Une IA responsable et digne de confiance

Boris Barraud

Boris Barraud

Une IA responsable et digne de confiance

Les clés pour évaluer la conformité éthique d'un système d'IA

Ce rapport de recherche a été soutenu par le projet Chaire IA Responsable (ANR-19-CHIA-0008) de l'Agence Nationale de la Recherche

2021

Sommaire

Préambule

La compliance (ou conformité) est, selon les dictionnaires, « l'état de ce qui présente un accord complet, une adaptation totale ». C'est un terme utilisé notamment en droit et en gestion. Dans le monde de l'entreprise, on utilise le terme « compliance » pour désigner cette recherche de la conformité des activités au droit — éventuellement un droit souple, non réduit aux lois des États et prenant en compte des usages, des codes de bonne conduite etc.

La Chaire IA Responsable de Nathalie Nevejans (Université d'Artois) a pour ambition de contribuer au déploiement sur le territoire européen, et spécialement en France, d'une IA éthique et respectueuse des droits fondamentaux dans le cadre d'une innovation responsable garantissant la protection de l'humain et le respect de l'environnement. Il s'agit ainsi de promouvoir en même temps le développement économique autour des IA et la protection des personnes contre tous effets indésirables. C'est une question de confiance, mais une confiance qui doit être méritée et non factice. Il est donc important de bien identifier les conditions d'une IA responsable (bornée par et conforme au droit et à l'éthique), considérant que l'IA sera responsable ou elle ne sera pas.

L'Union européenne et, plus précisément, la Commission européenne a déjà proposé une « liste d'évaluation pour une IA digne de confiance » à l'occasion des « Lignes directrices en matière d'éthique pour une IA digne de confiance » publiées par la Commission en avril 2019. Pour la Commission européenne, les clés de la confiance dans l'IA sont : une IA licite (respectueuse du droit et en premier lieu des droits fondamentaux) ; une IA éthique (respectueuse des valeurs humaines au-delà du droit, visant le bien et le juste) ; une IA robuste (présentant suffisamment de gages de sécurité). La liste d'évaluation souffre du défaut de rechercher l'exhaustivité. Ainsi les exigences essentielles, relevant de la protection des droits fondamentaux des individus, y côtoient les besoins plus accessoires (par exemple, « avez-vous mis en place des processus visant à décrire certains réglages susceptibles d'entraîner une défaillance du système d'IA ? »). Le tout semble trop complexe, trop exigeant pour inciter des entreprises à utiliser cette liste — répondre pertinemment à toutes les questions prendrait de nombreuses heures. Entrer trop dans les détails, avoir des attentes trop fines risque de cacher les enjeux essentiels. Peut-être mieux vaut-il se concentrer sur les exigences les plus impérieuses juridiquement et éthiquement, plutôt que de rechercher l'exhaustivité, de vouloir dénombrer absolument toutes les conditions d'une IA digne de confiance.

Par ailleurs, la Commission recourt parfois à des notions floues difficilement compréhensibles même pour le service juridique d'une entreprise (par exemple, « avez-vous mis en place une gouvernance appropriée des données ? »). Il semble nécessaire de prêter la plus grande attention à la formulation des questions, afin que les répondants comprennent précisément ce qui est attendu.

Le présent rapport vise à présenter la cinquantaine de questions (correspondant aux différentes conditions d'une IA responsable et digne de confiance) que toute entreprise recourant à l'IA devrait se poser afin d'évaluer la légitimité de son outil et, par suite, son acceptabilité. Cette liste concerne en premier lieu les systèmes d'IA qui interagissent directement avec les utilisateurs, mais elle peut aussi être appliquée à des IA qui emportent des conséquences pour des individus sans qu'ils soient en relation directe. Elle est surtout destinée aux développeurs et aux prestataires dont la mission est de mettre en œuvre et d'utiliser des IA. L'objectif est de proposer un cadre transversal, utilisable à l'égard de toutes les techniques, toutes les applications et tous les secteurs.

Cette liste d'évaluation prend en compte pour partie le droit mais s'ouvre aussi largement à l'éthique. Elle ne permet donc pas de juger de la légalité d'une IA. L'objectif est d'obtenir un cadre pouvant être appliqué de manière transversale à l'ensemble des applications et donc de jeter les bases d'une IA digne de confiance dans tous les domaines. Les informaticiens et les entrepreneurs, ayant à l'esprit le questionnaire, seraient de plus en plus en mesure de développer des IA éthiques *by design*.

Il s'agit d'une version provisoire destinée à être largement amendée. Au-delà, les technologies et les usages évoluent si vite qu'il faudrait pouvoir intégrer dans le futur de nouveaux besoins non encore évidents aujourd'hui. Et le processus pourrait être ouvert, permettant à toutes les parties prenantes de formuler des commentaires.

Le problème est qu'il est difficile de ne pas demander aux entreprises des choses impossibles. Souvent, cela revient à exiger que des humains contrôlent ce que fait l'IA, alors qu'on recourt justement à l'IA afin de réaliser des opérations qu'un humain ne saurait réaliser, même en prenant beaucoup de temps. Le contrôle humain des IA est dans bien des cas impossible — ou alors on répondra « oui » à la question de l'existence de ce contrôle bien qu'il soit très partiel et imparfait.

Enfin, peut-être une piste à explorer serait-elle de proposer un questionnaire non aux développeurs ou aux entreprises mais aux consommateurs qui se retrouvent confrontés à une IA. On pourrait alors imaginer une forme de notation (comme pour les sites d'e-commerce) dépendant des retours de la population qui subit les effets des IA.

Introduction

Jean-Jacques Rousseau a montré, il y a déjà longtemps dans son Discours sur les sciences et les arts, combien on peut être savant, habile, efficace, et faire de ces capacités le plus mauvais usage, les employer afin de dominer, tromper, exploiter les autres. Rousseau n'était pas un théoricien du progrès mais de la contingence d'une histoire qui n'est pas écrite à l'avance et qui ne va pas automatiquement dans le sens de l'amélioration de la condition humaine. Tout dépend de l'usage bon ou mauvais que les hommes font des savoirs qu'ils constituent. Ce qui peut apparaître comme un pessimisme est plutôt une lucidité aiguë, qui nous invite à prendre la mesure de l'audace critique et des multiples efforts que l'amélioration de la condition humaine exige.

« L'IA peut aider les êtres humains à être plus libres et à s'épanouir. Mais elle risque aussi de nous entraîner vers une société dystopique. Il est donc urgent de trouver le juste équilibre entre les progrès technologiques et la protection des droits de l'homme. C'est un choix de société dont dépend notre avenir », déclarait en 2018 la Commissaire aux droits de l'homme Dunja Mijatović[1]. Le développement mondial et accéléré des nouvelles technologies oblige les juristes, légistes, jurisconsultes et tous ceux qui pensent ou font le droit à concevoir des garde-fous sous la forme de droits et libertés nouveaux ou déduits de ceux déjà existants. Toute technologie devrait obéir à une double condition : sa maturité, autrement dit la possibilité de sa mise en œuvre concrète, et sa légitimité, soit son utilité pour les individus comme pour la société. Comme l'a écrit il y a bien longtemps Rabelais, « science sans conscience n'est que ruine de l'âme ».

Pour qu'on accepte et qu'on utilise les IA, on doit leur faire confiance. Pour leur faire confiance, on doit être certain que la vie privée est protégée, que l'autonomie est sauvegardée, que l'impartialité des robots est garantie etc. Le risque lié à un manque de confiance envers la technologie, ses développeurs ou son environnement (notamment juridique) est que finissent par être entièrement rejetés des outils qui pourraient pourtant nous apporter de grands bénéfices. Les succès de l'intelligence artificielle dépendront donc de la confiance que le grand public lui accordera. Cette confiance suppose notamment la transparence et la traçabilité — impératifs éthiques et politiques — dès lors qu'une décision est prise sur la base d'une indication algorithmique, cela afin d'éviter sa non-explicabilité. Si une science en devenir est la rétro-ingénierie, consistant à déconstruire les raisonnements des machines, créer des technologies plus lisibles et traçables permettrait de mieux les comprendre et donc mieux les accepter. Mais il est très incertain que de telles technologies soient concevables en matière d'IA. Peut-être faudrait-il surtout que davantage de sociologues, d'anthropologues et de philosophes accompagnent les projets d'intelligences artificielles, de leurs prémisses scientifiques à leurs conséquences pratiques ; et que les futurs ingénieurs et mathématiciens soient formés aux enjeux éthiques de leurs travaux, formés à la « roboéthique »[2]. L'autorégulation est par nature limitée par la défense des intérêts de chaque partie. Des règles du jeu claires et précises doivent être imposées à un niveau supérieur afin de générer un climat de confiance à l'égard de la technologie, tout en s'assurant que cette technologie produise des effets humainement acceptables.

[1] D. Mijatović, « Protéger les droits de l'homme à l'ère de l'intelligence artificielle », Strasbourg, 3 juill. 2018.

[2] N. Nevejans, *Traité de droit et d'éthique de la robotique civile*, LEH Édition, 2017, p. 709.

Or les droits de l'homme numérique, tant qu'ils demeurent à l'état de principes d'inspiration, de sources matérielles du droit, ne sont pas autre chose que cette « roboéthique ». Ce sont ces vecteurs de confiance sans lesquels la société des IA ne saurait se développer autrement qu'en produisant méfiance et même chaos. Alors que les spécialistes de la technologie, dans la Silicon Valley comme ailleurs, peuvent se laisser absorber par les seuls enjeux des performances et de l'efficacité des outils qu'ils élaborent et qu'ils cherchent à améliorer, d'autres doivent interroger les conséquences sociales, économiques, politiques, éducatives, morales, scientifiques et interscientifiques (conséquences des progrès d'une science sur les autres sciences). L'IA, en soi, est bien sûr une affaire de mathématiciens, d'informaticiens, de neuroscientifiques et de roboticiens. L'économie, le droit, la sociologie ou l'anthropologie sont des sciences périphériques à l'intelligence artificielle. Elles n'en sont pas moins indispensables, cela afin de la comprendre mais aussi afin de la guider et même la diriger, pour que la société des IA soit une société de la confiance et de la responsabilité.

La chose est désormais entendue : les données personnelles sont le pétrole du XXIe siècle et les nouveaux « rois du pétrole » sont les GAFAM. Dans ce contexte, il serait très peu pertinent de ne voir de droit, et donc de n'étudier, que les lois, règlements et jurisprudences des États ou, pire, d'un État en particulier. Le droit des IA, ce n'est pas que la loi applicable aux IA. Il s'agit aussi du droit produit par les IA ou par les multinationales du numérique. Ne pas s'y intéresser, en tant que juristes, conduirait à manquer l'essentiel de ce qui se joue dans l'univers numérique. Les IA régissent chaque jour un peu plus les vies des hommes connectés. La loi qu'elles produisent et appliquent s'oppose à la loi du Parlement en ce qu'elle n'est pas, comme cette dernière, issue d'une délibération, mode de décision supposant de prendre le temps de la réflexion et de peser les divers arguments en présence, et laissant une place importante aux compromis politiques et aux exigences politiciennes — une humanité qui est peut-être indispensable à l'élaboration des règles régissant les conduites et relations sociales. Cela pose la question de la capacité de la société numérique à être une société démocratique et ouverte. La loi des IA ne favorise-t-elle pas plutôt une société numérique tyrannique et fermée ? Alors que les hommes et les sociétés, assez paradoxalement, sont pris à la fois dans un mouvement de division et de désunion et dans un mouvement de standardisation qui font l'un et l'autre craindre le pire, le fonctionnement de nombre de plateformes numériques amène à redouter une large érosion du pouvoir de ces hommes et de ces sociétés sur leurs destins individuels et collectifs. Il n'est pas certain que l' « homme augmenté » le soit y compris dans sa capacité juridique. Sur le plan du droit, il pourrait être davantage un « homme diminué », réduit en données, écrasé par son ombre numérique et soumis à la loi de ses objets connectés. Croyant encore opérer librement des choix, il n'agirait en réalité qu'en raison d'un savant formatage opéré depuis l'enfance.

Le développement de l'IA est porteur de progrès et de dangers. Les conditions de l'IA responsable ont vocation à prévenir ces dangers. Parmi eux, il y a les atteintes à la vie privée et donc le besoin de protéger la vie privée. Mais les menaces que l'IA fait peser sur l'homme ne s'y réduisent pas. On parlera de « loi des IA » au sujet de ces menaces. Une telle expression renvoie aux effets normatifs des indications fournies par des systèmes informatiques. De plus en plus, on préfère se laisser « nudger », se laisser « driver », ce qui est source de confort mais emporte aussi des effets déshumanisants en réduisant notre liberté effective. En réalité, on se trompe de combat lorsqu'on dit qu'il faut protéger les données personnelles au nom du droit à la vie privée. Ce qu'il faudrait davantage invoquer, c'est le droit à la souveraineté individuelle, le droit au libre arbitre, le droit de ne pas être « nudgé ». Nous sommes manipulés comme jamais l'homme ne l'a été dans l'histoire de l'humanité, même au temps des grandes religions et des grandes idéologies collectives.

L'intelligence artificielle menace notre humanité en laissant croire que l'on pourrait maîtriser l'avenir grâce à des calculs et des statistiques basées sur le passé. Des hommes souhaitent réduire les vies individuelles et sociales à des nombres et les gouverner par des procédés algorithmiques, poursuivant un vieux rêve. Il s'agit d'investir un nouveau marché, celui de la vie téléguidée, potentiellement source de profits gigantesques. Pendant ce temps, d'autres hommes sont prêts à s'abandonner à ces systèmes au nom d'une foi aveugle dans la technique et au nom du besoin d'être « in ». D'autres, enfin, suivent le mouvement sans s'en rendre compte, se soumettent aux techniques par suivisme et conformisme, sans voir que d'autres choix restent possibles. Penser les conditions de l'IA responsable, c'est aussi et peut-être surtout un travail à l'égard de la prise de conscience des hommes quant à la nouvelle ère dans laquelle ils entrent. Il s'agit peut-être moins de consacrer ces droits de l'homme numérique dans des lois ou des conventions internationales que d'inculquer à la population la substance de ces droits — mais ces aspects sont bien évidemment liés. La loi des IA est naturellement insidieuse, non explicite, cachée, et ses effets sont pourtant forts. C'est pourquoi en prendre conscience pour décider librement et souverainement de s'y soumettre ou pas devrait être essentiel dans nos existences.

S'intéresser aux conditions de l'IA responsable conduit à s'intéresser seulement à certaines IA : celles qui possèdent des effets performatifs sur nos modes de vie, nos actes et nos décisions au quotidien. On n'ignorera pas qu'il existe beaucoup d'outils informatiques qui rendent de très grands services et qui, sans aucune discussion possible, sont des progrès en ce qu'ils augmentent l'homme et ne portent jamais atteinte à l'humanité. Mais d'autres, au contraire, diminuent l'homme et menacent l'humanité en la remplaçant par la machinité. De plus en plus, entourés par des enregistreurs et par des calculateurs, on s'en remet à leurs indications qui confinent ainsi aux injonctions. L'humanité diminue forcément lorsque les hommes ne pensent plus, ne décident plus, ne critiquent plus. Ces IA, souvent façonnées par des multinationales du numérique, qui plus ou moins discrètement modèlent les comportements, sont les plus concernées par les conditions de l'IA responsable. Les algorithmes de recommandation, s'ils ne sont techniquement qu'un exemple parmi différents types d'IA, constituent ainsi une partie importante de la problématique. En devinant les besoins des consommateurs, ils orientent et même produisent dans une mesure non négligeable ces besoins. Leur prévision des comportements et des préférences est aussi une construction de ces comportements et préférences.

Ce qui rend possible la loi des IA, c'est la numérisation de nos vies et de notre société dans tous leurs aspects. Il n'est plus un fait ou geste qui ne donne instantanément lieu à une trace numérique. Les capteurs, les objets connectés, les procédures et transactions dématérialisées sont partout. À l'ère des big data, la loi des IA règne en maître. Si le mouvement n'en est encore qu'à ses prémices, on constate déjà les conséquences importantes qu'il produit et qu'il produira. Les données sont une nouvelle ressource et un nouveau défi. Alors que de trop nombreuses informations nous perdent, il est commode et même rassurant de s'en remettre à la loi des algorithmes. Face à des informations pléthoriques, difficile de ne pas suivre les recommandations des « ordinateurs » — au sens premier du mot : ce qui ordonne.

Plutôt qu'une méfiance de principe stérile, on peut s'engager dans des démarches positives à l'égard du numérique, comme le fait par exemple le ministère de la justice lorsqu'il mise largement sur les nouvelles technologies de l'information et de la communication dans le projet de loi de programmation pour la justice 2018-2022. Tel est aussi le cas de la Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique lorsqu'elle « envisag[e] le numérique

comme un levier d'accélération démocratique et d'approfondissement des droits et libertés »[3]. Or cette commission n'a pu que reconnaître combien « la teneur des débats politiques et l'actualité législative l'ont obligée à adopter une démarche largement défensive »[4]. Il ne faut donc pas oublier que le numérique en général et l'intelligence artificielle en particulier ne sont que des instruments et que, en tant que tels, ils dépendent totalement des utilisations qu'en font les hommes qui les tiennent entre leurs mains. Si l'on imagine que ces instruments ne pourraient servir qu'à des agissements dangereux, on les abandonnera à ceux qui procèdent à ces agissements dangereux et l'on oubliera de les utiliser à des fins positives. Comme internet, l'IA est évidemment un progrès, il importe de s'en souvenir.

Il faut pouvoir penser les conditions de l'IA responsable en dépassant la seule problématique de la vie privée et des données personnelles. Beaucoup d'associations de « défense d'un internet libre » ne parviennent pas à aller plus loin, ce qui est emblématique de l'égoïsme généralisé de l'époque. La liberté peut aussi être collective, pas uniquement individuelle. On peut se préoccuper des modes d'organisation collective promus par les IA, de l'utilitarisme grandissant et des logiques de pouvoir qu'elles accompagnent. La protection des données à caractère personnel et de la vie privée est devenue l'unique objet des inquiétudes entourant les technologies numériques. Tout tend à s'analyser à travers ce prisme, comme si la liberté, l'égalité, la dignité ou la solidarité dépendaient du respect du droit à la vie privée, devenu la clé de voûte des droits de l'homme. Cela est le reflet du libéralisme politique des démocraties actuelles, qui réduit toute chose à sa dimension personnelle. Des institutions publiques ont pour mission de veiller sur le monde numérique, mais aucun autre motif que le respect de la vie privée ne les préoccupe. On ne voit pas, par exemple, ces institutions se mobiliser face à la marchandisation intégrale de la vie, au recul de la faculté de jugement ou à l'extrême rationalisation des sociétés. Au contraire, elles souhaitent surtout encourager l'essor de l'économie numérique en garantissant la confiance des consommateurs numériques.

L'enjeu est de construire une IA positive et éthique, une IA responsable et respectueuse de l'humain, une IA qui donne confiance. Or l'Europe semble, plus que toute autre région dans le monde, disposer des ressources pour cela. L'IA ne saurait se développer loin du droit, mais le cadre juridique de l'IA est en construction. Le pari pris ici est que ce cadre et que les conditions d'une IA responsable ne sauraient se passer des droits de l'homme numérique. On peut vouloir promouvoir une économie de la donnée, mais sans jamais franchir les limites de ces droits. Dès lors, l'Union européenne, héritière de la Renaissance et des Lumières, qui dispose de sa Charte des droits fondamentaux depuis 2000 et qui fréquente de près le Conseil de l'Europe et sa Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales, semble mieux que quiconque pouvoir promouvoir une IA équilibrée, une IA mature, capitalisant sur les hauts standards de protection imposés par la législation européenne sur les données. Comme l'a écrit Cédric Villani dans son rapport de 2018, « plutôt que de fragiliser nos trajectoires individuelles et nos systèmes de solidarités, l'IA doit prioritairement nous aider à activer nos droits fondamentaux, augmenter le lien social et renforcer les solidarités »[5]. À ces conditions, on pourrait dire que l'IA est un progrès.

Cependant, la Commission européenne et les experts qui la conseillent souhaitent depuis longtemps inscrire l'IA dans un cadre éthique plus que juridique. Ils estiment qu'il serait trop tôt pour

[3] Ch. Féral-Schuhl, Ch. Paul, Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique, *Numérique et libertés : un nouvel âge démocratique*, Assemblée Nationale, rapport n° 3119, 2015, p. 62.

[4] *Ibid.*

[5] C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 12.

réglementer le secteur de l'IA et que seule l'éthique, souple et vague, permettrait de façonner des IA responsables. Tel est le sens du label « IA de confiance » voulu par la Commission, reposant sur de nombreux critères éthiques détachés du droit positif — même s'il n'est pas rare qu'ils se rejoignent. Sans s'engager dans l'immédiat vers un projet de réglementation, l'Union européenne n'en oublie pas moins de promouvoir une approche « centrée sur l'humain », seule capable d'optimiser les bienfaits des IA tout en en diminuant les menaces. Mais, en refusant toute contrainte juridique, n'est-ce pas une manière de privilégier avant tout la compétitivité des acteurs économiques européens tout en s'assurant une bonne conscience grâce aux nombreux comités et rapports éthiques ? Peut-être s'agit-il d'une stratégie extrêmement ingénieuse afin d'obtenir, si ce n'est un équilibre effectif, du moins un équilibre apparent, sans trop entraver les efforts soutenus pour tenter de rejoindre l'Amérique du Nord et l'Asie dans la course au développement de l'IA. Et puis l'éthique est une vraie arme face à des concurrents qu'il est aisé de renvoyer à leur manque de conscience et alors que, de plus en plus, la confiance devient bel et bien l'enjeu numéro un pour toute entreprise proposant au public des services fonctionnant à base d'IA ou de numérique plus généralement. Le respect de certains principes éthiques deviendrait un avantage concurrentiel. Ce qui était hier un frein se transformerait ainsi en gain.

Telles qu'identifiées en ces pages, les conditions d'une IA responsable s'ordonnent autour de cinq axes : la vie privée (I), la liberté (II), l'égalité (III) et le contrôle (IV).

I. Vie privée

Question 1. Le système d'IA a-t-il été conçu avec pour objectif de protéger la vie privée et les données personnelles, l'identité et l'intégrité des utilisateurs ?

Question 2. Les données utilisées par le système d'IA sont-elles des données à caractère personnel ?

Question 3. Les utilisateurs dont les données personnelles sont utilisées y ont-ils consenti de façon explicite et éclairée ?

Question 4. Les utilisateurs dont les données personnelles sont utilisées sont-ils mis en mesure de révoquer ce consentement s'ils le souhaitent ?

Question 5. Est-ce que des mesures permettant de renforcer la protection de la vie privée, par exemple des mesures de cryptage, d'anonymisation ou d'agrégation, ont été prises ?

Question 6. Des mécanismes de contrôle de la qualité et de l'intégrité des données ont-ils été mis en place ?

Question 7. L'accès aux données par des membres de l'entreprise ou des tiers est-il strictement limité et les éventuelles consultations de ces données sont-elles enregistrées (qui, quand, pourquoi) ?

A. Le droit à l'identité numérique

1. La carte d'identité numérique de l'homme numérique

L'identité numérique est le résultat de l'agrégation, au fil du temps, des traces laissées par un individu au fur et à mesure de ses activités en ligne, des fragments de personnalité enregistrés sur les réseaux sociaux, forums, blogs et autres sites contributifs, des historiques de toutes ses activités numériques. Libérée du sol et du sang, de la généalogie, elle est l'ensemble des données et caractéristiques qui permettent de reconnaître une personne et d'établir son individualité[6]. Se construit ainsi une « nouvelle *polis* numérique, qui a sa rationalité, son territoire, ses classes et enfin ses barbares »[7]. L'identité numérique est, par suite, la persistance du « moi » dans l'espace immatériel et la conscience de cette persistance.

L'identité est tout pour l'homme qui vit en société. Sans elle, il devient à la fois aveugle et invisible. L'absence d'identité rend incapable, inutile, sauf éventuellement pour soi-même si l'on est un solitaire vivant tel un ermite. « L'identité, observe Raphaël Enthoven, est une instance séparatrice, une existence qui se prend pour une essence, un "empire dans un empire" selon Spinoza, qui rétrograde le vivant au rang d'observateur et fait passer du monde comme désir au monde comme divertissement »[8]. L'identité numérique, c'est le droit d'exister, le droit d'être dans l'univers numérique, le droit d'être reconnu comme individu. Alors que cette identité numérique est soumise à une traçabilité continue qui nous échappe, elle appelle la nécessaire « reconnaissance d'un droit à l'autodétermination informationnelle permettant à chaque individu de décider de la communication

[6] M. Doueihi, « Un humanisme numérique », *Communication & langages* 2011, n° 167, p. 3.

[7] M. Doueihi, *La grande conversion numérique – Suivi de Rêveries d'un promeneur numérique*, Le Seuil, 2011, p. 137.

[8] R. Enthoven, « Sagesses de l'amour », in A. Camus, *Œuvres*, Gallimard, coll. Quarto, 2013, p. 18.

de ses données et de garder la maîtrise de leur utilisation, afin de s'épanouir librement dans l'univers numérique »[9].

Il faut donc affirmer que tout homme numérique a droit à une identité numérique et a droit à sa protection. La reconnaissance de cette identité numérique est la base indispensable à la structure des droits de l'homme numérique, en prolongeant l'identité physique et les droits de l'homme biologique. Le droit au respect de la vie privée, le droit au secret des communications électroniques ou le droit au respect de la dignité humaine supposent la reconnaissance d'une identité numérique. L'exercice des libertés publiques numériques individuelles et collectives qui permettent le développement d'une véritable citoyenneté numérique ne va pas sans consécration d'une identité numérique. Celle-ci a déjà connu quelques développements concrets, à l'image de la reconnaissance de la valeur juridique de la signature électronique dans le Code civil par la loi du 13 mars 2000, qui a constitué un progrès en matière de sécurisation et de facilitation des transactions informatiques.

Le droit à une identité numérique implique certains droits tels que le droit d'accéder aux réseaux numériques, donc l'interdiction de toute sanction privative d'accès à internet. Il suppose surtout le droit à l'intégrité de cette identité numérique, qui s'exprime dans le droit des données personnelles. Le commissaire européen pour l'économie et la société numérique plaide pour « un code civil pour l'ère du numérique qui clarifie en détails les questions qui peuvent se poser en matière de droits relatifs aux données »[10]. Déjà en 2011 un délit spécifique a été créé s'agissant de l'usurpation d'identité numérique[11]. Introduit à l'article 226-4-1 du Code pénal par la loi du 14 mars 2011, dite « Loppsi 2 », il dispose que « le fait d'usurper l'identité d'un tiers ou de faire usage d'une ou plusieurs données de toute nature permettant de l'identifier en vue de troubler sa tranquillité ou celle d'autrui, ou de porter atteinte à son honneur ou à sa considération, est puni d'un an d'emprisonnement et de 15 000 € d'amende ». Y compris le titulaire d'une marque peut se prévaloir de cette disposition. Même s'il a fallu attendre 2014 pour que soit prononcée la première condamnation sur ce fondement[12], l'usurpation d'identité numérique est évidemment interdite. Il est donc défendu de créer un faux profil sur un réseau social ou tout autre service en empruntant son identité à un tiers, cela portant une grave atteinte à ses droits de la personnalité[13].

En outre, l'identité physique doit aussi être préservée. L'identification automatisée sur la base de données biométriques (empreinte digitale, détection de la voix, reconnaissance faciale) suppose un consentement valide à une telle surveillance ciblée, éclairé et réitéré. Et les IA ne sauraient être utilisées afin d'imiter ou déformer l'apparence physique, la voix ou d'autres caractéristiques individuelles dans le but de nuire à la réputation d'un individu ou de manipuler d'autres personnes[14]. Il semble ainsi indispensable de sanctionner très sévèrement les « *deep fake* », ces techniques de synthèse d'images basées sur l'intelligence artificielle et qui permettent, à des fins de canular, de manipulation de l'opinion ou pour nuire à quelqu'un, de superposer des fichiers audios et vidéos afin de faire dire ou faire faire quelque-chose à quelqu'un qui ne l'a jamais dit ou fait. Les utilisations de

---

[9] Déclaration commune de la Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et la Commission sur les droits et devoirs sur internet de la Chambre des députés italienne, 28 sept. 2015.

[10] G. H. Oettinger, « Pour un code civil des données numériques », *Le Monde* 15 oct. 2016.

[11] L. n° 2011-267, 14 mars 2011, *LOPPSI 2*.

[12] TGI Paris, 13e ch. cor., 18 déc. 2014, *X c. Rachida Dati*.

[13] TGI Paris, 17e ch., 20 mai 2015, *V. P. et autres c. A. S. et autres*.

[14] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

toutes ces données par les pouvoirs publics, à l'image des systèmes mis en place en Inde ou en Chine, posent bien sûr d'autres questions. Sur ce point, il est significatif que le Kenya, après avoir envisagé sérieusement de créer un système national intégré de gestion de l'identité combinant les documents d'identité des personnes à leurs données biométriques, aux registres du cadastre, aux registres scolaires et à d'autres informations personnelles détenues par divers organismes gouvernementaux, a finalement dû renoncer à ce projet sous la pression de sa Haute Cour. Cette dernière a jugé qu'une loi sur la protection des données personnelles devait au préalable être mise en place[15].

2. La protection des données personnelles, un droit de l'homme numérique

La Convention 108 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, conclue dans le cadre du Conseil de l'Europe le 28 janvier 1981 et qui a constitué le premier instrument international dans le domaine de la vie privée et des données en entrant en vigueur dès le 1er octobre 1985, soulignait déjà que, « dans certaines conditions, l'exercice d'une complète liberté de traiter les informations risque de nuire à la jouissance d'autres droits fondamentaux (par exemple les droits à la vie privée, à la non-discrimination et à un procès équitable) ou à d'autres intérêts personnels légitimes (par exemple en matière d'emploi ou de crédit à la consommation). C'est pour maintenir un juste équilibre entre les différents droits et intérêts des personnes que la Convention impose certaines conditions ou restrictions au traitement d'informations ». Le règlement du 27 avril 2016, n° 2016/679, du Parlement européen et du Conseil 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, dit « règlement général sur la protection des données » (RGPD) — adopté le 27 avril 2016, après plus de quatre années de discussions, entré en vigueur le 25 mai 2018 et qui constitue désormais le cadre de référence pour les européens —, au premier point de ses considérants, consacre également le fait que « la protection des personnes physiques à l'égard du traitement des données à caractère personnel est un droit fondamental ». Et l'article 4-2 du RGPD définit le traitement de données comme « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction ».

L'article 8§1 de la Charte des droits fondamentaux de l'Union européenne, comme l'article 16§1 du Traité sur le fonctionnement de l'Union européenne, affirment que « toute personne a droit à la protection des données à caractère personnel la concernant ». Cette protection des données personnelles est donc d'ores et déjà considérée comme un droit de l'homme numérique ô combien nécessaire face à l'évolution des modes d'interaction des individus avec les outils numériques connectés. L'usage du web participatif, en particulier des réseaux sociaux, ne pose pas véritablement de nouvelles questions concernant la protection des droits de la personnalité. Mais il démultiplie le nombre et la gravité des atteintes. Et la dimension transnationale des réseaux de communication rend périlleuse la bonne application d'un droit dont le caractère indispensable ne fait guère de doute mais qui demeure par trop national.

[15] A. Faivre, « Données personnelles : comment "l'industrie de l'influence" traverse l'Afrique ? », lepoint.fr, 3 oct. 2020.

Le droit à l'identité et à l'intégrité numérique se traduit essentiellement dans le droit des données personnelles, dans le droit de contrôler en amont les données que l'on accepte de partager et le droit d'en conserver en permanence la maîtrise. L'identité numérique, normalement, devrait impliquer que ce soit aux plateformes d'accepter nos conditions et non à nous d'accepter les leurs. L'anonymat devrait être l'option par défaut, avec possibilité de décider, en pleine conscience, de révéler son identité. Des instruments tels que des portefeuilles numériques permettraient de centraliser nos données personnelles, les sécuriser et les partager selon nos préférences de vie privée. Ainsi, plutôt que de devoir rechercher dans les paramètres les plus souterrains de Google ou Facebook les *privacy settings*, on pourrait automatiquement imposer nos exigences. Bien sûr, un tel système nous appauvrirait en services gratuits, en utilité, mais il nous enrichirait en droits sur notre individualité. N'est-ce pas dans ce sens que va le modèle européen fondé sur le droit et la confiance ? Jumbo, par exemple, est un « assistant de vie privée » qui règle, en fonction des souhaits de l'utilisateur, tous les paramètres de confidentialité disponibles sur les plateformes. Jumbo se charge donc de gérer la protection de notre intimité à notre place, car il serait trop fastidieux de paramétrer chaque service, parfois à chaque utilisation. Pour Mark Zuckerberg, les Occidentaux, comme les Chinois, auraient renoncé à leurs vies privées. Peut-être les y a-t-on surtout forcé, à grand renfort de communication-manipulation et de nudge, et suffirait-il de leur fournir quelques outils comme Jumbo pour qu'ils se préoccupent à nouveau de la protection de leurs informations personnelles.

