirigée contre une personne reconnaissable. Or, tel est

100 S. TRECHSEL/M.-J. LEHMKUHL (n. 96), art. 177 N 9 ; M. DEPUIS *et al.* (n. 86), art. 177 N 33.

101 L. RIEBEN/M. MAZOU (n. 97), art. 173 N 1 ss.

102 M. DEPUIS *et al.* (n. 86), art. 173 N 1 ss.

103 L. RIEBEN/M. MAZOU (n. 97), intro. aux art. 173-178 N 29.

104 B. CORBOZ (n. 80), art. 173 N 43-44.

105 Cf. *supra* III.B.2.bb)aaa). Une partie de la doctrine admet cependant un concours parfait entre la diffamation et l'injure dans ce scénario ; voir : L. RIEBEN/M. MAZOU (n. 97), art. 173 N 55. Toutefois, on voit mal pourquoi un concours serait admissible avec la diffamation et pas avec la calomnie dans la mesure où le bien juridiquement protégé (l'honneur) est identique pour les trois infractions et que le texte de l'art. 177 CP indique bien qu'il est subsidiaire aux deux autres normes.

généralement le cas dans la mesure où c'est son visage et/ou sa voix qui est superposé dans un film ou sur une photo. Il n'est d'ailleurs pas exclu qu'une combinaison d'éléments permette clairement d'identifier la personne, par exemple si un *deepfake porn* comporte une légende qui indique l'identité de la victime.

En mettant en ligne un *deepfake* à caractère pornographique, il est vraisemblable que l'auteur agisse avec la conscience et la volonté de nuire à l'honneur de la victime ou du moins l'accepte dans le cas où cela se produirait (art. 12 al. 2 CP). Se faisant, l'auteur agit généralement par intention.

Il reste encore à déterminer si de tels *deepfakes* constituent une atteinte à l'honneur pénal de la victime. Tout d'abord, l'atteinte peut prendre diverses formes qui ne résument pas uniquement à la forme verbale et écrite. En outre, la diffamation et la calomnie peuvent être réalisées par la parole, l'écriture, l'image, le geste, ou par tout autre moyen (art. 176 CP). La doctrine considère à cet égard que cela inclut les atteintes réalisées sur Internet et les réseaux sociaux[106], mais aussi sous la forme de photomontages, de films ou de vidéos[107]. Les *deepfakes porn* et *deepnudes* peuvent ainsi réaliser ces infractions.

Sans se livrer à un examen exhaustif de l'honneur pénalement protégé, on retiendra qu'il s'agit d'un droit au respect[108] de la réputation et du « sentiment d'être un homme honorable, de se comporter, en d'autres termes, comme un homme digne a coutume de le faire selon les idées généralement reçues »[109]. En outre, le Tribunal fédéral a jugé que le fait de désigner une personne comme expéditrice d'un message à caractère pornographique[110] ou le fait de simuler une fellation de sorte à réduire une personne au rang d'objet sexuel[111] marquent tous deux un mépris certain pour la victime et sont constitutifs d'une atteinte à l'honneur. Par extension, un *deepfake porn* réduit aussi la victime au simple statut d'objet sexuel, de sorte que nous pensons qu'il constitue une atteinte à l'honneur. La situation peut être plus nuancée pour les *deepnudes*, mais il reste vraisemblable à notre avis qu'ils violent généralement le sentiment de la victime d'être une personne honorable, qui se comporte comme une personne digne a coutume de le faire.

Par conséquent, nous retenons que les *deepfakes* à caractère pornographique peuvent être constitutifs d'une diffamation (art. 173 CP) ou d'une calomnie (art. 174 CP). En principe, l'auteur qui crée ou commandite le *deepfake* sait ou doit savoir que ce dernier est faux. Comme la connaissance de la fausseté de

106 L. RIEBEN/M. MAZOU (n. 97), art. 176 N 1.
107 F. RIKLIN (n. 97), art. 176 N 1.
108 ATF 137 IV 313, consid. 2.1.1
109 L. RIEBEN/M. MAZOU (n. 97), intro. aux art. 173-178 N 18 et les arrêts cités.
110 ATF 6S_147/2002 du 21 août 2002, consid. 3.2.
111 ATF 6B_492/2013 du 18 juin 2013, consid. 1.

l'allégation est un élément constitutif propre à l'art. 174 CP, l'auteur d'un *deepfake* sera en général punissable sous l'angle de la calomnie.

ccc) La propagation des *deepfakes* (art. 173 et 174 ch. 1 al. 2 CP)

La personne qui commandite ou met en ligne le *deepfake* n'est toutefois pas la seule personne susceptible d'être punissable pour calomnie ou diffamation. Les art. 173 et 174 ch. 1 al. 2 CP disposent que « celui qui aura propagé une telle accusation ou un tel soupçon » est également punissable sur plainte.

Le Tribunal fédéral a récemment jugé qu'une personne qui *like* ou partage un post diffamatoire sur Facebook est punissable selon l'art. 173 ch. 1 al. 2 CP[112]. On peut en déduire que toute personne qui contribue à propager un *deepfake* en ligne, par exemple en partageant le *deepnude* ou la vidéo, en *likant* un tel contenu, en le *retweetant*, etc., peut se rendre coupable pénalement.

Il convient ici de différencier la situation dans laquelle une personne partage le *deepfake* en reconnaissant qu'il s'agit d'un faux, ce qui serait constitutif de l'art. 174 ch. 1 al. 2 CP, de celle de la personne qui partage le *deepfake* en ignorant la fausseté de l'allégation, ce qui remplirait les éléments constitutifs de l'art. 173 ch. 1 al. 2 CP.

Ainsi, quelqu'un qui *likerait* sur un site dédié un *deepfake porn* d'une actrice de cinéma notoirement connue[113] sait ou doit savoir qu'il s'agit d'un faux, et pourrait par conséquent être punissable conformément à l'art. 174 ch. 1 al. 2 CP.

cc) Infractions contre le domaine secret ou privé

Il convient également d'aborder l'art. 179^{decies} CP réprimant l'usurpation d'identité qui entrera en vigueur en même temps que la future loi sur la protection des données, vraisemblablement en septembre 2023[114]. Cette disposition, qui s'insère sous le titre « Infractions contre l'honneur et contre le domaine secret ou domaine privé », protège la personnalité, à savoir le droit de la

112 ATF 146 IV 23. La propagation au sens de l'art. 173 ch. 1 al. 2 CP est considérée comme une variante distincte de l'infraction. Elle exige que la déclaration déjà faite par une autre partie soit communiquée à un tiers. Ce n'est que lorsque le contenu diffamatoire de l'auteur initial, auquel la personne (*Weiterverbreiter*) réagit par un « *like* » ou un « *share* », devient visible pour un tiers et que celui-ci l'a remarqué, que l'infraction est consommée (ATF 146 IV 23, consid. 2.2.4).
113 Cf. R. SPIVAK (n. 3), p. 345 s., pour une liste de célébrités concernées.
114 www.bj.admin.ch/bj/fr/home/staat/gesetzgebung/datenschutzstaerkung.html (consulté le 30.08.2022).

personne au respect de son identité, et punit toute usurpation qui cause une grave atteinte à la personnalité[115].

Cet article mérite une attention particulière en matière de *deepfakes* dans la mesure où il permet de sanctionner quiconque utilise l'identité d'un tiers (1), sans son consentement (2), avec l'intention (3) et le dessein de nuire ou de procurer un avantage (4).

Tout d'abord, c'est le fait d'utiliser l'identité d'un tiers qui est sanctionné, ce qui couvre toutes les variantes possibles de l'infraction[116]. Le Message du Conseil fédéral liste (de manière non exhaustive) un certain nombre d'éléments permettant de déterminer l'identité d'une personne, comme son nom, et sa photo[117]. La photo couvre les *deepnudes*, mais il convient de retenir que ça inclut plus généralement toute représentation visuelle d'une personne, que ce soit une image, un dessin, une vidéo, ou toute autre reconstitution synthétique de la personne (ou d'une partie caractéristique de celle-ci, comme son visage et sa voix).

Toutefois, certaines utilisations de *deepfakes* peuvent être problématiques, notamment lorsque le média synthétique est le résultat d'un mélange de plusieurs identités (par exemple un *deepfake porn* qui superpose le visage d'une victime A, avec la voix d'une personne B, et avec un sous-titre mentionnant C). En effet, le recours à une identité inventée n'est pas sanctionné par le nouvel article[118]. Dans ce cas où le média synthétique est créé sur la base de plusieurs éléments appartenant à des identités différentes, il faudra examiner concrètement si les personnes concernées sont reconnaissables[119]. Dans l'affirmative, l'infraction sera réalisée[120].

Concernant les *deepfakes*, l'absence de consentement de la victime et l'intention de l'auteur d'usurper l'identité de celle-ci ne posent pas de difficulté particulière. Toutefois, la simple création d'un *deepfake* ne suffit pas encore pour réaliser l'infraction, car il faut encore que l'auteur agisse avec le dessein particulier de nuire ou de procurer un avantage illicite[121]. Ce dernier peut notamment se matérialiser par le fait de rendre accessible à d'autres personnes des contenus

115 Message concernant la loi fédérale sur la révision totale de la loi fédérale sur la protection des données et sur la modification d'autres lois fédérales, FF 2017 6741.

116 S. MÉTILLE, *in* A. MACALUSO/L. MOREILLON/N. QUELOZ (édit.), *CR Code pénal II*, 2e éd., (à paraître), art. 179decies.

117 Voir : FF 2017 6741 (n. 115), tout en précisant que l'identité peut aussi être déterminée au moyen d'une combinaison de ces différents éléments. S. MÉTILLE (n. 116).

118 FF 2017 6741 (n. 115).

119 Y. REBER, « Der neue Tatbestand des Identitätsmissbrauchs nach Art. 179decies E-StGB », *ex ante* 2/2020, p. 33 ss, p. 35.

120 S. MÉTILLE (n. 116).

121 FF 2017 6741 (n. 115).

qui ridiculisent la victime ou lui donnent une mauvaise image, ce qui est indubitablement le cas pour les *deepfakes*[122]. Le Conseil fédéral voulait, en exigeant un dessein particulier, éviter de punir les actes commis par exubérance ou espièglerie[123].

Dans le cas des *deepfakes*, surtout à caractère pornographique, on ne saurait retenir un agissement par espièglerie, étant donné les répercussions que peut avoir la publication de tels contenus sur la victime. Ainsi, nous retiendrons que l'auteur qui met en ligne ou rend accessible des *deepfakes* pourra être punissable à l'avenir d'usurpation d'identité pour autant qu'au moins une victime de l'usurpation soit clairement reconnaissable.

dd) Infractions contre le droit d'auteur

L'examen des infractions contre le droit d'auteur pouvant résulter des *deepfakes* aurait pu faire l'objet d'une contribution à part entière, si bien que nous ne présentons ici que des pistes de réflexion en la matière.

La Loi sur le droit d'auteur et les droits voisins (LDA)[124] contient une disposition pénale punissant notamment quiconque, intentionnellement et sans droit modifie une œuvre ou utilise une œuvre pour créer une œuvre dérivée (art. 67 let. c et d LDA).

Conformément à l'art. 2 al. 1 LDA, est une œuvre protégée par la Loi sur le droit d'auteur et les droits voisins « toute création de l'esprit, littéraire ou artistique, qui a un caractère individuel ». En fonction des circonstances, les *deepfakes porn* et les *deepnudes* peuvent constituer des créations de l'esprit (art. 2 al. 2 let. g LDA) entrant dans le domaine artistique et présentant un caractère individuel[125]. Si tel est le cas, ce sont des œuvres protégées par le droit d'auteur sur lesquelles l'auteur a certains droits exclusifs. Parmi ceux-ci, il peut notamment décider comment son œuvre peut être utilisée (art. 10 LDA), si, quand et de quelle manière l'œuvre peut être modifiée, ou être utilisée pour la création d'une œuvre dérivée (art. 11 al. 1 LDA).

À défaut de consentement de l'auteur de l'œuvre, qui peut être le producteur de la vidéo à caractère pornographique altérée ou la victime d'un *deepnude*

122 Y. REBER (n. 119), p. 36.

123 FF 2017 6741 (n. 115).

124 Loi fédérale sur le droit d'auteur et les droits voisins du 9 octobre 1992 (LDA ; RS 231.1).

125 Sur ce point, l'art. 2 al. 3bis LDA protège également les photos dépourvues de caractère individuel, ce qui inclut donc les *deepnudes*.

pour une photo habillée qu'elle avait prise elle-même, l'auteur du *deepfake* qui le met en ligne peut, selon nous, être punissable en vertu de l'art. 67 LDA[126].

b) *Deepfakes news*

Le droit pénal n'a pas pour but d'instaurer un devoir de vérité généralisé, et dès lors ne saurait punir tout mensonge ou manipulation de l'information[127]. Cependant, le législateur a considéré que la volonté populaire et l'ordre constitutionnel suisse constituaient des biens juridiques dignes de protection, si bien qu'une atteinte à ceux-ci est susceptible d'entraîner une responsabilité pénale.