Le principe qui doit être proclamé, et qui l'est déjà dans une large mesure, au niveau national comme au niveau supranational, est que tout être humain a droit à la confidentialité et au contrôle de ses données personnelles, y compris celles produites par ses comportements en ligne et ses objets connectés. Chacun devrait pouvoir retrouver l'anonymat à tout moment s'il le souhaite. Quant aux utilisateurs de données personnelles, puissances publiques ou privées, ils devraient être entièrement transparents dans la collecte et l'usage des données de tout être humain et en faciliter l'accès, la traçabilité, la confidentialité et la sécurité[16]. Comme l'a affirmé la « Déclaration d'Avignon », « toute exploitation des données comme créations de tout être humain suppose son consentement préalable, libre, éclairé, limité dans le temps et réversible »[17]. Quant à la Déclaration de Montréal, elle peut juger que « les services d'intelligence artificielle ne doivent pas construire de profils de préférences individuelles pour influencer le comportement des personnes concernées sans leur consentement libre et éclairé »[18]. Et, pour la Commission européenne, « une IA digne de confiance suppose d'offrir aux citoyens la maîtrise sur leurs données personnelles et d'éviter que ces dernières ne soient utilisées à leur encontre à des fins préjudiciables ou discriminatoires »[19]. L'exemple du site <note2be.com>, qui invitait les élèves à noter leurs enseignants (identifiés grâce à des données personnelles), est significatif : la CNIL a condamné le site[20]. Ensuite, il s'agit de trouver les moyens de rendre cette maîtrise effective, réelle, qu'elle ne reste pas une pétition de principe. Car le droit à la protection des données personnelles existe depuis longtemps, surtout en France, bien avant que l'internet et les IA n'envahissent nos vies, mais il peine de plus en plus à produire ses effets en pratique.

---

[16] « Déclaration préliminaire des Droits de l'Homme Numérique », Forum d'Avignon, 2014, art. 6.

[17] *Ibid.*, art. 5.

[18] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

[19] Commission européenne, « Lignes directrices en matière d'éthique pour le développement et l'utilisation d'une IA », 8 avr. 2019.

[20] CNIL, 25 juin 2008.

À l'ère des données massives, il devient de plus en plus compliqué de savoir à l'avance comment et pourquoi des calculs seront opérés à partir de nos données. La conception contractuelle d'une collecte des données finalisée et proportionnée, telle qu'imaginée par la loi Informatique et libertés de 1978, a perdu son sens[21]. Il devient nécessaire d'envisager la régulation des données personnelles à travers un contrôle *ex post* de la régularité des traitements et une gestion libre et éclairée de chacun. De nouvelles orientations doivent être données au droit des données personnelles afin d'éviter qu'il ne sombre dans l'archaïsme[22]. On peut, en premier lieu, suivre la Convention modernisée du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (« Convention 108+ »[23]) qui prévoit que le traitement des données à caractère personnel, à toute étape du cycle de vie d'un système d'IA, devrait reposer sur les principes suivants :

- le traitement des données à caractère personnel aux étapes pertinentes du cycle de vie du système d'IA doit s'effectuer en vertu d'un fondement légitime prévu par la loi ;
- les données à caractère personnel doivent être traitées licitement, loyalement et de manière transparente ;
- les données à caractère personnel doivent être collectées pour des finalités explicites, déterminées et légitimes, et ne doivent pas être traitées de manière incompatible avec ces finalités ;
- les données à caractère personnel doivent être adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont traitées ;
- les données à caractère personnel doivent être exactes et, si nécessaire, mises à jour ;
- les données à caractère personnel doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire aux finalités pour lesquelles elles sont traitées.

3. Données et données personnelles

Circulantes, fluides, et bien souvent coconstruites, les données posent question du point de vue de leur définition et surtout s'agissant de leur caractère personnel ou non. La grande majorité des données produites par les individus semblent inoffensives, insignifiantes. Il ne paraît pas opportun de les intégrer dans le régime des données personnelles. Mais les IA sont de plus en plus capables de déduire des caractéristiques individuelles à partir de ces traces anonymes laissées directement (un « clic » ou un « like ») ou indirectement (sous forme de métadonnées), lesquelles peuvent être utilisées afin de contrôler les individus[24]. Les capacités de corrélation des procédés d'analyse statistique font entrer dans l'espace des données personnelles un ensemble de données fragmentées, en apparence anodines, qui en étaient jusqu'alors exclues. Lorsque des paramètres latents, des traits

---

[21] D. Cardon, *À quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Le Seuil, coll. La République des idées, 2015, p. 79.

[22] Au-delà des nombreux travaux de la doctrine, le Conseil d'État, en 2014, a produit une étude sur « Le numérique et les droits fondamentaux » dans laquelle il insiste sur la nécessité de définir un droit des algorithmes. En mai 2016, la Maison-Blanche a publié un rapport intitulé « On Algorithmic Systems: Opportunity and Civil Rights ». Et le 15 décembre 2016, la Secrétaire d'État au numérique et à l'innovation a publié les recommandations du rapport du Conseil général de l'économie concernant les « Modalités de régulation des algorithmes de traitement des contenus ».

[23] La « Convention 108 » est ratifiée par les 47 États membres du Conseil de l'Europe et 6 États non-membres (Cap-Vert, Maurice, Mexique, Sénégal, Tunisie, Uruguay). Sa version modernisée, dite « Convention 108+ », est déjà signée par 24 États membres et un État non membre (Uruguay).

[24] J.-M. Deltorn, « La protection des données personnelles face aux algorithmes prédictifs », *RDLF* 2017, chron. n° 12.

secondaires, qui par recoupement permettent de suivre un individu à la trace, de le réidentifier, sont produits dans un modèle, ils forment alors un faisceau identifiant, une donnée personnelle. Telle était déjà la position retenue par le G29, organe consultatif de l'Union européenne sur la protection des données personnelles, dans un avis du 20 juin 2007 : l'identité d'une personne ne passe pas nécessairement par la connaissance d'éléments d'identité avérés mais peut ressortir d'un faisceau d'autres éléments[25].

L'adresse IP, par exemple, doit-elle être qualifiée de donnée personnelle protégeable à ce titre ? Lorsqu'on visite un site web de e-commerce, celui-ci peut recueillir diverses informations, en particulier l'adresse IP à partir de laquelle on s'est connecté au web. Mais ce numéro d'identification attribué par le fournisseur d'accès à internet ne permet pas, seul, de remonter au nom et au prénom de l'internaute. Cela permet simplement une localisation. L'adresse IP est plus utile en matière pénale puisque des dispositifs comme HADOPI qui lutte contre le téléchargement illégal ainsi que les autorités responsables de la lutte contre le terrorisme ou la pédopornographie ont recours à un traçage des adresses IP. Ils sont en droit d'exiger du fournisseur d'accès à internet qu'il révèle l'identité qui se cache derrière une adresse IP. Alors cette dernière permet bien d'identifier indirectement un individu donné. Quant à la multinationale du web, l'adresse IP lui permet de suivre les activités d'un internaute sans savoir qui est cet internaute, une information de toute façon sans valeur car l'identité numérique se passe des noms et prénoms des personnes. Sans doute faut-il donc protéger l'adresse IP en tant que donnée personnelle. Les conditions de conservation, de stockage et d'exploitation des données sont rarement précisées clairement, tandis que les paramètres de confidentialité peuvent être difficiles à maîtriser. Les informations seraient systématiquement anonymisées : les publicités ciblées seraient adressées à des adresses IP et non à des individus identifiés. Il est difficile de suivre ce raisonnement dès lors qu'une donnée personnelle, en droit, est toute information concernant une personne physique identifiée ou identifiable. Une adresse IP ou des données de connexion semblent rendre identifiable la personne concernée. Inclure des identifiants tels que les adresses SSID, MAC, ou encore IP dans le domaine des données personnelles semble le minimum requis lorsque les appareils auxquels ces données s'attachent révèlent nécessairement d'autres paramètres (par exemple de géolocalisation) qui suffiront sans peine à identifier une personne unique[26].

Les données personnelles sont celles qui permettent d'identifier directement ou indirectement un individu. Elles sont au cœur du web participatif, tant du point de vue des internautes (qui les produisent et les abandonnent) que du point de vue des plateformes (qui les utilisent et les revendent). Toute information au sujet d'un individu ne saurait être qualifiée de « personnelle ». Aux termes de l'article 2 de la loi n° 78-17 du 6 janvier 1978, dite « Informatique et libertés », « constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne ». En des termes proches, l'article 4 du RGPD dispose qu'est une donnée personnelle

[25] G29, « Avis 4/2007 sur le concept de données à caractère personnel », document 01248/07/FR – WP 136, 20 juin 2007.

[26] L'adresse MAC (Media Access Control) est un numéro unique identifiant une carte réseau. L'identifiant SSID (Service Set Identifier) identifie quant à lui un réseau sans fil Wi-Fi. Enfin, l'adresse IP (Internet Protocol) constitue un numéro d'identification unique attribué à chaque appareil connecté au réseau internet.

« toute information se rapportant à une personne physique identifiée ou identifiable », précisant qu' « est réputée être une "personne physique identifiable" une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale ». Il peut s'agir de l'état civil d'une personne, une image la représentant ou un enregistrement vocal, son numéro de téléphone, son adresse postale, ses informations de localisation, ses identifiants bancaires, son numéro de sécurité sociale, son empreinte, des informations relatives à sa vie personnelle, ses habitudes de consommation, ses données informatiques (adresse IP, adresse mail, pseudo de réseaux sociaux, codes d'accès) etc. Seul le traitement de ces données, c'est à dire « toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé » (article 2 de la loi du 6 janvier 1978), est soumis aux impératifs de contrôle juridique. Ensuite, selon la définition de l'article 3-I de la même loi, les responsables de traitements de données à caractère personnel sont ceux qui en déterminent les finalités et les moyens. Nul doute que les plateformes du web participatif peuvent être — et sont très souvent — de tels responsables de traitement. Elles doivent dès lors veiller à respecter les droits des internautes relatifs à leurs informations personnelles, ce qui entre de façon récurrente en contradiction avec leurs modèles économiques.

L'IA entraîne un éclatement de la notion de données personnelles. Les capacités de corrélation des algorithmes d'apprentissage obligent à étendre le champ des données personnelles bien au-delà des identifiants classiques permettant de reconnaître une personne physique. De multiples signes, individuellement inaptes à identifier un individu, mais qui, une fois associés, créent une empreinte unique, permettent de caractériser un individu. Les profils établis sur la base de ces traces permettent par agrégation et recoupement de parvenir à une identification[27]. La frontière entre données et données personnelles tend donc à se brouiller. Toute donnée brute transmise par l'ordinateur pourrait s'avérer être une trace permettant de préciser un profil. Cependant, les « données à caractère personnel » au sens de la loi Informatique et libertés et du RGPD ne sauraient être assimilées à toutes les données. Ces textes tracent un périmètre à géométrie variable dont la condition déterminante est l'existence d' « éléments spécifiques propres à l'identité » d'un individu. On peut donc s'appuyer sur le critère de l'objectif du traitement de l'information pour distinguer les données et les données personnelles : si une information est collectée dans le but de permettre l'identification d'un individu, il est fort probable qu'elle permette d'une manière ou d'une autre de l'identifier et donc que ce soit une donnée personnelle. En 2011, la CNIL pouvait déjà retenir que la collecte conjointe de données de localisation et d'une adresse MAC « permet de déterminer *in fine* la position géographique des utilisateurs du système » et qu'en conséquence « la finalité de la collecte des adresses MAC combinée aux autres informations collectées conduit […] à considérer que ces données combinées entre elles constituent des données à caractère personnel »[28]. Mais il existera toujours des collectes de données brutes qui ne seront pas des collectes de données personnelles. Quant aux données enregistrées afin de dresser des profils plus ou moins précis grâce au travail des IA, afin de suivre, retrouver ou cibler un individu, elles doivent nécessairement être protégées en tant que données à caractère personnel.

[27] J.-M. Deltorn, « La protection des données personnelles face aux algorithmes prédictifs », *RDLF* 2017, chron. n° 12.
[28] CNIL, déc. n° 2011-035.

4. Le droit des données personnelles

Face aux manipulations généralisées de données personnelles et aux enjeux qui s'attachent à leur détention, il est nécessaire que le droit intervienne. Concernant le dépôt de cookies, il s'agit d'un accès indu à un système informatique, pénalement sanctionné[29]. Il peut donner lieu à une violation de la vie privée, également sanctionnée par le droit pénal dans la mesure où les cookies débouchent nécessairement sur un traitement automatisé de données nominatives. C'est pourquoi les plateformes doivent annoncer clairement qu'elles font usage de cookies et indiquer que ceux-ci peuvent être refusés. Une section du Code pénal porte sur les « atteintes aux droits de la personne résultant des fichiers ou des traitements informatiques ». Sont ainsi réprimés le fait « de procéder ou de faire procéder à des traitements de données à caractère personnel sans qu'aient été respectées les formalités préalables à leur mise en œuvre prévues par la loi » (article 226-16) ; le fait « de collecter des données à caractère personnel par un moyen frauduleux, déloyal ou illicite » (article 226-18) ; le fait « de procéder à un traitement de données à caractère personnel concernant une personne physique malgré l'opposition de cette personne, lorsque ce traitement répond à des fins de prospection, notamment commerciale » (article 226-18-1) ; le fait de traiter des données sensibles (article 226-19) ; le fait « de conserver des données à caractère personnel au-delà de la durée prévue par la loi ou le règlement » (article 226-20) ; le fait de ne pas respecter le besoin d'une finalité déterminée du traitement (article 226-21) ; et le fait de dévoiler des informations pouvant porter atteinte à la considération de la personne ou à sa vie privée (article 226-22).

La loi, nationale ou européenne, encadre l'utilisation des données personnelles nécessaires au fonctionnement des IA. Elle encadre les conditions de collecte et de conservation des données à travers les principes de finalité, de proportionnalité, de sécurité, de limitation de la durée de conservation des données. Elle régit aussi l'exercice de leurs droits par les personnes pour protéger leur vie privée et leur liberté : droit à l'information, droit d'opposition, droit d'accès, droit de rectification. Dans la lignée de la jurisprudence de la Cour de justice de l'Union européenne visant à favoriser le développement du numérique tout en protégeant les libertés fondamentales[30], et dans la veine de la loi française pour une République numérique[31], le Règlement européen intéresse l'IA en ce qu'il encadre la collecte, la conservation et l'utilisation de son carburant : les données personnelles — même si beaucoup d'IA fonctionnent à base de données qui, n'étant pas personnelles, ne sont pas concernées par le RGPD. Il insiste dans son introduction sur le besoin de permettre aux personnes physiques « d'avoir le contrôle des données à caractère personnel les concernant ». Il a modifié et approfondi le droit de la protection des données à caractère personnel, spécialement afin de renforcer la protection des personnes dont les informations sont l'objet d'un traitement et responsabiliser les acteurs de ce traitement.

Le régime juridique réservé au traitement de ces données à caractère personnel a largement repris le droit français déjà applicable. Cependant, les nouveaux principes de protection des données consacrés ont induit un changement de logique notable : le glissement d'une logique de formalités préalables (déclaration, autorisation, avis, normes simplifiées) vers une logique de responsabilisation renforcée et de conformité continue dont les acteurs sont comptables.

---

[29] C. pén., art. L. 323.

[30] CJUE, 8 avr. 2014, *Digital Rights Ireland*. D. Simon, « La révolution numérique et le juge de l'Union : les premiers pas vers la cybercitoyenneté », *Europe* 2014, n° 7, p. 9 s.

[31] L. Cluzel-Métayer, « La loi pour une République numérique : l'écosystème de la donnée saisi par le droit », *AJDA* 2017, p. 340.

L'introduction du concept d' « accountability » a sensiblement renouvelé l'environnement juridique des entreprises. À présent, elles ne doivent plus réaliser certaines formalités préalables à la mise en œuvre d'un traitement de données personnelles, mais il leur incombe d'être en mesure de prouver, le cas échéant, qu'elles respectent le RGPD et particulièrement les grands principes suivants : finalité, proportionnalité, pertinence, durée de conservation limitée, sécurité et confidentialité. En vertu de l'article 5 du RGPD, les données personnelles doivent être :

- « traitées de manière licite, loyale et transparente au regard de la personne concernée » ;
- « collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités » ;
- « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées » ;
- « exactes et, si nécessaire, tenues à jour » ;
- « conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées » ;
- « traitées de façon à garantir une sécurité appropriée ».

Le traitement des données à travers des IA doit donc être proportionné à la finalité légitime poursuivie. À chaque étape du traitement, il convient de poursuivre un juste équilibre entre, d'une part, les intérêts visés par le développement et le déploiement de l'IA et, d'autre part, les droits et libertés à préserver. Les entreprises ou administrations doivent donc examiner si la collecte de données est véritablement nécessaire et utile. Le big data suppose la collecte et le traitement d'autant de données que possible. Comment ménager cet impératif technique avec le principe de minimisation des données issu de l'article 5 du RGPD ? Ce dernier entend en tout cas offrir aux citoyens un droit à l'autodétermination informationnelle[32].

Ensuite, les principales dispositions du RGPD permettent son application extra-territoriale (article 3), donc son application aux entreprises établies en dehors de l'Union européenne dès lors qu'elles ciblent des européens. Le règlement régit tous les traitements de données effectués sur le territoire de l'Union européenne ou visant un résident européen, peu important que l'acteur concerné soit établi en dehors de l'Union. Le RGPD devient ainsi une référence en dehors des frontières européennes. Il a d'ailleurs été repris volontairement par certains acteurs privés, à l'image de Microsoft qui a décidé d'appliquer ses règles à l'ensemble de ses clients, européens ou non. Le RGPD et, par suite, le modèle européen qu'il incarne tendent progressivement à s'imposer au-delà de l'Europe. Google a lui-aussi fait le choix d'appliquer les règles du RGPD à l'ensemble de ses utilisateurs, ce qui a conduit à la modification de sa politique de confidentialité pour l'ensemble de ses clients.

S'agissant des obligations imposées par le RGPD, un consentement « explicite » et « positif » doit être obtenu (article 4), en particulier s'agissant de l'installation de cookies durant la navigation sur le web et l'envoi de formulaires de contact.

Est imposée une obligation d'information et de protection des données personnelles traitées (sécurité physique des serveurs, des lieux et des dispositifs informatiques, contrôle de l'accès aux données (article 32)).

[32] B. Ancel, « La vie privée dans un monde digitalement connecté : la démocratie en danger ? », *RLDI* 2019, n° 159, p. 34.

Il est obligatoire de notifier à la CNIL toute violation de ces données dans les 72 heures, si la violation est susceptible d'engendrer un risque pour les droits et libertés d'une personne physique.

En vertu des articles 12 et 13, il est obligatoire de communiquer de façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples un certain nombre d'informations aux utilisateurs (les données personnelles qui sont collectées, les finalités ou buts d'utilisation des données, les destinataires des données, la base juridique des traitements, les durées de conservation etc.).

Le « droit à l'oubli » est garanti (article 17), soit le droit d'obtenir d'un responsable de traitement l'effacement, dans les meilleurs délais, de ses informations à caractère personnel.

Est consacré le droit à la portabilité des données personnelles : « les personnes concernées ont le droit de recevoir les données à caractère personnel les concernant qu'elles ont fournies à un responsable de traitement » (article 20). Concrètement, tout individu peut exercer ce droit pour migrer d'un écosystème de services à l'autre sans pour autant abandonner son historique numérique[33].

Toute personne peut s'opposer au traitement de ses données (article 21), mais aussi y accéder (article 15) et exiger leur rectification (article 16), la limitation de leur traitement (article 18) ou leur suppression (article 17).

En outre, est prohibé le profilage (article 22), il est possible de refuser qu'une décision produisant des effets juridiques soit prise par un traitement automatisé de données personnelles.

L'article 13-2-f prévoit que le responsable de traitement doit informer la personne concernée de « l'existence d'une prise de décision automatisée, y compris un profilage, et, au moins en pareils cas, [lui délivrer] les informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement ».

Il est obligatoire de prendre en considération la protection des données personnelles dès la conception des services et systèmes (article 25, en anglais : *privacy by design*). Quant au *privacy by default*, il signifie que, par défaut, seules les données à caractère personnel nécessaires eu égard à la finalité des traitements peuvent être collectées.

Il est, selon les cas, obligatoire ou possible, en tant que responsable de traitement, de désigner un délégué à la protection des données (article 37-1) et de créer et conserver un registre des traitements (article 30).

Et il devient obligatoire de réaliser une étude d'impact sur la vie privée dès lors qu'une activité est susceptible d'impacter les droits et libertés des personnes concernées (article 35), cette étude d'impact devant prévoir les moyens de réduire les dommages potentiels. Son objet est de mesurer la nécessité et la proportionnalité des opérations de traitement par rapport à leurs finalités afin d'identifier et d'atténuer le risque élevé pour les personnes concernées. La réalisation d'une telle

[33] La loi pour une République numérique va plus loin en permettant la récupération de toutes les données associées à un compte utilisateur. Son article 48 introduit un droit pour le consommateur de récupérer en toutes circonstances l'ensemble de ses données. Ce texte confère aux individus un droit à la portée plus large que le droit à la récupération consacré par le RGPD dans la mesure où il couvre l'ensemble des données et pas seulement les données personnelles. Les fournisseurs de services doivent ainsi proposer une fonctionnalité gratuite permettant la récupération de tous les fichiers qu'il a mis en ligne ainsi que de « toutes les données résultant de l'utilisation du compte d'utilisateur du consommateur et consultables en ligne par celui-ci, à l'exception de celles ayant fait l'objet d'un enrichissement significatif par le fournisseur en cause ».

analyse est obligatoire dans de nombreuses situations (traitement de données sensibles au sens de l'article 9 ou profilage de personnes sur la base de données provenant de sources externes).

Par ailleurs, un principe de responsabilité a vocation à responsabiliser les entreprises exploitant des informations personnelles. Celles-ci ne doivent plus s'adresser à une autorité de contrôle centrale afin d'obtenir une autorisation d'utilisation de données. En contrepartie, elles doivent à tout moment pouvoir justifier de leur respect du RGPD. Et des sanctions lourdes peuvent être prononcées en cas d'infraction (article 83-6) : jusqu'à 4 % du chiffre d'affaires mondial annuel ou 20 millions d'euros (le montant le plus élevé étant retenu).

Dans l'Union européenne, un groupe de travail, le G29, a été formé autour du contrôleur européen de la protection des données afin d'organiser un contrôle coordonné du respect du droit des données personnelles. Ce groupe enclenche des actions, y compris contre les multinationales du web.

5. Le RGPD et ses limites

Avec le RGPD, nous sommes entrés dans l'ère des bannières — ces petits bandeaux qui clament « Le respect de votre vie privée est notre priorité ! », alors qu'ils ne sont que le respect d'une obligation légale que l'on cherche à détourner en faisant croire qu'il s'agirait d'une initiative volontaire. Ces bannières sont un casse-tête que les webdesigners s'évertuent à faire disparaître en un clic, notamment parce qu'elles ruinent l'expérience utilisateur et la mise en valeur des contenus[34]. Le grand public se rend compte de la présence du RGPD essentiellement lorsque, en naviguant sur le web, on est sommé d'accepter les cookies ou de gérer les paramètres. Machinalement, neuf individus sur dix[35], pour gagner du temps et parce qu'ils ne comprennent que mal quels sont les enjeux, cliquent sur « Accepter les cookies ». Le réflexe est de se débarrasser du bandeau au plus vite, car il ralentit la navigation et perturbe l'accès au contenu. La situation est finalement proche de celle qui existait avant l'ère du RGPD, à la différence qu'il faut désormais cliquer cinquante fois par jour sur « J'accepte les cookies ». Cette réglementation est donc paradoxale puisque, faite pour protéger la vie privée, elle incite l'internaute à laisser capter encore plus ses données personnelles[36]. Les bannières offrent, il est vrai, un choix délicat : accepter les cookies en un clic ou bien se perdre dans un marathon du consentement en cliquant sur « Personnaliser les cookies ». La « *consent fatigue* », comme on l'a baptisée aux États-Unis, fait que les utilisateurs préfèrent très majoritairement qu'on collecte et utilise leurs données personnelles.

Par ailleurs, beaucoup d'acteurs se sont mis en conformité avec les exigences du RGPD du point de vue de la transparence : ils ont élaboré des politiques de confidentialité très complètes pour expliquer quelles données sont traitées, pour quelle finalité, combien de temps elles seront conservées etc. Mais on sait que pratiquement aucun internaute ne les consulte. Cette exigence de transparence et d'information est donc remplie en théorie mais pas en pratique. On peut y voir une certaine forme d'hypocrisie ; ou bien une solution faisant figure de pis-aller. On rapporte que si un Américain moyen devait lire toutes les politiques de confidentialité auxquelles il a souscrit, cela l'occuperait durant 25 jours[37]. Les entreprises n'ignorent pas qu'elles sont contraintes par la loi de

[34] A. Remay, « Le RGPD offre paradoxalement une immense opportunité de collecte pour les publicitaires et autres sociétés de services marketing », lemonde.fr, 15 mars 2020.

[35] *Ibid.* (Quantcast (deuxième solution la plus utilisée aux États-Unis et au Royaume-Uni) affiche en effet un taux d'acceptation de 90 %).

[36] *Ibid.*

[37] J.-G. de Ruffray, « Les (vraies) faiblesses du RGPD », lebigdata.fr, 11 mars 2020.

penser et publier ces actes et elles respectent cette obligation, mais elles savent bien que très peu d'utilisateurs les liront.

Un autre effet regrettable de l'entrée en vigueur du RGPD, heureusement passager, a été l'envoi massif de spams au moment de sa mise en place car toutes les entreprises ont dû prévenir les personnes dont elles détenaient l'adresse mail et leur demander si elles avaient le droit de leur écrire. Chaque demande d'autorisation était elle-même illégale car non sollicitée.

Le but affiché des institutions européennes est de « redonner aux citoyens le contrôle de leurs données personnelles, tout en simplifiant l'environnement réglementaire des entreprises »[38]. Sur l'un et l'autre point, des craintes sont permises, tant s'agissant de l'effectivité de la protection de ces nouveaux droits des citoyens européens que concernant la simplicité de cette nouvelle réglementation que les responsables de traitements devront appliquer. Le RGPD, comme tout règlement de l'Union européenne, est directement applicable sans besoin de transposition de la part des États membres. Néanmoins, il renvoie aux droits nationaux sur différents aspects. Et certaines dispositions nationales sont sans doute incompatibles avec le nouveau texte européen. En 2020, le Règlement général sur la protection des données demeure fort difficile à appliquer pour les entreprises[39]. Son grand défi est bien celui de son effectivité.

Reste que la collecte de données est fréquemment clandestine sur le web, ce qui oblige à utiliser avec la plus grande prudence les réseaux sociaux et autres services participatifs ; et ce qui amène à relativiser le contenu du droit des données personnelles tant celui-ci est souvent ignoré ou contourné. Il entre largement en confrontation avec les objets, pratiques, activités et modèles économiques des entreprises du numérique, qui ne peuvent donc les appliquer qu'à regret. Dès lors, le RGPD est dans une grande partie inefficace vis-à-vis des applications des pays tiers, car il est impossible de vérifier qu'il est appliqué. Les entreprises américaines ou chinoises sont forcément tentées d'ignorer ce texte, au risque d'amendes légères par rapport aux enjeux auxquelles elles sont attachées. Pour certains commentateurs, « nous avons encore enfanté un monstre juridique, nous nous sommes encore tiré une balle dans le pied avec cette loi »[40].

Le 12 décembre 2019, l'Institut Montaigne a rendu publique une étude sur le RGPD dans laquelle il déplore les faiblesses du règlement européen et plaide pour des règles différentes en fonction des secteurs, considérant que la santé, l'assurance ou la banque ne posent pas les mêmes difficultés que le e-commerce ou les jeux en ligne. On pourrait ainsi produire des régimes plus lisibles et plus concrets pour les acteurs. Sont notamment regrettés le caractère trop général du texte, la longueur des considérants ou les errements légistiques consistant à inclure des normes dans les considérants[41]. Le RGPD, insuffisamment concret, serait une directive cachée, à tel point que la France s'est sentie obligée de légiférer et de « transposer » ce texte en modifiant la loi Informatique et libertés de 1979. Les acteurs économiques, même de bonne foi et volontaires, sont souvent confrontés à des problématiques d'interprétation et de mise en œuvre pratique. Alors qu'il s'agissait surtout de lutter contre les manipulations de données opérées par les multinationales telles que les GAFAM, ce sont finalement elles qui sont les moins durement impactées, tandis que nombre de PME se retrouvent dans le flou et gênées dans leurs activités. S'agissant de la durée de conservation, par exemple, le

[38] Conseil européen, « Le règlement général sur la protection des données », consilium.europa.eu, 2018.

[39] F. Debès, « Après deux ans, des milliards investis et quelques progrès, le RGPD patine », lesechos.fr, 28 mai 2020.

[40] Th. Klein, « Déconfinement : "Nous nous sommes encore tiré une balle dans le pied avec le RGPD" », lemonde.fr, 24 avr. 2020.

[41] J.-G. de Ruffray, « Les (vraies) faiblesses du RGPD », lebigdata.fr, 11 mars 2020.

RGPD a établi une règle très générale, sur le modèle de celle figurant dans la loi française : les données ne doivent pas être conservées plus longtemps que nécessaire par rapport à la finalité pour laquelle elles ont été collectées. Plus généralement, tout le texte définit des notions vagues dans un esprit d'auto-contrôle et d'auto-compliance. Ainsi beaucoup d'acteurs ne savent-ils pas précisément ce qu'ils peuvent et ne peuvent pas faire, tandis qu'ils se savent menacés par de lourdes sanctions.

L'exploitation raisonnable des données est une opportunité pour le développement de la recherche et de l'intérêt général. Le libre accès aux données favorise la science et l'innovation ouvertes. La protection des données est alors un frein. Le RGPD doit protéger à la fois la vie privée et les données personnelles et la liberté d'entreprise, la liberté d'exploiter et vendre des données au nom des libertés économiques qui constituent l'essence de l'Union européenne. Les auteurs du RGPD ont donc recherché un subtil équilibre, quasi-impossible à trouver, et il n'est guère surprenant qu'il soit l'objet de vives critiques tant de la part des entrepreneurs que de la part des défenseurs des libertés individuelles.

6. Conceptions américaine et européenne des données personnelles

Le RGPD est aussi l'objet de commentaires positifs et même une source d'inspiration pour des pays non européens, y compris les États-Unis. L'entrée en vigueur du RGPD a donné lieu à de grandes asymétries en matière de flux de données entre les États-Unis et l'Union européenne. Berceau de beaucoup d'entreprises du numérique, qui ont plutôt intérêt à ce que les collectes, conservations et exploitations de données personnelles soient le moins encadrées possible, la Californie a adopté, le 28 juin 2018, une loi en partie inspirée du RGPD européen. Le California Consumer Privacy Act impose aux entreprises de rendre publics les types de données qu'elles collectent à travers leurs activités. Cette loi les oblige également à permettre à leurs utilisateurs de refuser que leurs données soient utilisées à des fins commerciales. Et ceux-ci peuvent exiger la suppression des informations déjà recueillies. Enfin, le texte interdit toute communication de données relatives à des enfants de moins de 16 ans. Alors que cette législation est entrée en application le 1er janvier 2020, ses conséquences pour les acteurs de la Silicon Valley sont dans tous les cas limitées puisque leurs services s'adressent à tous les américains et même au monde entier — la Californie est cependant l'État le plus peuplé des États-Unis. Sans compter que des dispositions très favorables aux entreprises ont été insérées dans la loi. Un article, par exemple, autorise celles-ci à appliquer des tarifs différents selon que l'utilisateur accepte ou non que ses informations soient récoltées et exploitées ou à ne rendre le service payant que pour les consommateurs ne souhaitant pas divulguer leurs données — ce qui est cependant parfaitement en cohérence avec le modèle économique de ces services, ordonné autour de ce qui est devenu l'adage des activités en ligne : « Quand c'est gratuit, c'est vous le produit ». C'est ainsi à un véritable jeu de loi autour des données personnelles qu'on assiste en Californie.

Aux États-Unis, la Privacy Law est sectorielle. Il n'y a pas de réglementation fédérale générale applicable à l'ensemble des secteurs d'activité[42]. Le Privacy Act de 1974 ne concerne que les traitements de données des agences fédérales, non les activités des acteurs privés. Ce sont les États et les villes qui peuvent édicter les textes relatifs à la protection des données personnelles. Dans le millefeuille réglementaire qui en résulte, l'absence de protection des données personnelles est l'impression qui domine. Ainsi, par exemple, le Personal Privacy Protection Law adopté en 1984

[42] D. Solove, P. Schwartz, *Information Privacy Law*, 6e éd., Wolters Kluwer, 2018.

par l'État de New York vise lui-aussi les données personnelles traitées par les agences de sécurité étatiques mais pas les acteurs privés. Mais les choses bougent, suivant le modèle du RGPD européen, et la Californie n'a donc pas attendu une hypothétique loi fédérale sur la protection des données aux États-Unis. Elle a adopté son propre texte, avec des intentions qui rappellent celles du RGPD. Parmi les nouvelles dispositions juridiques contraignantes pour les entreprises en même temps que rassurantes pour les consommateurs, on trouve notamment le fait que tout résident de l'État de Californie peut, s'il le souhaite, exiger un accès à toutes les données détenues par une entreprise et permettant de l'identifier. Chaque citoyen peut, à la suite de cette requête, demander quel usage est fait de ces données et éventuellement leur suppression. Il est également prévu que chacun puisse refuser explicitement que ses données soient revendues à des tiers par le biais d'un bouton dédié et que les sites sont tenus de proposer. Mais ce texte ne s'applique qu'aux entreprises dont le chiffre d'affaires excède 25 millions de dollars, qui détiennent les données de 50 000 personnes ou qui génèrent plus de 50 % de leurs revenus en vendant des données. Il écarte également de son champ d'application les traitements des données par le gouvernement fédéral, les États fédérés ou les administrations locales — il ne s'attaque donc pas aux collectes de données par les pouvoirs publics américains, lesquelles avaient entraîné quelques scandales suite aux révélations d'Edward Snowden ou Julian Assange.