Tout d'abord, la manipulation d'une information par le biais de *deepfakes* pourrait tomber sous le coup des art. 279 ss CP qui figurent sous le Titre 14 du Code pénal intitulé « Infractions contre la volonté populaire ». Il s'agit ici de criminaliser les comportements qui menacent la volonté populaire, car il est capital dans une démocratie que les droits politiques conférés aux citoyens puissent être exercés librement « en vue d'assurer l'expression de la réelle volonté du peuple »[128].

Toutefois, l'art. 279 CP (violences), l'art. 280 CP (atteinte au droit de vote), l'art. 281 CP (corruption électorale), l'art. 282 CP (fraude électorale), l'art. 282bis CP (captation de suffrages) et l'art. 283 CP (violation du secret de vote) sont inapplicables aux *deepfakes news*, car ces normes ne visent à protéger la volonté populaire que dans le cadre de l'exercice du droit de vote, soit au moment du scrutin populaire, mais pas au moment de la formation de la volonté populaire[129].

Il reste, sous l'angle du droit pénal, les infractions relatives à la mise en danger de l'ordre constitutionnel (art. 275 ss CP). L'art. 275ter CP (groupements illicites) n'est pas pertinent pour notre analyse, et l'art. 275bis CP (propagande subversive) nécessite l'usage de la violence, ce qui ne concerne pas les *deepfakes news*. Ainsi, il convient d'écarter ces normes et d'examiner le seul

126 Autrement dit, le titulaire des droits lésés a qualité pour porter plainte au sens de cette disposition. N. TISSOT/D. KRAUS/V. SALVADÉ, *Propriété intellectuelle : Marques, brevets, droit d'auteur*, Berne 2019, N 1219.

127 K. LUBISHTANI/M. FLATTET, « La démocratie directe face à la manipulation de l'information par des particuliers », *PJA* 2019, p. 710 ss, p. 715.

128 K. LUBISHTANI/M. FLATTET (n. 127), p. 716 ; ATF 121 I 138, consid. 3, JdT 1997 I 74 ; S. WEHRLE, *in* M.-A. NIGGLI/H. WIPRÄCHTIGER (édit.), *Basler Kommentar Strafrecht II*, 4e éd., Bâle 2019, Vor Art. 279 N 5.

129 K. LUBISHTANI/M. FLATTET (n. 127), p. 716 ; P. EGLI/D. RECHSTEINER, « Social Bots und Meinungsbildung in der Demokratie », *PJA* 2017, p. 249 ss, p. 253.

article potentiellement applicable aux manipulations de l'information en ligne : l'art. 275 CP, qui réprime tout atteinte à l'ordre constitutionnel.

Cet article appréhende les actes (1) intentionnels (2) tendant à troubler ou à modifier d'une manière illicite (3) l'ordre fondé sur la Constitution (4)[130]. Certains auteurs considèrent que cette infraction peut s'appliquer aux manipulations d'information en ligne, mais de manière restrictive[131]. En effet, ils considèrent que l'ordre constitutionnel doit être compris comme la tenue libre et sans entrave du processus démocratique lors d'un scrutin populaire, qu'il s'agisse d'une votation ou d'une élection[132]. De plus, seul le comportement adopté délibérément à des fins manipulatoires dans le contexte politique pour nuire au débat démocratique, manipuler l'opinion publique ou pour déstabiliser la démocratie doit être incriminé[133].

Ainsi, le champ d'application de l'art. 275 CP est relativement restreint et exclut un certain nombre de *deepfakes news*. Par exemple, la vidéo du premier ministre britannique Boris Johnson qui encourage son électorat à voter pour son concurrent Jeremy Corbyn et vice-versa[134], ne réaliserait pas l'infraction, car il manque l'intention de fausser l'opinion publique. En revanche, si cette vidéo avait vraiment pour but d'influencer l'électorat afin de promouvoir l'un des deux candidats, l'infraction aurait été réalisée[135]. En outre, dès lors que l'auteur du *deepfake* a réellement pour objectif de manipuler le résultat d'une votation (fédérale ou cantonale) ou d'une élection, l'infraction est réalisée.

Ainsi, le contexte dans lequel intervient le *deepfake* est déterminant pour cette infraction. Par exemple, si l'on imagine un *deepfake* similaire à celui de Nancy Pelosi[136] en Suisse, qui discréditerait un candidat au Conseil des États en période d'élection, l'art. 275 CP devrait pouvoir s'appliquer. En revanche, si ce même *deepfake* est publié en dehors de toute campagne électorale, il s'agira tout au plus d'une atteinte à l'honneur dans la mesure où le but n'est plus de troubler une élection populaire.

De plus, si l'on considère que l'ordre constitutionnel selon cet article ne concerne que la liberté de vote, il faudrait également exclure le *deepfake* du président ukrainien encourageant sa population à rendre les armes. Cette situation

130 N. LANDSHUT, *in* M.-A. NIGGLI/H. WIPRÄCHTIGER (édit.), *Basler Kommentar Strafrecht II*, 4e éd., Bâle 2019, art. 275 N 4 ; B. CORBOZ, *Les infractions en droit suisse*, Volume II, 3e éd., Berne 2010, art. 275 N 1 ss.

131 K. LUBISHTANI/M. FLATTET (n. 127), p. 720.

132 K. LUBISHTANI/M. FLATTET (n. 127), p. 719. Cependant, B. CORBOZ (n. 130) considère que l'ordre constitutionnel inclut tous les principes fondamentaux.

133 K. LUBISHTANI/M. FLATTET (n. 127), p. 720.

134 www.bbc.com/news/av/technology-50381728 (consulté le 12.08.2022).

135 Dans le cadre d'une votation ou d'une élection en Suisse.

136 www.nytimes.com/2019/05/24/us/politics/pelosi-doctored-video.html (consulté le 12.08.2022)

devrait *a priori* constituer une propagande subversive (art. 275bis CP), mais cet article ne peut pas s'appliquer car le moyen utilisé doit être la violence. Or, un tel *deepfake* est à notre sens de nature à troubler l'intégrité d'un pays, si bien qu'il nous paraît peu souhaitable qu'une vidéo *deepfake* du président de la Suisse encourageant sa population à prendre ou déposer les armes ne soit pas punissable, à tout le moins sous l'angle de l'art. 275 CP[137]. Pour pallier cette lacune, il serait selon nous judicieux de supprimer la composante « violence » des infractions à caractère politique.

c) Autres infractions réalisées par l'utilisation de *deepfakes* (aperçu)

Nous avons abordé jusqu'à présent les deux types de *deepfakes* les plus courants, tout en nous concentrant sur les infractions *per se*, c'est-à-dire celles qui peuvent être réalisées du seul fait que ces *deepfakes* sont diffusés. Toutefois, ces technologies permettant la manipulation et la création de contenus synthétiques peuvent être utilisées[138] pour commettre un grand nombre d'infractions plus « classiques », dont nous ne donnerons qu'un aperçu à titre exemplatif.

Les *deepfakes*, notamment des *deepnudes* ou des *deepfakes porn*, peuvent être utilisées afin de faire chanter une victime (art. 156 CP), par exemple en demandant une certaine somme d'argent pour éviter une publication[139]. On pourrait également imaginer des cas où les *deepfakes* sont utilisés pour contraindre la victime à adopter un certain comportement, par exemple pour éviter que la compagne de l'auteur ne le quitte pour un autre (art. 181 CP), ou pour obtenir de la victime des avances sexuelles (art. 189 CP).

Des *deepfakes* peuvent également permettre d'escroquer des personnes (art. 146 CP), pour autant qu'ils soient suffisamment bien faits pour réaliser l'astuce au vu des circonstances, par exemple dans un scénario de *romance scam*, de *phishing* ou de fraude au président[140]. Enfin, cette technologie permet

137 Toutefois, il y a matière à interprétation dans la mesure où d'autres auteurs, comme B. CORBOZ (n. 130), ne limitent pas l'interprétation de « l'ordre constitutionnel » aux votations et aux élections.

138 C'est donc l'utilisation des *deepfakes* qui réalise ces infractions et non plus la simple existence en ligne de ceux-ci.

139 Il y a eu une vague massive de chantages de ce genre en Inde. Voir : www.thehindu.com/news/national/law-enforcers-worried-as-deep-nude-makes-a-return/article61657718.ece (consulté le 25.08.2022).

140 Cf. *supra* n. 68.

aussi de créer de faux documents d'identité ou de faux titres (art. 251 CP), comme une photo passeport créée synthétiquement[141].

3. *Le droit civil suisse permet-il à la victime d'agir ?*

Le droit civil offre un certain nombre de possibilités aux victimes de *deepfakes* sous l'angle de la protection de la personnalité (art. 28 ss CC[142]) et du droit d'auteur (art. 39d, 61 et 62 LDA). Dans les deux cas, la victime dispose d'actions défensives (art. 28a CC, art. 61 *cum* 62 LDA) et réparatrices (art. 28a al. 3 CC et art. 62 al. 2 LDA). Dans la mesure où ces actions ont été amplement traitées en doctrine, il n'est pas nécessaire de les développer ici.

Nous nous contentons simplement de relever que la victime peut agir contre toute personne qui participe à l'atteinte illicite à sa personnalité (art. 28 al. 1 CC), ce qui inclut tant l'auteur qui met en ligne le *deepfake* que la plateforme qui l'héberge[143]. En matière de droit d'auteur, le fournisseur d'hébergement – par exemple un site Internet ou un média social comme Facebook, Instagram, TikTok, etc. – a une obligation supplémentaire d'empêcher qu'un contenu illicite soit à nouveau rendu accessible par le biais de son service d'hébergement lorsqu'il en a déjà été averti (art. 39d LDA).

En résumé, il convient d'examiner *in casu* si le *deepfake* porte atteinte aux droits de la personnalité de la victime ou à un droit d'auteur. Si tel est le cas, il est possible d'exiger que l'auteur ou la plateforme sur laquelle le contenu se trouve le retire (art. 28a al. 1 let. b CC et art. 62 al. 1 let. b LDA). De plus, si la victime subit un dommage, elle peut également se prévaloir des actions réparatrices en dommages-intérêts (art. 41 ss CO[144]) et en réparation du tort moral (art. 49 CO), ainsi que réclamer la remise du gain selon les dispositions sur la gestion d'affaires (art. 423 CO).

141 Sur ce point, voir : D.-J. ROBERTSON/A. MUNGALL *et al.*, « Detecting Morphed Passport Photos : A Training and Individual Differences Approach », *Cogn. Research* 3:27, 2018, https://doi.org/10.1186/s41235-018-0113-8 (consulté le 25.08.2022) ; EUROPOL (n. 8), p. 12.

142 Code civil suisse du 10 décembre 1907 (CC ; RS 210).

143 Le Tribunal fédéral a en effet considéré que la Tribune de Genève, qui agissait en tant que fournisseur d'hébergement, avait la légitimation passive selon l'art. 28 CC pour une atteinte résultant de la publication d'un article rédigé par un internaute, car en permettant le stockage de cet article, elle a contribué à sa diffusion sur le web (ATF 5A_792/2011 du 14 janvier 2011, consid. 6.3). J. FRANCEY, *La responsabilité délictuelle des fournisseurs d'hébergement et d'accès Internet*, thèse (Université de Fribourg), Genève/Zurich/Bâle 2017, N 287.

144 Loi fédérale complétant le Code civil suisse du 30 mars 1911 (CO ; RS 220).

4. *Responsabiliser les plateformes ?*

La responsabilité des plateformes numériques pour les contenus générés par des tiers est un sujet qui est de plus en plus au cœur des préoccupations politiques et juridiques actuelles.

Cela peut notamment s'expliquer par le fait qu'il s'agit aujourd'hui des acteurs principaux dans le monde des médias et qu'ils peuvent grandement participer à la propagation d'infractions en ligne. De plus, il peut être intéressant pour les victimes d'infractions en ligne de rechercher la plateforme plutôt que l'auteur, notamment lorsque ce dernier est introuvable ou insolvable[145].

Cette partie vise à donner un aperçu des principales approches sur ce sujet, à savoir l'approche américaine (a), européenne (b) et suisse (c).

a) L'approche américaine

Le droit américain codifie la responsabilité civile des plateformes et plus généralement des fournisseurs de services Internet à la section 230 § 47 du *Communication Decency Act*[146] qui dispose :

> *« No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider ».*

Cela signifie en substance que comme les plateformes ne sont pas considérées comme l'auteur des publications faites par leurs utilisateurs ou par des tiers, elles ne peuvent pas en être tenues responsables[147]. Cet article octroie ainsi une quasi-immunité aux plateformes numériques pour héberger des contenus illicites, à l'exception de certains contenus violant notamment le droit pénal fédéral américain et le droit d'auteur[148].

145 P. GILLIÉRON, « La responsabilité des fournisseurs d'accès et d'hébergement », *RDS* 2002, p. 387 ss, p. 387.

146 Section 230 du Titre 47 du *Communication Decency Act* intitulée « *Protection for private blocking and screening of offensive material* », disponible sous : www.law.cornell.edu/uscode/text/47/230 (consulté le 18.08.2022).