Le California Consumer Privacy Act, malgré le lobbying intense dont il a été l'objet, n'a pas cédé aux exigences de la puissante Internet Association à laquelle coopèrent Amazon, Facebook et Google. Ce lobby a cependant obtenu la suppression de l'interdiction par défaut de la vente des informations personnelles des utilisateurs — c'est pourquoi les sites devront proposer un bouton permettant aux internautes de demander au cas par cas cette interdiction, ce qui rappelle le mécanisme anti-cookies européen et laisse supposer que sa portée sera limitée. La loi californienne butte par ailleurs, comme le RGPD européen, sur le problème de son effectivité. La mise en conformité des entreprises s'avère longue et parfois chaotique. En 2020, près de la moitié d'entre elles ne se seraient toujours pas adaptées aux nouvelles règles[43].

Le California Consumer Privacy Act, s'il se rapproche du RGPD, n'en abandonne pas pour autant la conception américaine qui fait des données personnelles des biens commerciaux comme les autres ou presque. Ainsi la loi sur la protection des données de Californie ne permet-elle pas aux citoyens américains de reprendre le contrôle de leurs vies privées. Elle demeure fidèle à la vision américaine dans laquelle l'approche économique est l'essentiel, même si cela implique de permettre aux GAFAM de s'introduire dans l'intimité des individus[44]. Sous l'angle économique, le traitement des données personnelles est un avantage concurrentiel pour les entreprises que le législateur californien a pris soin de ne pas trop malmener. Les entreprises peuvent ainsi fournir divers niveaux de service en fonction des données que l'individu accepte de leur transmettre. Refuser que l'on collecte ses données risque donc de diminuer la qualité des services ou même d'entraîner un refus de l'accès à ces services. Tandis que, en Europe, les données personnelles sont l'objet d'un droit fondamental des citoyens, auquel ils ne peuvent pas renoncer, les Américains se contentent de protéger des consommateurs.

La libre circulation des fichiers informatiques est par ailleurs un objectif de l'OCDE, comme en témoignent les lignes directrices adoptées le 23 octobre 1980, qui exhortent certes à la conciliation

[43] M. Chartier, « La Californie se dote de sa loi “RGPD”, en attendant une législation fédérale », lesnumeriques.com, 4 janv. 2020.

[44] A. Lazarègue, « Les Américains considèrent la donnée personnelle comme un simple bien commercialisable », lemonde.fr, 20 janv. 2020.

entre « des valeurs à la fois primordiales et antagonistes, telles que le respect de la vie privée et la libre circulation de l'information », mais insistent surtout sur le fait que les États doivent « s'efforcer de supprimer ou d'éviter de créer, au nom de la protection de la vie privée, des obstacles injustifiés aux flux transfrontières de données à caractère personnel ». En Europe, on a d'abord envisagé les données sous l'angle de leur marché et de leur commerce. L'approche initiale insistait sur leur libre circulation et la directive n° 95/46 du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation des données et la directive n° 2002/58/CE, dite « vie privée et communications électroniques », du 12 juillet 2002 trouvent leur justification dans les dispositions du traité relatives au rapprochement des législations dans le but de réaliser le marché intérieur. On voyait alors dans la liberté de circulation des données la cinquième liberté de circulation — après celles relatives aux biens, aux capitaux, aux services et aux personnes. Cependant, sans renier les objectifs économiques, l'Union européenne s'est érigée en fer de lance de la défense de la vie privée, surtout après l'entrée en vigueur du RGPD.

Par un arrêt du 6 octobre 2015, la CJUE a invalidé l'accord « Safe Harbor » conclu entre l'Union européenne et les États-Unis, autorisant et encadrant les transferts de données personnelles de l'un vers l'autre côté de l'Atlantique, où sont basés les datacenters des multinationales du web. Dans cette affaire, un étudiant en droit autrichien avait porté plainte contre Facebook qui transférait sans le consentement de ses abonnées des données vers des serveurs américains. Les juges européens, rappelant les révélations d'Edward Snowden, ont souligné combien était contradictoire l'attitude des entreprises américaines acceptant les conditions du « Safe Harbor » tout en donnant à la NSA accès aux données en leur possession dans le cadre du programme « Prism ». Et de considérer que « n'est pas limitée au strict nécessaire une réglementation qui autorise de manière généralisée la conservation de l'intégralité des données à caractère personnel de toutes les personnes dont les données ont été transférées depuis l'Union vers les Etats-Unis sans qu'aucune différenciation, limitation ou exception soit opérée en fonction de l'objectif poursuivi et sans que soit prévu un critère objectif permettant de délimiter l'accès des autorités publiques aux données et leur utilisation ultérieure à des fins précises, strictement restreintes et susceptibles de justifier l'ingérence que comportent tant l'accès que l'utilisation de ces données ». La Cour de justice a donc censuré cet accord qui, en outre, écartait les autorités nationales de protection des données et méconnaissait le droit au respect de la vie privée et au respect des données à caractère personnel garanti par la Charte des droits fondamentaux[45]. Le « Safe Harbor » comprenait des principes de protection des données à caractère personnel théoriquement suffisants et permettant aux entreprises américaines d'accéder au marché européen et donc aux données des européens. Parmi ces principes, on trouvait celui du consentement explicite pour les données sensibles ou celui du droit d'accès et de rectification — des règles qui restent très étrangères à l'approche nord-américaine. L'idée du Safe Harbor était donc de créer un niveau de protection minimal des données comme condition à leur libre circulation.

Européens et américains, très liés par leurs économies numériques, devaient pourtant trouver les moyens juridiques de concilier leurs approches largement antagonistes. Le 2 février 2016, la Commission européenne et le Département américain du commerce ont annoncé la conclusion d'un accord concernant la mise en place d'un nouveau cadre juridique pour les exportations de données personnelles de l'Union européenne vers les Etats-Unis. Le « Bouclier vie privée » ou « Privacy Shield » a ainsi remplacé le « Safe Harbor ». La secrétaire au commerce des États-Unis, Penny Pritzker, s'est réjouit de la conclusion de ce qu'elle a qualifié d' « accord historique [qui] va aider à

[45] CJUE, 6 oct. 2015, aff. C-362/14, *Maximillian Schrems c. Data Protection Commissioner*.

la croissance de l'économie numérique en garantissant que des milliers d'entreprises européennes et américaines, et des millions de particuliers, continuent à avoir accès aux services en ligne ». Il n'en demeure pas moins que cet accord, dont les GAFAM et autres multinationales du web se félicitent, continue d'être largement critiqué et même dénoncé par les défenseurs du droit au respect des données personnelles et de la vie privée. En France, la CNIL s'est rapidement fendue d'un communiqué dans lequel elle exprime son inquiétude. Elle rappelle que le G29, groupe des instances européennes de protection des données personnelles, avait le 13 avril 2016 exprimé de graves réserves face aux trop faibles progrès du « Privacy Shield » par rapport au « Safe Harbor ». Le 29 juillet 2016, le G29 a à nouveau fait part de sa position. S'il salue poliment quelques améliorations, il demeure préoccupé sur différents points, notamment concernant l'accès des pouvoirs publics aux données transférées vers les États-Unis. Pour ce qui est de la collecte en vrac de données personnelles, le G29 ne se satisfait pas des engagements pris en raison du « manque de garanties concrètes ». Par ailleurs, le G29 regrette que les informations relatives à l'application effective des principes du « Privacy Shield » soient trop rares. Ce texte a été diminué par l'attribution de pouvoirs supplémentaires à la NSA qui peut partager plus facilement les données de surveillance brutes avec plus d'une douzaine d'agences gouvernementales. Et un décret du Président Donald Trump a réduit sa portée en exigeant des agences qu'elles veillent à ce que leur politique de confidentialité exclut les personnes qui ne sont pas citoyens américains de la protection de la vie privée.

En Europe, tout responsable de traitement de données personnelles doit assurer la sécurité et la confidentialité de ces données vis-à-vis des tiers non-autorisés. En particulier, en vertu de la législation européenne, il ne peut procéder à un transfert d'informations à caractère personnel vers un Etat n'appartenant pas à l'Union européenne. Seuls font exception les pays qui disposent de réglementations nationales offrant une protection au moins équivalente à celle assurée par le droit européen. Une dizaine de pays tels que la Suisse, le Canada ou Israël ont ainsi été qualifiés de « pays adéquats ». Les États-Unis, pour leur part, ne sont pas éligibles au rang de pays adéquat puisqu'ils n'offrent guère de législation fédérale satisfaisante en matière de protection des données personnelles. Dans le même temps, à l'heure de la globalisation et d'internet, il est difficile, et même impossible, de poser le principe d'une interdiction des échanges d'informations avec le berceau des Google, Facebook et autres Twitter. C'est pourquoi, dès juillet 2000, la Commission européenne avait conclu avec le Département du commerce des États-Unis un programme d'autorégulation, purement déclaratif, incitant les entreprises et organisations y adhérant à assurer aux traitements d'informations provenant d'Europe une protection équivalente à celle accordée dans l'Union européenne. Il s'agissait du « Safe Harbor », littéralement « sphère de sécurité ».

Avec le nouveau « bouclier », différentes voies de recours, tant en Europe qu'aux Etats-Unis, sont consacrées. La Federal Trade Commission (FTC) pourra sanctionner et même exclure pour pratiques commerciales déloyales et trompeuses les entreprises participant au « Privacy Shield » qui n'en respecteraient pas les dispositions. Sous l'empire du « Safe Harbor », la FTC pouvait tout au plus demander à une société fautive de ne pas récidiver et les victimes ne pouvaient pas obtenir réparation pour la violation de leurs droits. De plus, l'accord prévoit la création d'un poste de médiateur (ombudsman) chargé de traiter les dossiers les plus sensibles. Et il comporte une innovante clause de révision annuelle censée permettre de suivre son application et, éventuellement, de l'adapter. Si la surveillance générale et inconditionnée des données personnelles des citoyens européens pratiquée par les services de renseignement américains est en principe interdite, il reste que de nombreuses incertitudes pèsent sur ce « Privacy Shield » : en premier lieu, il repose sur un engagement écrit du gouvernement américain à limiter la surveillance de masse à ce qui est «

nécessaire et proportionné ». Or les notions de « nécessité » et de « proportionnalité » ne se présentent pas nécessairement sous le même jour en Europe et aux Etats-Unis. Elles requièrent d'être interprétées et cette interprétation peut être très extensive. Sur le fond, le « Privacy Shield » repose en effet sur le même raisonnement juridique que le « Safe Harbor » : les Etats-Unis n'ont pas à modifier leurs lois fédérales et les sociétés privées s'engagent individuellement à fournir une protection « adéquate » aux données transférées depuis l'Europe. Il s'agit d'un programme d'autorégulation et déclaratif comparable à celui du « Safe Harbor ». C'est pourquoi de nombreuses voix se sont élevées pour dénoncer ce qui serait un accord purement politique très insuffisant du point de vue de la sauvegarde des droits et libertés fondamentaux des Européens, en premier lieu en ce qu'il n'exige aucune modification de la loi des États-Unis. Du manque général de clarté du texte aux imprécisions quant aux notions clés et aux incertitudes quant à l'efficacité et à l'indépendance du médiateur, en passant par les incompatibilités entre certains principes américains et leurs équivalents européens ou par l'excessive complexité des voies de recours ouvertes aux citoyens européens, les opposants au « Privacy Shield » semblent ne pas manquer d'arguments.

Le « rapport Villani » pouvait dès lors se prononcer en faveur d'une réforme du cadre international applicable aux transferts de données, soulignant combien, « s'il est indispensable de constituer en France et en Europe de véritables écosystèmes autour de la donnée nécessaire au développement de l'IA, cette condition ne doit pas, pour autant, conduire à simplifier le transfert de données hors de l'Union européenne. Ce principe dit de libre circulation des données (*free flow of data*) est réclamé de longue date – dans le cadre d'un important lobbying – par les géants américains, qui y voient un intérêt stratégique si l'on considère l'asymétrie actuelle des flux de données. Un tel dispositif, intégré aux traités de libre-échange, marquerait un recul fort en termes de souveraineté, de compétitivité et de protection des consommateurs européens »[46]. Cela porterait en effet gravement atteinte à la capacité de l'Europe de négocier de futurs cadres à la circulation des données. Pour la mission Villani, il faudrait négocier et conclure un accord plus robuste juridiquement, pour garantir la protection des données personnelles de tous les Européens, dans un cadre suffisamment stable pour les entreprises.

Par ailleurs, il est significatif que l'Union européenne a signé avec le Japon, le 17 juillet 2018, un accord de libre-échange (baptisé « Jefta », pour « Japan-EU Free Trade Agreement »). Un de ses volets porte sur les transferts de données. Ainsi les règles relatives à la protection des informations personnelles sont-elles harmonisées entre l'Union européenne et le Japon, cela sur la base du RGPD. Alors que les relations entre l'Europe et les États-Unis sont loin d'être simples, leurs philosophies relatives au besoin de protéger les données personnelles plus ou moins profondément étant difficiles à concilier — si bien que la Commission européenne a par exemple averti les États-Unis, le 26 juillet 2018, qu'elle envisageait la rupture de l'accord Privacy Shield si les Américains échouent à tenir tous leurs engagements —, le Japon est, pour les Européens, un partenaire plus compréhensif et coopératif. L'accord passé doit amener le Japon à s'aligner sur les standards offerts par l'Union européenne dans le RGPD. D'après le communiqué commun publié, devrait être créée « la plus grande zone sécurisée pour les transferts de données au monde ». L'ouverture de cette nouvelle autoroute des informations personnelles est aussi l'occasion pour le Japon et l'Union européenne de « réaffirm[er] leur engagement à créer des valeurs partagées concernant la protection des données

[46] C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 38.

personnelles, renforc[er] leur coopération et démontr[er] leur leadership en façonnant des normes mondiales basées sur un haut niveau de protection des données personnelles ».

7. Le principe d'effectivité du consentement

Confier ses données revient à donner un pouvoir fort à des entreprises. Cela est parfaitement envisageable tant qu'il y a une confiance du consommateur et une responsabilisation du commerce, et tant que le contrat et les conditions générales d'utilisation reflètent cette relation. Le principal défi que le droit des données personnelles doit relever pour permettre au droit fondamental à l'identité et à l'intégrité numériques de ne pas demeurer une pure pétition de principe est de parvenir à rendre le consentement de la personne à la collecte de ses informations personnelles réel et éclairé. Cela suppose tout d'abord de consacrer un principe de transparence numérique : chacun doit pouvoir être informé de l'existence et du contenu des données numériques à son sujet. L'exercice de la citoyenneté numérique suppose la possibilité pour tous de conserver à tout instant la maîtrise de ses informations à caractère personnel, ce qui est impossible lorsque l'information est opaque ou partielle. Cette transparence ne saurait aller sans bonne foi des acteurs qui manipulent les données, car la mauvaise foi les conduit forcément à remplacer la transparence par l'opacité, à manipuler les utilisateurs plutôt que les éclairer. Cette démarche de transparence peut être en outre un élément de compétitivité et de réputation des entreprises face à celles qui s'y refuseraient. L'information doit concerner, en plus de l'existence et du contenu des données collectées, les droits de l'individu. Le secret professionnel ayant vocation à protéger la personne, il ne saurait être utilisé à son encontre pour s'opposer à l'application du principe de transparence.

Ensuite, chacun doit pouvoir exprimer un consentement préalable à la gestion de ses données personnelles qui ne soit pas systématique et contraint mais libre et pleinement conscient. Ce n'est qu'à cette condition que le développement, l'apprentissage et l'utilisation des IA pourra être respectueux du droit à la protection de la vie privée et du droit à l'autodétermination informationnelle, donc de la souveraineté individuelle[47]. L'article 7 de la directive du 24 octobre 1995[48], tel qu'interprété par la Cour de justice[49], faisait du consentement un fondement possible parmi cinq d'un traitement automatisé de données. L'existence d'un intérêt légitime du responsable du traitement, notamment, en était un autre. Aujourd'hui, le RGPD prévoit, à son article 6, que « le traitement n'est licite que si, et dans la mesure où, au moins une des conditions suivantes est remplie : la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques (a) ; le traitement est nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles (b) ; au respect d'une obligation légale à laquelle le responsable du traitement est soumis (c) ; à la sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique (d) ; à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement (e) ; ou le traitement est nécessaire aux fins des intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère

[47] L. C. Berthet, C. Zolynski, N. Anciaux, P. Pucheral, « Contenus numériques, récupération des données et empouvoirement du consommateur », *Dalloz IP/IT* 2017 p. 29 s.

[48] Directive 95/46/CE du Parlement européen et du Conseil, 24 oct. 1995, *Relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données*.

[49] CJUE, 24 nov. 2011, aff. jointes C-468/10 et C-469/10, *ASNEF, FECEMD c. Administracion del Estad*.

personnel, notamment lorsque la personne concernée est un enfant (f) ». Le consentement n'est donc pas la seule justification des traitements de données, mais sa place est renforcée et les autres causes possibles doivent être comprises de manière restrictive[50]. Suite à l'entrée en vigueur du RGPD, beaucoup d'entreprises se sont empressées d'obtenir le consentement de leurs clients alors même que celui-ci ne leur était pas toujours nécessaire.

Le RGPD oblige à obtenir un consentement « explicite » et « positif » de l'utilisateur (article 4), spécialement s'agissant de l'installation de cookies. On ajoutera que ce consentement doit aussi être « effectif », « réel » et « éclairé ». Or une étude germano-américaine de la Ruhr-Universität Bochum et de l'University of Michigan a montré qu'au-delà des apparences et de la forme la grande majorité des utilisateurs donnent leur accord de façon machinale, sans comprendre les enjeux, et la plupart des sites ne laissent pas le choix quant à la récolte et l'utilisation des cookies[51]. Le Comité européen de la protection des données (EDPB) a publié de nouvelles lignes directrices le 4 mai 2020. Celles-ci considèrent notamment que les sites protégeant leur accès par un mur de cookies (« cookies wall ») ne sont pas compatibles avec le RGPD[52]. Si l'utilisateur peut le plus souvent paramétrer l'utilisation des cookies lorsqu'il navigue sur un site, il arrive qu'il soit obligé d'accepter en bloc la récolte de ses données et son traçage pour accéder au contenu d'une page. Pour l'EDPB, le principe cardinal est celui du consentement de l'utilisateur. Or le système des murs de cookies aboutit à forcer la main, à obliger à accepter les cookies, sous peine de ne plus pouvoir jouir des services. Pour l'organe européen, le consentement doit toujours demeurer libre, spécifique, éclairé et univoque.

Quant à la CNIL, elle a déjà souvent mis en lumière le manque de transparence et d'information des sites, au point que cela caractérise des violations du RGPD. Celui-ci, en effet, impose de communiquer de façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples diverses informations aux utilisateurs (les données personnelles qui sont collectées, les finalités ou buts d'utilisation des données, les destinataires des données, la base juridique des traitements, les durées de conservation etc.) (articles 12 et 13). Les 25 et 28 mai 2018, la CNIL a été saisie de deux plaintes collectives par les associations « None Of Your Business » et « La Quadrature du Net ». Ensemble, ces plaintes cumulaient les réclamations de près de 10 000 personnes. Il était reproché à Google le fait que les utilisateurs de smartphones et tablettes Android sont obligés d'accepter la politique de confidentialité et les conditions générales d'utilisation des services Google pour utiliser leurs appareils. Dans ces conditions, il est difficile de refuser les conditions de la multinationale. Or Google utilise ensuite les données personnelles collectées pour personnaliser la publicité. La CNIL a alors pu constater, d'une part, que les informations fournies par Google aux utilisateurs ne sont pas facilement accessibles et, d'autre part, que ces informations ne sont pas toujours claires et compréhensibles pour un utilisateur moyen. Google fait en sorte de perdre les rares utilisateurs qui seraient tentés de consulter et comprendre ses conditions d'utilisation en éparpillant dans plusieurs documents les informations pertinentes, en utilisant des boutons et des liens qu'il est nécessaire d'activer pour prendre connaissance d'informations complémentaires, donc en divisant l'accès à l'information en plusieurs étapes fastidieuses, nécessitant jusqu'à cinq ou six actions ou clics. Tel est le cas, par exemple, de l'utilisateur qui, pour savoir ce que Google fait de ses données de géolocalisation, doit : (1) consulter « les Règles de confidentialité et conditions d'utilisation » ; (2) cliquer sur « Plus d'options » ; (3) cliquer sur le lien « En savoir plus » pour que

[50] S. Merabet, *Vers un droit de l'intelligence artificielle*, th., Université d'Aix-Marseille, 2018, p. 237.

[51] « (Un)informed Consent: Studying GDPR Consent Notices in the Field », 22 oct. 2019.

[52] E. Luquet, « Les "cookies walls" jugés incompatibles avec le RGPD », clubic.com, 7 mai 2020.

soit affichée la page « Historique des positions » ; (4) retourner au document « Règles de confidentialité » et consulter la rubrique « Informations relatives à votre position géographique » pour accéder au reste de l'information ; (5) cliquer sur les liens relatifs aux différentes sources utilisées pour le géolocaliser. Pour ce qui est de l'information donnée dans la rubrique « Personnalisation des annonces », elle empêche l'utilisateur de comprendre que ses données sont collectées chaque fois qu'il utilise un service, va sur un site ou une application Google (comme Google Search, YouTube, Google Maps, Google Photos ou Google Analytics) et que ces informations sont ensuite associées pour mieux établir son profil.

Quant aux traitements de données effectués par Google, la CNIL souligne qu'ils sont particulièrement massifs et intrusifs : les données (historique de navigation web, historique d'usage des applications, carnets d'adresses, géolocalisation, etc.) sont collectées non seulement à partir de l'utilisation du téléphone, de l'utilisation des services comme Gmail et YouTube, mais aussi en fonction des activités des utilisateurs lorsqu'ils se rendent sur des sites tiers utilisant les services Google (via les cookies Google Analytics déposés sur ces sites). Ainsi Google accède-t-il à des nombreuses données qui révèlent des aspects intimes de la vie privée des individus comme les habitudes de vie, les goûts, les contacts, les opinions ou encore les déplacements, cela sur la base d'un consentement très peu conscient et éclairé. Pratiquement personne n'est en capacité de prendre connaissance et de bien comprendre les « Règles de confidentialité et conditions d'utilisation », ni de saisir l'ampleur des traitements de données auxquels il sera procédé et leurs conséquences pour la vie privée. Les finalités sont ainsi présentées de manière bien trop générale, floue et faussement positive : « proposer des services personnalisés en matière de contenu et d'annonces, assurer la sécurité des produits et services, fournir et développer des services etc. » ou « les informations que nous collectons servent à améliorer les services proposés à tous nos utilisateurs. […] Les informations que nous collectons et l'usage que nous en faisons dépendent de la manière dont vous utilisez nos services et dont vous gérez vos paramètres de confidentialité ». L'information donnée devrait permettre à l'utilisateur de se faire une idée claire du nombre et de la portée des traitements mis en œuvre. Or, autres exemples, l'information délivrée est trop floue pour pouvoir comprendre que la base juridique des publicités personnalisées est le consentement et non l'intérêt de Google, tandis que la durée de conservation de certaines données n'est pas indiquée.

Dans cet exemple fort significatif, la CNIL estime que le consentement des utilisateurs n'est pas valablement recueilli pour la personnalisation de la publicité. Elle condamne Google à une amende administrative de 50 millions d'euros ainsi qu'à une sanction complémentaire de publicité. La CNIL justifie sa sévérité par la gravité des manquements constatés, concernant des principes fondamentaux du RGPD : la transparence, l'information et le consentement. Comme dans ce cas, le consentement des utilisateurs est trop souvent extorqué en ce qu'il n'est pas suffisamment éclairé. Les informations communiquées doivent être plus accessibles, sur la forme (en n'étant pas dissimulées dans plusieurs documents et en n'obligeant pas à fouiller les sites par de multiples clics) et sur le fond (en permettant de bien comprendre l'ampleur des enregistrements de données et les conséquences de leur réutilisation). Le consentement recueilli est trop peu « spécifique » et « univoque ». Dans le cas de Google, on peut modifier certains paramètres en cliquant sur « plus d'options ». Dans ces paramètres de personnalisation du compte, l'affichage d'annonces personnalisées est pré-coché par défaut. Or le consentement doit être donné par le biais d'un acte positif (cocher une case). Comme l'a rappelé la CNIL, l'utilisateur doit consentir spécifiquement et distinctement au traitement de ses données à des fins de personnalisation de la publicité. Dans le cas de Google, l'utilisateur est invité à cocher les cases « j'accepte les conditions d'utilisation de Google

» et « j'accepte que mes informations soient utilisées telles que décrit ci-dessus et détaillées dans les règles de confidentialité ». Comme le souligne la CNIL, « un tel procédé conduit l'utilisateur à consentir en bloc, pour toutes les finalités poursuivies par Google sur la base de cet accord (personnalisation de la publicité, reconnaissance vocale etc.) ». L'utilisateur devrait pouvoir donner un consentement spécifique pour chaque finalité. De la même manière, dans un arrêt du 1er octobre 2019, la Cour de justice de l'Union européenne a jugé que le consentement que l'utilisateur d'un site web doit donner pour le placement et la consultation de cookies sur son terminal n'est pas valablement donné au moyen d'une case cochée par défaut que cet utilisateur doit décocher pour refuser de donner son consentement. Pour la Cour, « l'exigence d'une "manifestation" de volonté de la personne concernée évoque clairement un comportement actif et non pas passif. Or un consentement donné au moyen d'une case cochée par défaut n'implique pas un comportement actif de la part de l'utilisateur d'un site internet »[53].

En 2020, dans la version définitive de ses lignes directrices sur les cookies publicitaires, la CNIL considère que la simple poursuite de la navigation sur un site web ne peut plus être considérée comme une « expression valide » du consentement de l'internaute. Partant, les utilisateurs doivent consentir par un acte « positif clair », comme le fait de cliquer sur « j'accepte » dans une bannière informative. Par ailleurs, aucun traceur « non essentiel » ne pourra être déposé sur leurs terminaux. Surtout, la CNIL recommande que l'interface de recueil du consentement ne comprenne pas seulement un bouton « tout accepter » mais aussi un bouton « tout refuser ». Et l'internaute doit pouvoir retirer son consentement facilement et à tout moment. Refuser les traceurs doit être aussi aisé que de les accepter, indique la Commission[54]. S'agissant de l'information des internautes, la CNIL note que doivent leur être précisées les finalités des traceurs avant de consentir et les conséquences découlant d'une acceptation ou d'un refus de traceur. Ils doivent également être informés de l'identité de tous les acteurs utilisant des traceurs. Enfin, les organismes recourant à des traceurs doivent pouvoir, à tout moment, apporter la preuve du recueil valable du consentement. La CNIL ajoute que certains traceurs doivent être exemptés de ces règles relatives au recueil du consentement : il s'agit des traceurs destinés à l'authentification auprès d'un service, ceux destinés à garder en mémoire le contenu d'un panier d'achat sur un site marchand, certains traceurs visant à générer des statistiques de fréquentation, ou encore ceux permettant aux sites payants de limiter l'accès gratuit à un échantillon de contenu demandé par les utilisateurs.

Les services du web 2.0, spécialement les réseaux sociaux — et surtout Facebook —, sont régulièrement critiqués en raison des manipulations d'informations personnelles qu'ils opèrent. Et la justice a déjà souvent eu l'occasion de condamner ces usages, même si ces services arguent qu'ils anonymisent les données ou encore que les législations nationales ne leur seraient pas applicables. Facebook a ainsi pu être condamné en raison de l'utilisation de plug-ins et de cookies permettant de suivre les non-membres du réseau social et d'enregistrer leurs activités. Or, par définition, ces non-membres ne sauraient accepter de manière éclairée de telles pratiques. Facebook a aussi été condamné du fait de nombreux manquements à la loi Informatique et libertés et du fait du manque d'informations claires et explicites apportées aux internautes souscrivant au service, ceux-ci n'étant finalement guère plus éclairés que les non-membres[55]. Certes, les internautes acceptent les conditions générales d'utilisation lorsqu'ils s'inscrivent, mais, d'une part, ils ne se préoccupent que

[53] CJUE, 1er oct. 2019, aff. C-673/17, *Bundesverband der Verbraucherzentralen und Verbraucherverbände*.

[54] A. Vitard, « La Cnil dévoile sa nouvelle doctrine en matière de cookies », usine-digitale.fr, 1er oct. 2020.

[55] CA Paris, 2e ch., 12 févr. 2016, n° 15/08624, *Facebook Inc. c. M. X*.

rarement de leur contenu qui est volontairement rendu abstrait et complexe et, d'autre part, il s'en faut de beaucoup que ces clauses respectent toutes le droit national des données personnelles[56]. Aussi la CNIL a-t-elle dû intervenir et condamner Facebook[57]. La simple mise en demeure de se conformer aux règles applicables adressée au réseau social par la CNIL ne pouvait suffire dès lors que c'est le modèle économique même d'un tel acteur qui semble difficilement conciliable avec le droit français de la protection des données personnelles[58].

8. Le principe de finalité des traitements

Alors que le Safe Harbor, invalidé par la Cour de justice de l'Union européenne, a été remplacé par un Privacy Shield censé mieux protéger les données des européens rapatriées aux États-Unis, l'utilisation faite par les plateformes du web participatif des informations personnelles de leurs utilisateurs suscite nombre de contentieux[59]. Ceux-ci se rapportent en particulier au principe de finalité, délicat à interpréter, dont les contours sont par nature incertains. Le Big Data et l'IA fonctionnent à partir d'une exploitation des données dont il est souvent difficile de définir la finalité au moment où elles sont collectées. En effet, le processus d'apprentissage entraîne des calculs et des corrélations non anticipés, autant d'opérations risquant de réorienter les finalités du traitement[60]. Cette caractéristique ne paraît pas pouvoir être conciliée avec le principe de finalité. L'IA a besoin de grandes masses de données collectées et sauvegardées dans l'attente d'une utilisation future ou pour nourrir un algorithme dont les résultats sont en partie inattendus. Or l'article 5 du RGPD dispose que les données à caractère personnel doivent être « collectées pour des finalités déterminées, explicites et légitimes » et que les traitements ultérieurs doivent être compatibles avec ces finalités. Cet article impose donc de déterminer à l'avance, de manière spécifique, l'objectif ou les objectifs du traitement afin de le ou les porter à la connaissance des personnes concernées (articles 13 et 14) dès le stade de la collecte, comme l'indique le considérant 39 du RGPD. Cela est un gage de transparence, de loyauté et de prévisibilité du traitement et de contrôle de l'utilisation de leurs données par les individus. Il est impossible de collecter des données à caractère personnel sans finalité particulière, « au cas où », pour une finalité future non encore connue. Cette explication des finalités qui s'effectuait auparavant dans le cadre des formalités préalables auprès de la CNIL doit être opérée dans le registre des activités de traitements et surtout à travers l'information à délivrer aux utilisateurs des services, les articles 13 et 14 du RGPD énumérant, parmi les informations à fournir, une description des finalités du traitement. Le risque est alors de voir certains afficher des finalités factices, afin de respecter cette obligation, et dissimuler les finalités réelles.

L'exigence de légitimité est la plus délicate à apprécier. Elle renvoie certainement à la question de la licéité du traitement des données, donc aux articles 7 et 9 du RGPD qui listent les fondements

---

[56] *Cf.* J.-M. Bruguière, « La stigmatisation des clauses abusives dans les conditions générales d'utilisation des réseaux sociaux », *RLDC* 2015, n° 125, p. 69 s.

[57] CNIL, 27 avr. 2017, délibération n° 2017-006 prononçant une sanction pécuniaire à l'encontre des sociétés Facebook Inc. et Facebook Ireland.

[58] CNIL, 26 janv. 2016, déc. n° 2016-007 mettant en demeure les sociétés Facebook Inc. et Facebook Ireland ; CNIL, 4 févr. 2016, délibération n° 2016-026 décidant de rendre publique la mise en demeure numéro 2016-007 du 26 janvier 2016.

[59] *Cf.* B. Haftel, « Transferts transatlantiques de données personnelles : la Cour de justice invalide le Safe Harbour et consacre un principe de défiance mutuelle », *D.* 2016, p. 111 s.