147 B. CHESNEY/D. CITRON (n. 12), p. 1795.

148 K. DHRUVA (n. 5), p. 35.

b) L'approche européenne

L'Union européenne dispose pour sa part, depuis 2002, d'un cadre horizontal de responsabilité[149], c'est-à-dire applicable à toutes les catégories de contenus, de produits, de services et d'activités[150]. Actuellement prévu dans la Directive sur le commerce électronique[151], ce cadre figurera dès 2024 dans le Règlement sur les services numériques[152].

Que ce soit dans la directive ou dans le règlement, il est impératif aux yeux du législateur européen de ne pas imposer aux fournisseurs de services Internet une obligation générale de surveiller les contenus qu'ils hébergent ou transmettent[153]. Cela signifie qu'une plateforme n'a pas d'obligation proactive de monitorer les *deepfakes* criminels. En revanche, le droit européen prévoit un mécanisme de « *notice and take down* » qui peut entraîner la responsabilité de la plateforme en cas de non-respect.

De manière simplifiée, le mécanisme mis en place est le suivant : dès lors qu'il est signifié à la plateforme qu'elle héberge du contenu illicite (*notice*), elle se doit d'agir promptement pour le supprimer (*take down*)[154]. De plus, si une autorité l'enjoint d'agir contre certains contenus, elle doit coopérer sous peine d'entraîner sa responsabilité[155]. Le futur Règlement sur les services numériques contient d'autres obligations qui concernent spécialement les plateformes numériques, comme celle de devoir mettre en place un mécanisme de notification et d'action (art. 14 DSA) qui priorise les notifications émanant de signaleurs de confiance (art. 19 DSA). L'art. 21 DSA introduit une nouvelle obligation d'informer les services répressifs nationaux en cas de soupçon d'infractions pénales graves, c'est-à-dire impliquant une menace pour la vie ou la sécurité

149 Par rapport aux États-Unis qui ont adopté une approche verticale, c'est-à-dire qui n'est pas applicable à tous les domaines du droit.

150 PARLEMENT EUROPÉEN ET CONSEIL DE L'UNION EUROPÉENNE, *Proposition de règlement du 15 décembre 2020 relatif à un marché intérieur des services numériques (Législation sur les services numériques) du 15 décembre 2020* (COM/2020/825 final ; cité : DSA), p. 5.

151 PARLEMENT EUROPÉEN ET CONSEIL DE L'UNION EUROPÉENNE, *Directive 2000/31/CE du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« directive sur le commerce électronique »)* (JO L 178, 17.7.2000, p. 1 ; cité : DR-CE).

152 Le Parlement européen et le Conseil de l'Union européenne ont adopté le 6 juillet la Proposition de règlement du 15 décembre 2020 relatif à un marché intérieur des services numériques (Législation sur les services numériques) (COM/2020/825 final). Ce Règlement sera directement applicable à tous les membres de l'UE au plus tard le 1er janvier 2024, https://ec.europa.eu/commission/presscorner/detail/fr/IP_22_2545 (consulté le 18.08.2022).

153 Art. 14 DR-CE et art. 7 DSA.

154 Art. 14 DR-CE et art. 5 DSA.

155 Art. 14 par. 3 DR-CE et art. 8 DSA.

des personnes. En matière de *deepfakes*, il est peu probable que cet article s'appliquera, mais on pourrait toutefois imaginer qu'un *deepfake* similaire à celui du président ukrainien incitant sa population à se rendre puisse rentrer dans le champ d'application.

Ainsi, il convient de retenir une responsabilité de la plateforme lorsqu'on lui a signalé l'existence d'un *deepfake* illicite et qu'elle ne réagit pas pour le supprimer. Dans ce cas, elle est susceptible d'une amende qui pourra se monter jusqu'à 6 % de son chiffre d'affaires mondial (art. 42 par. 3 DSA).

c) L'approche helvétique

La Suisse n'a adopté aucune loi sur la responsabilité des plateformes numériques et compte sur les règles générales des domaines du droit concernés pour sanctionner les plateformes numériques (approche verticale). Contrairement aux États-Unis, la Suisse ne souhaite pas immuniser les plateformes contre toute responsabilité, mais estime que sa législation actuelle est suffisante[156].

Néanmoins, les auteurs qui se sont intéressés au sujet soulèvent l'incertitude qui règne lorsqu'il s'agit de transposer et d'appliquer des normes générales de responsabilité civile et pénale à des acteurs tels que les plateformes numériques[157]. Outre l'incertitude juridique insatisfaisante et les controverses doctrinales, il convient de retenir que les auteurs tendent à s'aligner sur la conception européenne. Ainsi, une plateforme pourra engager sa responsabilité si elle ne procède pas à la suppression des *deepfakes* illicites alors qu'elle en a été dûment informée[158]. Pour le surplus et parmi les rares arrêts sur la question, le Tribunal fédéral estime – à l'instar du droit européen – que l'on ne peut pas imposer aux plateformes une obligation générale de surveillance[159].

156 Cela ressort clairement des rapports du Conseil fédéral pour estimer le cadre juridique actuel et son application aux acteurs numériques. Voir par exemple : CONSEIL FÉDÉRAL, *Cadre juridique pour les médias sociaux. Rapport en réponse au postulat Amherd 11.3912 du 29 septembre 2011*, Berne 2013, p. 82 ; CONSEIL FÉDÉRAL, *La responsabilité civile des fournisseurs de services Internet du 11 décembre 2015*, Berne 2015 ; CONSEIL FÉDÉRAL, *Un cadre juridique pour les médias sociaux : Nouvel état des lieux, Rapport complémentaire sur le postulat Amherd 11.3912 « Cadre juridique pour les médias sociaux »*, Berne 2017, p. 52.

157 Cf. D. EQUEY, *La responsabilité pénale des fournisseurs de services Internet*, thèse (Université de Lausanne), Berne 2016, N 1235, et J. FRANCEY (n. 143), N 789 et N 798.

158 Sur le plan civil, voir par exemple : J. FRANCEY (n. 143), N 568 ss. Sur le plan pénal, voir : D. EQUEY (n. 157), N 1051 *cum* N 907.

159 ATF 6B_645/2007 du 02.05.2008, consid. 7.3.4.4.2.

C. Solutions communautaires

Les plateformes numériques comme Facebook, Twitter, Instagram et Tiktok constituent de nos jours les vecteurs principaux de communication en ligne, où leurs utilisateurs se chiffrent parfois en milliards. Il convient donc de prendre en considération l'importance que peuvent avoir leurs règles communautaires pour réguler les phénomènes criminels qui se produisent sur celles-ci[160].

L'importance des règles communautaires s'explique selon nous de deux manières : premièrement, les plateformes créent une relation verticale avec leurs utilisateurs qui est comparable à celle existant entre les citoyens et l'État[161]. En effet, elles fixent unilatéralement les conditions générales d'utilisation (y inclus les contenus autorisés), ce qui ne laisse guère le choix à leurs utilisateurs s'ils veulent bénéficier des services fournis[162]. Secondement, l'application du droit dans le cyberespace comporte un certain nombre d'obstacles. Par exemple, une plateforme n'est pas obligée de tenir compte des spécificités juridiques propres à chaque État. Ainsi, si un pays peut contraindre ponctuellement une plateforme à supprimer un *deepfake* illicite, il ne dispose toutefois pas du pouvoir d'imposer à la plateforme de supprimer l'ensemble des *deepfakes* qu'elle héberge.

A contrario, les plateformes imposent « leur droit » à leurs utilisateurs en définissant ce qui est permis et ce qui ne l'est pas, tout en étant en mesure de réguler, modérer et supprimer les contenus qui contreviennent à ses règles communautaires, et de décider des mesures à prendre contre l'auteur d'un acte non autorisé ainsi que de leur mise en œuvre[163].

Parmi les exemples de mesures prises à l'encontre des *deepfakes* par les plateformes, on peut tout d'abord mentionner la suppression du canal *r/deepfakes* par Reddit en 2018. L'année 2020 fut décisive sur ce point, car c'est à cette période que la majorité des plateformes se sont dotées de règles communautaires pour lutter contre les *deepfakes*, qu'il s'agisse de règles interdisant ou limitant spécialement les *deepfakes* ou plus généralement la manipulation de contenus audiovisuels. On notera tout de même que toutes les plateformes ne

160 EUROPOL (n. 8), p. 19.

161 F. GUILLAUME/S. RIVA, « L'atteinte à l'intégrité numérique appréhendée par le droit international privé », *in* F. GUILLAUME/P. MAHON (édit.), *Le droit à l'intégrité numérique, réelle innovation ou simple évolution du droit ?*, Bâle 2021, p. 117 ss, N 142.

162 F. GUILLAUME/S. RIVA (n. 161), N 141.

163 F. GUILLAUME/S. RIVA (n. 161), N 140 ss.

luttent pas de la même manière contre les *deepfakes*. Par exemple, pour préserver sa réputation Pornhub bannit les vidéos incluant le terme « *deepfakes* » mais pas « *deep fakes* » ni « *deepfake* » en 2018[164].

Plus récemment (en juin 2022), Google a pris une décision qui risque de fortement limiter l'effervescence des *deepfakes*. En effet, la société américaine a décidé de bannir de Google Colab (ou Colaboratory)[165] tous les projets visant à créer des *deepfakes*, nonobstant que certains projets n'avaient pas un but illicite[166]. Ce choix risque d'avoir un fort impact dans la mesure où la création de *deepfakes* nécessite une puissance de calcul importante qui n'est pas facilement à la portée de tout un chacun. Google Colab présentait l'immense avantage d'être gratuit et de pouvoir exécuter un code d'un logiciel de *deepfakes* directement depuis le *cloud* en utilisant les ressources physiques de calcul (cartes graphiques et processeurs) de Google plutôt que celles de son propre ordinateur[167].

Même s'il existe toujours d'autres alternatives à Google Colab pour créer des *deepfakes* potentiellement illicites, comme Azur Notebooks, IBM DataPlatform Notebooks, Amazon Sagemaker et Kaggle, ces dernières sont toutefois payantes et avec une marge de manœuvre plus restreinte[168]. Il s'agit ici de la démonstration de l'impact que peut avoir une décision émanant d'une entité privée comparée à une mesure étatique.

IV. Conclusion

On relèvera finalement que les *deepfakes* soulèvent d'autres problématiques que purement juridiques. Tout d'abord, les *deepfakes* peuvent causer d'importants dégâts réputationnels, financiers et sur le plan psychologique[169]. On peut notamment penser aux victimes de *deepfakes porn* qui se voient publiées et critiquées sur toutes les plateformes en ligne. C'est pourquoi il est important de pouvoir agir contre ces contenus, pour éviter une plus grosse diffusion, que ce soit en agissant contre l'auteur ou contre la plateforme. Néanmoins, lorsqu'un *deepfake* est mis en ligne, il est pratiquement impossible de

164 S. MADDOCKS (n. 1), p. 419.
165 Google Colab est un service *cloud* gratuit permettant d'entraîner des modèles de *machine learning* directement dans le *cloud*.
166 https://techcrunch.com/2022/06/01/2328459/ (consulté le 19.08.2022).
167 Voir par exemple : https://ledatascientist.com/google-colab-le-guide-ultime/ (consulté le 19.08.2022).
168 D. MISAL, « 5 Alternatives To Google Colab For Data Scientists », 5.07.2019, https://analyticsindiamag.com/5-alternatives-to-google-colab-for-data-scientists/, (consulté le 30.08.2022).
169 K. DHRUVA (n. 5), p. 17 ; B. CHESNEY/D. CITRON (n. 12), p. 1773.

supprimer définitivement l'atteinte qui en résulte. Qu'il soit accessible via le *darknet* ou stocké sur le périphérique privé d'un tiers (ordinateur, disque dur, smartphone, etc.), il est concrètement difficile de pouvoir garantir un *stay down* du contenu illicite. Pour cette raison, il ne faudrait pas perdre de vue les moyens non juridiques, comme des suivis psychologiques, qui peuvent parfois s'avérer plus pertinents[170].

La prise de conscience de cette technologie devrait également susciter une plus grande prudence dans la manière d'aborder les preuves dans un procès. Par exemple, une vidéo montrant clairement l'auteur braquer une banque aura implicitement une forte valeur probante. Or, il convient de garder à l'esprit que tout autant réaliste qu'apparaisse une vidéo ou une image, il peut s'agir d'un *deepfake*.

Pour conclure, nous avons vu que cette technologie présente un certain nombre d'usages bénéfiques pour la société, ce pourquoi il ne serait pas souhaitable de la bannir totalement, mais plutôt de limiter son utilisation à des fins criminelles.

Même s'il n'existe actuellement pas de solution miracle sur le plan technique, il n'est pas pour autant inutile de continuer à chercher des contre-mesures. En effet, une des problématiques du *deepfake* est qu'il est possible d'en créer en utilisant des ressources librement accessibles et faciles d'utilisation[171]. La continuation des recherches en matière de détection permettra déjà de trier les *deepfakes* les moins sophistiqués. En outre, des mesures telles que la limitation d'accès aux *deepfakes as a service* et aux ressources permettant d'en créer (comme Google Colab) évitent que tout un chacun puisse trop facilement utiliser cette technologie à des fins criminelles.

Sur le plan juridique, nous avons vu que la Suisse est dotée de règles suffisantes pour réprimer les comportements criminels relatifs à l'utilisation des *deepfakes*, principalement par le biais des infractions contre l'honneur. L'usurpation d'identité est un ajout bienvenu qui viendra compléter l'arsenal du droit pénal en 2023. Le droit civil semble également offrir aux victimes une protection satisfaisante par le biais des actions défensives et réparatrices de la protection de la personnalité et du droit d'auteur.

Même si la question de la responsabilité (pénale et civile) des plateformes reste encore incertaine en droit suisse, la démarche spontanée des plateformes de supprimer ou limiter certains contenus joue un rôle déterminant pour endiguer la criminalité en ligne. Cela montre l'intérêt de continuer à promouvoir des partenariats publics-privés en matière de cybercriminalité[172].

170 EUROPOL (n. 8), p. 14.
171 K. KOBRIGER *et al.* (n. 2), p. 206 ; B. CHESNEY/D. CITRON, (n. 12), p. 1753 ; R. SPIVAK (n. 3), p. 345 ; M. BODI (n. 10), p. 146.
172 Sur les bienfaits des partenariats publics-privés : K. DHRUVA (n. 5), p. 63 ss.

La répression des crimes internationaux commis dans le cyberespace par la Cour pénale internationale (CPI)

ELENA VOLKOVA
Doctorante à l'Université Paris II Panthéon-Assas | École doctorale de droit privé

Table des matières

I. Introduction

« Si un jour il y avait une confrontation majeure, elle commencerait par une cyberattaque massive, non seulement sur les installations militaires, mais aussi sur certaines infrastructures civiles »[1]. Cette pensée d'António Guterres, Secrétaire général des Nations Unies, prouve que les conflits armés modernes ne se limitent plus au sol, à la mer et à l'air, car l'espace numérique est devenu un nouveau théâtre d'affrontement. Des entreprises de cybersécurité et le Comité international de la Croix-Rouge, qui suivent les dernières évolutions technologiques sur les champs de bataille, parviennent aux mêmes conclusions

[1] N. THOMPSON, « UN Secretary-General : US-China Tech Divide Could Cause More Havoc Than the Cold War », *The Wired*, 15 janvier 2020, www.wired.com/story/un-secretary-general-antonio-guterres-internet-risks/ (consulté le 19.07.2022).

en s'appuyant sur une expérience réelle[2]. Dans le même temps, la liste des infrastructures à la fois civiles et militaires qui sont devenues victimes des cyberattaques s'allonge : la centrale nucléaire iranienne en 2010[3] et en 2021[4], le National Health Service au Royaume-Uni en 2017[5], les systèmes de lancement de missiles iraniens en 2019[6], l'usine d'approvisionnement en eau aux États-Unis[7]. La réalité d'aujourd'hui montre en effet clairement que ces actes malveillants commis dans le cyberespace pourraient prendre la forme de crimes internationaux et dépassent le cadre de la théorie : crimes de guerre, crimes de génocide, crimes contre l'humanité, crimes d'agression. Cela signifie que l'influence grandissante du cyberespace et d'Internet a conduit à la possibilité de commettre des crimes dans l'espace numérique qui pourraient tomber sous le coup de la juridiction de la Cour pénale internationale (ci-après : CPI). Cependant, parallèlement à la prolifération considérable des cyberattaques et à l'essor de la dépendance aux technologies de l'information et de la communication (ci-après : TIC), la question de la réglementation du cyberespace d'un point de vue militaire et juridique, ainsi que l'adaptation des règles et mécanismes juridiques à cette nouvelle réalité restent au fil du temps problématiques.

II. Cybercrimes : le droit international à la poursuite des nouvelles technologies

L'un des plus grands défis du droit pénal international est de suivre le rythme rapide des changements technologiques. Les modes de commission des

2 D. BURKHALTER, « Une cyberattaque peut-elle être un crime de guerre ? », *Swissinfo*, 06.05.2022, www.swissinfo.ch/fre/une-cyberattaque-peut-elle-%C3%AAtre-un-crime-de-guerre-/47569096 (consulté le 08.08.2022).

3 B. FERRAN, « Stuxnet : l'Iran se dit victime de « guerre électronique », *Le Figaro*, 29.09.2010, www.lefigaro.fr/sciences-technologies/2010/09/27/01030-20100927ART FIG00417-stuxnet-l-iran-se-dit-victime-de-guerre-electronique.php (consulté le 19.07.2022).

4 D. FILIPPONE, « En Iran, une cyberattaque vise la centrale de Natanz », *Le Monde informatique*, 12.04.2021, www.lemondeinformatique.fr/actualites/lire-en-iran-une-cyberattaque-vise-la-centrale-de-natanz-82569.html (consulté le 19.07.2022).

5 PIXELS, « 200'000 victimes, 150 pays : le premier bilan de la cyberattaque mondiale », *Le Monde*, 14.05.2017, www.lemonde.fr/pixels/article/2017/05/14/cyberattaque-200-000-victimes-essentiellement-des-entreprises-dans-150-pays-assure-europol_5127506_4408996.html (consulté le 19.07.2022).

6 AFP, « Les États-Unis ont lancé des cyberattaques contre l'Iran, selon des médias », *Le Point*, 23.06.2019, www.lepoint.fr/monde/les-États-unis-ont-lance-des-cyberattaques-contre-l-iran-selon-des-medias-23-06-2019-2320465_24.php (consulté le 19.07.2022).

7 AFP, « Un réseau d'eau potable victime d'un piratage informatique en Floride », *Le Monde*, 09.02.2021, www.lemonde.fr/pixels/article/2021/02/09/un-reseau-d-eau-potable-victime-d-un-piratage-informatique-en-floride_6069274_4408996.html (consulté le 19.07.2022).

crimes et les instruments utilisés progressent et ont de plus en plus de liens avec un environnement virtuel. Les crimes internationaux ne font pas exception et peuvent également être commis dans le cyberespace (B) ce qui, inévitablement, pose la question de l'application des normes juridiques déjà existantes et, dans le même temps, du respect du principe de légalité en droit pénal (A).

A. L'état actuel du droit

Au fil des ans, de nombreux États et ONGs ont déployé des efforts pour élargir, concrétiser et mettre à jour les doctrines juridiques traditionnelles afin de relever les défis posés par le développement de nouvelles armes. Et à chaque fois, la communauté internationale s'est heurtée à l'impossibilité de prévoir de nouvelles règles à chaque développement technologique. Le potentiel militaire dont disposent aujourd'hui les États dans le cyberespace n'était pas envisagé au XX^e^ siècle, lorsque la communauté internationale a élaboré les principaux documents régissant le droit de la guerre. Cependant, le vide juridique apparent n'a pas permis de laisser les coudées franches aux États et acteurs non étatiques. Au lieu de recourir au *Lotus principle* en vertu duquel tout ce qui n'est pas expressément interdit est par nature autorisé, la communauté internationale a choisi la voie responsable et a déclaré que ce principe ne peut pas être appliqué aux lois de la guerre. Au contraire, dans toutes les situations incertaines qui concernent l'utilisation des armes nouvelles, les États doivent respecter la clause de Martens selon laquelle tout ce qui n'est pas expressément interdit par un traité n'est pas pour autant autorisé[8].

1. Les premiers pas vers la répression des crimes dans le cyberespace

Étant préoccupée par la militarisation croissante du cyberespace et également consciente de la nécessité pour les États d'adopter un comportement responsable en élaborant des règles propres à l'espace numérique, l'Organisation des Nations Unies (ci-après : ONU) a créé en 2010 le Groupe d'experts gouvernementaux (ci-après : GEG) chargé d'examiner les progrès de la téléinformatique dans le contexte de la sécurité internationale. Après trois ans de travail, le GEG a conclu que le droit international ainsi que la Charte des

8 Voir Y. SANDOZ *et al.*, *Commentaire des Protocoles additionnels du 8 juin 1977 aux Conventions de Genève du 12 août 1949*, CICR, Genève, 1986, par. 55 ; CIJ, *Avis consultatif, Licéité de la menace ou de l'emploi d'armes nucléaires*, 8.07.1996, par. 87, www.icj-cij.org/public/files/case-related/95/095-19960708-ADV-01-00-FR.pdf (consulté le 05.08.2022).

Nations Unies « sont applicables et essentiels au maintien de la paix et de la stabilité »[9]. La publication de ce rapport a constitué un point de départ essentiel pour les États et les Organisations non gouvernementales (ci-après : ONG). Suite à la reconnaissance de l'applicabilité du droit international dans ce domaine, le Comité international de la Croix-Rouge (ci-après : CICR) a déclaré que le droit international humanitaire devait être respecté par les parties au conflit dans le cyberespace[10] et a mis à jour les Commentaires aux Conventions de Genève qui couvrent les cyberopérations ayant des « effets similaires aux opérations cinétiques classiques » pendant les conflits armés internationaux[11]. Parallèlement à ces efforts, la communauté internationale a abordé les questions de l'applicabilité du droit international au cyberespace sur le plan pratique. Deux groupes d'experts internationaux en droit et en TIC ont élaboré le Manuel de Tallinn contenant l'ensemble des règles classiques transposées dans le domaine du cyber. L'apport important de ce document est qu'il examine les situations les plus graves (recours à la force, agression armée, actes ayant lieu à la fois dans le cadre d'un conflit armé et en dehors de celui-ci) en tenant compte de la spécificité de l'espace numérique. Néanmoins, bien que le Manuel de Tallinn ait permis de combler des lacunes juridiques, il n'a aucune valeur contraignante et sert plutôt de référence pour les États.

2. *L'application du Statut de Rome aux cybercrimes*

Les développements importants réalisés dans ce domaine au cours des dernières années ont permis de se concentrer sur un sujet plus précis : l'application du Statut de Rome aux crimes internationaux commis dans le cyberespace. Seulement l'année dernière, le *Council of Advisors* regroupant onze États[12] a présenté le rapport final qui analyse, article par article, comment le Statut de Rome peut être appliqué à la cyberguerre. Le principal résultat de ce

9 Groupe d'experts gouvernementaux chargé d'examiner les progrès de la téléinformatique dans le contexte de la sécurité internationale, *Rapport*, Résolution A/68/98, 24 juin 2013, par. 19, https://documents-dds-ny.un.org/doc/UNDOC/GEN/N13/371/67/PDF/N1337167.pdf?OpenElement (consulté le 21.07.2022).

10 CICR, *Le droit international humanitaire et les défis posés par les conflits armés contemporains, Rapport*, XXXII[e] Conférence internationale de la Croix-Rouge et du Croissant-Rouge, 8-10 décembre 2015, Genève, p. 48, www.icrc.org/fr/download/file/15110/32ic-report-on-ihl-and-challenges-of-armed-conflicts-fre.pdf (consulté le 19.07.2022).

11 CICR, *Les Commentaires sur la Convention de Genève pour l'amélioration du sort des blessés et des malades dans les forces armées en campagne de 1949, art. 2, 2020, par. 255*, https://ihl-databases.icrc.org/applic/ihl/dih.nsf/Comment.xsp?action=openDocument&documentId=C7D5DC96DB5B3109C1257F7D0060524A (consulté le 21.07.2022).

12 Argentine, Autriche, Belgique, Costa Rica, République tchèque, Estonie, Liechtenstein, Luxembourg, Portugal, Espagne et Suisse, ainsi que le Bureau du Procureur de la CPI.

travail est la prise de conscience qu'aujourd'hui une cyberattaque suffisamment grave a le potentiel de s'inscrire dans chacun des crimes – crime d'agression, crime de guerre, crime contre l'humanité et génocide – et que le Statut de Rome doit être appliqué dans le cyberespace[13].

De la même manière que la compétence de la Cour pénale internationale s'étend « aux crimes les plus graves qui touchent l'ensemble de la communauté internationale »[14], nous nous concentrerons dans notre étude seulement sur les cyberattaques et crimes commis dans le cyberespace qui sont conformes aux exigences de la nature, la gravité et l'ampleur requis au sens du Statut de Rome et qui s'inscrivent dans la liste des crimes dits du « noyau dur ». Cette précision est importante à faire car il existe de nombreux auteurs qui ajoutent à la liste des crimes internationaux des actes cybercriminels commis pour des raisons purement politiques ou économiques[15]. Par contre, nous limiterons notre analyse aux crimes internationaux *stricto sensu* et qui suivent un régime spécial distinct du droit pénal général[16].

3. *Quand les cybercrimes peuvent-ils faire l'objet d'un procès devant la CPI ?*

L'incertitude et le manque de clarté qui entourent la définition du seuil de nuisance requis au sens du Statut de Rome[17] pour les actes commis dans le cyberespace représentent aujourd'hui le problème le plus débattu, parce qu'à

13 COUNCIL OF ADVISORS, *Report On The Application Of The Rome Statute Of The International Criminal Court To Cyberwarfare*, 2021, www.regierung.li/files/medienarchiv/The-Council-of-Advisers-Report-on-the-Application-of-the-Rome-Statute-of-the-International-Criminal-Court-to-Cyberwarfare.pdf (consulté le 23.07.2022).