[60] N. Forgo, S. Hänold, B. Schütze, *The Principle of Purpose Limitation and Big Data, in New Technology, Big Data and the Law*, Springer, 2017.

juridiques sur lesquels peuvent reposer des traitements de données : consentement de la personne concernée, exécution d'un contrat, obligation légale etc. La légitimité du traitement doit aussi se mesurer à l'aune des autres législations applicables au traitement et des droits et libertés fondamentaux comme la non-discrimination. Ainsi des finalités peuvent-elles être jugées légitimes si l'on se contentait de prendre l'article 6 du RGPD comme référence mais pas dès lors qu'on étend le champ d'investigation à l'ensemble du droit et même à l'éthique ou à la déontologie. Allant plus loin encore, le G29 affirme que la notion de légitimité suppose de se mettre à la place de la personne concernée et de vérifier si les finalités poursuivies ne vont pas dépasser ses « attentes raisonnables » par rapport au contexte initial de la collecte de ses données[61]. La légitimité des finalités signifie en effet qu'il faut prendre garde à ne pas abuser de la confiance des individus ou les prendre par surprise. Cela s'applique notamment à la transmission de données par le responsable de traitement à des tiers.

Quant aux traitements ultérieurs (les opérations sur les données subséquentes, secondaires ou nouvelles), qui doivent être compatibles avec les finalités initiales, la logique est de protéger en même temps les individus contre tout détournement des finalités des traitements, leur permettre de garder une certaine maîtrise sur leurs données, et de conserver suffisamment de flexibilité pour que les responsables de traitements puissent procéder à des opérations au-delà de la collecte des données initiale, à condition donc qu'elles présentent un lien suffisant avec la finalité première de la collecte. Selon l'article 6-4 du RGPD, les traitements ultérieurs incompatibles avec les finalités initialement prévues sont illicites, sauf s'ils reposent sur le consentement de la personne concernée ou sur une obligation légale. Le test de compatibilité qu'impose l'article 5 du RGPD est nécessaire dès lors qu'un traitement ultérieur répond à une finalité différente de la finalité initialement prévue lors de la collecte. Des finalités différentes ne sont donc pas forcément incompatibles. Ensuite, le RGPD impose que les personnes concernées soient informées du traitement ultérieur de leurs données : « Lorsqu'il a l'intention d'effectuer un traitement ultérieur des données à caractère personnel pour une finalité autre que celle pour laquelle les données à caractère personnel ont été collectées, le responsable du traitement fournit au préalable à la personne concernée des informations au sujet de cette autre finalité et toute autre information pertinente visée au paragraphe 2 ». Ces personnes doivent aussi être informées de leur droit de s'opposer à ce nouveau traitement de leurs données[62]. En pratique, toutefois, on peut douter que ces obligations soient tout le temps respectées.

Le principe de finalité est présenté depuis longtemps comme la « pierre angulaire »[63] ou la « colonne vertébrale »[64] du droit de la protection des données personnelles en ce qu'il constitue le prérequis de tous les autres principes. Il est donc au centre du RGPD comme de tous les textes visant à protéger

---

[61] G29, WP 203, « Opinion on Purpose Limitation », 2 avr. 2013, p. 4.

[62] Les articles 13 et 14 du RGPD prévoient que l'information délivrée doit porter sur la finalité du traitement ultérieur et sur toute information pertinente du paragraphe 2 des articles 13 et 14, lequel vise la durée de conservation des données, l'existence des droits des personnes concernées (dont le droit de retirer son consentement lorsque c'est la base juridique du traitement initial), l'existence du droit d'introduire une réclamation auprès d'une autorité de contrôle, les informations sur la question de savoir si l'exigence de fourniture de données à caractère personnel a un caractère réglementaire ou contractuel ou si elle conditionne la conclusion d'un contrat et si la personne concernée est tenue de fournir les données à caractère personnel, ainsi que sur les conséquences éventuelles de la non-fourniture de ces données et l'existence d'une prise de décision automatisée et les informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée.

[63] G29, WP 203, « Opinion on Purpose Limitation », 2 avr. 2013, p. 4.

[64] J. Frayssinet, *Informatique, fichiers et libertés*, Litec, 1992, n° 172.

les données à caractère personnel depuis la « Convention 108 »[65] du Conseil de l'Europe et depuis la loi Informatique et libertés du 6 janvier 1978 qui prévoyait à son article 44 la nécessité de préciser les finalités du traitement dans le cadre des formalités préalables et la sanction pénale du détournement des finalités d'un traitement de données[66]. Le principe de finalité a aussi été consacré par l'article 8 de la Charte des droits fondamentaux de l'Union européenne selon lequel « toute personne a droit à la protection des données à caractère personnel la concernant. Ces données doivent être traitées loyalement, à des fins déterminées et sur la base du consentement de la personne concernée ou en vertu d'un autre fondement légitime prévu par la loi ». Cela témoigne de l'importance du principe de finalité au cœur du droit à la protection des données personnelles. C'est à l'aune des finalités du traitement, que le responsable du traitement doit déterminer, que les autres notions, principes et obligations de la législation doivent s'analyser[67]. Ces finalités, dans chaque cas, servent de référence au moment de contrôler la licéité d'un traitement, l'adéquation, l'exactitude et la pertinence des données, le respect des principes de transparence ou de minimisation des données, la durée de conservation des données etc. Et le non-respect de l'obligation de définir les finalités du traitement entraîne le non-respect des autres principes. Le principe de finalité apparaît essentiel afin de prévenir les risques de dérives que l'usage des données personnelles fait courir à la vie privée ou à l'égalité et la non-discrimination. Le respect du principe de finalité a été caractérisé par le Conseil constitutionnel comme une garantie légale du droit à la vie privée[68]. On ne peut maîtriser physiquement des données qui, parce qu'elles sont immatérielles, sont fluides ou vaporeuses. Le principe de finalité permet alors d'imposer au responsable de traitement de limiter les usages des données à un périmètre précis afin d'en maîtriser et d'en contrôler la destination. C'est pourquoi il serait plus exact de parler de « principe de limitation des finalités », comme le fait le RGPD[69].

Les finalités doivent être déterminées et connues dès le début du traitement, dès la collecte, cela d'autant plus avec l'obligation de protection des données dès la conception (*privacy by design*) prévue par l'article 25 du RGPD. Le principe de finalité oblige les concepteurs de systèmes d'IA à envisager très en amont des traitements de données la détermination de leurs finalités. Il faut aussi songer aux finalités dérivées, susceptibles d'apparaître à mesure de l'entraînement de l'IA, et leur compatibilité avec la finalité initiale. Pour ce faire, l'article 6-4 du RGPD prévoit un test de compatibilité : il s'agit d'établir un lien suffisant avec la finalité initiale à partir de différents critères. Ainsi cet article 6-4 dispose que « le responsable du traitement, afin de déterminer si le traitement à une autre fin est compatible avec la finalité pour laquelle les données à caractère personnel ont été initialement collectées, tient compte, entre autres : a) de l'existence éventuelle d'un lien entre les

---

[65] Conseil de l'Europe, Convention 108 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, 28 janv. 1981. Ce texte a consacré formellement le principe de finalité en posant les grands principes encadrant le traitement automatisé de données personnelles et notamment celui selon lequel les données personnelles « faisant l'objet d'un traitement automatisé sont enregistrées pour des finalités déterminées et légitimes et ne sont pas utilisées de manière incompatible avec ces finalités » (art. 5).

[66] Cet article est devenu l'actuel article 226-21 du Code pénal : « Le fait, par toute personne détentrice de données à caractère personnel à l'occasion de leur enregistrement, de leur classement, de leur transmission ou de toute autre forme de traitement, de détourner ces informations de leur finalité telle que définie par la disposition législative, l'acte réglementaire ou la décision de la Commission nationale de l'informatique et des libertés autorisant le traitement automatisé, ou par les déclarations préalables à la mise en œuvre de ce traitement, est puni de cinq ans d'emprisonnement et de 300 000 euros d'amende ».

[67] F. Gaullier, « Le principe de finalité dans le RGPD : beaucoup d'ancien et un peu de nouveau », *Comm. com. électr.* 2018, n° 4.

[68] Cons. const., déc. n° 2008-562 DC, 21 févr. 2008.

[69] F. Gaullier, « Le principe de finalité dans le RGPD : beaucoup d'ancien et un peu de nouveau », *Comm. com. électr.* 2018, n° 4.

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

Le principe de finalité n'interdit donc pas le développement des IA, mais il l'encadre et le freine fortement. La notion de finalités désigne les objectifs poursuivis par un traitement de données. Pour définir une finalité, il faut déjà avoir une vision claire et globale du traitement. Il faut être capable de distinguer des finalités proches mais différentes pour un même traitement de données, car une caractérisation générale et vague des finalités n'est pas acceptable. Selon le G29, les définitions suivantes sont trop incertaines : « amélioration de l'expérience utilisateur », « offre de publicité personnalisée », « future recherche », « prospection commerciale », « sécurité informatique », « développer de nouveaux services », « à des fins de recherche », sans autres précisions[70]. Sont en revanche des finalités suffisamment précises, par exemple, les suivantes extraites de la norme simplifiée n° 46 de la CNIL consacrée à la gestion du personnel : « gestion des carrières », « mise à disposition des personnels d'outils informatiques », « gestion administrative des personnels » et « gestion de la formation des personnels ». Ne sont par ailleurs pas des finalités les opérations techniques ou les fonctionnalités logicielles comme l'enregistrement, le tri, la numérisation de documents ou la gestion de bases de données. Ce qui importe est la raison de cet enregistrement, tri, numérisation ou gestion. Toutefois, la notion de traitement et la notion de finalité sont intimement liées. Ainsi, pour tenir le registre des « activités de traitements » prévu par l'article 30 du RGPD, les traitements sont identifiés par finalité et non par opération.

La définition des finalités du traitement de données personnelles doit être précise, non ambiguë. Elle doit permettre de comprendre pourquoi on a besoin de procéder à ce traitement, quels en sont les objectifs réels. Les finalités doivent être exprimées de manière à être comprises non seulement par le responsable de traitement et son personnel, mais aussi par les autorités de contrôle et les personnes concernées, indépendamment de leurs différences culturelles ou linguistiques, de leur niveau de compréhension ou de leurs besoins spécifiques[71]. La CNIL a par exemple sanctionné Google qui avait défini les finalités poursuivies par l'ensemble des traitements de données de l'ensemble de ses services de la manière suivante : « Les données que nous collectons par le biais de nos services nous permettent de les fournir, les entretenir, les protéger et les améliorer, tout en développant de nouveaux services et en protégeant Google ainsi que nos utilisateurs ». La société ne mentionnait ainsi que la fourniture d'un service global, sans préciser les finalités des catégories de traitements opérés sur ses données. Pour la CNIL, cette présentation ne répond pas aux exigences du principe de finalité car « cette information générale ne permet pas à l'utilisateur de prendre conscience des finalités réelles, et par conséquent de l'ampleur de la collecte des données le concernant. Par conséquent, elle ne lui permet pas davantage de mesurer l'intérêt que peut revêtir, pour lui, tant la

[70] G29, WP 148, Avis 1/2008, « Sur les aspects de protection des données liées aux moteurs de recherche », 4 avr. 2008.
[71] G29, WP 203, « Opinion on Purpose Limitation », 2 avr. 2013, p. 17.

recherche d'informations complémentaires quant à la manière dont ses données sont traitées et/ou combinées que l'exercice de ses droits, afin de maîtriser l'usage de ses données »[72].

Par ailleurs, le contrôle des finalités des recueils de données peut être contrarié par le recours non aux données mais aux modèles qui en sont tirés et qui ne concernent pas directement l'individu mais un ensemble agrégé, implicitement anonyme. On ne peut alors plus établir un lien univoque entre la donnée initialement prélevée et la finalité de son utilisation[73]. Les profils de navigation, constitués à partir des données individuelles, de la liste des préférences ou des lieux fréquemment visités, pourraient dès lors être librement échangés. Cela *a fortiori* dès lors qu'on considère que le principe de finalité n'interdit pas la liberté de réutilisation statistique, que la finalité statistique est toujours présumée compatible avec la finalité du traitement[74]. Pourtant, on sait bien qu'il est possible d'identifier quelqu'un à partir d'informations individuellement anonymes et insignifiantes. C'est pourquoi il est important de combiner le principe de finalité avec le droit « de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage » (article 22 du RGPD).

9. Le principe d'intégrité des données

Une autre problématique est le besoin de protéger les internautes-éditeurs contre les fuites de données. À l'ère de la dématérialisation et des flux transfrontaliers, le risque est très élevé. D'ailleurs, conscientes de ce risque et de l'opportunité économique qui s'y attache, des compagnies d'assurance proposent de couvrir les éventuelles pertes de données. Quant au droit, il doit au maximum prévenir les failles de sécurité que les plateformes et notamment les réseaux sociaux peuvent subir ou déclencher involontairement en leur imposant de strictes obligations de surveillance et de vigilance. Dans l'actualité récente, la presse a à plusieurs reprises révélé que des réseaux sociaux avaient laissé « s'échapper » des millions de données personnelles. Et la CNIL a pu procéder à des investigations qui ont, par exemple, abouti à mettre les réseaux sociaux en demeure de garantir des paramètres par défaut protecteurs de la vie privée — cela à la suite d'un dysfonctionnement, le 24 septembre 2012, ayant provoqué la publication de messages pourtant privés puisque issus de la messagerie des utilisateurs français de Facebook.

Cependant, pour l'heure, les plateformes du web participatif échappent à toute réglementation spécifique en matière de violation des données à caractère personnel, à l'inverse des fournisseurs d'accès à internet et des opérateurs de télécommunications. Si elles ne sont pas soumises aux obligations telles que la notification en cas de violation des données traitées[75], elles sont toutefois soumises aux obligations générales incombant à tout responsable de traitement. Et l'article 17 de la directive n° 95/46 du 24 octobre 1995 *Relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données* prévoit que « le responsable du traitement doit mettre en œuvre les mesures techniques et d'organisation

[72] CNIL, délib. n° 2013-420, 3 janv. 2014, prononçant une sanction pécuniaire à l'encontre de la société Google.

[73] J.-M. Deltorn, « La protection des données personnelles face aux algorithmes prédictifs », *RDLF* 2017, chron. n° 12.

[74] La directive 95/46/CE du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, à son article 6, a prévu que les données personnelles doivent être « collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement de manière incompatible avec ces finalités. Un traitement ultérieur à des fins historiques, statistiques ou scientifiques n'est pas réputé incompatible pour autant que les États membres prévoient des garanties appropriées ».

[75] Règl. UE, 25 août 2013, n° 611/2013, *Concernant les mesures relatives à la notification des violations de données à caractère personnel en vertu de la directive n° 2002/58*.

appropriées pour protéger les données à caractère personnel contre la destruction accidentelle ou illicite, la perte accidentelle, l'altération, la diffusion ou l'accès non autorisés » ; tandis que la loi « Informatique et libertés », à son article 34, dispose que « le responsable du traitement est tenu de prendre toutes précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données et, notamment, empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès ». Quant au RGPD, il impose la notification immédiate de l'autorité nationale de protection en cas de violation et/ou de fuite de données (article 33), *a fortiori* dès lors que la garantie des droits et libertés des personnes concernées risque d'être touchée. Et les entreprises doivent signaler à l'autorité compétente et aux personnes concernées tout piratage de données à caractère personnel dans un délai de 72 heures au maximum.

En 2020, au terme d'une longue enquête du FBI et de six autres départements de police européens, le site WeLeakInfo a été interdit[76]. Celui-ci vendait de grandes masses de données personnelles amassées par le biais de brèches et de failles exploitées sur différents services en ligne. Au total, 12 milliards d'identifiants et de mots de passe étaient concernés. Ils étaient revendus à partir de deux dollars, une somme qui permettait d'accéder à des informations diverses comme des numéros de téléphone, des adresses IP etc. Toutes les données personnelles volées étaient bien organisées, dans une base régulièrement entretenue et mise à jour. La même année, Weibo, l'équivalent de Twitter en Chine, qui compte beaucoup plus d'utilisateurs, a vu un hacker dérober et mettre en vente sur le darkweb sa base de données. Les informations personnelles de 538 millions de personnes se sont ainsi retrouvées dans la nature[77]. Et d'autres affaires, comme celle qui a touché Sony et les joueurs de Playstation, illustrent les difficultés à protéger l'intégrité des bases de données, lesquelles sont aux yeux des pirates des trésors. Aujourd'hui, on ne cambriole plus les banques mais les serveurs.

Au-delà, des mesures préventives et proactives peuvent être (et devraient être) prises. Différentes plateformes ont ainsi amélioré leurs politiques de cryptage des données personnelles afin de rendre plus difficile l'accès des tiers à ces données. Et il est remarquable que Facebook a mis en place une procédure d'alerte visant à mieux prévenir les fuites de données. Toute personne qui signale un défaut dans le fonctionnement du réseau social de nature à compromettre l'intégrité des données ou à contourner la protection de leur confidentialité peut ainsi être récompensée par une prime de 500 dollars minimum. Malgré le retard du droit, les plateformes du web participatif s'efforcent donc à protéger au mieux l'intégrité des données personnelles qu'elles conservent. L'enjeu est évidemment pour elles de préserver leur image et leur réputation, sachant qu'un éventuel scandale pourrait leur porter un grave préjudice, altérer irrémédiablement la confiance les liant à leurs utilisateurs et profiter à leurs concurrentes. Quand les intérêts des plateformes et des internautes concordent, tout devient plus facile ; et l'on peut alors se passer du droit.

[76] A. Siméon, « Le FBI ferme un site qui vendait des données personnelles à partir de... 2 dollars », 01net.com, 20 janv. 2020.

[77] A. Vera, « Les données personnelles de 538 millions d'internautes en vente sur le dark web », presse-citron.net, 23 mars 2020.

B. Le droit à la vie privée numérique

1. La protection des données intimes ou sensibles

Le droit à la vie privée est consacré par le Code civil, dont l'article 9 prévoit que « chacun a droit au respect de sa vie privée ». La jurisprudence, notamment de la Cour de cassation, a notamment retenu qu'il y a une atteinte à cet article dans l'hypothèse de révélations d'informations concernant la vie sentimentale ou sexuelle[78], l'état de santé[79], les opinions religieuses[80], l'engagement syndical[81] ou encore la situation patrimoniale[82] des sujets de droit. Les atteintes à la vie privée qui peuvent être le fait de systèmes d'IA ne sont pas nouvelles. En revanche, l'univers numérique modifie l'intensité, la fréquence voire la nature des atteintes. Jusqu'à ces dernières années, les atteintes à la vie privée provenaient de la révélation d'informations sans le consentement de l'intéressé. Cela signifiait que l'information avait été rendue publique contre la volonté de la personne concernée, donc en capturant cette information de manière frauduleuse, à la manière de la photographie d'un paparazzi captant l'intimité d'une personnalité. Avec le web, énormément d'internautes livrent des informations sensibles les concernant sans bien s'en rendre compte. La fréquentation de sites tels que les réseaux sociaux peut mettre en évidence les préférences politiques, religieuses ou sexuelles d'un individu, alors pourtant qu'il souhaite les garder secrètes. Même sans révéler par des actes positifs ces informations, la seule consultation de certains sites extériorise possiblement des convictions personnelles, au risque qu'elles soient ensuite publiées ou du moins conservées.

Le droit au respect de la vie privée numérique se traduit par la protection des données personnelles. « Les données personnelles doivent être protégées pour garantir le respect de la dignité de chacun et de sa vie privée », comme l'a affirmé en 2015 la « Déclaration commune de la Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et la Commission sur les droits et devoirs sur internet de la Chambre des députés italienne »[83]. Le meilleur moyen d'assurer une protection de la vie privée efficace consiste donc à limiter les collectes et traitements de données. Seules devraient être autorisées les récoltes de données strictement nécessaires au fonctionnement du service. Un réseau social n'a pas besoin d'enregistrer les données de géolocalisation. Il n'est pas nécessaire à un service de streaming de conserver les données relatives aux heures d'écoute ou de visionnage de ses contenus. Cela existe en droit avec le principe de minimisation des données retenu par l'article 5 du RGPD : « Les données à caractère personnel doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées ». La difficulté est qu'il revient au service de déterminer la frontière entre les informations qui lui sont utiles pour assurer le service et celles qui ne le sont pas ; il le fera généralement de manière large, à son profit. Il n'est pas difficile à un réseau social d'avancer que la géolocalisation lui est nécessaire afin de proposer des contenus pertinents à ses utilisateurs. Pour les GAFAM, qui proposent tous les services possibles et imaginables, toute

[78] Cass. civ. 2e, n° 01-01.186, 24 avr. 2003.

[79] Cass. civ. 1ère, n° 86-16.185, 10 juin 1987.

[80] Cass. civ. 1ère, n° 99-10.928, 6 mars 2001.

[81] Cass. soc., n° 09-60.011, 8 juill. 2009.

[82] Cass. civ. 1ère, 12 oct. 1976.

[83] Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et la Commission sur les droits et devoirs sur internet de la Chambre des députés italienne, déclaration commune, 28 sept. 2015.

information présente, d'une manière ou d'une autre, une utilité. Facebook propose par exemple, en cas d'attentat ou de catastrophe naturelle, un service baptisé *Safety Check* qui permet, en cas de localisation à proximité des lieux, d'informer les « amis » que tout va bien. En échange de ce gadget que la plupart n'utiliseront jamais, il faut autoriser la géolocalisation.

L'article 38 de la loi Informatique et libertés dispose que « toute personne physique a le droit de s'opposer, pour des motifs légitimes, à ce que des données à caractère personnel la concernant fassent l'objet d'un traitement » et à ce qu'elles « soient utilisées à des fins de prospection, notamment commerciale ». Le droit au respect de la vie privée numérique se traduit *a fortiori* par la protection des données intimes ou sensibles, sortes de données hyper-personnelles[84]. Celles-ci font l'objet d'un régime plus strict : elles sont définies aux articles 8 de la loi du 6 janvier 1978 et 9§1 du RGPD et leur utilisation est interdite, sauf consentement exprès des personnes concernées ou intérêt public impérieux. Ainsi, les traitements relatifs à l'origine raciale ou ethnique, aux opinions politiques, convictions religieuses ou philosophiques ou à l'appartenance syndicale sont prohibés. Sont également protégées les données relatives à la vie et à la santé, les informations sur l'orientation sexuelle, le numéro de sécurité sociale (ou INSEE), ainsi que les données génétiques et les données biométriques lorsqu'elles servent à identifier une personne physique de manière unique. Déjà la « Convention 108 », dès les années 1980, interdisait le traitement informatique de données sensibles telles que les opinions politiques ou l'orientation sexuelle. En 2008, la CNIL a ainsi sanctionné d'amendes à hauteur de 15 000 euros deux sociétés commerciales qui avaient constitué des fichiers ethniques de prospection commerciale afin de proposer leurs services de « rapatriement vers le pays d'origine ».

Faisant preuve d'un équilibre et d'un pragmatisme exemplaires, dont on pourrait s'inspirer dans d'autres cas, la loi autorise toutefois ces traitements dans une série de cas limitatifs :

- en cas de consentement de la personne ;
- en matière de sécurité sociale, protection sociale ou droit du travail ;
- lorsque des intérêts vitaux sont en jeu ;
- pour les associations à finalité politique, philosophique, religieuse ou syndicale ;
- quand les données ont été rendues publiques par la personne ;
- dans le secteur de la justice ;
- si sont en jeu des « motifs d'intérêt public importants » qui doivent néanmoins respecter « l'essence du droit à la protection de la donnée » ;
- dans le secteur de la santé (médecine, etc.) ;
- pour l'archivage, la recherche scientifique ou historique etc.

Cela évite donc d'empêcher la constitution de fichiers par les associations ou les partis politiques, la création de dossiers médicaux par les hôpitaux et professionnels de santé ou les traitements statistiques opérés par des instituts tels que l'INSEE.

Les données biométriques sont de plus en plus régulièrement utilisées afin d'identifier les personnes. En Chine, la reconnaissance faciale est très développée : outre le système de crédit social qui repose sur un suivi quasi-permanent des personnes au moyen de caméras, on paye ou on se connecte à diverses applications avec son visage. On peut de la même manière être automatiquement verbalisé si on traverse la route en dehors des passages cloutés. En Europe, on se garde, pour l'heure, d'autoriser de tels systèmes, cela au nom de la préservation de la vie privée.

---

[84] Ch. Koumpli, *Les données personnelles sensibles – Contribution à l'évolution du droit fondamental à la protection des données personnelles*, th., Université Paris 1, 2019.

La législation européenne pourrait même devenir encore plus exigeante s'agissant de la protection des données sensibles et prévoir des garanties fortes lorsque les systèmes d'IA traitent des données telles que les données génétiques ou les données policières ou judiciaires, concernant des infractions, des procédures et des condamnations. De telles garanties doivent également offrir une protection contre le traitement discriminatoire ou biaisé de ces) données. Pour réduire les risques d'atteinte à la vie privée par les IA, la meilleure solution est d'intervenir au stade de la récolte des données personnelles[85]. Pour être efficace, il suffit de prohiber toute collecte de certaines données à raison de leur objet, donc des données sensibles, notamment parce qu'elles pourraient engendrer une discrimination. On peut interdire par principe la récolte de données relatives à la santé, à la confession religieuse ou à l'orientation sexuelle. La difficulté est que, même en posant une telle règle, de pareilles informations sont souvent révélées indirectement, non explicitement, simplement à travers le comportement en ligne, les « clics »[86]. Pris isolément, ceux-ci sont insignifiants et ne présentent aucun risque. Combinés entre eux, ils acquièrent un sens relativement à la vie privée de la personne, et donc y portent atteinte. Le danger de l'intelligence artificielle se situe dans sa capacité à établir des corrélations entre des données éparses.

2. Origine et avenir du droit à la vie privée

Au-delà des données sensibles ou intimes, c'est bien de la vie privée en général dont on peut se demander, à force d'être malmenée par les géants du web et par les velléités sécuritaires des États, si elle ne serait plus qu'une vue de l'esprit[87]. Aujourd'hui, le droit de mener en toute discrétion une vie retirée a-t-il encore un sens ? En pratique, il devient chaque jour un peu plus difficile de faire le choix d'une vie déconnectée ou simplement protégée de toute immixtion de tiers. Économiquement, la vie privée et les *business models* des entreprises du secteur du numérique entrent quasi-frontalement en conflit. Parfaitement respecté, le RGPD, protecteur surtout de la vie privée des individus, pourrait ruiner les entreprises européennes face à la concurrence américaine et chinoise. Toute la gratuité apparente du web dépend d'utilisations des données personnelles toujours à la limite de la l'illégalité, et souvent au-delà. Les exigences de plus en plus fortes de respect de l'intimité de chacun inquiètent forcément les acteurs de l'économie numérique. Si le législateur européen en venait à rendre obligatoire le paramétrage par défaut du navigateur, pour ne pas collecter de cookies, ce serait une révolution remettant en cause les fondements du web, cela au nom de la « *privacy* ».

Sous l'angle du droit positif, les données personnelles et la vie privée des internautes utilisateurs des services du web 2.0 sont théoriquement très bien protégées. Il s'en faut de beaucoup que cela soit également le cas dans les faits tant ces services, qui ne sont généralement pas domiciliés en France mais aux États-Unis, n'hésitent pas à prendre de grandes libertés avec les droits nationaux dès lors que ces derniers se veulent très protecteurs des informations personnelles. C'est pourquoi le respect des droits de la personnalité et de la vie privée est un long et difficile combat qui est mené en premier lieu devant les tribunaux — et par les tribunaux.

---

[85] A. Bensamoun, C. Zolynski, « Big data et privacy : comment concilier nouveaux modèles d'affaires et droits des utilisateurs ? », *LPA* 2014, n°164, p. 8 s. ; F. Rochelandet, C. Zolynski, « De la Privacy by Design à la Privacy by Using », *Réseaux* 2015, n° 1, p. 15 s.

[86] S. Merabet, *Vers un droit de l'intelligence artificielle*, th., Université d'Aix-Marseille, 2018, p. 236.

[87] K. Benyekhlef, « L'IA et nos principes de justice fondamentale », 15 févr. 2018.

Or défendre la vie privée en tant qu'objet d'un droit de l'homme numérique ne va pas de soi. L'idée même de vie privée est à la fois récente et géographiquement située. Elle a prospéré en même temps que la bourgeoisie dans les sociétés monothéistes, au travers d'un processus que le sociologue allemand Norbert Elias a appelé « la civilisation des mœurs »[88]. La consécration d'un droit au respect de cette vie privée est logiquement plus récente encore. Aux États-Unis, un article publié en 1890 par Samuel Warren et Louis Brandeis dans la *Harvard Law Review* est classiquement considéré comme l'une des dates de naissance du droit à la vie privée (« *right to privacy* »). Selon ces auteurs, tout homme devrait jouir d'un « droit à être laissé tranquille » (« *right to be left alone* »), un droit pensé en réaction au développement naissant de la photographie.

Au XXe siècle, avec la prise de pouvoir de la presse et des médias, les consécrations formelles du droit à la vie privée se sont multipliées. Ce droit se retrouve consacré tant dans plusieurs instruments internationaux, à l'instar de la Déclaration universelle des droits de l'Homme (article 12) et du Pacte international relatif aux droits civils et politiques (article 17), que dans des textes régionaux de protection des droits de l'homme, comme la Convention européenne de sauvegarde des droits de l'homme (article 8), la Charte des droits fondamentaux de l'Union européenne (article 7) ou la Convention interaméricaine des droits de l'homme (article 11). En 1998, l'UNESCO, par sa déclaration de Monaco, mettait en garde solennellement les États quant à la nécessité de protéger la vie privée et d'empêcher la diffusion de n'importe quelle information. D'autres déclarations des droits, comme la Charte africaine des droits de l'homme et des peuples, ignorent tout droit au respect de la vie privée. En droit français, ce n'est qu'en 1970 qu'un droit au respect de la vie privée a été consacré par la loi[89], à l'article 9 du Code civil selon lequel « chacun a droit au respect de sa vie privée ». En revanche, ni le Préambule de la Constitution de 1946, ni la Constitution de 1958 n'ont affirmé expressément la valeur constitutionnelle du respect dû à la vie privée. Face au manque de base textuelle, le Conseil constitutionnel a déduit un tel droit de l'article 2 de la Déclaration des droits de l'homme et du citoyen (« le but de toute association politique est la conservation des droits naturels et imprescriptibles de l'homme. Ces droits sont la liberté, la propriété, la sûreté, et la résistance à l'oppression »). Il a ainsi fait de la vie privée une composante de la liberté personnelle[90].

Auparavant, on trouve dans le secret des correspondances, auquel l'internet redonne une grande actualité, l'ancêtre du droit à la vie privée. La censure des correspondances écrites était apparue avec la création des postes par l'édit de Louis XI de 1464 créant le service régulier des postes royales et prévoyant une possibilité de contrôle par l'administration, donc une possibilité d'immixtion dans l'intimité de la personne, ce qui revient à nier cette intimité. Le contrôle des correspondances privées, en cachette, sans le dire, est depuis longtemps une obsession de tous les pouvoirs en ce que cela permet de mieux connaître les « sujets », qui se dévoilent davantage en privé qu'en public. Or ce n'est que la loi du 15 juin 1922 qui consacra le droit au secret des correspondances écrites. Ensuite, avec l'apparition des communications électroniques, ce droit a été adapté. Le droit au secret des correspondances postales et électroniques est désormais protégé par l'article L. 3-2 du Code des postes et des communications électroniques et l'article L. 241-1 du Code de la sécurité intérieure. En outre, des dispositions pénales prohibent l'interception des correspondances (articles 226-15 et 432-9 du Code pénal).

---

[88] S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 513.

[89] L. n° 70-643, 17 juill. 1970, *Tendant à renforcer la garantie des droits individuels des citoyens*.

[90] Cons. const., déc. n° 94-352 DC, 18 janv. 1995, *Sécurité* ; Cons. const., déc. n° 99-419 DC, 9 nov. 1999, *Pacs*.

3. Contenu du droit à la vie privée numérique

La protection de l'intimité de la personne amène à consacrer à son profit une « sphère secrète de vie d'où [elle] aura le pouvoir d'écarter les tiers », selon l'expression de Jean Carbonnier[91]. La vie privée est cet espace dans lequel on peut se replier, à l'abri des regards et des écoutes, tranquille. C'est un « droit de se voiler »[92]. Ce droit est évidemment largement menacé par l'utilisation des services numériques, qui sont autant d'yeux et d'oreilles qui pénètrent sans mal les domiciles des individus. C'est pourquoi il semble indispensable d'affirmer le pan numérique du droit à la vie privée. Y compris en ligne, on doit pouvoir, si on le veut, être laissé en paix et se placer à l'abri de tout enregistrement de ses données de navigation. Il s'agit d'abord de permettre à l'individu de s'opposer à toute intrusion non consentie dans sa sphère intime. Cette protection d'un droit à la tranquillité vaut aussi bien pour les éléments immatériels, comme la vie amoureuse ou l'état de santé, que pour les éléments matériels, comme le domicile ou les correspondances, qui tous peuvent être les objets d'enregistrement de données permettant de les connaître avec plus ou moins de précision.