14 Art. 5 du Statut de Rome.

15 Par exemple, David Scheffer, avocat et diplomate américain qui a été le premier ambassadeur itinérant des États-Unis pour les questions de crimes de guerre, considère que les cybermesures visant à saper gravement les processus démocratiques d'un État et influencer de manière significative le résultat des élections doivent relever de la compétence de la CPI en tant qu'acte d'agression : D. SCHEFFER, « The Missing Pieces in Article 8 bis (Aggression) of the Rome Statute », *Harvard International Law Journal*, vol. 58, 2017, p. 84, https://harvardilj.org/wp-content/uploads/sites/15/Scheffer-Formatted.pdf (consulté le 05.08.2022). Sur les cybercrimes économiques, voir S. RUHLAND, « Economic Cyber Crimes and the Rome Statute », *ICC Forum*, 4 mars 2022, https://iccforum.com/forum/permalink/131/38980 (consulté le 06.08.2022). Le *Council of Advisors*, qui a examiné l'application du Statut de Rome au cyberespace, à l'inverse, considère que les actes dans le cyberespace qui provoquent des impacts purement économiques, financiers ou politiques (p.ex. ingérence électorale) ne sont pas couverts par le Statut : COUNCIL OF ADVISORS (n. 13), par. 16, p. 14.

16 Sur ce sujet voir D. REBUT, *Droit pénal international*, Dalloz, 3[e] éd., 2019, p. 575-581.

17 Art. 17-1-d du Statut de Rome.

ce jour il n'y a toujours pas de consensus parmi les États et les ONGs à ce sujet. La grande diversité des opinions des experts peut être classée en trois catégories : approche axée sur les moyens, approche axée sur les effets, approche mixte.

D'après la première approche (*means approach*) pour que les crimes internationaux relèvent de la compétence de la CPI, ils doivent être commis par des armes conventionnelles ou non conventionnelles : armes chimiques, armes biologiques, armes à laser aveuglantes, mines terrestres, etc. L'élargissement de la compétence de la CPI pour y inclure les crimes commis dans le cyberespace violerait le principe de légalité en droit pénal (*nullum crimen sine lege*)[18]. Cette position est fondée sur l'idée qu'aujourd'hui il n'existe aucune convention qui mentionne le cyberespace et que ni le Statut de Rome, ni les Conventions de Genève, ni les Conventions de La Haye ne contiennent aucun amendement sur ce sujet. Il en va de même pour les règles du droit humanitaire coutumier : bien qu'elles représentent une base de réglementation pour la conduite des hostilités, ces règles ne prennent pas en compte l'aspect cyber des crimes en question qui rend ces infractions si spécifiques[19]. Les outils informatiques qui peuvent être employés lors de la commission de crimes internationaux – rançongiciels, cheval de Troie, vers – n'existent pas en dehors du cyberespace ; ils ne peuvent être ni vus ni touchés et diffèrent sensiblement des armes traditionnelles. Ces éléments justifient une régulation totalement différente.

Un autre raisonnement théorique convaincant pour expliquer cette approche figure à l'art. 22 du Statut de Rome qui souligne que « [l]a définition d'un crime est d'interprétation stricte et ne peut être étendue par analogie ». Ceux qui soutiennent ce point de vue considèrent, par conséquent, qu'il faut soit compléter les Conventions de Genève et le Statut de Rome[20], soit adopter une nouvelle convention pour la répression des crimes internationaux commis dans le

18 Le Professeur de droit international pénal, Kenneth Gallant, en particulier, pose comme postulat que les juges qui interprètent les infractions de manière si large créent de nouveaux crimes et appliquent, de ce point de vue, de nouvelles lois de manière rétroactive : K. GALLANT, *The Principle Of Legality In International And Comparative Criminal Law*, juin 2011, p. 52-62, www.researchgate.net/publication/228210851_The_Principle_of_Legality_in_International_and_Comparative_Criminal_Law (consulté le 10.08.2022).

19 D. HOLLIS, « Why states need an international law for information operations », *Lewis and Clark Law Review*, vol. 11, 2007, p. 1041 s., https://law.lclark.edu/live/files/9551-lcb114art7hollis.pdf (consulté le 06.08.2022).

20 Voir p.ex. D. SCHEFFER, Amending the Crime of Aggression under the Rome Statute, *in* C. KRESS/S. BARRIGA (édit.), *The Crime of Aggression : A Commentary*, Cambridge University Press, 2016, p. 1480 s. ; SCHEFFER (n. 15).

cyberespace[21] à l'instar de la Convention de Budapest qui traite de la cybercriminalité.

Néanmoins, la majorité des ONGs et des États se sont largement prononcés en faveur d'une approche axée sur les effets (*effects approach*). L'idée principale est que si les actes commis dans le cyberespace provoquent des effets comparables à ceux engendrés par les attaques cinétiques, c'est-à-dire des dommages matériels, des blessures ou des morts, ils tombent sous le coup du Statut de Rome.

> *« [K]inetic and cyber-attacks may be comparable in their effects or consequences. The conventional physical bombardment of a military base or causing a complete loss of function through a cyber-attack may have the same effect, that is, the temporary suspension of the use of the military base. »*[22]

Cette position est appuyée par l'ONU et, notamment, par le Groupe des experts gouvernementaux[23], par les experts qui ont élaboré le Manuel de Tallinn sur l'application du droit international aux cyberattaques, par le Comité International de la Croix Rouge, par des États européens[24], et même par le *Council of Advisors* qui a confirmé l'application du Statut de Rome aux cyberarmes. Un argument principal en faveur de cette position est l'Avis consultatif rendu par la Cour internationale de justice (ci-après : CIJ) sur la licéité de l'utilisation des armes nucléaires, dont les dispositions s'appliquent à n'importe quel emploi de la force, indépendamment des armes employées[25]. Un avantage significatif de cette approche est qu'elle permet de couvrir les armes qui n'étaient pas utilisées à l'époque où les documents juridiques ont été élaborés.

21 Voir p.ex. D. BROWN, « A Proposal for an International Convention to Regulate the Use of Information Systems in Armed Conflict », *Harvard International Law Journal*, vol. 47, n° 1, 2006, p. 212 ; HOLLIS (n. 19), p. 1023.

22 K. AMBOS, « Cyber-Attacks as International Crimes under the Rome Statute of the International Criminal Court ? », *ICC Forum*, 4 mars 2022, https://iccforum.com/forum/permalink/131/38980 (consulté le 06.08.2022).

23 ONU, Assemblée générale, *Rapport du Groupe d'experts gouvernementaux chargé d'examiner les progrès de l'informatique et des télécommunications dans le contexte de la sécurité internationale, par. 24, 22 juillet 2015*, www.un.org/ga/search/view_doc.asp?symbol=A/70/174&Lang=F (consulté le 10.08.2022).

24 Voir, p.ex., la position de l'Allemagne, FEDERAL GOVERNMENT OF GERMANY, *On the Application of International Law in Cyberspace, Position Paper*, mars 2021, p. 4, www.auswaertiges-amt.de/blob/2446304/32e7b2498e10b74fb17204c54665bdf0/on-the-application-of-international-law-in-cyberspace-data.pdf (consulté le 06.08.2022). Sur la position de la France, voir MINISTÈRE DES ARMÉES, *Droit international appliqué aux opérations dans le cyberespace*, p. 7, 2019, www.justsecurity.org/wp-content/uploads/2019/09/droit-internat-appliqu%C3%A9-aux-op%C3%A9rations-cyberespace-france.pdf (consulté le 06.08.2022).

25 CIJ (n. 8), par. 39.

Cependant, même parmi les spécialistes qui soutiennent l'approche basée sur les effets, il y a désaccord : que faut-il entendre par des « effets comparables à ceux engendrés par des attaques cinétiques » ? Est-ce qu'il y a une définition du seuil de gravité applicable au cyberespace ?

La grande majorité des universitaires, spécialistes des TIC, des juristes et des États s'accordent sur le fait qu'un crime international commis dans l'espace numérique doit provoquer des dommages matériels aux biens, des blessures et des morts au sens du droit international humanitaire[26]. Autrement dit, les conséquences de l'infraction doivent aller au-delà de l'espace virtuel et se manifester dans le monde physique. Si une cyberattaque prennant pour cible les données médicales stockées sur les ordinateurs des hôpitaux provoque la mort de civils, cela pourrait constituer un crime de guerre ou un crime contre l'humanité en fonction du contexte. Dans ce sens, le cyberespionnage ne constituerait pas un crime international au sens du Statut de Rome.

Les experts du Manuel de Tallinn ajoutent à la liste des conséquences qui tombent sous le coup du droit international humanitaire la perte de fonctionnalité d'un système informatique qui nécessite un remplacement de l'un de ses éléments et affirment qu'un tel acte doit être couvert par les dispositions des Conventions de Genève et, par conséquent, par le Statut de Rome[27]. Il est indispensable de noter que la création d'une catégorie distincte pour ce type de conséquences est dictée par le fait qu'il est possible de mettre un système hors service sans l'endommager physiquement. Toutefois, nous pouvons reprocher à cette analyse d'étendre excessivement la catégorie des conséquences qui doivent être couvertes par les dispositions du droit international humanitaire. D'ailleurs, la proposition de considérer un acte qui vise à mettre hors service un système informatique comme un crime international suscite inexorablement la question suivante : « La perte de fonctionnalité concerne-t-elle tous les systèmes informatiques sans exception ou seulement ceux liés aux infrastructures critiques ? ». D'un autre côté, si la perte de fonctionnalité ne nécessite pas de remplacement d'éléments physiques du système, mais pourrait prendre un temps considérable pour être réparée, le non-fonctionnement du système pourrait entraîner beaucoup plus d'effets nuisibles pour la population civile. Dans ce cas, est-ce que la perte de fonctionnalité pourrait être qualifiée de crime international ? Par exemple, lorsqu'un rançongiciel attaque un système d'approvisionnement et de distribution d'eau en cryptant des données, avec pour conséquence que la population civile ne peut pas avoir l'accès à l'eau potable, aucun remplacement d'éléments physiques n'est requis sur le plan technique. Donc, d'après le Manuel de Tallinn, cet acte ne représente pas un crime international.

26 Voir p.ex. AMBOS (n. 22).

27 M. SCHMITT, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Cambridge University Press, 2017, par. 10, p. 417.

Cependant, le décryptage des données du système pour assurer son bon fonctionnement peut prendre beaucoup de temps, alors que la population civile reste privée d'eau dans l'intervalle. Dans ce contexte-là, cet acte pourrait-il constituer un crime international ? Le CICR a essayé de mettre en lumière cette situation et a précisé qu'une telle perte de fonctionnalité ne pourrait constituer une violation du droit international que si elle vise à tuer ou blesser des civils ou endommager des biens[28].

Finalement, on peut mentionner une troisième approche (*mixed approach*) qui permettrait de couvrir les cyberattaques ou crimes commis dans le cyberespace même s'ils ne produisent pas des effets graves. D'après cette approche, pour atteindre le seuil nécessaire, les cyberactes peuvent simplement faire partie d'une attaque cinétique ou d'une attaque plus large impliquant clairement l'utilisation de la force physique[29]. On a déjà été témoins des cyberattaques qui accompagnaient des attaques classiques pendant le conflit entre la Russie et la Géorgie. Ces cyberattaques n'ont pas été assez graves, car elles visaient seulement le système bancaire et les opérateurs mobiles, cependant elles ont été lancées dans le cadre d'un conflit armé. Même si la CPI ne les a pas examinées et ne les a pas prises en considération à l'époque, sa position sur cette question n'est pas claire.

Ainsi, la seule position qui est partagée par la majorité des experts aujourd'hui est celle de la nécessité de réprimer les actes dans le cyberespace, lesquels constituent, en raison de leur nature, leur gravité et leur ampleur, une violation du Statut de Rome, et font ainsi partie des actes les plus graves qui heurtent profondément la conscience humaine[30]. Nous allons prendre des exemples concrets pour illustrer ce point.

28 Par exemple, un simple brouillage radio ne tomberait pas dans le champ d'application du droit international humanitaire, tandis que la cyberattaque qui coupe le courant des unités de soins intensifs et provoque le décès de patients constituerait un crime ; voir CICR, *Le droit international humanitaire et les défis posés par les conflits armés contemporains, Rapport*, XXXII[e] Conférence internationale de la Croix-Rouge et du Croissant-Rouge, 8-10 décembre 2015, Genève, p. 50, www.icrc.org/fr/document/le-droit-international-humanitaire-et-les-defis-poses-par-les-conflits-armes-contemporains (consulté le 06.08.2022).

29 Voir, p.ex., M. ROSCINI, « Cyber Operations Can Constitute War Crimes Under the ICC Jurisdiction Without Need to Amend the Rome Statute », *ICC Forum*, Cyber Operations and Cyberwarfare Question, 2022, https://iccforum.com/cyberwar (consulté le 06.08.2022).

30 Préambule du Statut de Rome.

B. Des risques hypothétiques aux menaces réelles

La réalité d'aujourd'hui montre clairement que les actes malveillants commis dans le cyberespace peuvent dépasser le cadre de la théorie et prendre la forme de crimes internationaux : crimes de guerre, crimes d'agression, crimes contre l'humanité et crimes de génocide. Pour mieux le comprendre, examinons successivement chaque infraction.