Les systèmes d'information personnels faisant l'objet d'un accès personnalisé et sécurisé (messageries électroniques, outils mobiles, comptes ouverts sur des sites web et réseaux sociaux) font partie intégrante du domicile privé. « La forteresse d'un individu, c'est sa maison », affirmait Jean Carbonnier[93]. Les espaces personnels numériques des individus doivent donc être défendus comme des forteresses. Ils sont devenus indispensables ou presque à l'épanouissement de l'intimité de la personne. L'inviolabilité du domicile est garantie en tant que principe à valeur constitutionnelle[94]. Elle est protégée aussi bien contre les atteintes de l'autorité publique que contre les abus des puissances privées, ce qui justifie à la fois la réglementation des perquisitions faites par les autorités de police et l'incrimination de la violation du domicile d'autrui à l'aide de manœuvres, menaces, voies de fait ou contraintes[95]. Techniquement, cela peut justifier la mise en place de systèmes de cryptage et d'accès protégé par mot de passe. Le droit protège l'hospitalité du domicile, c'est-à-dire le défend contre les nuisances environnementales et les troubles du voisinage. De telles nuisances et troubles peuvent exister s'agissant du domicile numérique. Celui-ci peut être concerné, comme le domicile physique, par des formes d'ingérences matérielles ou immatérielles. Si celles-ci atteignent un certain niveau de gravité, la Cour européenne des droits de l'homme estime qu'elles entravent la jouissance du domicile, affectent le bien-être de la personne et la privent de la sorte de son droit au respect de son domicile[96]. Le domicile numérique doit donc être protégé, en tant qu'espace d'intimité, contre toute surveillance, intrusion ou gêne.

Le « droit d'être laissé en paix » concerne tous les aspects de la vie d'une personne : son image, son corps, sa vie spirituelle, ses relations amicales et affectives etc. Il s'applique aussi dans le cadre de la vie professionnelle. Tout salarié a droit au respect de sa vie privée au travail, y compris lorsqu'il utilise les dernières technologies de communication. Une abondante jurisprudence a interdit à un employeur de s'immiscer dans la vie privée de ses salariés et plus encore de les sanctionner en raison de faits appartenant à leurs vies privées. L'article L. 1121-1 du Code du travail dispose ainsi que «

---

91 J. Carbonnier, *Droit civil*, vol. 1, Puf, coll. Quadrige, 2004, p. 518.

92 S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 514.

93 J. Carbonnier, *Droit civil*, vol. 1, Puf, coll. Quadrige, 2004, p. 514

94 Cons. const., déc. n° 83- 164 DC, 29 déc. 1983, *Perquisitions fiscales*.

95 S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 521.

96 CEDH, n° 4143/02, 16 nov. 2004, *Moreno Gomez c. Espagne*.

nul ne peut apporter aux droits des personnes et aux libertés individuelles et collectives des restrictions qui ne seraient pas justifiées par la nature de la tâche à accomplir ni proportionnées au but recherché ».

Bien que plutôt récente, la notion de vie privée a évolué, surtout sous l'influence de la jurisprudence de la Cour européenne des droits de l'homme. Celle-ci a en effet fait de la vie privée une « notion large qui englobe, entre autres, des aspects de l'identité physique et sociale d'un individu, notamment le droit à l'autonomie personnelle, le droit au développement personnel et le droit d'établir et entretenir des rapports avec d'autres êtres humains et le monde extérieur »[97]. La vie privée s'est ainsi étendue et sa nouvelle dimension, une dimension sociale permettant de nouer des relations et d'opérer des choix de vie librement, est particulièrement interrogée par l'univers numérique dans lequel elle peut aussi bien se déployer qu'être bafouée et ignorée.

Certains, de manière révélatrice, renoncent au concept de « vie privée » et parlent de « désidentification » en reconnaissant que l'anonymat réel est un idéal inatteignable dans le monde de l'intelligence artificielle. Thomas Drake, ancien cadre de la NSA et lanceur d'alerte au sujet du projet de surveillance de masse Trailblazer au début des années 2010, explique : « Si vous dites : "dans mes communications électroniques, je n'ai aucune attente en matière de vie privée", alors où espérez-vous avoir une vie privée ? ». Nos vies sont aujourd'hui en grande partie des vies numériques. Il est parfaitement incompréhensible d'opposer vie numérique — qui pourrait se passer de toute protection de l'intimité — et vie physique. L'une est le reflet de l'autre. Vie privée et vie privée numérique se confondent. En 2020, nous sommes ce sur quoi nous cliquons. Les plateformes du web participatif en général et les réseaux sociaux en particulier donnent lieu à de nouvelles formes d'atteinte aux droits de la personnalité et à la vie privée, plus nombreuses et plus profondes[98]. L'article 9 du Code civil et ses équivalents semblent d'autant plus nécessaires à l'ère de l'internet, qui permet des diffusions mondiales et instantanées de contenus potentiellement nuisibles. Reste que les garanties en matière de vie privée sont défaillantes dans de trop nombreux cas. Rejoignant la question de la protection des données personnelles, celle du respect de la vie privée doit relever des défis inédits à l'heure des IA. Beaucoup d'utilisateurs de services fonctionnant à base d'IA ne savent pas, ou du moins pas dans quelle mesure et comment, que celle-ci ne fonctionne que grâce à des atteintes portées à la vie privée.

Avec le RGPD, pratiquement tout type de projet d'IA oblige à procéder à une étude d'impact sur la vie privée car il implique une utilisation massive de données, dont des données personnelles voire sensibles. Cependant, rares sont les États à avoir protégé expressément les données personnelles circulant grâce à internet. Des textes protègent les dossiers médicaux, les informations financières ou les données relatives aux mineurs, mais pas la vie privée en ligne. La régulation des réseaux est rendue délicate par le fait que l'internaute peut faire varier l'intensité du caractère privé des données révélées en ligne, en fonction des paramètres qu'il retient : informations publiques, privées, semi-publiques, accessibles seulement aux amis ou seulement aux amis des amis. Aux États-Unis, quand quelqu'un diffuse des informations sur un réseau social, son droit à la vie privée n'est plus protégé par le quatrième amendement de la Constitution. Ainsi, lorsqu'un utilisateur de Facebook permet à ses « amis » de consulter ses informations, le gouvernement peut-il également y accéder sans

[97] CEDH, n° 2346/02, 29 avr. 2002, *Pretty c. Royaume-Uni*.

[98] E. Derieux, « Réseaux sociaux et responsabilité des atteintes aux droits de la personnalité », *RLDI* 2014, n° 100, p. 77 s. ; M. Dupuis, « La vie privée à l'épreuve des réseaux sociaux », *RLDC* 2013, n° 102, p. 39 s. ; L. Marino, « Notre vie privée : des little data aux big data », *JCP G* 2012, NS 47, p. 14 s.

enfreindre le quatrième amendement[99]. De même en France et en Europe, on distingue les comptes accessibles seulement aux amis (privés) et ceux accessibles au moins aux amis des amis (publics)[100]. Cela a de grandes implications si l'on vous poursuit en raison de propos publiés en ligne : selon qu'il y avait ou non un public, il s'agissait de communication publique ou privée et les conséquences ne sont pas les mêmes. Le droit au respect de la vie privée s'épuise avec la première divulgation d'un fait à caractère privé[101]. La loi du 6 janvier 1978 prévoit, à son article 8-II, que le fait de rendre volontairement publiques des informations écarte l'interdiction d'un traitement des données. Mais la CNIL interprète restrictivement cette exception. Pour elle, la protection des données personnelles ne disparaît pas en raison de leur divulgation publique dès lors que les intéressés n'ont pas conscience de leur traitement[102]. Selon la CNIL, un internaute peut donc s'opposer à ce que ses données soient réutilisées bien qu'ils les aient volontairement diffusées grâce à un service du web participatif. La plateforme qui récolte des données personnelles est ainsi soumise à une obligation de veille consistant à vérifier régulièrement que les données dont elle se sert présentent toujours un caractère public. Celles-ci peuvent en effet avoir été reprivatisées par l'internaute — s'il les a déplacées dans la partie privée de son profil, elles perdent leur caractère public. Et il est indifférent que ces données aient été indexées au moment où elles étaient situées sur la partie ouverte du profil, car seul importe le statut des données au moment de leur traitement.

Il faudrait, comme l'article 1er de la « Déclaration préliminaire des droits de l'homme numérique » du Forum d'Avignon, déclarer que « les données personnelles en particulier numériques de tout être humain traduisent des valeurs culturelles et sa vie privée. Elles ne peuvent être réduites à une marchandise »[103]. Toute personne a droit au respect de sa vie privée numérique et au secret de ses échanges numériques. Dans son arrêt du 6 octobre 2015, invalidant l'accord « *Safe Harbor* » conclu entre l'Union européenne et le gouvernement des États-Unis, la Cour de justice de l'Union européenne a considéré que « la réglementation permettant aux autorités publiques d'accéder de manière généralisée au contenu de communications électroniques doit être considérée comme portant atteinte au contenu essentiel du droit fondamental au respect de la vie privée »[104]. Désormais, avec le nouvel accord dit « *Privacy Shield* », entré en vigueur le 1er août 2016, les principes de consentement, droit d'information, d'accès ou de rectification doivent être respectés. Les entreprises européennes qui envoient des données auprès de sociétés américaines sont tenues de vérifier que ces dernières disposent d'une certification.

Autoriser un réseau social à accéder à votre carnet d'adresses constitue une autre infraction au droit à la vie privée, en l'occurrence de toutes les personnes y figurant. Même une communication sur une plateforme qui n'utilise pas le chiffrement de bout en bout, où des interceptions sont donc possibles, peut faire craindre quelques intrusions dans la vie privée. Par ailleurs, la diffusion sur un blog d'un article relatif à la compagne d'un défunt peut être sanctionnée dès lors que les informations contenues dans ce texte portent atteinte à son intimité[105]. Il est donc indispensable de choisir les informations à utiliser et celles à garder secrètes afin de ménager au maximum la vie privée des

---

[99] B. Ancel, « La vie privée dans un monde digitalement connecté : la démocratie en danger ? », *RLDI* 2019, n° 159, p. 34.

[100] TGI Paris, 17e ch., 17 déc. 2014, *J. P. c. Edwy Plenel*.

[101] Cass. 1ère civ., 3 avr. 2002, n° 99-19.852.

[102] CNIL, délib. 1er juin 2012, n° 2012-156, *Portant avertissement à l'encontre de la société Yatedo France*.

[103] « Déclaration préliminaire des Droits de l'Homme Numérique », Forum d'Avignon, 2014.

[104] CJUE, 6 oct. 2015, aff. C-362/14, *Maximillian Schrems c. Data Protection Commissioner*.

[105] CA Nîmes, 1ère ch. civ., 10 janv. 2013, n° 12/00466.

personnes concernées. Et les articles L. 226-1 et suivants du Code pénal renforcent la protection de l'identité et de l'intimité. En particulier, l'article L. 226-19 prohibe les traitements non-autorisés d'informations nominatives. Il incrimine plus exactement le fait de conserver ou diffuser des données révélant les « origines raciales ou les opinions politiques, philosophiques ou religieuses ou les appartenances syndicales ou les mœurs des personnes ». Quant à l'article 226-1, il punit « le fait pour toute personne de porter volontairement atteinte à l'intimité de la vie privée d'autrui en captant, enregistrant ou transmettant, sans le consentement de leur auteur, des paroles prononcées à titre privé ou confidentiel ».

En outre, la loi *Pour une République numérique* du 7 octobre 2016, modifiant la loi n° 70-643 du 17 juillet 1970 *Tendant à renforcer la garantie des droits individuels des citoyens*, a institué un délit d'atteinte à l'intimité sexuelle, mesure ô combien nécessaire alors que se multiplient les diffusions, sur les plateformes de partage spécialisées, de vidéos et photographies intimes — notamment avec le phénomène du « *revenge porn* ». Et divers articles de loi interdisent de diffuser des informations touchant à une situation fiscale individuelle, à une adoption, au procès d'un mineur, au suicide d'un mineur etc. Toutes ces dispositions étant d'ordre public, les conditions générales d'utilisation des services ne sauraient poser des normes plus permissives en matière d'atteinte à la vie privée[106].

Or les exemples de technologies à base d'IA malmenant toutes ces dispositions ne manquent pas. Ainsi la start-up Fidzup a-t-elle été poursuivie parce qu'elle utilisait des SDK, c'est-à-dire des petits programmes intégrés dans des applications populaires et permettant d'enregistrer des informations relatives au profil et à la localisation de l'internaute. En juillet 2018, quatre sociétés de ciblage publicitaire de ce type (Fidzup, Teemo, SingleSpot, Vectaury) ont été mises publiquement en demeure par la CNIL, avant de trouver un accord avec le régulateur. Selon la CNIL, la mise en demeure de ces sociétés s'imposait « pour alerter les millions de personnes dont les données étaient collectées et traitées à leur insu. […] Un écosystème était en train de se construire sur la base de telles pratiques, il est apparu nécessaire d'envoyer rapidement une alerte collective pour toutes les entreprises susceptibles de les mettre en œuvre ». Mais Fidzup a fait faillite. Et ses responsables ont accusé la CNIL de défendre excessivement la vie privée des Français, au point de nuire gravement à l'économie numérique nationale. En France, on préfère donc interdire à une telle société d'exploiter, comme elle le faisait, les données d'utilisateurs de smartphones, sans accord clair de leur part, pour afficher des publicités ciblées pour des magasins qui se trouvaient près d'eux. En tant que droit de l'homme numérique, on ne se plaindra pas que le droit à la vie privée puisse au moins rivaliser avec la liberté d'entreprise.

L'affaire Cambridge Analytica a montré combien le monde numérique peut être liberticide du point de vue de la vie privée. Cette firme britannique avait utilisé les données personnelles de dizaines de millions d'utilisateurs de Facebook à leur insu, pour influencer l'issue de l'élection présidentielle américaine de 2016, en faveur de Donald Trump. À des fins de manipulation de l'électorat, il s'agissait de connaître le plus précisément possible les opinions politiques des citoyens américains. En 2020, le commissaire à la protection de la vie privée du Canada a demandé, le 6 février, à un tribunal fédéral de déclarer que Facebook a enfreint les lois canadiennes sur la protection de la vie privée. Le Canada vient ainsi s'ajouter à la longue liste des pays qui ont intenté des actions en justice contre le réseau social de la Silicon Valley à la suite du scandale Cambridge Analytica — dont en premier lieu les États-Unis eux-mêmes, où le réseau a été condamné à une amende record de 5 milliards de dollars pour ne pas avoir protégé les données personnelles de ses utilisateurs. Selon son

[106] Cass. 1ère civ., 15 janv. 2015, n° 13-25.634.

communiqué, le commissaire canadien « exige que Facebook mette en place des mesures efficaces, précises et facilement accessibles pour obtenir le consentement valable de tous les utilisateurs et s'assurer de le conserver ». Il demande également que le réseau social se voit interdire « de continuer à recueillir, à utiliser et à communiquer les renseignements personnels des utilisateurs » en violation des lois canadiennes. Cela fait suite à une enquête qui a révélé de graves lacunes dans les pratiques de Facebook en matière de traitement des renseignements personnels. Pour Facebook, la vie privée n'a pas lieu d'être. Tout doit être visible, car sa prospérité en dépend. Heureusement, partout dans le monde, des pouvoirs publics se chargent de faire respecter le droit à la vie privée contre ces forces pour lesquelles la transparence devrait prévaloir.

Peut-être est-ce la Commission de la protection de la vie privée belge (CPVP, équivalent de la CNIL française) qui a le plus strictement surveillé les activités et les procédés de l'entreprise californienne. En tout cas, cette commission a engagé un bras de fer autour du cookie « Datr » et de la pratique consistant à pister tous les internautes, y compris ceux qui sont déconnectés du service et même ceux qui ne sont pas inscrits, qui ne possèdent pas de comptes personnels. Facebook ne se conformant pas aux recommandations de l'autorité belge, celle-ci l'a attaqué en justice, laquelle a lourdement sanctionné la société américaine. Devant le refus de coopérer de la société californienne, la « CNIL belge » l'a assignée en justice, en juin 2015, dénonçant un traçage « invasif » par Facebook des habitudes de navigation des internautes forcément incompatible avec le droit des données personnelles et, en premier lieu, avec la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel. Ce n'est toutefois que le problème du suivi des activités en ligne des personnes non inscrites au réseau social qui a focalisé toute l'attention et qui a conduit à la lourde condamnation de Facebook par la justice belge. Les juges ont donc donné raison à la CPVP selon laquelle il n'est pas légal de « collecter systématiquement des données relatives à la consultation de sites internet externes mais qui contiennent des modules sociaux […], alors même que l'internaute n'interagit pas avec ces modules sociaux ». Le 9 novembre 2015, le tribunal de première instance néerlandophone de Bruxelles a suivi les griefs de la CPVP et a imposé à Facebook Inc., Facebook Belgium SPRL et Facebook Ireland Limited de cesser de déposer le cookie « Datr » lorsque des internautes non-inscrits visitent le site <facebook.com> sans les en informer au préalable et de manière claire. Facebook devra à l'avenir demander expressément l'accord des internautes belges non membres du réseau social et leur fournir des explications suffisantes quant à l'usage fait des données recueillies. Or il est évidemment difficile pour un service d'obtenir quelque autorisation expresse de la part de personnes qui ne sont pas inscrites à ce service. Les juges ont donc requis de Facebook qu'il arrête de collecter des informations personnelles relativement à ses « utilisateurs passifs » par l'intermédiaire de plug-ins placés sur des sites web tiers.

4. Le droit à l'image

La question du droit à l'image est significative. L'explosion des réseaux sociaux s'est accompagnée d'innombrables atteintes à ce droit, commises par d'autres individus, souvent des amis ou des membres de la famille, parfois des ennemis cherchant à porter un préjudice à celui ou celle dont on publie l'image[107]. Avec le web 2.0, s'est développée toute une économie de l'image numérique qui évolue très en marge du droit à la vie privée. Le droit à l'image, qui découle de l'article 9 du Code

[107] M. Doueihi, *Pour un humanisme numérique*, Le Seuil, 2011, p. 67.

civil, interdit de diffuser en ligne la représentation d'une personne sans son autorisation expresse — sauf exceptions : personnalités publiques ou personnes exerçant dans le cadre de leurs activités professionnelles, illustration d'un évènement d'actualité immédiate ou d'un sujet d'intérêt général, groupe de personnes dans un lieu public. Il a ainsi été jugé que la diffusion sur un blog ou un réseau social d'un selfie représentant un homme politique en train de dormir dans un avion porte atteinte à son droit à l'image[108]. Lorsque la personne se trouve dans un lieu privé, l'article 226-1 du Code pénal réprime le fait de porter volontairement atteinte à l'intimité de sa vie privée en fixant, enregistrant ou transmettant, sans son consentement, son image.

Le principe de spécialité implique de demander une autorisation *ad hoc* chaque fois qu'une nouvelle publication de l'image d'une personne est envisagée[109]. Que cette image ait été une première fois téléchargée sur un blog ou un autre service ne signifie donc pas qu'il serait possible de la reproduire sans obtenir l'accord de la personne concernée. Les images ne deviennent pas de libre parcours une fois en ligne. Les tribunaux ont en particulier insisté sur l'impossibilité de reproduire librement sur des réseaux sociaux des photographies représentant des mineurs, quoi qu'en disent leurs conditions générales d'utilisation[110].

Au-delà de la publication de la photographie de quelqu'un, le simple fait de « taguer » des amis sur des photos publiées sur Facebook où ils apparaissent porte atteinte à leur vie privée. En la matière, l'État de l'Illinois a obligé Facebook à verser 550 millions de dollars aux parties impliquées dans une action de groupe visant à dénoncer son utilisation frauduleuse du logiciel Tag Suggest — un programme de reconnaissance faciale qui invite à « taguer » ses connaissances sur les photos du réseau social, cela loin de tout consentement des personnes concernées[111]. Cela constitue une nouvelle victoire pour les défenseurs de la vie privée. La reconnaissance faciale renouvelle évidemment en profondeur le droit à l'image. Une technologie qui analyse en détails les photographies postées par des utilisateurs afin d'identifier les personnes qui s'y trouvent est une manipulation de données sensibles — d'autant plus dans l'Illinois, qui possède les règles les plus strictes en matière de protection des données biométriques aux États-Unis. Depuis 2010, année de lancement de Tag Suggestion, Facebook doit redoubler d'ingéniosité pour utiliser ce service sans outrepasser les lois en vigueur dans les différents pays où l'entreprise est présente. Initialement activée par défaut au début des années 2010, Tag Suggestion avait été condamnée par l'Union européenne, aboutissant à sa disparition pure et simple en 2012 pour les pays européens. Ce n'est qu'en 2018 que cette technologie, revue et corrigée, est réapparue en Europe, cette fois en étant optionnelle et désactivée par défaut.

Par ailleurs, loin du système de crédit social reposant sur l'espionnage permanent de la population, on voit de plus en plus de remises en cause de la vidéosurveillance-vidéoprotection ou, du moins, de l'analyse des images ainsi récoltées par des IA, cela afin d'éviter des biais ou simplement pour protéger la vie privée des individus. Ainsi, aux États-Unis, en 2020, la ville de Boston a-t-elle interdit la reconnaissance faciale. Cette mesure vise l'utilisation de cette technologie par tout employé municipal, qui ne peut pas non plus demander à une partie tierce de s'en servir[112]. Boston est devenue la deuxième plus grosse ville américaine à interdire l'usage de cette technologie,

---

[108] TGI Paris, réf., 10 févr. 2016, *J.-M. Le Pen c. B. Zaibat*.

[109] C. civ., art. 371 s.

[110] CA Versailles, 2e ch., 25 juin 2015, n° 13/08349.

[111] P. Crochart, « Reconnaissance faciale : Facebook accepte de payer 550 M$ pour régler un contentieux», clubic.com, 30 janv. 2020.

[112] G. Renouard, « La ville de Boston interdit la reconnaissance faciale », clubic.com, 25 juin 2020.

derrière San Francisco. D'autres villes ont pris la même décision : Oakland, en Californie, ou Cambridge, dans le Massachusetts. Et trois états américains (la Californie, l'Oregon et le New Hampshire) ont interdit l'usage de la reconnaissance faciale par les caméras de police. En janvier 2020, l'Union européenne a pour sa part fait savoir qu'elle envisageait d'interdire son usage dans les espaces publics pour une durée de cinq ans, avant de se raviser et de finalement proposer l'instauration de règles encadrant son utilisation.

C. Le droit à l'honneur numérique

1. Le droit à la réputation numérique

Aujourd'hui, c'est davantage dans le monde numérique que dans le monde physique que la dignité des personnes peut être malmenée. Il faut alors affirmer que la dignité numérique est un droit fondamental. Celle-ci, tout d'abord, s'oppose à la rationalité qui tend à faire de chacun de nos faits et gestes l'objet d'une transaction marchande et à inclure tout instant de la vie dans un marché toujours plus étendu. La dignité numérique, c'est le droit de se placer à l'abri de ces ambitions intégrales et de rendre l'acte de consommation facultatif, non obligatoire, et surtout demeurer soi-même hors du marché, n'être vendu qu'avec notre accord éclairé. « À l'opposé d'une rationalité qui s'obstine à réduire tout élément ou chacun de nos gestes à des codes, procédant d'un misérable réductionnisme devant dorénavant régir le rapport au réel, écrit Éric Sadin, nous comptons plus que jamais user des pouvoirs offerts par notre sensibilité, seuls à même de nous mettre pleinement en contact avec les palpitations les plus indéfinissables de la vie »[113].

Ensuite, la protection de l'honneur des personnes est particulièrement concernée par les nouvelles technologies de communication. Le respect de la considération sociale explique l'incrimination pénale de la diffamation et de l'injure, que le web et particulièrement les réseaux sociaux facilitent. Si la critique est libre, l'honneur est protégé[114]. La loi du 29 juillet 1881 définit, à son article 29, la diffamation comme « toute allégation ou imputation d'un fait qui porte atteinte à l'honneur ou à la considération de la personne ou du corps auquel le fait est imputé ». Le même article précise que constitue une injure « toute expression outrageante, terme de mépris ou invective qui ne renferme l'imputation d'aucun fait ». La frontière entre injure et diffamation est parfois difficile à établir : la distinction repose sur l'imputation d'un fait précis, vérifiable, avec la possibilité d'en rapporter la preuve. Injure et diffamation peuvent aisément être commises grâce aux services du web participatif. Et les algorithmes qui sélectionnent les contenus à mettre en avant risquent fort d'avoir un faible pour les diffamations et injures, qui plus que des informations objectives sont susceptibles de générer de l'intérêt de la part du public.

La protection de l'enfance justifie une vigilance particulière. Le respect de la dignité humaine numérique interdit les injures, diffamations et outrages en ligne. Il défend de porter atteinte à l'ordre public et aux bonnes mœurs. Les IA devraient dès lors non mettre en avant ces contenus illicites mais les trier et les supprimer. Il faut en particulier prendre garde aux contenus qui portent atteinte à l'e-réputation. Ceux-ci peuvent faire des ravages. Des « *fake news* personnelles » aux vidéos de

[113] É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'échappée, coll. Pour en finir avec, 2018, p. 249.

[114] B. Beignier, *L'honneur et le droit*, LGDJ, 1995.

« *revenge porn* », les occasions ne manquent pas de nuire aux personnes grâce aux nouveaux moyens de communication. Trop de personnes sont victimes de « mauvais buzz ». Le droit de la presse, avec toutes ses infractions limitant la liberté d'expression, s'applique au web et aux contenus publiés par tout internaute. Internet ne doit pas permettre de diffuser des contenus qui seraient interdits à la radio ou à la télévision. Dans l'univers numérique comme dans l'espace physique, les hommes ont le droit à la dignité, qui est la première des limites à la liberté d'expression. Aux États-Unis, y compris les informations les plus sensibles (policières ou judiciaires par exemple) peuvent être publiées et circuler, sans que les personnes concernées ne puissent s'y opposer. Des sites web reproduisent ainsi les mugshots (photos d'identité judiciaire) des personnes qui, une fois dans leurs vies, ont été soupçonnées d'avoir commis un délit. C'est pourquoi la notion d'e-réputation a prospéré, en même temps que les entreprises spécialisées dans l'enfouissement ou le remplacement des informations négatives circulant en ligne au sujet des individus.

Les hommes doivent être protégés contre le « bourdonnement », qui, surtout sur les réseaux sociaux, permet de répandre progressivement de fausses informations au sujet d'une personne[115]. Or il est difficile de lutter juridiquement contre ce phénomène, même si la calomnie et la diffamation peuvent être sanctionnées, car l'identification de l'individu ou des individus ayant lancé en premier lieu la rumeur est le plus souvent impossible. Des poursuites contre les auteurs de fausses nouvelles peuvent aussi être engagées sur le terrain de la diffamation. L'article 29 de la loi réprime « toute allégation ou imputation d'un fait qui porte atteinte à l'honneur ou à la considération de la personne ou du corps auquel le fait est imputé ». Il arrive souvent que tel soit le cas avec les fausses informations. D'ailleurs, les poursuites pour fausses nouvelles se doublent souvent de poursuites en diffamation. Et on a pu parler d' « *ingeniering* procédural » concernant des poursuites combinées, diffusion de fausses nouvelles et diffamation[116]. S'agissant de fausses informations visant, avec une dimension déshonorante, une personne précise (ou un groupe de personnes à condition qu'il soit recevable à agir), une action sur le fondement de la diffamation est envisageable[117]. La preuve de la « vérité des faits diffamatoires » est, selon l'article 35 de la loi, le seul justificatif qui permet à la personne poursuivie d'échapper à la condamnation. La pratique jurisprudentielle y a ajouté la condition, plus souple et moins exigeante, de la bonne foi. Celle-ci implique cependant, entre autres, le sérieux et la vérification des sources. Cela exclut que le bénéfice de la bonne foi soit accordé face à une information manifestement erronée.

Dès lors, « l'infraction de diffamation est tout à fait apte à sanctionner n'importe quelle fausse information de nature à porter atteinte à l'honneur et à la considération d'une personne, ce qui est généralement le cas d'une "*fake news*" »[118]. Des journalistes professionnels ne doivent pas franchir la limite de la mauvaise foi et de l'imprudence professionnelle ; et les tribunaux ne peuvent que sanctionner au titre de la diffamation envers des particuliers visée par les articles 32, alinéa 1, 23, alinéa 1, 29, alinéa 1, et 42 de la loi du 29 juillet 1881 les constructions diffamatoires par l'accumulation convergente d'imputations diffamatoires afin d'accréditer une thèse dubitative particulièrement subjective. Or le choix de présentation, l'angle d'interprétation, l'absence de

---

[115] Ch. Caron, « De la calomnie au bourdonnement sur les réseaux sociaux », *Comm. com. électr.* 2016, n° 12, p. 1 s.

[116] C. Lienhard, « Catastrophe, diffusion de fausses nouvelles et diffamation », *D.* 2002, p. 2972. Une exception de nullité soulevée par la défense, fondée sur la double qualification des faits, a été écartée par le tribunal selon qui « la loi n'interdit pas de donner à un fait unique plusieurs qualifications lorsque ces qualifications ne sont pas inconciliables entre elles » (trib. corr. Toulouse, 27 juin 2002).

[117] *Cf.* G. Sauvage, « Quel(s) outil(s) juridique(s) contre la diffusion de "fake news" ? », *Légip.* 2017, p. 428 s.

[118] Ch. Bigot, « Légiférer sur les fausses informations en ligne, un projet inutile et dangereux », *D.* 2018, p. 344.

précaution, notamment dans la désignation nominative, peuvent servir tant une fausse nouvelle qu'une diffamation[119].

2. Le droit à l'oubli numérique

Tout comme le corps, l'esprit et le cœur doivent être protégés. L'intimité concerne aussi les idées et les pensées. La Cour européenne des droits de l'homme a ainsi affirmé que l'article 8 de la Convention confère aux hommes un droit à la protection de la réputation[120]. Ainsi, les opinions politiques d'un citoyen, qui sont protégées par le secret du vote, ne peuvent-elles être divulguées sans son consentement. Les convictions religieuses ou l'appartenance à des cercles sociaux (comme la franc-maçonnerie) sont de la même manière des informations qui relèvent de la vie privée des personnes.

Si de telles informations ont cependant été révélées à l'initiative d'un individu, la dignité et l'honneur numériques justifient le droit à l'oubli. Celui-ci est indispensable pour protéger la personnalité numérique. Déjà en 2001 Jean Frayssinet proposait d' « étudier la mise en place d'un droit à la tranquillité du consommateur ou droit à l'oubli numérique qui doit pouvoir se défendre efficacement contre le harcèlement informationnel qui prend des formes de plus en plus sophistiquées »[121]. En premier lieu, la question de la possibilité de faire valoir son droit à l'oubli numérique se pose à l'égard des réseaux sociaux[122]. Ce droit était déjà l'une des principales recommandations du rapport d'information sur la vie privée à l'heure des mémoires numériques rendu public le 27 mai 2009. Le droit à l'oubli est d'ores et déjà consacré explicitement par l'article 17 du Règlement général sur la protection des données : « La personne concernée a le droit d'obtenir du responsable du traitement l'effacement dans les meilleurs délais de données à caractère personnel la concernant et le responsable du traitement à l'obligation d'effacer ces données à caractère personnel dans les meilleurs délais, lorsque l'un des motifs suivants s'applique : les données à caractère personnel ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées ou traitées d'une autre manière ». Ce droit signifie donc que toute personne doit pouvoir exiger le retrait de toute information ou autre contenu la concernant accessible en ligne, cela sans besoin de motiver cette demande. On évoque même la possibilité pour des entreprises, des personnes morales, de profiter d'un « droit à l'oubli économique »[123]. Le droit à l'oubli ne peut toutefois s'exercer qu'à condition de ne pas porter un préjudice grave aux personnes physiques ou morales détenant les informations ou les contenus en cause et de ne pas nuire au devoir d'information générale, la conservation de données pouvant se justifier par des traitements à des fins historiques, statistiques ou scientifiques[124]. Ce droit va donc de pair avec un devoir de loyauté : ne pas porter préjudice aux personnes physiques ou morales détenant les données.

---

[119] C. Lienhard, « Catastrophe, diffusion de fausses nouvelles et diffamation », *D.* 2002, p. 2972 s. ; J.-L. Matheu, « Halte à la désinformation des victimes d'AZF : pas de faire-savoir sans savoir-faire », *Gaz. Pal.* 2002, p. 781 s.

[120] CEDH, 7 févr. 2012, n° 39954/08, *Axel Springer c. Allemagne*.

[121] J. Frayssinet, « Nouvelles technologies de l'information et de la communication et protection des libertés des consommateurs », in J. Frayssinet et *alii*, *L'individu à l'épreuve des NTIC*, PUL, 2001, p. 31.

[122] M. Boizard, « Facebook forever, les réseaux sociaux peuvent-ils être contraints de nous oublier ? », *Comm. com. électr.* 2015, n° 4, p. 10 s.

[123] H. Oberdorff, *Droits de l'homme et libertés fondamentales*, 7e éd., LGDJ, coll. Manuel, 2019, p. 430.

[124] A. Auger, « L'Union européenne et le droit à l'oubli sur internet », *RDP* 2016, p. 1841.

La Cour de justice de l'Union européenne a consacré ce droit à l'oubli numérique dès 2014. Pour les juges européens, la personne titulaire des données peut exiger de la part de l'exploitant d'un moteur de recherche la suppression des liens vers des pages web contenant des informations la concernant, et cela sans qu'il soit besoin de rechercher si la présence de ces informations en ligne cause ou non un préjudice à la personne concernée[125]. M. Costeja Gonzalez s'était plaint devant les autorités espagnoles de ce qu'une recherche de son nom sur Google révélait une mesure de saisie et de vente aux enchères de ses biens dont il avait fait l'objet des années plus tôt. Il estimait que ce référencement lui causait préjudice et n'était plus justifié, du fait de la clôture de la procédure. Répondant à une question préjudicielle de la justice espagnole, la CJUE a rendu cet arrêt important obligeant Google, sur le fondement des articles 7 (droit au respect de la vie privée) et 8 (droit à la protection des données à caractère personnel) de la Charte des droits fondamentaux de l'UE, à faire droit aux demandes de déréférencement dont l'entreprise serait saisie.

Suite à cet arrêt, Google a créé un formulaire à l'attention des citoyens européens souhaitant obtenir le retrait de résultats de recherche jugés inappropriés. Les plateformes du web participatif, et en premier lieu les réseaux sociaux, seraient bien avisés de prendre une même initiative. Pour la seule année 2014, Google a été saisi de 170 000 demandes de déréférencement ; et il en a accepté 50 % environ. Le droit au déréférencement ne valait en effet qu'à un certain nombre de conditions — concernant tant la personne (seules les personnes physiques sont concernées, et pour autant qu'elles ne soient pas des personnalités médiatiques, auquel cas leur droit à l'oubli est moindre) que le contenu en cause (les contenus professionnels ont moins vocation à être déréférencés que des contenus sensibles, relatifs à l'orientation sexuelle d'une personne par exemple) et le contexte (diffusion à l'initiative ou à l'insu de l'intéressé). À présent, avec l'article 17 du RGPD, il n'est plus nécessaire de justifier par un motif légitime les demandes de déréférencement. Par ailleurs, la demande de déréférencement doit viser des liens identifiés et signalés, elle ne peut pas consister en une demande générale de suppression de tous les liens apparaissant au terme d'une recherche[126]. Dans le cadre de leurs publications de contenus en ligne, les internautes peuvent dévoiler certaines de leurs informations personnelles. Celles-ci deviennent alors publiques à leur initiative. Dans ce cas, le traitement qui peut en être fait n'est néanmoins pas entièrement libre puisque les utilisateurs des plateformes bénéficient d'un droit d'opposition que lesdites plateformes doivent respecter en leur offrant les moyens de le mettre en œuvre. Et, dès lors que tel est le cas, les éléments comportant des informations privées doivent être corrigés afin de les faire disparaître.

Le droit à l'oubli consiste à imposer aux détenteurs de données personnelles de ne pas les conserver au-delà de la finalité d'origine. Face à la capacité quasi-infinie de la mémoire numérique, il s'agit de protéger l'individu par rapport à son passé. Chacun doit pouvoir être son propre archiviste, pouvoir déterminer les éléments de son histoire et de ses activités passées rendus publics et ceux qui ne le sont pas. On ne saurait accepter qu'il existe un casier numérique parfois plus pénalisant qu'un casier judiciaire — qui prévoit pour sa part un effacement périodique des données — pour les personnes cherchant un emploi, par exemple, ou souhaitant simplement avoir une vie sociale non impactée par des actes ou pensées passés. Avec internet, est apparu le problème de l'imprescriptibilité des données. Il semble difficile de croire qu'une mise en ligne d'informations puisse être réversible. Il est impossible de maîtriser, à l'échelle mondiale qui est celle du web, le sort réservé aux données personnelles une fois qu'elles ont été rendues publiques face aux capacités des serveurs et à la

[125] CJUE, 13 mai 2014, aff. C-131/12, *Google Spain*.

[126] CA Aix-en-Provence, 15 sept. 2016, n° 2016/842.

puissance des robots de recherche et d'extraction d'informations. Les services du web participatif présentent des dimensions internationales et une vocation universelle qui s'opposent au caractère essentiellement national ou régional du droit des données personnelles. Aussi est-il difficile d'obtenir le retrait d'un contenu au nom du droit à l'oubli, y compris en exécution d'une décision de justice. On peut mettre en ligne un contenu puis, plus tard, regretter cette publication et vouloir la supprimer. C'est alors un véritable parcours du combattant. Qu'il s'agisse d'une photographie de soirée alcoolisée, d'une pose trop sexy ou d'un article ou commentaire trop véhément ou exprimant une opinion avec laquelle on n'est plus d'accord, difficile d'en obtenir la disparition.

L'effectivité de ce droit à l'oubli est difficile à garantir. Il peut être fort compliqué de contraindre des services situés à l'autre bout du monde, lesquels peuvent mettre de la mauvaise foi dans leur respect de règles qui leur coûte cher. Au-delà, l'application du droit à l'oubli est délicate s'agissant des données dérivées, c'est-à-dire des informations construites à partir de données primaires concernant un individu. En résultent de nouveaux modèles qui contiennent indirectement une image de l'individu. Il peut donc être compliqué de faire valoir son droit d'opposition auprès de services du web participatif. Ce droit n'en paraît pas moins indispensable tant les consentements aux traitements de données peuvent être viciés ou simplement influencés par des incitations commerciales ou une information inexacte ou insuffisante. Il faut une grande maturité et une certaine clairvoyance pour refuser de s'inscrire à un service afin de protéger ses informations personnelles, ce dont beaucoup d'utilisateurs des réseaux sociaux, des forums, des blogs ou des plateformes de partage ne disposent pas — *a fortiori* lorsqu'ils sont mineurs. Or cela est d'autant plus essentiel que le droit à l'oubli numérique fait apparaître d'autres droits insoupçonnés dans une société où le souvenir est érigé en devoir : le droit de repentir, le droit au regret, le droit à la rédemption, le droit au déni, le droit à la mémoire, le droit à la suppression, le droit d'être trouvé sur Google et même le droit de ne pas être cherché sur Google.

3. Le droit à la mort numérique

Avec l'essor du numérique et du web, la question de la mort — certes peu stimulante — ne peut plus être éludée. Les hommes meurent et disparaissent. Qu'en est-il des hommes numériques, de ces ensembles de données relatives à des personnes décédées ? Les héritiers peuvent-ils accéder aux données numériques du défunt (photos, vidéos, messages, mots de passe etc.) et en disposer ? La protection de l'e-réputation s'étend-elle aux individus disparus ? Surtout, peut-on exiger la disparition de nos pages et de nos contenus à notre mort ? La loi du 7 octobre 2016 *Pour une République numérique* a prévu en ce sens la possibilité pour une personne de laisser des directives relatives à la conservation, à la transmission ou à l'effacement de ses données après son décès. La situation reste néanmoins délicate et chaque internaute doit se montrer vigilant dans la préparation de sa « mort numérique ».

Le droit à la mort numérique, selon la loi *Pour une République numérique*, permet à chacun d'organiser, de son vivant, les conditions de conservation et de communication de ses données à caractère personnel après son décès. En vertu du nouvel article 40-1 de la loi *Informatique et libertés*, les droits à la protection des données et notamment les droits d'opposition, d'accès, de rectification et de suppression deviennent caducs à la mort de l'intéressé. Cependant, « toute personne peut définir des directives relatives à la conservation, à l'effacement et à la communication de ses données à caractère personnel après son décès. Ces directives sont générales ou particulières ». Le législateur a ainsi divisé en deux parties ces directives, complexifiant la situation.

Tout homme numérique devrait, en théorie, définir des directives relatives au sort de ses données après sa mort ou bien désigner une personne afin d'exécuter et de mettre en œuvre ces directives. Selon l'alinéa 2 de l'article 40-1 de la loi de 1978, « les directives générales concernent l'ensemble des données à caractère personnel se rapportant à la personne concernée et peuvent être enregistrées auprès d'un tiers de confiance numérique certifié par la Commission nationale de l'informatique et des libertés (CNIL) ».

Plus encore qu'un droit à la mort numérique, on pourrait imaginer un droit au suicide numérique, rejoignant le droit à l'oubli. Si la mort de l'individu physique peut faire disparaître en tout ou partie l'individu numérique, on pourrait décider de faire disparaître l'homme numérique pour que ne subsiste que l'homme physique, sans pendant informatique.

4. Le droit à l'anonymat numérique

Toute personne devrait pouvoir utiliser les services numériques sans y être identifiée ou identifiable. Le droit à l'anonymat numérique devrait bénéficier à tous, partout dans le monde. Et il devrait être gratuit, ne générer aucun coût supplémentaire — on trouve encore beaucoup d'exemples dans lesquels l'anonymisation revêt un caractère payant. Ce serait un gage de démocratie, une garantie des autres droits et libertés fondamentaux. Alors que la révolution numérique et les techniques de géolocalisation permettent de suivre les hommes partout et tout le temps, avec ou sans leur consentement, le risque est celui du nudge permanent et aussi celui de l'autocensure et, finalement, de la perte de toute individualité. Or, dans une démocratie, l'anonymat — incarné par le secret du vote — constitue une condition essentielle du plein exercice des libertés : liberté d'aller et venir, liberté d'expression, liberté de la presse, liberté d'opinion, liberté de culte. En ligne, le droit à l'anonymat peut se traduire par un droit au pseudonymat. Utiliser des pseudonymes permet de « maquiller » son identité réelle et donc de pouvoir s'exprimer librement sans craindre les représailles.

Les services de communication numérique doivent garantir la confidentialité des données et l'anonymisation des profils personnels. En ce sens, la déclaration commune franco-italienne de 2015 pouvait affirmer que « le secret des correspondances s'applique également aux communications sur internet et reconnaissent la possibilité de publier des contenus sur internet en usant d'un pseudonyme ou en intervenant sous forme anonyme pour exercer les libertés civiles et politiques sans subir des discriminations ou des censures. La liberté de développer et utiliser des technologies d'anonymisation et de chiffrement est une condition concrète d'exercice de ces droits »[127].

L'anonymat et le pseudonymat sont malheureusement souvent détournés de leurs fonctions premières. D'aucuns les utilisent non à des fins de liberté et de démocratie mais à des fins de violence, d'injure, d'attaque en tous genres ou pour diffuser des discours extrémistes. À tel point que l'on en vient à se demander si l'anonymat ne serait pas moins la garantie des droits et libertés qu'un vecteur d'atteintes aux droits et libertés. Il faut donc s'assurer de pouvoir identifier toute personne qui se servirait de ces services pour porter atteinte à l'ordre public ou aux droits et libertés d'autrui. Le droit à l'anonymat numérique ne saurait donc être absolu. À l'ère des attentats de masse et de la propagande extrémiste en ligne, on comprend que les États souhaitent conserver la possibilité

[127] Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et Commission sur les droits et devoirs sur internet de la Chambre des députés italienne, déclaration commune, 28 sept. 2015.

d'arrêter les desseins meurtriers de certains. Pour des raisons d'ordre public ou d'atteinte aux droits d'autrui, il doit demeurer possible d'établir un lien entre l'apparence numérique et l'identité biologique. Mais le risque est alors de voir les pouvoirs publics de certains États à tendance autoritaire profiter de cette possibilité. C'est pourquoi le secret des communications et l'anonymat ne sauraient être levés que sur décision d'une autorité judiciaire indépendante motivée et contestable devant un tribunal statuant en public selon une procédure contradictoire équitable. Les utilisateurs de réseaux sociaux, plateformes de partage, blogs, forums et wikis opèrent souvent derrière des pseudonymes. Quel que soit le mérite de leurs publications, ces utilisateurs ont droit au respect de leur anonymat comme de leur vie privée. Aucune information personnelle les concernant ne peut donc être dévoilée contre leur gré. Il est ainsi défendu de révéler les noms et prénoms des participants à un forum qui avaient choisi d'utiliser des pseudonymes[128]. Que le forum soit public ne changeait rien à l'affaire.

En même temps, la responsabilité doit être pleine et entière pour celui qui s'exprime ou agit en son nom comme pour celui qui s'exprime ou agit derrière un pseudonyme — même si des juges allemands ont pu consacrer un droit à l'anonymat quasi-absolu pour protéger la liberté d'expression[129]. Comme l'alcool, l'anonymat désinhibe. « Toute la dignité de l'homme est dans la pensée », disait Blaise Pascal[130]. Or, avec le web, l'expression l'emporte sur la pensée. Elle devient une fin qui se suffit à elle-même alors qu'elle devrait n'être qu'un moyen au service de la pensée. Il n'est pas rare que les discours tenus sur les plateformes participatives soient très débridés et irresponsables, à tel point qu'ils peuvent contrevenir à certaines prescriptions de la loi. Les réseaux sociaux montrent la part la plus noire des hommes qui s'expriment derrière l'anonymat. L'usage très fréquent de pseudonymes en lieu et place des véritables identités par les utilisateurs de blogs, forums, réseaux sociaux, wikis et plateformes de contenus les fait baigner dans un sentiment d'impunité erroné puisque cela n'atténue aucunement leur responsabilité — même si, dans les faits, il est difficile de poursuivre et de condamner tous les auteurs d'infractions du web 2.0. L'anonymat n'est d'ailleurs que de façade puisque les intermédiaires techniques peuvent être sommés par les autorités judiciaires de leur transmettre les véritables identités des auteurs des messages incriminés, qu'ils sont obligés d'enregistrer au préalable — mais de faux renseignements peuvent leur avoir été donnés.

Entre protection de la démocratie, des droits et libertés fondamentaux, de l'ordre public et des droits d'autrui, l'équilibre entre anonymat et transparence est donc difficile à trouver. Le droit peut difficilement proposer autre chose qu'un précaire compromis. La possibilité de retrouver l'identité de quelqu'un derrière son pseudonyme risque d'amener à l'autocensure, du moins dans les pays en proie à des régimes autoritaires. Et le véritable anonymat est un cadeau fait à tous les criminels. Mais n'est-ce pas là le lot de tous les droits de l'homme : pouvoir être détournés à des fins malveillantes ? C'est bien pourquoi un pays comme la Chine connaît moins de délits que les pays des droits de l'homme.

---

[128] CA Montpellier, 5e ch., 15 déc. 2011.

[129] Cour d'appel de Hamm, 3 août 2011, I-3U196/10.

[130] Cité par *Dictionnaire des citations littéraires*, Larousse, coll. Références, 2014, p. 149.

II. Liberté

Question 1. L'IA a-t-elle été conçue afin de protéger l'autonomie, la souveraineté individuelle et la diversité de ses utilisateurs ?

Question 2. L'IA se contente-t-elle de lui fournir des informations objectives à l'utilisateur sans influencer ses décisions afin de défendre des intérêts particuliers ?

Question 3. Si l'IA impacte les décisions de l'individu et diminue son autonomie et son indépendance, est-il informé du fait qu'une décision, un résultat ou un conseil est produit par un système d'IA ?

Question 4. Lorsque l'IA est utilisée dans le cadre de l'entreprise, permet-elle de renforcer ou d'augmenter les capacités humaines ?

Question 5. Que ce soit dans l'entreprise ou à l'égard des utilisateurs, est-ce que des mécanismes permettent d'éviter qu'une trop grande confiance à l'égard des IA ne se développe ?

Question 6. L'IA protège-t-elle la diversité de ses utilisateurs en évitant de les astreindre à des formes de standardisation, de comportements stéréotypés ?

Question 7. Lorsque l'utilisateur interagit avec un chatbot (agent conversationnel), est-il informé de sa nature non humaine ?

Question 8. L'IA a-t-elle été conçue pour ne pas simuler des interactions sociales et susciter de l'attachement et de l'empathie de la part des utilisateurs ?

A. Le droit à la souveraineté individuelle

1. Ne pas craindre la liberté

Sans doute « la liberté est[-elle] dangereuse, dure à vivre autant que exaltante »[131] car elle implique le poids énorme de la responsabilité à l'égard d'autrui et à l'égard de soi. Elle n'en est pas moins la plus grande richesse de la vie et le cœur battant de l'individu, sans laquelle il n'existe pas. Contre le déterminisme de Leibniz, pour qui il appartient à Dieu de déterminer l'avenir de chaque être humain et « tout est toujours au mieux dans le meilleur des mondes possibles »[132], ou d'Arthur Schopenhauer[133] selon qui le libre arbitre est fatalement une illusion, contre, donc, l'idée que nos actions seraient prédéterminées, l'existentialisme, dont Jean-Paul Sartre a livré un traité systématique[134], nous enseigne combien l'affirmation de la liberté humaine n'est pas une exaltation souriante de cette liberté, puisqu'elle implique la responsabilité et l'incertitude de l'action dans un monde privé à la foi de dieu et de valeurs morales absolues ou consensuelles. La liberté est source d'angoisse à l'épreuve de la contingence du monde. Elle donne « la nausée »[135]. L'homme peut la

[131] A. Camus, « Discours de Suède, 10 décembre 1957 », in *Œuvres*, Gallimard, coll. Quarto, 2013, p. 82

[132] G. W. Leibniz, *Essais de Théodicée sur la bonté de Dieu, la liberté de l'homme et l'origine du mal*, 1710.

[133] A. Schopenhauer, *Essai sur le libre arbitre*, 1839.

[134] J.-P. Sartre, *L'Être et le néant*, Gallimard, 1943.

[135] J.-P. Sartre, *La Nausée*, Gallimard, 1938.

ressentir comme un fardeau et chercher à y échapper en se livrant à diverses forces extérieures ou en se donnant quelques prothèses le libérant de la liberté dans certains aspects de l'existence. Pourtant, comme l'écrivait Sartre, « on ne fait pas toujours ce que l'on veut, mais on est toujours responsable de ce que l'on fait. Le propre de la réalité humaine, c'est qu'elle est sans excuse »[136]. L'homme est « condamné à être libre »[137], il ne peut échapper à la nécessité de choisir et le refus de tout choix est déjà un choix. À l'ère des IA, cependant, la condamnation à être libre risque d'être commuté en condamnation à suivre son tuteur numérique, ce qui est tellement moins angoissant, tellement plus rassurant, comme s'il s'agissait d'un dieu et d'une religion numériques, d'un refuge informatique. Les hommes ont l'habitude de se reposer sur des rôles auxquels, par mauvaise foi, ils s'identifient et souhaitent être identifiés (garçon de café, philosophe, juriste, bourgeois etc.). À présent, le statut d'homme numérique suffit à les rassurer.

Refusant de se dérober à l'histoire, il faut défendre son droit à la souveraineté individuelle. Est souverain celui qui se gouverne lui-même, qui possède « la compétence de la compétence » au sujet de sa vie, qui a le pouvoir suprême impliquant l'exclusivité de la compétence concernant ses agissements et ses pensées. Bien sûr, la souveraineté individuelle ne saurait être totale ou parfaite. Il s'agit d'un idéal vers lequel il faut tendre afin de défendre l'individu, afin de défendre la liberté. Comme le risque n'est pas et ne peut pas être celui d'un excès de liberté mais celui d'un manque de liberté, la souveraineté individuelle — souveraineté numérique dans l'espace numérique — peut être consacrée et défendue par le droit. Il s'agit forcément d'une méta-norme, d'une règle de droit transversale destinée à influencer de nombreuses branches du droit.

Cette souveraineté individuelle est l'héritière directe de l'humanisme, la lanterne qui éclaire les droits de l'homme depuis toujours. Elle suppose l'éducation, l'information, la culture et la démocratie — qui toutes sont remises en cause à l'époque actuelle. Elle implique le libre arbitre, l'autonomie, l'indépendance d'esprit, la tolérance, l'indépendance, l'ouverture, la curiosité et la responsabilité — qui, eux-aussi, sont tous remis en question aujourd'hui. La confrontation de l'humanisme juridique et de l'intelligence artificielle semble ainsi se cristalliser autour de la souveraineté numérique. L'humanisme de la Renaissance, selon la formule de Pic de la Mirandole, empruntée à Plotin, appelle l'homme à « sculpter sa propre statut ». Cette formule, comme plus récemment le « surhomme » de Friedrich Nietzsche, évoque un épanouissement par un effort personnel plutôt qu'une transformation technologique. Or, de nos jours, on est davantage tenté par la seconde que par la première voie. L'IA doit éclairer et non dicter les choix. Elle doit favoriser une liberté dont le plein accomplissement serait l'autodétermination active à l'abri des contraintes extérieures et subies, « l'attirance quasi invincible de la volonté vers le vrai ou le bien clairement connu »[138]. Même si les sciences du comportement et les neurosciences indiquent le contraire, l'individu peut s'imaginer souverain, exercer son libre arbitre, délibérer avec soi-même et uniquement avec soi-même et constituer ainsi son originalité. D'où l'importance de ne pas déléguer cette capacité à des machines qui subrogeraient leur propre souveraineté à celle des hommes.

L'individu se définit par les décisions qu'ils prend, qu'elles soient conscientes ou non[139]. Il se fabrique ainsi son identité réelle, alors que de plus en plus d'humains se satisfont d'une identité civile sans identité réelle. Or ce sont l'une et l'autre qui permettent l'attribution d'une responsabilité,

---

136 J.-P. Sartre, *L'Être et le néant*, Gallimard, 1943.

137 *Ibid.*

138 V° « Liberté », in F. Worms, dir., *Les 100 mots de la philosophie*, Puf, coll. Que sais-je ?, 2019, p. 24.

139 J. Lehrer, *The Decisive Moment – How the Brain Makes Up its Mind*, Canongate Books, 2009.

indépendamment des déterminismes biochimiques. On peut offrir un droit à la souveraineté individuelle à chaque individu afin d'encourager son libre arbitre, mais celui-ci, contrairement à la formule de la Déclaration des droits de l'homme selon laquelle nous serions « nés libres », doit se cultiver et se fortifier tout au long de la vie. Le droit peut bien sûr l'y aider, mais, comme souvent, ce qui importe est l'effectivité de la liberté plus que sa juridicité. Le libre arbitre, l'autonomie, l'indépendance d'esprit ne sont ni donnés biologiquement ni présupposés intellectuellement. Ils se forgent lentement, à mesure de l'expérimentation de la vie. C'est d'ailleurs pourquoi les « *digital natives* », accompagnés dès leur plus tendre enfance par des IA et qui n'ont pas connu l'ouverture ni l'insouciance des années 1960 à 2000, risquent beaucoup plus que leurs aînés de perdre leur souveraineté. Plus un acte ou une décision dépend de circuits de décision complexes, faisant appel à nos capacités réflexives, plus il témoigne de ce que nous sommes. Mais moins nous réfléchissons et délibérons, plus nous obéissons à des réflexes et des signaux extérieurs, plus nous perdons le contrôle sur nous-mêmes. Une éducation morale semble nécessaire afin d'apprendre à faire des choix libres et éclairés, afin de croire en la possibilité du jugement personnel. Et la souveraineté individuelle et le droit en général ont un rôle à jouer dans cette éducation qui permettrait de « retrouver l'individu sous les datas »[140], dans la lignée humaniste d'Érasme qui s'opposait au cerf arbitre de Luther.

Être libre, cela s'apprend. Décider est une capacité qui s'éduque. Or, sans cette souveraineté individuelle effective, il devient difficile, ou même impossible, de résister au confort des recommandations. Qui est encore capable de dire « je refuse d'acheter ce produit qui m'est suggéré par l'IA, qui me plaît et qui me serait utile, parce que je suis moi et je veux décider de ma vie » ? L'appétence pour l'autonomie et la fierté personnelle peuvent résister à la pression des nudges. Contre la thèse selon laquelle, l'IA nous connaissant mieux que nous-mêmes, nous devrions lui déléguer nos choix et nos actions, on peut soutenir que l'intelligence artificielle est un obstacle à l'avènement de l'individu, empêchant de devenir soi-même et de constituer son libre arbitre. Cela paraît d'autant plus important dès lors que la loi des IA, contrairement aux autres modes de gouvernement des hommes, interdit toute négociation, toute discussion, toute opposition, toute délibération collective, donc toute marge d'action afin, par des concessions réciproques, de trouver un juste équilibre entre la pression normative nécessaire et la préservation de soi tout autant nécessaire.

En psychologie, la théorie de la motivation et du besoin d'Abraham Maslow (connue sous le nom de « pyramide des besoins de Maslow ») montre combien les comportements des hommes sont les réactions à différents besoins : besoins physiologiques élémentaires, tout d'abord, puis besoins de sécurité ; besoin d'amour et d'amitié, ensuite, ainsi que de reconnaissance ; besoin d'accomplissement, d'épanouissement, enfin[141]. Chaque besoin assouvi conduit les humains à aspirer à la satisfaction d'un besoin supérieur. Et la psychologie humaniste de Maslow introduit le postulat de l'autodétermination et s'appuie sur l'expérience consciente du patient : il s'agit de développer chez lui la capacité de faire des choix personnels. Que reste-t-il de ce volontarisme dans le monde numérique. Quels besoins peuvent encore être assouvis réellement et non artificiellement ? Quels hommes sont encore capables de suivre leurs propres expériences et de se débarrasser des conditionnements qui limitent leurs libertés, comme le pensait le chercheur américain ? Si l'humain peut peiner à emprunter naturellement ce chemin, on peut l'y aider par le droit.

---

[140] G. Kœnig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de l'observatoire, coll. De facto, 2019, p. 334.

[141] A. Maslow, « A Theory of Human Motivation », *Psychological Review* 1943.

2. La souveraineté individuelle comme droit de l'homme

La souveraineté individuelle est la source des libertés si on considère ces dernières comme des résultats. Elle est la condition des libertés individuelles. Elle exprime l'individualisme présent, notamment, dans la Déclaration des droits de l'homme et du citoyen de 1789. Rompant avec la structure de l'Ancien régime où la société était faite d'ordres et de corps, la Déclaration a fait passer le pouvoir des groupes vers les individus. Les droits déclarés en 1789 ne sont susceptibles que d'un exercice individuel — on comprend alors pourquoi la liberté syndicale ou la liberté d'association n'ont pas été consacrées, voire ont été combattues à l'image de la loi Le Chapelier de 1791 qui a interdit toute action syndicale. C'est à l'individu de se construire librement, souverainement. Être hauteur de ses actes est une question de liberté, mais aussi de dignité car l'exercice du libre arbitre est source d'estime de la part d'autrui comme de soi-même. En outre, il ne saurait y avoir de démocratie sans souveraineté individuelle. Dès lors que le ciblage comportemental et la manipulation algorithmique empêchent d'élaborer une opinion et de construire une identité, le vote perd son sens. L'encadrement du filtrage personnalisé des informations constitue donc un enjeu majeur pour la conservation d'une société pluraliste et de la liberté d'opinion. La « Déclaration de Montréal » peut donc affirmer que « les systèmes d'intelligence artificielle doivent permettre aux individus de réaliser leurs propres objectifs moraux et leur conception de la vie digne d'être vécue »[142]. La souveraineté numérique est une condition *sine qua non* du bien-être numérique. L'individu autonome, libre de ses choix et responsable de ses actions, est la condition des droits individuels, des mécanismes des marchés, de la justice et de la démocratie. La Cour suprême des États-Unis ne s'y est d'ailleurs pas trompée puisqu'elle a consacré le libre arbitre comme la condition même du système juridique américain, permettant à l'individu de choisir entre le bien et le mal, le souhaitable et le condamnable, le permis et l'interdit.

La souveraineté individuelle, dans la tradition libérale, est aussi une propriété de soi. Pour John Locke, l'État n'a de sens que s'il se donne pour raison d'être et pour mission de préserver les droits naturels de l'homme et surtout le premier d'entre eux : la propriété, qu'il définit comme propriété des biens mobiliers et immobiliers, mais aussi propriété de soi-même, ce qui inclut et dépasse la liberté. Pour le philosophe britannique, les individus doivent être « parfaitement libres d'ordonner leurs actions, de disposer de leurs biens et de leurs personnes comme ils l'entendent, sans demander l'autorisation à un autre homme ni dépendre de sa volonté »[143]. On ajoutera que les individus doivent de la même manière pouvoir décider de leurs sorts sans besoin de s'en remettre à quelques objets informatiques. Cette propriété de soi détermine les limites de l'action publique : en son nom, chacun est d'abord maître d'une sphère personnelle dans laquelle il ne saurait y avoir d'intrusion légitime. Cela vaut identiquement et *a fortiori* à l'égard de toute puissance privée. Comme la souveraineté régit les relations entre les États, la souveraineté individuelle d'autrui vient seule borner celle de chacun. La liberté des modernes est donc cette sphère personnelle indérogeable de liberté, cet espace privé à l'abri de toute intrusion de la puissance publique ou d'une puissance privée, « la jouissance paisible de l'indépendance privée »[144].

La Déclaration de 1789, cependant, place l'individu privé sous le citoyen public et attache les droits de l'homme aux droits politiques du citoyen. La souveraineté individuelle vise, au contraire, à placer

---

[142] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

[143] J. Locke, *Second traité sur le gouvernement civil*, 1690.

[144] B. Constant, *De la liberté des Anciens comparée à celle des Modernes*, 1819.

l'individu personne privée au fondement de l'ordre juridique. « Toute souveraineté réside essentiellement dans la nation. Nul corps, nul individu ne peut exercer d'autorité qui n'en émane expressément », pose l'article 3 de la DDHC. Serait-il inconséquent de convenir plutôt que « toute souveraineté réside essentiellement dans l'individu. Nul corps, nulle nation ne peut exercer d'autorité qui n'en émane expressément » ? Avec la souveraineté individuelle, l'individu est la seule réalité irréductible dans l'ordre éthique, dans l'ordre politique et donc dans l'ordre juridique. Il est le seul absolu de ces questions, pour lesquelles par conséquent « toutes les propriétés du groupe peuvent se ramener à des combinaisons quantitatives des propriétés de ses éléments individuels »[145]. Contrairement à une approche nationaliste ou socialiste, l'approche individualiste consiste donc à tout mesurer à l'échelle de l'individu plutôt qu'à l'échelle du groupe ou de la nation. On retrouve ici l'opposition entre les droits de l'individu occidentaux et les droits de la société chinois.

Au XVe siècle, Jean Pic de la Mirandole pouvait exprimer son enseignement en ces termes : « Je ne t'ai donné ni place déterminée, ni visage propre, ni don particulier, afin que ta place, ton visage et tes dons, tu les veuilles, les conquières et les possèdes par toi-même. Toi, que ne limite aucune borne, par ton arbitre, entre les mains duquel je t'ai placé, tu te définis toi-même. Je t'ai mis au milieu du monde, afin que tu puisses mieux contempler autour de toi ce que le monde contient »[146]. Sans doute la souveraineté individuelle peut-elle s'inspirer de cet héritage humaniste. Les IA ne doivent pas nous empêcher de « contempler le monde », de « n'être limité par aucune borne » ou de « conquérir son visage et ses dons par soi-même ». On peut ainsi « définir l'humanisme que nous devrions défendre, non pas comme celui qui ambitionne de nous doter d'une puissance sans limite sur les choses, procédant d'un anthropocentrisme dominateur et dévastateur, mais comme celui qui nous enjoint de cultiver nos capacités, seules à même de nous rendre pleinement maîtres de nos destins, de favoriser l'éclosion d'une infinité de possibles, n'empiétant sur les droits de personne et donnant voix au chant polyphonique et ininterrompu des divergences »[147].

3. Le droit à la déconnexion

Des études neuroscientifiques estiment qu'un être humain procède à environ 35 000 décisions par jour, la plupart inconscientes ou très peu conscientes[148]. Les IA contribuent à réduire sensiblement ce chiffre en décidant à la place de l'homme — ce qui peut potentiellement constituer un grand progrès : l'informatique servirait à désengorger nos facultés cognitives, nous permettant de nous concentrer sur les choix les plus importants. Personnaliser son intelligence artificielle serait une solution afin de ne sacrifier ni la souveraineté individuelle ni les formidables potentialités de ces technologies. Chaque utilisateur établirait ses propres normes dans une sorte de *machine learning* interactif, où l'individu est le maître des paramètres de la machine[149]. La machine pourrait dès lors se voir confier des décisions de la vie courante, en fonction d'orientations générales arrêtées par

---

[145] M. Bernès, in A. Lalande, *Vocabulaire technique et critique de la philosophie*, 10e éd., Puf, coll. Quadrige dicos poche, 2010, p. 500.

[146] J. Pic de la Mirandole, *Discours sur la dignité de l'homme*, 1486.

[147] É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'échappée, coll. Pour en finir avec, 2018, p. 272.

[148] B. Sahakian, J. N. LaBuzetta, *Bad Moves – How Decision Making Goes Wrong, and the Ethics of Smart Drugs*, Oxford University Press, 2013.

[149] A. Holzinger et *alii*, « Towards Interactive Machine Learning: Applying ant Colony Algorithms to Solve the Travelling Salesman Problem with the Human in-the-loop Approach », in *Proceedings of the International Conference on Availability, Reliability and Security*, Springer, 2016.

l'humain et modifiable à tout moment, avec effet immédiat. Le nudge serait présent mais dépendrait des considérations éthiques de l'individu. Celui-ci se servirait de l'IA pour prolonger sa morale personnelle et le rapport de force entre homme et machine serait renversé. Les systèmes de recommandation pourraient laisser la porte ouverte à la divergence. En réalité, de tels mécanismes sont largement utopiques, car même techniquement réalisables encore faudrait-il que les individus soient concrètement décidés à reprendre le contrôle, ce dont on peut douter, surtout après des années d'accoutumance à la tutelle informatique.