1. Crime de guerre

L'art. 8 du Statut de Rome définit les crimes de guerre comme les infractions graves aux Conventions de Genève du 12 août 1949 ainsi que les autres violations graves des lois et coutumes applicables aux conflits armés dans le cadre établi par le droit international[31]. Si nous examinons les éléments constitutifs d'un crime de guerre[32] et les appliquons au cyberespace, nous pouvons identifier les éléments suivants : un cyberacte doit avoir un lien avec un conflit armé (international ou non international), son auteur doit avoir connaissance de l'existence de ce conflit, et une cyberopération doit relever de l'un des actes énumérés à l'art. 8 du Statut de Rome. Le *Council of Advisors* considère que si ces critères sont réunis, même un acte unique commis dans le cyberespace peut constituer un crime de guerre[33]. En ce qui concerne le territoire du crime, il est important de noter que la CPI peut réprimer des actes criminels même s'ils se produisent au-delà du territoire des parties en conflit[34]. Cette précision est particulièrement importante pour les crimes internationaux commis dans le cyberespace, parce que ces infractions sont commises à distance dans la plupart des cas.

31 Art. 8 du Statut de Rome.

32 CPI, Éléments des crimes, art. 8, www.icc-cpi.int/sites/default/files/ElementsOfCrimesFra.pdf (consulté le 25.07.2022).

33 COUNCIL OF ADVISORS (n. 13), p. 26.

34 Voir, p.ex., CPI, Jugement en appel contre la décision d'ouvrir une enquête sur l'Afghanistan, Situation en République islamique d'Afghanistan, Chambre d'appel, le 5 mars 2020, par. 74, 76 : « [T]he text of Common Article 3 read in its totality does not suggest that the requisite nexus with the armed conflict in Afghanistan cannot exist if the criminal conduct occurred outside Afghanistan and the victim was not captured in Afghanistan. Importantly, such a conclusion would also be contrary to the purpose of Common Article 3, which is to provide minimum guarantees in relation to armed conflicts. [...] Thus, in the view of the Appeals Chamber, it is incorrect to assume that, merely because the alleged capture of the victim did not take place in Afghanistan and the alleged criminal act also occurred outside Afghanistan, the conduct cannot possibly have taken place in the context of, and have been associated with, the armed conflict in that State », www.icc-cpi.int/sites/default/files/CourtRecords/CR2020_00828.PDF (consulté le 25.07.2022).

Le Statut de Rome comprend plusieurs dizaines d'actes qui peuvent être considérés comme crimes de guerre. Pour mieux comprendre si des actes cybernétiques peuvent violer les lois de la guerre, prenons l'exemple d'une attaque contre des hôpitaux[35]. En vertu du droit international humanitaire coutumier, les hôpitaux et le matériel sanitaire doivent être respectés et protégés en toutes circonstances[36] ; ce principe a été consacré aux art. 8-2-b-xxiv et 8-2-e-ii du Statut de Rome, ainsi que dans les décisions de la CPI[37]. Depuis le début de la pandémie du Covid-19, le monde a été le témoin de plusieurs cyberattaques contre des hôpitaux. Par exemple, le nombre de cyberattaques à l'encontre du secteur de la santé en Tchéquie a connu une hausse de 267 % en 2020 selon le rapport du Ministère public suprême[38]. L'une des nombreuses attaques de ce type a été lancée contre l'hôpital de Brno[39]. Suite à cette cyberopération, l'établissement de santé a dû reporter des interventions chirurgicales et rediriger de nouveaux patients vers un autre hôpital. L'hôpital a été forcé de fermer tout son réseau informatique pendant l'incident, ainsi que de suspendre les travaux de deux autres établissements qui dépendaient de lui (l'hôpital pour enfants et la maternité). Cette situation nous force à penser aux scénarios potentiellement catastrophiques : des morts parmi les civils. L'expérience allemande montre que la réalité a rejoint la fiction : en septembre 2020, une femme est décédée à cause d'un *ransomware* dans un hôpital universitaire de Düsseldorf[40].

Les exemples des attaques susmentionnées contre des établissements de santé ne constituent pas des crimes de guerre dans le cyberespace, dès lors qu'ils n'ont pas été commis dans le contexte d'un conflit armé. Néanmoins, ils montrent qu'il est déjà techniquement possible de lancer une cyberopération contre

35 Voir E. VOLKOVA, « Protection des infrastructures de santé contre les cyberattaques dans les conflits armés », *Calameo*, 2020, https://fr.calameo.com/books/006401546497064e46e94 (consulté le 25.07.2022).

36 J.-M. HENCKAERTS/L. DOSWALD-BECK, *Droit international humanitaire coutumier*, vol. I : règles, CICR, Cambridge University Press, 2006, Règle 28, p. 124, www.icrc.org/fr/doc/assets/files/other/icrc_001_pcustom.pdf (consulté le 24.07.2022).

37 Voir, p.ex., CPI, *Procureur c. Bosco Ntaganda*, Chambre de Première Instance VI, Jugement, par. 1146 s., www.icc-cpi.int/sites/default/files/CourtRecords/CR2020_06486.PDF (consulté le 25.07.2022).

38 NATIONAL CYBER AND INFORMATION SECURITY AGENCY, *Report On Cyber Security In The Czech Republic*, 2020, p. 13, www.nukib.cz/download/publications_en/2020_report_on_cyber_security_in_the_czech_republic.pdf (consulté le 24.07.2022).

39 C. CIMPANU, « Un hôpital tchèque frappé par une cyberattaque en pleine épidémie de Covid-19 », *ZDNet*, 14.03.2020, www.zdnet.fr/actualites/un-hopital-tcheque-frappe-par-une-cyberattaque-en-pleine-epidemie-de-covid-19-39900659.htm (consulté le 26.07.2022).

40 « En Allemagne, une attaque informatique contre une clinique provoque une mort, le 17 septembre 2020 », *Le Monde*, 12.11.2020, www.lemonde.fr/pixels/article/2020/09/17/en-allemagne-une-attaque-informatique-contre-une-clinique-provoque-une-mort_6052638_4408996.html (consulté le 24.07.2022).

des hôpitaux par l'une de partie en conflit, ce qui constituerait un crime de guerre.

2. *Crime d'agression*

Les nouveaux amendements adoptés en 2010 à Kampala liés au crime d'agression, ainsi que la décision des États parties d'activer la compétence de la CPI en la matière, ont permis d'élargir la liste des crimes poursuivis. À compter du 17 juillet 2018, la Cour est habilitée à réprimer « la planification, la préparation, le lancement ou l'exécution par une personne effectivement en mesure de contrôler ou de diriger l'action politique ou militaire d'un État, d'un acte d'agression qui, par sa nature, sa gravité et son ampleur, constitue une violation manifeste de la Charte des Nations Unies »[41]. Le Statut de Rome emprunte cette définition à l'art. 3 de la Résolution de l'ONU 3314 de 1974 sur l'agression et énumère sept actes entrant dans son champ d'application[42] :

1) L'invasion ou l'attaque du territoire, toute occupation militaire, ou toute annexion ;
2) Le bombardement ou l'emploi de toutes armes ;
3) Le blocus des ports ou des côtes ;
4) L'attaque contre les forces armées terrestres, navales ou aériennes, ou la marine ou l'aviation civiles ;
5) L'utilisation des forces armées d'un État qui sont stationnées contrairement aux conditions prévues dans l'accord, ou toute prolongation de leur présence sur le territoire en question ;
6) L'utilisation d'un territoire d'un État pour perpétrer un acte d'agression ;
7) L'envoi de bandes ou de groupes armés, de forces irrégulières ou de mercenaires pour commettre des actes de force armée contre un autre État.

Ces dernières années, la communauté internationale a beaucoup débattu de la question de savoir s'il est possible d'inclure les actes dans le cyberespace dans cette liste, parce que la Charte des Nations Unies ne propose aucun critère qui spécifie si un acte atteint le seuil de recours à la force. Les groupes d'experts et les ONGs ont apporté des clarifications sur ce point. En particulier, le *Council of Advisors* souligne que la notion de force armée qui est étroitement liée à un acte d'agression s'applique quelle que soit l'arme spécifique utilisée, qu'elle

41 Art. 8 *bis* du Statut de Rome.
42 ONU, Assemblée Générale, *Résolution 3314 (XXIX)*, 29 novembre 1974, art. 3, https://digitallibrary.un.org/record/190983/files/A_RES_3314%28XXIX%29-FR.pdf (consulté le 02.08.2022).

soit conventionnelle ou cybernétique[43], en se référant à l'Avis consultatif sur les armes nucléaires déjà cité.

Les conclusions du groupe d'experts du Manuel de Tallinn vont dans le même sens[44]. Ils proposent, entre autres choses, de se référer à l'affaire Nicaragua dans laquelle la CIJ a déclaré que les critères de l'ampleur et des effets doivent être pris en compte pour déterminer si les actes constituent une « attaque armée » ou pas[45]. Par conséquent, les experts ont convenu qu'il n'y a aucune raison d'exclure les cyberopérations du champ des actes constituant un recours à la force si l'ampleur et les effets de l'opération sont comparables à des opérations cinétiques qui seraient qualifiées d'attaques armées[46].

Dans le cadre de ces débats, une attention particulière doit être également portée à la position du Conseil de sécurité de l'ONU, lequel a le droit de déterminer si d'autres actes constituent une agression en vertu des dispositions de la Charte des Nations Unies[47]. Bien que le Conseil n'ait jamais identifié les cyberopérations comme un acte d'agression[48], nous pouvons noter qu'il condamne néanmoins fermement les attaques, y compris cybernétiques, contre des infrastructures civiles essentielles[49].

Un autre argument avancé par des experts consiste dans le fait que la Résolution 3314, sur la définition de l'agression, indique clairement que la liste des actes énumérés n'est pas limitative[50]. Cependant, contrairement à la Résolution, le paragraphe 2 de l'art. 8 *bis* du Statut de Rome, qui reproduit cette liste, ne précise pas que cette énumération a un caractère non limitatif, ce qui laisse penser que cette disposition a été rédigée de la sorte pour une raison bien particulière. *De jure*, cela signifie que les dispositions de l'art. 8 *bis* du Statut de Rome ne font pas l'objet d'une interprétation large et que la liste des actes d'agression ne peut pas être complétée par de nouveaux actes d'emploi des forces armées, y compris cybernétiques. Tout ceci nous conduit à conclure

43 COUNCIL OF ADVISORS (n. 13), p. 9.
44 SCHMITT (n. 27), Règle 68, par. 2, p. 329.
45 CIJ, *Nicaragua c. États-Unis*, Arrêt du 27 juin 1986, par. 1958.
46 SCHMITT (n. 27), Règle 69, par. 1, p. 331.
47 ONU, Charte de l'ONU, art. 39.
48 Voir à ce sujet le cas de la Géorgie qui a adressé une lettre au Conseil de sécurité des Nations Unies pour examiner la cyberattaque contre l'infrastructure géorgienne : SECURITY COUNCIL, *Identical letters dated 21 February 2020 from the Permanent Representative of Georgia to the United Nations addressed to the Secretary-General and the President of the Security Council*, 24 février 2020, p. 2, https://digitallibrary.un.org/record/3853090?ln=fr (consulté le 09.08.2022).
49 ONU, Conseil de Sécurité, *Security Council Strongly Condemns Attacks against Critical Civilian Infrastructure, Unanimously Adopting Resolution 2573 (2021)*, 27 avril 2021, https://press.un.org/en/2021/sc14506.doc.htm (consulté le 09.08.2022).
50 ONU, Assemblée générale, *Résolution 3314 des Nations Unies sur la définition de l'agression*, 14 décembre 1974, art. 4.

qu'un acte dans le cyberespace peut constituer un crime d'agression, à condition d'être réalisé sous la forme de l'un des sept actes énumérés dans le Statut de Rome.

Prenons un exemple qui est entièrement théorique. L'État A permet à l'État B d'utiliser son territoire pour héberger des serveurs. L'État B est conscient du fait que l'État A lance des cyberattaques qui font exploser les usines chimiques dans l'État C et provoquent des morts parmi la population civile. Dans pareil exemple, l'État B peut également être tenu responsable en vertu de l'art. 8 *bis* par. 2 f) du Statut de Rome, dès lors qu'il a connaissance du fait que son territoire sert à la commission, par un autre État, d'un acte d'agression contre un État tiers.

3. *Crime contre l'humanité*

L'art. 7 du Statut de Rome définit le crime contre l'humanité comme un acte commis dans le cadre d'une attaque généralisée ou systématique contre toute population civile, ou en connaissance de cette attaque, et prévoit une liste des actes qui pourraient constituer ce type de crime. Comme le crime de génocide, cette catégorie d'infractions peut être commise à la fois pendant un conflit armé et en temps de paix.

Maintenant, nous allons examiner comment un crime contre l'humanité peut se manifester dans le cyberspace. Le *Council of Advisors* donne comme exemple la cyberattaque qui vise à couper le courant électrique pendant l'hiver rigoureux pour une longue période[51]. L'un des cas les plus cités a eu lieu en Ukraine en 2015 lorsque la cyberattaque contre le réseau électrique a provoqué une coupure d'électricité à Ivano-Frankivsk[52]. Selon les estimations provenant de diverses sources, entre 200 000 et 1,5 million de civils ont été victimes de ce virus informatique, bien que le réseau ait été mis hors service seulement pendant quatre heures et n'ait pas provoqué de graves conséquences. Toutefois, une coupure d'électricité prolongée pourrait avoir des effets potentiellement néfastes pour les services civils, y compris les services de santé, ce qui, à son tour, pourrait entraîner des morts parmi les civils et constituer un crime contre l'humanité au sens du Statut de Rome.

51 COUNCIL OF ADVISORS (n. 13), par. 14, p. 65.

52 Ukraine : « Une cyberattaque coupe l'électricité », *Le Figaro*, 05.01.2016, www.lefigaro.fr/flash-actu/2016/01/05/97001-20160105FILWWW00381-ukraine-une-cyber attaque-coupe-l-electricite.php (consulté le 08.08.2022).

Prenons un autre exemple tiré de la vie réelle pour mieux comprendre comment un crime contre l'humanité peut être commis dans le cyberespace : la persécution des Ouïghours en Chine[53]. Le Consortium international des journalistes d'investigation a révélé que, dans le but de procéder au contrôle de la population ouïghoure, le gouvernement chinois utilise les technologies modernes pour surveiller les Ouïghours et les envoyer dans des centres dits « d'éducation et de formation »[54]. Les portables des Ouïghours ont des applications de reconnaissance faciale qui sont associées à des caméras sur les rues de la Chine : ces applications relayent l'information sur le nom du propriétaire du portable, son numéro, son adresse personnelle, sa photo, sa date de naissance, et même l'identité de son employeur[55] ; tout ça pour contrôler la vie de ce groupe ethnique dans le cyberespace.

> « La vie des Ouïghours consiste désormais à générer des données. Tout le monde sait que le smartphone est quelque chose que vous devez porter sur vous, et que si vous ne le portez pas, vous pouvez être détenu, ils savent que vous êtes suivi par lui. Et ils ont l'impression qu'il n'y a pas d'échappatoire. »[56]

Bien qu'il soit impossible de faire une demande d'enquête sur la situation auprès de la CPI pour l'instant[57], les actions du gouvernement chinois en relation avec les Ouïghours peuvent avoir des conséquences considérables et pourraient constituer un crime contre l'humanité.

53 Voir CPI, *Bureau du Procureur, Rapport sur les examens préliminaires 2020*, 14 décembre 2020, par. 71, www.icc-cpi.int/sites/default/files/itemsDocuments/2020-PE/2020-pe-report-eng.pdf (consulté le 04.08.2022).

54 S. ALECCI, « Uighur répression "turbocharged by technology", Confidential documents show », *ICIJ*, 14.12.2020, www.icij.org/investigations/china-cables/uighur-repression-turbocharged-by-technology-confidential-documents-show/ (consulté le 04.08.2022).

55 S. SEIBT, « Comment Pékin organise la surveillance 2.0 des Ouïghours », *France24*, 18.02.2019, www.france24.com/fr/20190218-chine-ouighour-surveillance-xinjiang-reconnaissance-faciale-qr-code-musulman (consulté le 31.07.2022).

56 J. WAKEFIELD, « Intelligence artificielle : un logiciel pour déchiffrer les émotions des Ouïghours », *BBC*, 02.06.2021, www.bbc.com/afrique/monde-57270581 (consulté le 08.08.2022).

57 En décembre 2020, la Procureure générale Fatou Bensouda a mentionné dans son rapport que les actes du gouvernement chinois pourraient constituer des crimes contre l'humanité, y compris la persécution, toutefois, elle a refusé d'enquêter sur les Ouïghours, parce que la Chine n'était pas État partie au Statut de Rome. À propos des crimes commis contre la minorité musulmane ouïghoure voir, par exemple, CPI, Bureau du Procureur, *Rapport sur les examens préliminaires 2020*, 14 décembre 2020, par. 71, www.icc-cpi.int/sites/default/files/itemsDocuments/2020-PE/2020-pe-report-eng.pdf (consulté le 04.08.2022).

4. *Génocide*

L'art. 6 du Statut de Rome définit le génocide comme un acte commis dans l'intention de détruire, en tout ou en partie, un groupe national, ethnique, racial ou religieux. Cette infraction nécessite que l'acte soit commis contre une ou plusieurs personnes qui appartiennent à l'un des groupe susmentionnés, que l'auteur ait l'intention de détruire ce groupe, et que son comportement s'est inscrit dans le cadre d'une série de comportements analogues dirigés contre ce groupe, ou pouvait en lui-même produire une telle destruction[58]. S'agissant du crime de génocide, le Statut de Rome s'est inspiré des dispositions de la Convention sur le génocide[59] et a prévu la responsabilité pénale individuelle, entre autres, pour l'incitation directe et publique à commettre ce crime[60]. Par définition, l'incitation est un crime inchoatif ; pour être poursuivi, ce crime ne nécessite pas qu'un génocide ait eu lieu : un simple encouragement à la violence génocidaire suffit. Les technologies modernes permettent aujourd'hui de se retrouver facilement dans une situation d'incitation au génocide, en particulier avec le recours aux réseaux sociaux qui sont capables d'atteindre des millions de personnes avec une seule publication. C'est ce qu'a montré le cas des Rohingyas qui mérite pour cette raison d'être examiné plus en détail.

Persécutée depuis des décennies au Myanmar, la communauté ethnique musulmane Rohingya a été victime d'une campagne meurtrière lancée par le gouvernement du pays. Des milliers de Rohingyas ont été tués en raison de leur appartenance à cette minorité ethnique et des centaines de milliers d'entre eux ont fui au Bangladesh en 2017. D'après le rapport du Mécanisme onusien d'enquête indépendant pour le Myanmar, l'utilisation de Facebook par le gouvernement du Myanmar a joué un rôle important pour répandre le génocide contre la communauté musulmane[61]. Les exemples suivants de posts publiés sur les réseaux sociaux sont parlants :

- Un représentant des forces armées du Myanmar (Tatmadaw) a publié un message qu'il avait hâte d'être déployé dans l'État de Rakhine, car les « chiens musulmans » constituent une menace pour les citoyens[62].

58 CPI (n. 32), art. 6.
59 Art. III de la Convention pour la prévention et la répression du crime de génocide.
60 Art. 25 par. 3 e) du Statut de Rome.
61 ONU, Comité des droits de l'homme, Mécanisme onusien d'enquête indépendant pour le Myanmar, *Rapport*, 17 septembre 2018, https://digitallibrary.un.org/record/1643079 (consulté le 24.07.2022).
62 *Idem*, par. 1378.

- Un officier de police impliqué dans les opérations contre les Rohingyas a écrit qu'il voulait « tuer ces ‹ Kalar › depuis si longtemps. Je n'ai qu'à les tuer tout à l'heure »[63].

Après une analyse approfondie, le Mécanisme d'enquête pour le Myanmar est arrivé à une grave conclusion : « Ces déclarations font partie d'une campagne de propagande plus large, diffusant des informations manifestement fausses et incitant [...] à la violence »[64]. Cette situation a également attiré l'attention du Bureau du Procureur de la CPI[65] et les juges de la Cour ont autorisé l'ouverture d'une enquête. Cet exemple est comparable à la campagne de haine préparée par la radio des Mille Collines qui fonctionnait comme un canal de propagande et un outil pour diaboliser les Tutsis en diffusant des discours incitant à l'exécution du génocide rwandais en 1994[66].

Bien qu'il faudra plusieurs années avant que la CPI se prononce sur la situation au Myanmar, nous pouvons déjà discerner une volonté de la Cour de réprimer les crimes commis dans le cyberespace, dès lors qu'elle a décidé de prendre en compte les actes commis sur Facebook.

III. La CPI face aux nouveaux défis posés par les cybercrimes

Les défis éventuels auxquels la Cour pénale doit faire face aujourd'hui sont de deux types : les obstacles propres à l'action de la CPI et ceux découlant de la nature spécifique du cyberespace (A). Certains experts considèrent que les crimes internationaux commis dans le cyberespace sont si graves qu'il est indispensable soit de compléter le champ des compétences de la CPI, soit de créer un nouveau tribunal qui ne serait compétent que pour les cybercrimes (B).

63 *Ibid.*
64 *Idem*, par. 1379.
65 CPI, Bureau du Procureur, *Requête aux fins d'autorisation d'ouvrir une enquête dans la situation au Myanmar*, 4 juillet 2019, par. 116, par. 176, www.icc-cpi.int/sites/default/files/CourtRecords/CR2019_03510.PDF (consulté le 08.08.2022).
66 A. HEFTI/L. JONAS, « From Hate Speech to Incitement to Genocide : The Role of the Media in the Rwandan Genocide », *Boston University International Law Journal*, vol. 38, 2020, p. 14, www.bu.edu/ilj/files/2020/08/Article_HeftiJonas.pdf (consulté le 19.07.2022).

A. Les obstacles à la répression des crimes internationaux commis dans le cyberespace

La nature inhérente du cyberespace crée une opportunité pour les criminels d'exploiter anonymement les vulnérabilités des cyberinfrastructures militaires et civiles. Les États et les acteurs non étatiques peuvent utiliser une vaste gamme de techniques et de logiciels afin de cacher leurs traces et l'origine d'un crime via plusieurs serveurs qui peuvent se trouver aux deux extrémités du globe, ce qui empêche la répression des crimes commis dans le cyberespace[67]. Le problème est tellement grave qu'aussi bien la littérature sur les cybercrimes que les entreprises de cybersécurité insistent sur la difficulté d'attribuer les actes commis dans le cyberespace[68].

> « Juger des crimes de guerre pour des atrocités commises dans le monde physique est un processus long et difficile pouvant prendre des années. La sphère numérique ajoute à cette complexité. »[69]

Ces difficultés techniques pourraient avoir des conséquences importantes sur le travail de la CPI. Certains experts soulignent également que les principaux défis auxquels la Cour fait face « ne sont pas liés à la définition des crimes ou aux règles du droit international humanitaire, mais à des obstacles techniques bien connus à l'identification des auteurs »[70]. Premièrement, le risque de mauvaise attribution est très élevé[71]. Même, si la Cour recourt à la technique de l'attribution de fichiers pour obtenir des informations sur des auteurs éventuels des crimes (les métadonnées du code, la langue utilisée par les auteurs, le fuseau horaire de l'auteur, etc.), cette technique peut ne pas être très fiable, car les éléments pris en considération peuvent être faussés, notamment les codes utilisés pour la commission de crimes sont systématiquement piratés et réutilisés par plusieurs auteurs.

67 ROSCINI (n. 29).

68 Voir p.ex. KASPERSKY, *The power of threat attribution, Kaspersky Threat Attribution Engine*, p. 2, https://media.kaspersky.com/en/business-security/enterprise/threat-attribution-engine-whitepaper.pdf (consulté le 07.08.2022) ; G. BROWN, « Some Nondestructive State Cyber Operations Probably Constitute the Crime of Aggression under the Rome Statute, but Attribution Difficulties and State Practice Make Effective Deterrence Unlikely », *ICC Forum*, Cyber Operations and Cyberwarfare Question, 2022, https://iccforum.com/cyberwar (consulté le 06.08.2022) ; P. RASCAGNERES, « Who Wasn't Responsible for Olympic Destroyer ? », *Talos*, 26.02.2018, https://blog.talosintelligence.com/2018/02/who-wasnt-responsible-for-olympic.html (consulté le 07.08.2022).

69 BURKHALTER (n. 2).

70 Voir p.ex. ROSCINI (n. 29).

71 AMBOS (n. 22).

À titre d'exemple, nous pouvons citer le cas *Lazarus* : les concepteurs du virus *Lazarus* l'ont lancé sous une fausse bannière et ont utilisé des mots russes exprès pour embrouiller les spécialistes des TIC. Ils ont également lié un logiciel malveillant à un protecteur commercial (*Enigma*) développé par un auteur russe pour projeter l'ombre sur la Russie. Toutefois, l'analyse a montré que ces mots ont été une mauvaise imitation apparemment réalisée au moyen de *Google Translator*[72].

Le second problème des investigations est lié au temps que l'enquête peut prendre pour détecter un vrai auteur. La durée moyenne d'une procédure devant la CPI (depuis la phase préalable au procès jusqu'à la phase de l'appel) varie de 51 à 72 mois, parfois plus[73] ; l'examen des crimes commis dans le cyberespace pourrait augmenter sensiblement le temps d'examen qui est déjà très long.

Nous ne pouvons pas ignorer également une autre triste réalité : la CPI a mentionné à maintes reprises qu'elle se heurtait au manque de ressources financières pour l'examen des cas qui lui sont soumis[74]. La Coalition pour la Cour pénale internationale a souligné que le sous-financement chronique des activités de la Cour entraîne le refus d'ouvrir de nouvelles enquêtes, affecte l'efficacité de celles en cours, et retarde donc l'accès des victimes à la justice[75]. Dans ces conditions, l'élargissement de la compétence de la CPI pour les crimes internationaux commis dans le cyberespace conduirait à accroître encore l'écart entre la charge de travail de la Cour et les ressources dont elle dispose dans son

72 KASPERSKY, *Lazarus Under the Hood*, 2017, p. 17, https://media.kaspersky.contenthub.com/wpcontent/uploads/sites/43/2018/03/07180244/Lazarus_Under_The_Hood_PDF_final.pdf (consulté le 07.08.2022).