Tout à l'inverse, plutôt que des IA omniprésentes mais personnalisées, la souveraineté numérique pourrait justifier la consécration d'un droit à la déconnexion, un droit de se placer à tout moment — et même, pourquoi pas, durant toute la vie — à l'abri des incitations numériques. Les services d'intelligence artificielle devraient explicitement offrir le choix de la déconnexion à intervalles réguliers, sans inciter à rester connecté. Il conviendrait, tout d'abord, de pouvoir échapper à la traçabilité en ligne, mais aussi IRL (*in the real life*) permise essentiellement par la géolocalisation. La géolocalisation peut se déployer aussi bien à des fins commerciales (analyse des comportements des consommateurs) que managériales ou disciplinaires (suivi des employés maquillé en « sécurité ») ou encore pénales (bracelet électronique pour les condamnés, balises GPS dans le cadre d'enquêtes judiciaires). Or, en droit français, bien que les enjeux en termes de préservation des droits individuels soient largement comparables à ceux qui accompagnent la biométrie, ni la loi de 2004 relative aux fichiers et aux libertés, ni celle de 2014 relative à la géolocalisation n'ont conféré de compétence particulière à la CNIL en matière de géolocalisation[150]. Celle-ci est dès lors uniquement soumise à un régime de déclaration préalable et nul ne peut s'opposer à la réalisation d'un projet de géolocalisation. Il est donc parfaitement possible, par exemple, que des dispositifs de géolocalisation soient installés dans les véhicules des employés afin de suivre ou facturer une prestation de transport, veiller sur l'employé, le véhicule ou la cargaison ou s'assurer du respect des règles d'utilisation du véhicule. La CNIL ne pouvant que donner des conseils, elle propose des fiches techniques qui invitent tout employeur à ne pas recourir à la géolocalisation pour contrôler le respect du Code de la route par ses employés ou dès lors que sont en cause des délégués du personnel. Mais il est aisé de maquiller la vraie raison d'un suivi derrière de fausses raisons. La Cour de cassation a toutefois estimé qu'un licenciement motivé par des données de géolocalisation était illicite, lesdites données ne pouvant être utilisées qu'en vue des finalités déclarées lors de la constitution du traitement[151].

La biométrie connaît de plus en plus d'usages privés qui posent la question de la possibilité de demeurer à l'abri des regards intrusifs et des bras tentaculaires. Concrètement, la géolocalisation prend une dimension nouvelle avec la généralisation des puces RFID (*radio frequency identification*), en plus de la multiplication des applications nécessitant l'activation de la géolocalisation — souvent sans que l'on comprenne bien pourquoi. Des discothèques à Mexico, Madrid, Barcelone et Rotterdam ont convaincu leurs clients que l'injection intra-corporelle d'une puce RFID leur permettrait de devenir des clients « VIP »[152]. Il s'agit bien entendu surtout de mieux les connaître afin de pouvoir leur offrir des incitations personnalisées et les amener à consommer plus et plus souvent dans ces lieux festifs. Juridiquement, la souveraineté individuelle peut déjà être relayée par l'exigence du consentement de la personne. Mais cela paraît fort insuffisant et le droit à

[150] S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 580.

[151] Cass. soc., 3 nov. 2011, n° 10-18.036.

[152] A. Türk, *La vie privée en péril – Des citoyens sous contrôle*, Odile Jacob, 2011, p. 57.

la déconnexion permettrait de prévenir quelques abus, à défaut de pouvoir tous les freiner. Plus de telles puces vont envahir discrètement le quotidien, renforçant la possibilité de téléguider à distance les conduites, plus cette question se posera. D'ores et déjà, beaucoup d'objets de consommation courante comportent des puces RFID. L'intérêt pour les entreprises est évident : si un client rentre chez lui avec des puces RFID dans les bagages, les responsables du marketing pourront engager des actions commerciales personnalisées en fonction des informations renvoyées par ces puces. Il faudrait au moins pouvoir « tuer » les puces, les désactiver à tout moment ou même imposer leur désactivation par défaut avec possibilité pour le client qui souhaite être traqué de les activer. Au travail, l'implantation d'une puce RFID sous la peau des salariés leur permet d'entrer dans les locaux et de « badger » automatiquement, sans besoin de gérer des mots de passe ou de prendre garde à ne pas oublier sa carte à puce. En contrepartie, l'atteinte aux libertés individuelles est évidemment disproportionnée. L'implantation de micropuces sur les travailleurs est aussi un moyen pour les employeurs de mieux surveiller les employés afin de renforcer leur productivité[153]. Le TGI de Paris a d'ailleurs pu retenir qu'un dispositif biométrique de pointage et de contrôle des horaires et des salariés était disproportionné en raison des « dangers d'atteinte aux libertés individuelles » et de la possibilité de recourir à des alternatives moins liberticides telles qu'une badgeuse classique avec identification par un code secret[154]. Aujourd'hui, au travail comme chez soi, le principal danger est celui de la mise en visibilité permanente. Le bureau et la maison risquent de se transformer en prisons panoptiques.

L'objectif affiché de l'industrie numérique est d'introduire l'intelligence artificielle partout. *In fine*, il deviendrait impossible de vivre sans être connecté. Il n'y a pas que la consommation intelligente et la vie sociale intelligente qui seraient impossibles sans connexion, tel serait également le cas de la consommation et de la vie sociale en soi. Là encore, la Chine offre des exemples révélateurs de ce qui pourrait attendre les pays occidentaux. La stratégie « Internet Plus », impliquant le développement rapide des infrastructures de communication et la démocratisation des smartphones, vise à intégrer toutes les activités, notamment industrielles, dans un vaste réseau auquel il serait impossible d'échapper. Modernisée autour de nouveaux moteurs de croissance, développant au maximum l'internet mobile, l'informatique en nuage, les big datas et l'internet des objets, la Chine est un pays où il est désormais impossible de vivre loin du commerce électronique ou des services bancaires en ligne. Par exemple, le smartphone devient indispensable pour faire ses courses. Tencent a ainsi développé WeChat, une messagerie qui remplit des fonctions quasi universelle. WeChat permet de communiquer, être actif sur les réseaux sociaux, mais aussi faire ses courses, régler ses achats, commander un taxi, prendre un rendez-vous avec un spécialiste, s'enregistrer à l'hôpital etc. On scanne des codes-barres, par exemple sur les marchés, et on n'utilise plus beaucoup les espèces ni les emails et encore moins les courriers papier. Même les entreprises fonctionnent autour de WeChat. En Chine, refuser WeChat implique la mort professionnelle, sociale et sentimentale — alors qu'en Europe on peut encore vivre à peu près normalement sans avoir un compte sur Facebook.

Un droit à la déconnexion préserverait de telles dérives totalitaires et serait une condition du bien-être numérique. Le « débranchement » serait aussi la seule véritable solution pour les personnes ayant développé une addiction, qui souffrent de la sursollicitation qu'entraînent les notifications et d'une volonté de reconnaissance sociale imposant de ne rien rater et de répondre rapidement et à

[153] S. Prévost, « Du développement du numérique aux droits de l'homme digital », *Dalloz IP/IT* 2019, p. 345.

[154] TGI Paris, 19 avr. 2005, n° 05/00382, *Comité d'entreprise d'Effia Services, Syndicat Sud Rail c. Effia Services*.

tout. Elles sont atteintes de « FOMO » (*fear of missing out*), de « stress digital »[155]. Il devrait donc toujours exister, pour ces personnes et les autres, une alternative matérielle et non connectée à un service offert en ligne. La première des résistances consiste à repasser en « mode manuel », à quitter volontairement les réseaux pour renouer avec un mode de décision 1.0, sans doute moins optimal mais subjectif et authentique. C'est ainsi que l'exercice concret de notre souveraineté permettrait de nous éviter de devenir des « zombies »[156]. En cessant d'émettre des données, on se place à l'abri du nudge. La Cour européenne des droits de l'homme elle-même a reconnu avec fatalisme que la déconnexion serait le seul moyen de préserver sa vie privée et qu'il est illusoire de prétendre à la même protection en ligne qu'hors ligne[157]. Des militants ou de simples individus avertis font ainsi le choix d'utiliser de vieux téléphones portables du début des années 2000, avec des cartes prépayées, d'occulter la webcam de l'ordinateur avec un autocollant, d'éteindre systématiquement la géolocalisation ou d'utiliser des applications en pair-à-pair pour éviter de partager ses données avec des tiers. On peut donc soit tirer les conséquences de l'avènement inexorable de l'homme numérique et chercher à en tirer le meilleur profit, soit défendre l'homme physique. On peut aussi contribuer aux deux mouvements et ne pas nier par principe tous les bienfaits du numérique et des IA. Et on peut encore utiliser les outils informatiques tout en demeurant souverain numériquement grâce aux messageries cryptées, VPN (réseaux privés virtuels qui permettent de créer un lien direct entre des ordinateurs distants, isolant leurs échanges du reste du trafic) et outils de navigation anonyme ne laissant pas de traces, comme Tor, chers aux hackers[158]. Mais vivre sur un tel *dark web* pose d'autres difficultés, avec notamment le risque d'y faire de mauvaises rencontres. Certains proposent d'autres moyens ingénieux pour vivre connecté librement : l' « obfuscation », qui consiste à produire délibérément des informations ambiguës, confuses ou trompeuses afin de rendre la collecte de données inutile[159]. On se camoufle pour pouvoir utiliser les nombreux services du web en cachette. Cela demande cependant beaucoup d'efforts en même temps que d'imagination. Pour être invisible, mieux vaut être absent du réseau qu'y être présent et s'efforcer à y enregistrer des données erronées pour brouiller les pistes. Par ailleurs, des services de clonage virtuel permettent de créer de fausses identités pour perturber les algorithmes. Mais, là encore, on ne vit plus vraiment lorsqu'on endosse une fausse identité. Reste que diminuer la qualité et la représentativité des jeux de données est effectivement un moyen de se protéger dans l'univers numérique.

On propose aussi que le droit consacre une « imprédictibilité d'ordre public »[160], c'est-à-dire le droit de se mettre à l'abri des recommandations personnalisées sans pour autant se déconnecter. La mise en pratique d'un tel droit semble toutefois délicate, mais son effet symbolique, ne serait pas nul. La première mesure de bon sens serait déjà d'encadrer et limiter les relations entre l'homme et l'ordinateur, de conditionner la connexion à internet afin de prévenir les atteintes irréversibles à la vie privée et surtout à l'autonomie et à la liberté[161]. Alors que déjà des corps humains accueillent des puces électroniques, pour l'instant surtout à des fins médicales, qui peuvent collecter des

---

155 M.-P. Fourquet-Courbet, D. Coubert, *Connectés et heureux ! Du stress digital au bien-être numérique*, Dunod, 2020.

156 J. Lanier, *Ten Arguments for Deleting your Social Media Accounts Right Now*, Henry Holt & Co., 2018.

157 CEDH, 13 nov. 2007, n° 31358/03, *Muscio c. Italie*.

158 G. Kœnig, *Les aventuriers de la liberté*, Plon, 2016.

159 F. Brunton, H. Nissenbaum, *Obfuscation: A User's Guide for Privacy and Protest*, MIT Press, 2016.

160 S. Merabet, *Vers un droit de l'intelligence artificielle*, th., Université d'Aix-Marseille, 2018, p. 233.

161 J. Le Gars, « Homme augmenté, transhumanisme en embuscade », *Dr. de la famille* 2018, n° 6.

données nombreuses et variées[162], le droit peut interdire la connexion permanente ou, au moins, accorder le droit à la déconnexion. Alors que la sécurité informatique totale ne peut être garantie et qu'il existe des risques de prise de contrôle par des personnes mal intentionnées, notamment des prises d'otage numérique, la seule solution peut être de quitter le réseau. L'effacement sélectif des souvenirs est d'ores et déjà pratiqué dans le cadre de certains traitements médicaux. Le droit pourrait-il autoriser de telles pratiques à une plus vaste échelle ? Les implants TIC, puces en silicone insérées dans le corps humain afin d'améliorer sa communication, pourraient-ils être déployés à grande échelle, sans possibilité de retour en arrière, mis à part dans les cas extrêmes tels que ceux des personnes souffrant de paralysie totale. Le cadre juridique actuel, en tout cas, est encore loin d'offrir toutes les réponses à l'apparition de telles prothèses technologiques.

Au-delà du monde des IA, la liberté, consistant à pouvoir faire tout ce qui ne nuit pas à autrui, est de plus en plus abandonnée au profit d'un hygiénisme qui semblera très excessif aux yeux d'un libertaire. Les messages des autorités sanitaires martelés partout invitent l'individu à se protéger contre lui-même et à se méfier de sa liberté comme de la peste[163]. Pour notre plus grand bien, ces messages influencent nos comportements et portent atteinte à notre souveraineté individuelle. Mi-parent bien intentionné, mi-directrice de conscience étouffante, la santé publique s'érige en nouvelle morale séculière dont les commandements seraient « perdez du poids, arrêtez de fumer, mangez sainement, faites de l'exercice, buvez modérément, soyez prudents, donnez votre sang de votre vivant et vos organes après votre mort, ne faites pas l'amour avec n'importe qui et n'importe comment etc. »[164]. Dans cette époque où la liberté individuelle des Lumières et des révolutionnaires de 1789 s'essouffle de plus en plus, on se rappellera que, pour Albert Camus, « en face de la société politique contemporaine, la seule attitude cohérente de l'artiste est le refus sans concession. Par sa fonction même, l'artiste est le témoin de la liberté »[165]. Tout homme devrait développer ses talents artistiques ou, s'il n'en a point, s'en inventer, car, ce faisant, il développerait son goût pour la liberté.

B. Le droit à l'autonomie numérique

1. Se donner sa propre loi

Le droit pourrait consacrer le principe selon lequel les systèmes d'intelligence artificielle doivent être développés et utilisés afin de favoriser ou, du moins, ne pas porter atteinte à l'autonomie des individus, dans le but de renforcer le contrôle de chacun sur sa vie et son environnement. La Commission européenne retient d'ailleurs en ce sens que « les systèmes d'IA devraient être les vecteurs de sociétés équitables en se mettant au service de l'humain et des droits fondamentaux, sans restreindre ou dévoyer l'autonomie humaine »[166]. La notion d'autonomie personnelle a notamment été consacrée par la Cour européenne des droits de l'Homme en tant que « principe important qui sous-tend l'interprétation des garanties de l'article 8 [de la Convention de sauvegarde des droits de

[162] D. Neerdael, *Une puce dans la tête – Les interfaces cerveau-machine qui augmentent l'humain pour dépasser ses limites*, Fyp, 2014.

[163] J.-M. Larralde, dir., *La libre disposition de son corps*, Bruylant, coll. Droit et justice, 2009.

[164] S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 63.

[165] A. Camus, « Prométhée aux enfers », in *Œuvres*, Gallimard, coll. Quarto, 2013.

[166] Commission européenne, « Lignes directrices en matière éthique pour le développement de l'IA », 8 avr. 2019.

l'homme] »[167] — qui protège la vie privée des personnes. L'autonomie occupe une place cardinale dans la jurisprudence européenne en ce qu'elle vise à préserver la capacité des personnes de faire des choix concernant leurs modes de vie et le sens qu'elles entendent donner à cette dernière[168]. De la sorte, il s'agit d'un aspect important de la figure du sujet à laquelle la CEDH garantit des droits. Le sujet est un individu libre et responsable dont « la dignité et la liberté sont l'essence même » de la Convention européenne de sauvegarde des droits de l'homme[169]. Cependant, l'autonomie individuelle ne fait pas l'unanimité. Certains considèrent qu'il s'agirait d'un excès d'individualisme aboutissant à conférer une force normative excessive aux droits de l'individu, contre ceux du citoyen ou de la personne, et à valoriser excessivement la volonté du sujet à laquelle il deviendrait impossible d'opposer de quelconques limites[170]. La protection juridique de l'autonomie et, plus généralement, de l'individualité vise surtout, au-delà de ses éventuelles limites, à protéger la dignité humaine, notamment en favorisant le développement personnel, ainsi que la démocratie. Même pour les petites choses de la vie, on ne peut admettre que nos conduites nous soient imposées. Suivant la formule d'Alexis de Tocqueville, « c'est surtout dans le détail qu'il est dangereux d'asservir les hommes ». Comment quelqu'un qui aurait perdu l'habitude de choisir son trajet, la musique qu'il écoute ou le contenu de ses repas pourrait-il choisir librement un métier, un partenaire ou un candidat lors d'une élection ? Si, par exemple, une IA pouvait attribuer à chacun l'emploi lui correspondant le mieux en fonction de ses caractéristiques personnelles, ses qualités et ses goûts, ceux-ci seraient tirés du monde réel, où ils ont été élaborés au hasard des événements et des rencontres. Mais, demain, concernant ceux qui n'auraient pas pu expérimenter la vie sans IA, ils seraient les produits de recommandations automatiques martelées dès le plus jeune âge et durant toute l'éducation et le développement personnel. Pareille prédétermination intégrale serait difficilement viable en même temps que vivable.

L'autonomie est un corollaire fondamental de la liberté. Elle est un espace où il est possible d'agir ou de ne pas agir et, en cas d'action, de lui donner des motifs et des modalités arbitrairement définis. Il s'agit d'un pouvoir d'autodétermination spontanée et volontaire, un espace libre utilisable au gré du choix des personnes[171]. Autrement dit, il s'agit de la capacité à être cause première et absolue de ses actes. Cette autonomie, ce droit de se donner à soi-même (*auto*) sa propre loi (*nomos*), ne va bien sûr pas sans pleine et entière responsabilité. Les lois des uns s'arrêtent là où commencent celles des autres. L'autonomie fonde le devoir de répondre de ses actes, cet impératif qui constitue l'honneur du genre humain. Et l'autonomie profite d'abord à l'individu, mais aussi à la collectivité des individus. Il n'y a pas d'autonomie collective sans autonomie individuelle.

Les dictionnaires définissent l'autonomie comme ceci : « Fait de se gouverner par ses propres lois ; faculté de se déterminer par soi-même, de choisir, d'agir librement ; liberté, indépendance morale ou intellectuelle »[172]. Cette autonomie doit être protégée, et pour cela être l'objet d'un droit, afin de ne pas se laisser gouverner par les lois des IA, ne pas être déterminé par des IA, ne pas se trouver dans une situation de dépendance morale ou intellectuelle à l'égard des IA. L'être humain, même

---

[167] CEDH, 29 avr. 2002, n° 2346/02, *Pretty c. Royaume-Uni*.

[168] S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 536.

[169] CEDH, 29 avr. 2002, n° 2346/02, *Pretty c. Royaume-Uni*.

[170] M. Fabre-Magnan, M. Levinet, J.-P. Marguénaud, F. Tulkens, « Controverse sur l'autonomie personnelle et la liberté du consentement », *Droits* 2008, n° 48, p. 3 s.

[171] M. Levinet, *Théorie générale des droits et des libertés*, Bruylant, 2010, p. 52.

[172] V° « Autonomie », in *Trésor de la langue française*.

lorsqu'il a le sentiment d'opérer des choix personnels, obéit à des influences, par exemple des habitudes routinières. Pour Pierre Bourdieu, l'*habitus* est cette disposition par laquelle la société et les membres qui la constituent font des choix réguliers et prévisibles. C'est pourquoi les calculateurs peuvent produire, à partir de masses de données pertinentes, des recommandations pertinentes. Mais il ne faudrait pas que le mouvement s'inverse, que ce ne soit plus le résultat informatique qui suive le comportement humain mais plutôt le comportement humain qui procède du résultat informatique. Car alors l'individu ne serait plus autonome, plus l'auteur de ses propres normes, plus un individu en bonne et due forme. Tant que les prédictions algorithmiques se bornent à confirmer des lois sociales et personnelles, l'autonomie est sauve. Dès lors qu'elles influencent les comportements, qu'elles les font dévier de leurs trajectoires naturelles, l'autonomie s'efface au profit de la loi des IA. Et, sans autonomie, l'individu devient un « dividu » en disparaissant dans les flux du contrôle machinique[173]. L'autonomie, forme normative de la liberté, est indispensable à l'individualité. Aucun homme ne peut se satisfaire de vivre replié sur le passé, enfermé dans ses caractères antérieurs, plutôt que de s'ouvrir à l'avenir et de le préparer. Or les IA fonctionnent à base de statistiques relatives aux événements passés, font constamment l'hypothèse que notre futur sera une reproduction de notre passé. Pour ne pas avoir seulement une histoire mais aussi une subjectivité, des représentations, des projets, une personnalité, il est important de pouvoir s'abriter du passé.

Le comportementalisme radical nous montre à quel point, alors que nous nous rêvons émancipés de toutes les déterminations, celles-ci sont en réalité irrésistibles et nous empêchent de devenir des singularités pleines et entières. L'autonomie pleine et entière est une utopie. Seule une semi-autonomie est concevable. Elle n'en doit pas moins être défendue afin de nous éviter de nous transformer en petites souris mécaniques prises dans les griffes des grands calculateurs[174]. Contre l'idée qu'il existerait dans l'homme et dans la société quelque chose de déterminé et de calculable, il faut défendre la liberté définie comme indépendance intérieure. Avec les stoïciens, il faut concevoir la liberté indépendamment de toute condition extérieure. L'autonomie, envisagée tel un idéal, est la capacité de se détacher de son environnement, de s'abstraire de toutes les formes de pression extérieure, de ne dépendre que de soi-même, de ne connaître aucune contrainte. Un tel détachement implique sans doute une force psychologique peu ordinaire. Après les stoïciens, nombreux sont les philosophes, de Spinoza à Jean-Paul Sartre, qui ont envisagé la liberté telle une indépendance intérieure et « la capacité morale de se déterminer en suivant les seuls conseils de la raison et de l'intelligence non dévoyée par la passion »[175].

2. La condition d'une vie heureuse, prospère et digne

Le libre arbitre peut susciter des conduites aléatoires et imprévisibles ou bien des conduites qui sont les résultats de délibérations. L'autonomie renvoie à ces dernières. Elle n'est pas le fruit du hasard mais le fruit du jugement indépendant. Elle compte davantage sur l'arbitre que sur la liberté dans le libre arbitre, qui devient, plus justement nommé, un « arbitre libre ». Celui-ci est capable de mettre dans la balance les différents arguments pour et contre avant de tirer une conclusion. Il est indifférent que son jugement soit prévisible d'un point de vue physique ou biochimique, l'important est qu'il soit plus ou moins mûri sous la forme d'une délibération intérieure. L'autonomie n'est donc pas un

[173] G. Deleuze, « Post-scriptum sur les sociétés de contrôle », in *Pourparlers – 1972-1990*, Minuit, 2003.
[174] A. Pentland, *Social Physics – How Good Ideas Spread – The Lessons from a New Science*, Penguin Press, 2014.
[175] V° « Liberté », in L. Hansen-Love, dir., *La Philosophie de A à Z*, Hatier, 2018, p. 265.

simple hasard brisant les chaînes de causes à effets. Elle est l'usage de cette conscience, que René Descartes installait dans la glande pinéale, permettant de parcourir le champ des possibles et d'y opérer un choix en pesant le « pour » et le « contre ». Elle est une intentionnalité non contrariée, par contraste avec ces gestes que l'on effectue sous la pression inconsciente d'un nudge, sans bien savoir pourquoi. Il y a autonomie lorsqu'on peut expliquer son action par d'autres motifs que « c'est comme ça » ou « le maître l'a dit ». L'être humain se distingue par sa réflexivité et sa capacité à délibérer, donc à justifier une action, à lui donner de bonnes raisons (subjectives). La reconnaissance de l'autonomie n'est pas celle des décisions finales mais celle des délibérations précédentes qui y conduisent. Or celles-ci sont au moins perturbées, si ce n'est effacées, par l'impact des IA sur les conduites. Plutôt que de se désespérer face à la finitude de l'existence, l'autonomie permet d'accéder à la richesse de la vie et à la dignité de la condition humaine. Les vertus du hasard et de l'improvisation sont essentielles dans la vie et, par exemple, un détour chez un disquaire ou un libraire offre des perspectives inattendues alors que Spotify ou Amazon ne proposent qu'une expérience cadrée. Mais cela suppose aussi d'accepter le lot de toute existence humaine, contre lequel les IA agissent : l'ouverture indéfinie du réel, l'absence de vérités absolues, l'incertitude et la persistance du doute qui empêchent d'arrêter en toutes circonstances la meilleure décision.

L'autonomie a constitué une grande avancée pour les hommes qui, en particulier après la Révolution française, se sont trouvés libérés des consentements sociaux qui caractérisaient les sociétés traditionnelles et l'Ancien Régime. Les individus, seuls ou en groupes, se sont vus reconnaître le droit de formuler leurs propres objectifs et de déterminer les moyens de les atteindre. Aux États-Unis, la « *privacy* », selon la jurisprudence de la Cour suprême, comporte le droit au respect de la vie privée, mais aussi la liberté de faire certains choix existentiels dans des domaines où l'État a perdu le droit de contrecarrer les libertés individuelles[176]. Or il serait sans doute regrettable que des puissances privées acquièrent une telle prérogative heureusement abandonnée par la puissance publique. En Europe, la CEDH a jugé de la même façon que la notion de vie privée ne permet pas seulement le « repli sur soi », mais est « une notion large qui englobe, entre autres, des aspects de l'identité physique et sociale d'un individu, notamment le droit à l'autonomie personnelle, le droit au développement personnel et le droit d'établir et entretenir des rapports avec d'autres êtres humains et le monde extérieur »[177]. Le droit à la vie privée est ainsi tout simplement un droit à la vie, un droit de vivre pleinement, comme on l'entend au plus profond de soi, sans se livrer corps et âme à des forces extérieures. L'équivalent contemporain du « destin à la turque » dont se moquait Leibniz — « *fatum mahometanum* », soit le fait d'accepter un destin de type fataliste selon lequel les événements se produiraient « quoi qu'on fasse », obligeant à adopter les comportements qui la feront advenir — serait de laisser l'intelligence artificielle et ses recommandations gérer entièrement nos vies, au motif que les algorithmes, comme Dieu autrefois, seraient en mesure d'identifier le « meilleur des mondes possibles ». Le droit à l'autonomie pourrait même muer en obligation d'autonomie, imposant de recourir à nos facultés de jugement.

La volonté humaine libre échappe aux régimes de causalité à l'œuvre dans les autres domaines de la nature. C'est un « empire dans un empire », selon la formule de Spinoza, qui désormais doit aussi échapper aux déterminismes numériques. Il faut croire en un tel pouvoir de l'esprit, même s'il connaît forcément des fragilités. Un travail rigoureux de l'esprit est nécessaire à l'être humain. Il faut croire, comme Descartes au XVIIe siècle, au cogito, c'est-à-dire le plein usage de la capacité de

[176] F. Rigaux, *La protection de la vie privée et des autres biens de la personnalité*, LGDJ-Bruylant, 1990, p. 20.
[177] CEDH 29 avr. 2002, n° 2346/02, *Pretty c. RU*.

discernement, permettant de rechercher les vérités en faisant preuve de méthode. Mieux vaut sans doute pouvoir dire « je pense donc je suis » que « je ne pense pas donc je ne suis pas ». Les réflexions sur l'acte gratuit, un acte voulu sans aucun motif, tentent de pousser aussi loin que possible l'idée d'une volonté absolument libre de ses choix et indifférente à tout détermination. L'homme numérique est-il capable d'actes gratuits ? À l'opposé de l'envie d'être secondé en toutes circonstances et de bénéficier d'un assistanat informatique au quotidien, on peut s'approprier la formule d'Emmanuel Kant « aies le courage de te servir de ton propre entendement ». Et, afin de favoriser cette autonomie et ce libre-arbitre, on peut exiger du droit qu'il pose des bornes afin de les protéger. L'autonomie passant du monde philosophique au monde juridique, elle pourra gagner en effectivité.

3. Le droit de disposer de soi-même

L'autonomie pourrait s'exprimer juridiquement sous la forme d'un droit des individus de disposer d'eux-mêmes. Il serait alors défendu aux services d'intelligence artificielle et à ceux qui les développent de prescrire aux individus des comportements, des modes de vie particuliers, directement ou indirectement en déployant des mécanismes de surveillance, d'évaluation ou d'incitation contraignants. En outre, les IA ne devraient pas augmenter le stress, l'anxiété ou le sentiment de harcèlement, qui sont autant de perturbateurs d'autonomie[178]. Surtout, elles ne doivent pas enfermer leurs utilisateurs dans des profils, dans des bulles, dans des identités fixées *a priori* par des traitements de données et immuables, réduisant drastiquement les choix, les options, les possibilités de développement personnel. En France, le rapport Touraine sur la révision de la loi relative à la bioéthique, dans son chapitre concernant l'intelligence artificielle, préconise en ce sens de faire prévaloir l'autonomie de l'homme au sens de l'autodétermination sur l'autonomie de décision des machines[179]. On pourrait aller jusqu'à décréter que tout instrument qui altère la capacité de jugement et de délibération personnels, se substituant à la conscience et à la liberté d'action, serait interdit, inutilisable. L'autonomie serait de la sorte le principal barrage aux mécanismes qui tendent à imposer un ordre unilatéral et infondé des choses et, pour Éric Sadin, ce serait là « une salutaire politique de légitime défense »[180].

Le droit de disposer de soi-même peut notamment prendre la forme d'un « droit à l'errance », comme le défend Gaspard Kœnig en reprenant cette expression du philosophe Isaiah Berlin commentant l'œuvre de John Stuart Mill[181]. En effet, la quête d'autonomie est au centre de la critique adressée par ce dernier à Jeremy Bentham — dont la vision utilitariste des choses influence largement le monde de l'IA industrielle, remplaçant la volonté générale en tant que cadre d'organisation du champ social. John Stuart Mill défend la capacité de chacun de chercher et tracer sa voie, diverger, changer, s'améliorer, ce qui engendre généralement des conduites sous-optimales, nuisant à l'utilité. Le bonheur et les valeurs doivent être personnels et non collectifs, résultant de l'exercice du jugement

---

178 « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

179 J.-L. Touraine, « Rapport de la mission d'information sur la révision de la loi relative à la bioéthique », Assemblée nationale, 15 janv. 2019.

180 É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'échappée, coll. Pour en finir avec, 2018, p. 106.

181 G. Kœnig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de l'observatoire, coll. De facto, 2019, p. 336 ; I. Berlin, *Liberty – Incorporating « Four Essays on Liberty »*, Oxford University Press, 2002.

et non de l'apprentissage social. On doit dès lors jouir d'un complet droit à l'autonomie ou d'un total « droit à l'errance ».

La médiation continue de l'expérience par des systèmes faisant des hommes autant de tours de contrôle continuellement informées de l'état des choses engendre une détérioration de la capacité à composer spontanément et activement avec le réel, ainsi qu'un oubli des sens à force de ne pas les solliciter. Heureux sont les ignorants, qui vivent encore en liberté quand les ultra-connectés sont en permanence informés de la misère de la condition humaine, de la nature humaine et de leur propre situation — tout cela impactant grandement l'équilibre physique et psychique général. L'homme autonome, l'homme errant, bien sûr, se cognera souvent contre le réel, car il lui manque de nombreuses clés pour pouvoir le décoder. Mais cela ne vaut-il pas mieux que de voir ce réel disparaître, remplacé par un monde artificiel dans lequel tout serait rigoureusement ordonné et extrêmement efficace ? Ne vaut-il pas mieux être vulnérable, fragile, mais souple et humain qu'omniscient, puissant, mais rigide et inhumain ? C'est bien ce qui se joue autour du droit à l'autonomie. Aristote ne disait-il pas qu' « une vie si vulnérable est pourtant la meilleure » ? Parce que je dispose du droit d'errer et que je compte bien m'en servir plutôt que me laisser asservir, je vais au devant de la vie et de ses aléas, j'expérimente, j'essaye, je me trompe, je tombe, je me relève, je commence, je recommence, je réussis, j'échoue, je ris, je pleure. C'est la vie, c'est ma vie. Doué de sensibilité et d'entendement, c'est moi qui décide.

Sans autonomie et sans droit à l'errance, il n'est pas de plaisirs et de joies véritables, de jubilations telles que celle liée à la découverte de l'inconnu ou à l'assouvissement de la curiosité, pas d'inattendu qui donne envie de « connaître la suite » et d'aventurer sa vie. Les contraires de l'autonomie et du droit à l'errance seraient la robotonomie et le droit au téléguidage. Si ces derniers triomphaient, c'est une bonne partie du sel de la vie qui serait dissoute dans une mer de pesanteur. Contre la vision du monde qui ne souffre pas l'incertitude et craint tout imprévu, une autre approche goûte au pouvoir créatif du hasard, sans lequel rien parmi la nature et parmi les hommes n'existerait, se félicite des surprises de la vie sans lesquelles celle-ci serait d'une fadeur étouffante. L'autonomie est le moyen de faire dérailler les habitudes et de s'aventurer dans de nouvelles expériences. Pour générer ces déviances sans lesquelles toute évolution, de l'évolution biologique à l'évolution technique et à l'évolution sociétale, serait impossible, il faut pouvoir assumer des choix profonds et délibérés, souvent sous-performants du point de vue de l'utilité, mais indispensables à la construction d'une individualité. C'est ainsi que l'on peut devenir soi-même.