73 SYRIA JUSTICE AND ACCOUNTABILITY CENTER, « Eight Questions about the International Criminal Court », 05.2014, https://syriaaccountability.org/eight-questions-about-the-icc/#note-1118-1 (consulté le 07.08.2022). Voir aussi HRW, « Une Cour pour l'Histoire. Les premières années de la Cour pénale internationale à l'examen », 2008, www.hrw.org/fr/report/2008/07/11/une-cour-pour-lhistoire/les-premieres-annees-de-la-cour-penale-internationale (consulté le 06.08.2022). La durée moyenne d'une procédure devant la Cour de justice de l'Union européenne, à titre de comparaison, est d'environ 15,6 mois. VIE PUBLIQUE, « Quel est le rôle de la Cour de justice de l'Union européenne (CJUE) ? », 2021, www.vie-publique.fr/fiches/38299-quel-est-le-role-de-la-cour-de-justice-de-lunion-europeenne-cjue (consulté le 07.08.2022).

74 Voir, p.ex., CPI, *Statement of the Prosecutor of the International Criminal Court, Karim A. A. Khan QC, following the application for an expedited order under article 18(2) seeking authorisation to resume investigations in the Situation in Afghanistan*, 27.09.2021, www.icc-cpi.int/news/statement-prosecutor-international-criminal-court-karim-khan-qc-following-application (consulté le 07.08.2022) ; CPI, *Rapport sur les activités menées en 2020 en matière d'examen préliminaire*, 14 décembre 2020, www.icc-cpi.int/sites/default/files/itemsDocuments/2020-PE/2020-pe-report-fra.pdf (consulté le 07.08.2022).

75 COALITION FOR THE INTERNATIONAL CRIMINAL COURT, « Victims could lose out with states' double-standard on International Criminal Court resources », 30.03.2022, https://coalitionfortheicc.org/news/20220330/OpenLetter_ICCresources (consulté le 07.08.2022).

budget, ce d'autant plus que la répression des cybercrimes nécessite des spécialistes des TIC de haut niveau.

Parmi les autres défis auxquels la CPI est confrontée, il importe de mentionner celui de la preuve de l'intention criminelle (*mens rea*) permettant à ce que le comportement de l'auteur puisse être juridiquement qualifié d'infraction criminelle. En particulier, la Fondation Carnegie pour la paix internationale, se rapportant au rapport de cybersécurité Microsoft Digital Defense, souligne qu'il y a eu de « nombreux cas dans lesquels les auteurs n'avaient pas tenté de masquer leurs actions, mais avaient tenté de dissimuler leurs véritables intentions »[76]. Par exemple, les criminels peuvent infecter des hôpitaux par un rançongiciel et les mettre hors service en faisant semblant d'être motivés par le gain financier, alors que leur intention réelle consiste à provoquer des morts parmi les civils, comportement qui serait qualifié de crime de guerre au sens du Statut de Rome. D'autre part, la nature du cyberespace est telle que les systèmes informatiques sont interconnectés, avec pour conséquence que les actes qui ciblent des infrastructures militaires peuvent avoir des effets sur des infrastructures civiles ou, dans le pire des cas, aller au-delà de la région visée en provoquant des effets nuisibles dans d'autres parties du monde[77]. En outre, l'autopropagation de la plupart des logiciels malveillants et l'imprévisibilité de leurs effets empêchent parfois aux auteurs de cybercrimes de prédire avec exactitude les conséquences de leurs actes[78]. Nous pouvons citer comme exemple le rançongiciel *NotPetya*, la cyberattaque la plus coûteuse de l'histoire, qui, d'après certains spécialistes, a visé initialement seulement l'Ukraine, mais s'est propagée dans d'autres pays[79].

76 A.E. LEVITE/J. LEE, « Attribution and Characterization of Cyber Attacks, Carnegie Endowment for International Peace », 28 mars 2022, https://carnegieendowment.org/2022/03/28/attribution-and-characterization-of-cyber-attacks-pub-86698#:~:text=Introduction,from%20another%20state's%20computer%20networks (consulté le 07.08.2022).

77 H. KOH, « International Law in Cyberspace », *Harvard International Law Journal*, vol. 54, 2012, p. 6.

78 ICRC, *Expert Meeting 14-16 novembre 2018 – Geneva, The potential human cost of cyber operations*, mai 2019, p. 32, p. 38, www.icrc.org/en/download/file/97346/the-potentialhuman-cost-of-cyber-operations.pdf (consulté le 07.08.2022).

79 M. BAEZNER, *Cyber and Information warfare in the Ukrainian conflict, Report, CSS Cyberdefense Hotspot Analyses*, octobre 2018, p. 35, https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/20181003_MB_HS_RUS-UKR%20V2_rev.pdf (consulté le 07.08.2022).

B. Rendre à César ce qui est à César : la CPI ou un nouveau tribunal ?

Les obstacles susmentionnés ne représentent qu'une petite partie des difficultés que devrait surmonter la CPI, ce qui fait réfléchir à la question de savoir si elle est bien capable de réprimer les crimes internationaux commis dans le cyberespace. À ce jour, la communauté internationale propose deux voies à suivre : créer un nouveau tribunal compétent pour les cybercrimes ou adapter la CPI aux nouvelles réalités de ces crimes.

La proposition de créer un Tribunal pénal international pour le cyberespace (*International Criminal Tribunal for Cyberspace*) a été faite pour la première fois en 2012 par le *Cybercrime Legal Working Group* qui travaillait sous la supervision de l'*EastWest Institute*. Son idée était de créer une juridiction distincte, sur la base d'une décision du Conseil de sécurité des Nations Unies, qui serait compétente seulement pour les cybercrimes « *of the most global concern* »[80]. Bien que cette initiative ait l'air assez ambitieuse, elle a été supportée par certains milieux universitaires[81] car elle permettrait de concentrer les efforts de la communauté internationale seulement sur les cybercrimes. Cette proposition n'est cependant pas à l'abri de critiques.

En premier lieu, le projet de Statut pour le futur Tribunal pénal international pour le cyberespace ne mentionne pas les crimes internationaux et se limite aux actes relevant plutôt de la cybercriminalité, comme la fraude, les infractions liées à la pédopornographie, le spam ou l'usurpation d'identité, ce qui ferait de ce tribunal un instrument indispensable pour l'application de la Convention de Budapest, mais pas du Statut de Rome. De plus, on ne voit pas non plus clairement comment enquêter sur les crimes internationaux si un cyberacte fait partie d'une attaque plus générale commise par des armes classiques qui tomberait déjà dans la compétence de la CPI. Enfin, la création d'un nouveau tribunal avec un nouveau Statut représenterait un processus long et ardu exigeant la volonté des États d'être liés par de nouvelles obligations, tandis que les auteurs des crimes internationaux resteraient impunis.

Dans ces circonstances, l'adaptation de la CPI aux nouvelles réalités du XXIe siècle pourrait être un choix rationnel. Avant tout, le partenariat de la Cour avec

80 S. SCHJOLBERG, *A paper for the EastWest Institute (EWI), Cybercrime Legal Working Group, Recommendations for potential new global legal mechanisms against global cyberattacks and other global cybercrimes, An International Criminal Tribunal for Cyberspace (ICTC), Prosecution for the Tribunal, Police investigation for the Tribunal*, mars 2012, p. 10, www.cybercrimelaw.net/documents/ICTC.pdf (consulté le 07.08.2022).

81 Voir à ce sujet : W. KRAFT, « The Best Way to Shape the Future is to Understand the Present », p. 5 s., https://iasl.space/wp-content/uploads/2021/04/International-Court-for-Cyber-Crime.pdf (consulté le 07.08.2022).

les entreprises privées spécialisées dans la cybersécurité semble nécessaire et prometteur, d'autant que ces derniers disposent de l'expérience nécessaire dans ce domaine. De plus, cela pourrait décongestionner la Cour et résoudrait le problème du manque de ressources humaines et financières. Cette solution est d'autant plus réalisable que le secteur privé est prêt à assumer des responsabilités dans la recherche des criminels, la collecte des preuves et l'attribution des cyberactes. À titre d'exemple, nous pouvons mentionner le travail de Microsoft en la matière. Cette organisation a déjà présenté une initiative en 2016 pour la création d'un organisme international non gouvernemental qui pourrait être responsable de juger des cyberattaques dépassant un certain seuil de gravité[82]. L'année suivante, en 2017, Microsoft a proposé la Convention de Genève digitale pour protéger les civils en temps de paix[83]. Bien qu'elles ne soient pas parfaites du point de vue juridique[84], ces initiatives démontrent la volonté des organisations privées de partager leurs compétences techniques, ainsi que de lutter contre l'impunité dans le cyberespace. Dans cette optique, la CPI pourrait suivre l'exemple d'Interpol qui a souvent eu recours aux entreprises privées et qui a arrêté, il y a plusieurs mois, un groupe de cybercriminels au Nigeria grâce à telle coopération[85]. Cette opération, dont le nom est Delilah, a été initialement basée sur les données fournies par les entreprises privées qui se spécialisent dans la cybersécurité. Elle rappelle aux hackers, États et militaires que les forces de l'ordre continuent de les poursuivre et montre que la collaboration entre les sociétés privées et la CPI est primordiale pour la justice.

IV. Conclusion

Aujourd'hui les actes malveillants dans le cyberespace sont de plus en plus courants : le nombre d'infrastructures militaires et civiles touchées augmente tant en temps de paix qu'en temps de guerre, et la nature de ces attaques est de plus en plus grave. Certains des cybercrimes peuvent avoir des effets comparables aux crimes produits avec des armes classiques et, par conséquent,

82 S. CHARNEY *et al.*, « From Articulation to Implementation: Enabling progress on cybersecurity norms », Microsoft Corporation, juin 2016.

83 MICROSOFT, *A Digital Geneva Convention to protect cyberspace Microsoft Policy Papers*, 2017, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW67QH (consulté le 07.08.2022).

84 Voir, p.ex., l'examen indépendant de chercheurs auprès de CCD COE : T. MINARIK/K. VAN DER MEIJ, *Geneva Conventions Apply to Cyberspace : No Need for a « Digital Geneva Convention »*, Centre d'excellence de cyberdéfense coopérative de l'OTAN, https://ccdcoe.org/news/2017/geneva-conventions-apply-to-cyberspace-no-need-for-a-digital-geneva-convention/ (consulté le 07.08.2022).

85 INTERPOL, « Suspected head of cybercrime gang arrested in Nigeria », 25 mai 2022, www.interpol.int/en/News-and-Events/News/2022/Suspected-head-of-cybercrime-gang-arrested-in-Nigeria (consulté le 07.08.2022).

cela pose la question si les actes eux-mêmes pourraient constituer des crimes internationaux relevant de la compétence de la CPI : crimes de guerre, crimes de génocide, crimes contre l'humanité, ou crimes d'agression.

Le problème est qu'aujourd'hui, la communauté internationale n'est parvenue à un consensus que sur des dispositions générales, à savoir sur le principe de l'applicabilité du droit international au cyberespace. Ce qui reste en suspens, c'est son application sur le plan pratique. À cet égard, les spécialistes proposent des approches différentes. Les partisans de l'approche axée sur les moyens estiment qu'avant que le droit pénal international, notamment le Statut de Rome, puisse être appliqué, il doit être complété par des dispositions qui tiennent compte de la nature particulière du cyberespace et mentionnent explicitement l'élément cybernétique, afin de ne pas violer le principe de légalité en droit pénal (*nullum crimen sine lege*). D'autres approches suggèrent que le Statut de Rome et d'autres documents juridiques fournissent déjà une base d'application et peuvent simplement être appliqués par analogie, sans qu'il soit nécessaire de les compléter.

Nous avons analysé systématiquement comment les actes dans le cyberespace peuvent être qualifiés et nous pouvons conclure que tous les actes commis dans le cyberespace doivent au moins satisfaire aux éléments constitutifs du crime et produire les mêmes effets que les actes cinétiques. Les actes qui ne répondent pas aux critères de nature, d'ampleur et de gravité ne devraient pas relever de la compétence de la CPI, notamment des actes purement politiques ou économiques. Exclure ces attaques du champ de compétence de la CPI serait non seulement cohérent avec la conclusion générale des experts selon laquelle seuls les crimes les plus graves sont traités par le Statut de Rome, mais pourrait également résoudre l'éternel problème du manque de ressources de la CPI. C'est d'autant plus important que certains experts émettent des doutes quant à la capacité de la CPI à traiter les cybercrimes internationaux et ne voient une issue que dans la création d'un organe entièrement nouveau qui serait responsable de toutes les atrocités dans le cyberespace. Toutefois, cette proposition n'est pas exempte de critiques, car elle n'est pas entièrement développée et prendrait beaucoup de temps, laissant les cybercrimes impunis. La variante que la majorité des chercheurs, pénalistes et spécialistes des TIC soutiennent, est qu'il serait préférable de suivre la voie choisie par Interpol en impliquant les organismes privés, d'autant plus qu'ils ont la volonté de le faire.