Le droit à l'errance, à l'ère des IA, est donc le droit de vivre sans les réseaux sociaux ou de les quitter, le droit de ne pas subir de recommandations ou du moins de ne pas les suivre, le droit de refuser les notifications, en somme le droit d'exister loin des outils à visée utilitaire qui envahissent la population, le droit de tracer sa route, suivre son propre chemin et s'abstraire de toute influence, y compris soi-même. La « *privacy* », la vie privée, prend ainsi bel et bien un tour nouveau : il ne s'agit plus de protéger ses informations personnelles mais de protéger son autonomie, plus de pouvoir se placer à l'abri des regards — on peut très bien anonymiser ou crypter toutes les données —, mais de pouvoir se déterminer soi-même hors du réseau et loin des nudges. Il faut dès lors assumer les conséquences négatives de l'autonomie à court terme sous l'angle de l'utilité, pour la collectivité, et rappeler qu'elle est la condition de tout progrès futur. Éradiquer la déviance reviendrait à consacrer l'inertie comme principe de fonctionnement du monde et de la vie, alors que jusqu'à présent ce principe était l'homéostasie[182], qui s'oppose à la rigidité des idées et de leurs représentations, qui

[182] A. Damasio, *L'ordre étrange des choses*, Odile Jacob, 2017.

s'appuie sur la créativité dans laquelle Henri Bergson voyait l' « élan vital » de l'être humain. Les sentiments et les sensations ne sont pas des barrières mais des conditions des processus intellectuels, de la délibération et du jugement. Ils forment la « musique de fond du vivant » sans laquelle il devient impossible de distinguer le bien du mal ou le plaisir de la souffrance, sans laquelle il ne resterait plus qu'à s'en remettre au travail infaillible des machines. On ne peut vivre sans ces réactions chimiques, seuls à même de produire du sens. La décision instinctive, la plus primaire de toutes, n'en est pas moins une décision pleine et entière digne de l'être humain. À l'opposé d'une conception de l'humain comme étant pétri de trop de défauts pour être fiable et utile, qu'il faudrait subroger par des machines hyperefficaces, faisant bientôt de nous des êtres superflus, le droit à l'autonomie célèbre l'éclosion et l'expression des pouvoirs virtuellement infinis de chaque être, même si naviguer à vue n'est pas la solution qui permet d'arriver le plus efficacement à bon port. Pleinement se réaliser à partir de ses propres facultés, au long de ses déambulations aléatoires, est un long et périlleux voyage, mais c'est le seul voyage que l'homme puisse entamer.

La technique pourrait d'ailleurs contribuer à solutionner le problème qu'elle a créé. En effet, on voit aujourd'hui des informaticiens développer des techniques de *machine learning* en incorporant un facteur de sérendipité. L'intelligence artificielle n'est donc pas vouée à demeurer toujours anti-individuelle. Conçue ou programmée autrement, elle pourrait parfaitement favoriser l'épanouissement de l'individu, de son autonomie et de son libre arbitre en lui donnant de nouveaux outils pour bien juger et bien délibérer plutôt que de le dessaisir de ses capacités. Dès lors, à l'instar de l'évolution biologique ou sociale, l'épanouissement de chacun dépendrait de processus volontairement rendus sous-optimaux. La limite — qui est de taille — se trouve dans l'intérêt pour l'industrie numérique de développer la servitude volontaire et l'addiction au confort des IA plutôt que le goût de la liberté. Pourtant, l'excès de rationalité nuit à la rationalité et les plus grandes découvertes ont été réalisées dans des régimes libéraux, où l'imprévisibilité et la sérendipité peuvent jouer leur rôle, où il est possible de laisser libre court à son imagination et à ses expériences, quitte à aboutir à des résultats fort éloignés des objectifs initiaux. La liberté n'est pas un défaut auquel il faudrait remédier, comme semblent le penser nombre de thuriféraires de l'IA, mais un élément essentiel à l'évolution des connaissances, des sciences et des techniques[183]. Il en va de même s'agissant de l'évolution d'un individu. Le progrès ne se déclare pas *ex ante* mais se constate *ex post*.

Sur ce point, Gaspard Kœnig observe, de manière imparable, que « l'évolution naturelle illustre de manière fondamentale le rôle du hasard. J'ai demandé à mes interlocuteurs d'imaginer une intelligence artificielle chargée d'optimiser les relations entre les premières bactéries issues de la soupe primitive il y a quelques trois ou quatre milliards d'années. En cherchant à améliorer le bien-être des bactéries, en tâchant d'éliminer les anomalies, l'intelligence artificielle n'aurait-elle pas barré la route qui devait aboutir à l'humanité ? Car l'évolution est une succession d'erreurs. La sélection naturelle, loin de suivre une ligne droite, bricole à partir des matériaux qu'elle trouve à sa portée. Elle ne poursuit aucun but, elle tâtonne à la recherche du nouveau. Ainsi, les pattes existaient chez certains vertébrés aquatiques bien avant la conquête des milieux terrestres, et n'ont été “exaptées” pour la marche que dans un second temps. Les poissons ne se sont pas dit “tiens, des pattes seraient bien pratiques pour sortir de l'eau”. Ils ont utilisé un organe modérément utile, sous optimal, pour se livrer à des expérimentations innovantes. Une intelligence artificielle leur aurait probablement conseillé de se débarrasser de ces pattes encombrantes, qui ralentissaient la nage et augmentaient le

[183] K. O. Stanley, *Why Greatness Cannot Be Planned – The Myth of the Objective*, Springer, 2015.

risque d'être victime d'un prédateur. Si donc la vie n'était pas profondément mue par des mécanismes aléatoires, aucun être humain ne serait jamais apparu »[184].

C. Le droit à la différenciation numérique

1. La diversité, richesse des hommes

Les Essais de Montaigne décrivent non un homme exceptionnel, mais un homme ordinaire. Et la richesse de cet homme ordinaire est sa diversité. Montaigne écrit ainsi : « Oui, je le confesse, la seule variété me paye, et la possession de la diversité, au moins si aucune chose me paye »[185]. L'infinie variabilité de l'humain constitue pour le penseur du XVIe siècle un sujet inépuisable, ce qu'il exprime en distinguant l'homme de l'animal, beaucoup moins divers : « Plutarque dit en quelque lieu qu'il ne trouve point si grande distance de bête à bête, comme il en trouve d'homme à homme. J'enchérirais volontiers sur Plutarque et dirais qu'il y a plus de distance de tel homme à tel homme qu'il n'y a de tel homme à telle bête »[186]. Or c'est moins physiquement que psychologiquement et culturellement que les hommes diffèrent. Chacun possède sa personnalité propre, son identité caractéristique, qui constituent toute sa fortune personnelle. Or ce particularisme de chaque individu et la diversité des hommes pris collectivement sont aujourd'hui remis en cause par le monde numérique et sa vocation à standardiser les personnalités. L'industrie numérique accélère encore un peu plus la mondialisation et l'américanisation du monde. À l'époque de Montaigne, les habitants d'Europe, d'Asie, d'Afrique, d'Amérique du nord et d'Amérique du sud se ressemblaient beaucoup moins qu'aujourd'hui ; et, à l'intérieur de l'Europe ou même à l'intérieur de la France, les individus étaient beaucoup moins semblables. Ce qu'une personnalité a d'intime et d'irremplaçable, éventuellement même de contradictoire, est ce qu'elle devrait préserver avant tout. L'homme se distingue nettement de l'animal — ou n'est pas un animal comme les autres — en raison de sa capacité à penser et à s'exprimer de manière très perfectionnée, mais aussi du fait de sa diversité, de sa capacité de différenciation. Aujourd'hui, Edgar Morin s'inscrit dans la veine d'un « humanisme concret » qui protège l'universalisme et l'unité des hommes et reconnaît « les diversités humaines qui sont des formes de richesse »[187]. Consacrer un droit à la diversité ou un droit à la différenciation serait un premier pas afin de protéger cet aspect essentiel de l'humanité des hommes.

La diversité et la différenciation sont des éléments clés de l'individualisme, qui désigne une vision des choses dans laquelle les différences individuelles sont très marquées[188]. L' « ordre technico-hygiénisto-libéral »[189] promu à travers les IA et se reflétant dans nos modes de vie tend à réduire la diversité, réduire la possibilité de se différencier et donc attenter à l'individualisme. Toute volonté

---

184 G. Kœnig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de l'observatoire, coll. De facto, 2019, p. 321-322.

185 M. De Montaigne, *Les Essais*, 1580, L. III, chap. IX.

186 *Ibid.*, L. III, chap. II.

187 E. Morin, « Repenser l'humanisme dans le sens d'un universalisme concret », *NonFiction* 10 avr. 2008.

188 V° « Individualisme », in A. Lalande, *Vocabulaire technique et critique de la philosophie*, 10e éd., Puf, coll. Quadrige dicos poche, 2010, p. 499.

189 É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'échappée, coll. Pour en finir avec, 2018, p. 197.

divergente est vouée à être étouffée avant de pouvoir s'exprimer par l'ordre autoritaire des IA et des multinationales qui se trouvent derrière, privilégiant la sécurité, le confort et le profit. Plus ce système se généralise, devient massif, incontournable, plus c'est jusqu'à la capacité de penser elle-même qui se voit entravée. Les différences entre les hommes s'amenuisent ainsi progressivement, mais rapidement à l'échelle de l'histoire, ce qui doit s'analyser tel un regrettable phénomène contre lequel le droit, y compris au niveau suprême des droits de l'homme, peut agir. Éric Sadin plaide en ce sens : « À l'opposé d'une rationalité qui prétend éradiquer tous désordre, lutter contre l'entropie et asseoir une maîtrise toujours plus étendue sur le cours des choses, ce sont les imperfections de la vie jamais résolues une fois pour toutes qui stimulent notre désir de nous réaliser »[190].

Il n'est par exemple pas admissible d'abandonner la cuisine, suivant la recette de Nathan Myhrvold, ancien ponte de Microsoft, à la science des datas, ou alors l'humanité tout entière finirait par consommer les mêmes mets, dans les mêmes quantités, aux mêmes horaires, et les goûts s'épuiseraient peu à peu. La défense de la liberté est donc forcément en même temps une défense de la diversité. Et cette dernière, à l'instar de la défense de la vie privée conçue comme autonomie et comme libre arbitre, permet de mettre l'accent sur des zones moins balisées, si ce n'est des angles morts, des discussions éthico-juridiques autour de l'IA. Le thème de la diversité humaine et sociale, jusqu'à présent, est resté dans l'ombre. Ce thème est pourtant un thème essentiel à une époque où le mode de fonctionnement des réseaux sociaux a tendance à la fois à enfermer les individus dans des « bulles de filtre » — sur Instagram, par exemple, les clichés proposés dans l'onglet de recherche ne sont pas choisis au hasard mais correspondent aux likes, interactions et recherches de l'utilisateur —, ainsi qu'à uniformiser les sociétés et les cultures en favorisant les conduites et les opinions « normales » car ce sont les plus aisées à monétiser. La normalisation est donc l'un des effets puissants de l'utilisation des technologies numériques. Et, en épousant les comportements des internautes, les algorithmes des réseaux sociaux tendent à maintenir les inégalités sociales et culturelles en offrant à chacun des contenus « à son niveau » plutôt que l'accès à de nouvelles ressources permettant de s'enrichir et d'entrevoir d'autres opportunités. Les compteurs du web social obligent chacun à construire son espace (im)personnel selon ses origines sociales et culturelles.

2. Lutter juridiquement contre l'uniformisation

L'article 1er de la Déclaration universelle de l'UNESCO sur la diversité culturelle, adoptée en novembre 2001 et reprise dans la Convention sur la protection et la promotion de la diversité des expressions culturelles de 2005, érige la diversité culturelle au rang de « patrimoine commun de l'humanité ». Une éventuelle déclaration des droits de l'homme numérique devrait en faire de même. Elle devrait, comme la Déclaration de Montréal, reconnaître que « le développement et l'utilisation de services d'intelligence artificielle doivent être compatibles avec le maintien de la diversité sociale et culturelle et ne doivent pas restreindre l'éventail des choix de vie et des expériences personnelles »[191]. Avec le droit à la diversité et à la différenciation et leurs déclinaisons en droit positif, il s'agirait donc de promouvoir un développement et une utilisation de l'IA ne conduisant pas à une uniformisation de la société par la normalisation des pensées et des comportements.

---

[190] *Ibid.*, p. 246.

[191] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

Très significativement, la Cour européenne des droits de l'homme a déduit du droit à la protection de la vie privée la reconnaissance au profit de chaque individu d'une « capacité à être soi-même ». Pour la Cour, « sur le terrain de l'article 8 de la Convention, où la notion d'autonomie personnelle reflète un principe important qui sous-tend l'interprétation des garanties de cette disposition, la sphère personnelle de chaque individu est protégée, y compris le droit pour chacun d'établir les détails de son identité d'être humain »[192]. Cette conception de la vie privée et de l'autonomie trouve tout spécialement à s'appliquer aujourd'hui, dans le lisse monde numérique. En tant que protection des choix identitaires de la personne, elle a déjà mené à des adaptations majeures de notions juridiques traditionnelles, qu'il s'agisse de l'état civil ou de la vie amoureuse, sexuelle et familiale[193]. En France, en 1977, une commission spéciale des libertés présidée par Edgar Faure avait d'ailleurs déjà rédigé une proposition de loi constitutionnelle qui consacrait notamment un « droit à la différence ». Ce texte retenait que « la République française, une et indivisible, reconnaît et protège la diversité des cultures, des mœurs et des genres de vie. Chacun a le droit d'être différent et de se manifester comme tel ». De nos jours, pour de nouvelles raisons, un tel projet mérite d'être remis sur la table et, au moins, débattu.

Il est possible d'utiliser les services numériques de manière à protéger son intimité et, par suite, demeurer à l'abri de l'influence des IA. On peut ainsi, avec quelques connaissances concernant le fonctionnement du web, changer de cookies pour plus de confidentialité — la fonctionnalité « cookies SameSite » de Google Chrome empêche les domaines tiers de créer des fichiers de cookies lorsque l'utilisateur n'est pas sur leur site web, ce qui préserve donc du risque d'être suivi partout en ligne. On peut aussi désinstaller un maximum d'applications et programmes superflus, désactiver les données de géolocalisation, réinitialiser les permissions accordées aux services, chiffrer les contenus et les communications, utiliser un VPN, paramétrer un pare-feu bloquant les connexions entrantes indésirables, utiliser un navigateur web et un moteur de recherche respectueux de la vie privée, installer des systèmes d'exploitation et des logiciels libres et ouverts comme LineageOS sans services Google sur Android et Linux sur ordinateur ou encore supprimer ses comptes sur Facebook et sur tous les réseaux sociaux[194]. Ainsi l'utilisateur des services numériques pourra-t-il reprendre le contrôle sur lui-même et, par suite, protéger la diversité tant en tant que contributeur à cette diversité qu'en tant que consommateur de diversité. Moins de données collectables et des données plus difficilement rattachables à la personne, c'est un profil plus délicat à établir, une personnalisation moins efficace ou même inexistante et, donc, une préservation de l'accès aux multiples visages de l'existence.

À moins de vivre totalement déconnecté, protéger son intimité sur internet demeure une tâche sisyphéenne. Il faudrait donc aller bien plus loin et, comme pour la plupart des droits de l'homme numérique, l'IA pourrait parfaitement être utilisée afin de rendre plus effectif le droit à la diversité et à la différenciation. Les préférences personnelles pourraient être des critères clés du fonctionnement d'une IA, qui pourrait régulièrement inviter l'utilisateur à les modifier au gré de ses envies et lui proposer des alternatives, des plus classiques aux plus rares. On disposerait ainsi d'un outil précieux pour forger sa personnalité. Le projet « Algodiv », par exemple, vise à lutter contre

---

[192] CEDH, 11 juill. 2002, n° 28957/95, *Goodwin c. Royaume-Uni*.

[193] S. Hennette-Vauchez, D. Roman, *Droits de l'homme et libertés fondamentales*, Dalloz, coll. Hypercours, 2017, p. 523.

[194] P. Crochart, « Quels outils pour protéger vos données personnelles ? », clubic.com, 13 mars 2020 (qui conseille de consulter le site prism-break.org : tirant son nom du célèbre programme de surveillance mondiale de la NSA, ce site liste un grand nombre d'alternatives sécurisées aux applications et services les plus populaires).

l'enfermement qu'induisent les systèmes de recommandation sur le web en promouvant la diversité et en stimulant la sérendipité. Mais encore faudrait-il qu'en amont les modèles statistiques n'aient pas déjà fait disparaître les goûts et les couleurs alternatifs. Le droit pourrait ainsi imposer que le développement des IA prenne en compte les multiples expressions des sociétés et des cultures, ainsi que des sous-sociétés et des sous-cultures, cela « *by design* », dès la conception des outils informatiques. Les développeurs d'IA, en particulier dans le secteur industriel et marchand, devraient se voir imposer de refléter la diversité des individus et des groupes et de permettre à cette diversité de s'exprimer. L'IA peut être une chance pour la diversité à condition de l'intégrer dès le départ en tant que principe essentiel. Les bénéfices de l'intelligence artificielle peuvent être importants s'ils prennent en compte les spécificités des multiples contextes culturels locaux. En même temps, l'IA doit être conçue pour ouvrir l'individu à la pluralité des cultures, des goûts et des natures humaines et ne pas fermer les identités personnelles sur elles-mêmes par le traitement des données passées et renforcer certains préjugés liés aux différences sociales, sexuelles, ethniques, religieuses ou autres. Plutôt que d'appauvrir l'humain, elle peut très bien fêter cette pluralité, cette multiplicité qui, selon Montaigne, constitue la plus grande richesse de l'être humain « ondoyant et divers ».

Mireille Delmas-Marty, constatant combien la puissance des nouvelles technologies numériques (et biologiques) pourrait engendrer un « homme profilé, augmenté, fabriqué, voire transformé » propose de retenir le principe de créativité parmi les quatre principes constitutifs de l'humanité qu'elle identifie[195]. L'indétermination, constitutive de l'humanité et devant à ce titre être prioritairement protégée, en dépend en effet. Les IA doivent permettre, non pas de nier la part irréductible de chacun, mais de la développer, de lui permettre de s'exprimer pleinement, renforçant la singularité et la richesse de tout homme. En 1984, lors du lancement du Mac, le clip publicitaire imaginé par Apple présentait un monde aux logiques inversées par rapport à la dystopie 1984 de George Orwell. « *Think different* », enjoignait la marque à ses clients, tandis que Steve Jobs, inspiré par le nouvel esprit du capitalisme, associait à sa présentation des portraits de Picasso, Einstein ou Bob Dylan. L'informatique devait alors servir la créativité, l'épanouissement, l'affranchissement personnel grâce à l'ordinateur. Et ce dernier était présenté comme un outil protégeant l'individu, le plaçant à l'abri de tout fichage à large échelle que l'informatisation des sociétés débutante faisait craindre. C'est avec cette conception de l'informatique — à condition qu'elle ait été sincère — que le droit à la diversité et à la différenciation invite à renouer. N'y aurait-il pas un public large pour des outils protecteurs des droits humains et en premier lieu de la liberté individuelle dans toutes ses déclinaisons ? N'est-il pas temps de renouer, y compris dans les textes juridiques, avec une forme d'humanisme romantique qui exalte les droits de la subjectivité créatrice et qui se passionne pour la diversité des peuples, ces forces qui donnent leur énergie à la nature comme à l'histoire des hommes ? Les IA sont en quelque sorte les concierges de l'époque actuelle, ces méticuleux gardiens de l'ordre à qui tout dysfonctionnement semble un crime et qu'Albert Camus n'appréciait guère, leur conférant souvent un mauvais rôle dans ses romans.

[195] M. Delmas-Marty, *Sortir du pot au noir – L'humanisme juridique comme boussole*, Buchet-Chastel, 2019, p. 90.

III. Égalité

Question 1. L'IA a-t-elle été conçue dans le but d'éviter toute forme de discrimination, d'être neutre et d'être transparente ?

Question 2. Des mesures ont-elles été prises afin d'assurer la traçabilité de l'IA (méthode de conception, programmation, données d'entraînement utilisées, méthode d'entraînement, méthode de test et de validation du système) ?

Question 3. Est-ce qu'une stratégie ou un ensemble de procédures a été mis en place afin d'éviter de créer ou de renforcer des biais injustes dans le système d'IA ?

Question 4. Des processus permettant de tester et de contrôler les biais éventuels du système sont-ils prévus ?

Question 5. Est-ce que l'absence de biais injustes dans les jeux de données utilisés a été vérifiée ?

Question 6. Les jeux de données utilisés sont-ils représentatifs de la diversité des utilisateurs ?

Question 7. Le système d'IA a-t-il été conçu dans le but de faciliter au maximum la compréhension et l'interprétation des résultats ?

Question 8. Les résultats fournis par l'IA et les décisions prises sur leur fondement peuvent-ils être expliqués et compris par les utilisateurs ?

A. Le droit à la non-discrimination numérique

1. L'égalité devant la loi des IA

« À l'opposé d'une rationalité qui génère une furie innovatrice concourant à l'extension de son empire et contribuant à l'instauration d'un utilitarisme généralisé, écrit Éric Sadin, nous nous refusons à compter indéfiniment un gain dans notre rapport au réel et aux autres, mais cultivons les pouvoirs de notre inventivité en vue d'expérimenter de multiples modes d'existence participant de notre épanouissement individuel et collectif »[196]. Le droit pourrait aller dans ce sens et, limitant les possibilités de personnalisation, il préviendrait en même temps les risques de discrimination. La personnalisation met en effet à l'épreuve le principe d'égalité, et inversement. L'intelligence artificielle entre souvent en conflit avec l'égalité. Certes, en France, le Conseil constitutionnel a souligné que l'égalité ne s'oppose pas à ce « que le législateur règle de façon différente des situations différentes ni à ce qu'il déroge à l'égalité pourvu que, dans l'un et l'autre cas, la différence de traitement qui en résulte soit en rapport direct avec l'objet de la loi qu'il établit »[197]. Le problème avec les IA est qu'elles risquent de reproduire des biais, donc non de traiter différemment des cas objectivement différents mais des cas pourtant semblables. Or, si « la France assure l'égalité devant la loi de tous les citoyens », comme le proclame l'article premier de la Constitution du 4 octobre 1958, il est temps qu'elle s'enquiert aussi de l'égalité devant la loi de l'IA. Ce principe est d'ailleurs

[196] É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'échappée, coll. Pour en finir avec, 2018, p. 247.

[197] Cons. const., déc. n° 96-375 DC, 9 avr. 1996, *Transferts d'entreprise du secteur privé*.

censé s'imposer aux personnes privées comme aux institutions publiques. Mais les pratiques discriminatoires peuvent cependant se justifier par des situations différentes. On traite alors différemment des personnes différentes afin de diminuer ces dernières différences. L'égalité des moyens laisse la place à l'égalité des résultats. Le contrôle opéré par le juge s'opère en deux temps[198]. En premier lieu, il observe si la norme ou pratique discriminatoire peut s'expliquer par une différence de situation. Ensuite, si tel est le cas, il vérifie que cette différence de traitement est proportionnée avec le but poursuivi. On peut donc très bien identifier au moyen d'une IA de nombreuses différences de situation entre des personnes et leur appliquer les traitements les plus appropriés. Si l'égalité s'oppose à ces traitements différents justifiés par des réalités différentes, l'équité prend son relais. Tel n'est bien sûr plus le cas lorsque ces traitements différents sont les résultats de manipulations ou d'erreurs, donc ne correspondent pas à de véritables écarts dans les situations des personnes. La nature des données en cause de même que les modalités de leur récolte et de leur traitement n'offrent trop souvent pas les garanties suffisantes pour justifier une différence de traitement. La problématique est ici celle de la fiabilité insuffisante de la technique pour pouvoir l'autoriser à produire de tels effets.

Loin de toute « égalité devant la loi », la loi des IA produit par principe des discriminations puisque l'objet même des algorithmes est de traiter différemment chaque utilisateur, de lui fournir des contenus personnalisés qui vont l'amener à adopter une conduite particulière. Ainsi l'automaticité et la rigidité s'associent-elles à la personnalisation ; et ce mélange peut potentiellement produire des résultats insatisfaisants qui ne seront l'objet d'aucun contrôle. En France, l'article 225-1 du Code pénal incrimine toute forme de discrimination fondée notamment sur l'origine, le sexe, la situation de famille, l'apparence physique, le patronyme ou le lieu de résidence. Il semble pourtant que le principe même de la loi des IA soit d'opérer des discriminations fondées sur l'origine, le sexe, la situation de famille, l'apparence physique, le patronyme ou le lieu de résidence. Dans les faits, les algorithmes n'ont de cesse de discriminer les utilisateurs des services sur la base de critères illicites[199]. Qu'on pense à l'exemple de la voiture autonome qui se retrouve dans une situation catastrophique, obligée de renverser soit à gauche deux enfants, soit à droite trois personnes âgées. L'ordinateur devra choisir en pondérant le nombre de victimes, leurs âges, leurs conditions physiques et mentales, et même, dans certains pays, sait-on jamais, leurs sexes. Un tel choix de l'IA est extrême, mais c'est en permanence qu'elle produit des résultats à l'aune des différences entre les personnes.

Les droits de l'homme numérique pourraient consacrer l'obligation de concevoir et entraîner les services d'IA de manière à ne pas créer, renforcer ou reproduire des discriminations fondées entre autres sur des différences sociales, sexuelles, ethniques, culturelles ou religieuses. Le recours à des algorithmes et des techniques probabilistes devrait toujours être porté à la connaissance des personnes concernées qui pourraient s'opposer à la construction et à la diffusion de profils les concernant[200]. Une IA digne de confiance suppose d'offrir aux citoyens la possibilité de refuser que leurs données soient utilisées à leur encontre à des fins préjudiciables ou discriminatoires[201]. Avec

198 L. Favoreu et *alii*, *Droit constitutionnel*, 17e éd., Dalloz, 2015, p. 1030.

199 Plus généralement, O. Itéanu, *Quand le digital défie l'État de droit*, Eyrolles, 2016.

200 Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et Commission sur les droits et devoirs sur internet de la Chambre des députés italienne, déclaration commune, 28 sept. 2015.

201 Commission européenne, « Lignes directrices en matière d'éthique pour le développement et l'utilisation d'une IA », 8 avr. 2019.

le principe d'égalité, tous les humains devraient être identiquement admissibles à toutes dignités, places et emplois, selon leurs capacités et sans autre distinction que celle de leurs vertus et de leurs talents. Aucun préjugé ni stéréotype ne devrait prospérer dans un système d'IA et finalement impacter le sens de vies humaines.

Historiquement, l'affirmation selon laquelle la loi doit être la même pour tous est importante tant on a connu nombre de législations et autres règles juridiques opérant précisément de telles distinctions, depuis l'esclavage jusqu'au Code de l'indigénat en passant par le statut des Juifs sous Vichy[202]. Alors que les distinctions fondées sur la race ont heureusement disparu de la loi, l'IA menace de les faire réapparaître. Si, comme l'enseigne depuis de nombreuses années le courant américain de la *Critical Race Theory*, la race demeure un concept éclairant en droit, par exemple s'agissant du droit de la propriété largement marqué par l'histoire de l'esclavage[203] ou, de nos jours, de la contraception, de l'avortement ou de la génétique qui sont autant de questions « racialisées »[204], rien ne saurait aujourd'hui justifier de traiter différemment des individus en fonction de leurs origines — comme l'a déclaré la Déclaration des Nations Unies sur l'élimination de toutes les formes de discrimination raciale du 20 novembre 1963. Pourtant, le risque est grand de voir des IA, souvent involontairement, opérer des choix « racistes ». Cela s'explique par le fait que, si les théories biologisantes racistes du début du XXe siècle ne sont désormais plus soutenables, les différences de races, comprises comme constructions sociales, continuent d'expliquer la réalité des sociétés contemporaines, si bien que les systèmes d'IA sont amenés à les refléter.

2. La prohibition de toutes les discriminations

En droit français, c'est la loi Pleven du 1er juillet 1972 relative à la lutte contre le racisme qui a constitué la première consécration explicite de la notion de discrimination. Elle a notamment modifié la loi du 29 juillet 1881 sur la presse, ainsi que le Code pénal, afin de sanctionner les discriminations raciales. Puis, dans les années 1990, notamment avec le nouveau Code pénal, l'arsenal pénal de la lutte contre les discriminations s'est sensiblement enrichi, à tel point qu'on a vu dans le droit de la non-discrimination une nouvelle branche du droit[205]. Que ce soit dans le commerce, dans l'emploi ou même dans les relations entre administration et administrés, le combat pour l'égalité entre les individus ou entre les citoyens s'est renforcé. Au niveau de la fonction publique, la loi du 16 novembre 2001 relative à la lutte contre les discriminations a ajouté dans le statut général de la fonction publique un dispositif antidiscriminatoire. Depuis les années 2000, c'est surtout dans le cadre de l'Union européenne que se joue le droit de la non-discrimination. La jurisprudence de la Cour de justice a joué un rôle majeur, faisant notamment du principe de non-discrimination à raison du sexe un principe général du droit communautaire[206]. Par exemple, la Cour a considéré que le fait que les travailleurs à temps partiel se voient refuser le bénéfice d'un régime complémentaire constitue une discrimination indirecte à l'égard des femmes dès lors que ce sont elles qui ont plus souvent recours au temps partiel[207]. Quant aux traités, alors que la non-discrimination concernait initialement l'égalité salariale entre les hommes et les femmes, en vertu

[202] D. Lochak, *Le droit et les paradoxes de l'universalité*, Puf, coll. Les voies du droit, 2010.
[203] C. Harris, « Whiteness as Property », *Harvard Law Review* 1993, vol. 8, n° 106, p. 1710 s.
[204] D. Roberts, *Killing the Black Body – Race, Reproduction and Liberty*, Random House-Pantheon, 1997.
[205] J.-L. Halpérin, E. Fassin, dir., *Discriminations : pratiques, savoirs, politiques*, La documentation française, 2009.
[206] CJCE, aff. C-43/75, 8 avr. 1976, *Defrenne*.
[207] CJCE, aff. C-170/84, 13 mai 1986, *Bilka Kaufhaus*.

de l'article 199 du Traité de Rome, le traité d'Amsterdam a ajouté une clause générale de non-discrimination à travers l'article 13 : le Conseil « peut prendre les mesures nécessaires en vue de combattre toute discrimination fondée sur le sexe, la race ou l'origine ethnique, la religion ou les convictions, un handicap, l'âge ou l'orientation sexuelle »[208]. Enfin, des directives européennes de l'an 2000 sont venues compléter cet important corpus juridique. La directive n° 2000/43/CE interdit les discriminations fondées sur la race ou l'origine ethnique, tant pour le secteur public que pour le secteur privé, en ce qui concerne l'accès à l'emploi, à la formation, aux prestations sociales, à l'éducation, à la fourniture de biens et services. Ce texte exige, en outre, des États qu'ils instituent des Equality Bodies, institutions nationales chargées de la lutte contre les discriminations — à l'image, en France, de la Haute Autorité de Lutte contre les Discriminations et pour l'Égalité (HALDE), créée par la loi n° 2004-1486 du 30 déc. 2004 et dont les compétences ont été transférées en 2008 au Défenseur des droits. Quant à la directive n° 2000/78/CE, elle interdit les discriminations en matière d'emploi à partir de huit critères : race, sexe, orientation sexuelle, convictions, origine ethnique, handicap, religion, âge. La lutte contre les discriminations donne donc lieu à un corpus juridique fourni et relativement précis. En France, il faut ajouter les nombreuses lois qui sont venues, dans les années 2000 et 2010, allonger la liste des critères de distinction prohibés : les caractéristiques génétiques[209], la grossesse[210], le lieu de résidence[211], l'identité sexuelle ou identité de genre[212], la vulnérabilité résultant de la situation économique[213] ou même l'incapacité à s'exprimer dans une langue autre que le français[214].

La non-discrimination suppose que l'on ne peut utiliser certaines distinctions pour opérer des choix entre des personnes. Si le droit s'en saisit, ces distinctions deviennent non seulement illégitimes mais aussi illégales. Les discriminations les plus absolument prohibées sont celles qui touchent à la race, l'origine et les croyances, comme l'exprime l'article premier de la Constitution. De nombreuses dispositions juridiques interdisent les discriminations, dans le sillage de l'article 7 de la Déclaration universelle des droits de l'homme selon lequel « Tous sont égaux devant la loi et ont droit sans distinction à une égale protection de la loi. Tous ont droit à une protection égale contre toute discrimination qui violerait la présente Déclaration et contre toute provocation à une telle discrimination ». Cela concerne naturellement les IA. Le droit positif s'oppose aux discriminations, surtout en matière d'emploi, de logement, d'éducation et d'accès aux biens et services. Dans ces cas, ce sont déjà des discriminations indirectes qui risquent de se produire, à savoir, selon l'article 1er de la loi du 27 mai 2008 *Portant diverses dispositions d'adaptation au droit communautaire dans le domaine de la lutte contre les discriminations*, « une disposition, un critère ou une pratique neutre en apparence, mais susceptible d'entraîner, pour l'un des motifs mentionnés au premier alinéa, un désavantage particulier pour des personnes par rapport à d'autres personnes, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifié par un but légitime et que les moyens pour réaliser ce but ne soient nécessaires et appropriés ». En France, les articles 225-1 et

---

[208] E. Dubout, *L'article 13 du traité CE – La clause communautaire de lutte contre les discriminations*, Bruylant, 2006.

[209] L. n° 2002-303, 4 mars 2002, *Relative aux droits des malades et à la qualité du système de santé*.

[210] L. n° 2006-340, 23 mars 2006, *Relative à l'égalité salariale entre les femmes et les hommes*.

[211] L. n° 2014-173, 21 févr. 2014, *De programmation pour la ville et la cohésion urbaine*.

[212] L. n° 2012-954, 6 août 2012, *Relative au harcèlement sexuel* ; L. n° 2016-1547, 18 nov. 2016, *De modernisation de la justice du XXIe siècle*.

[213] L. n° 2016-832, 24 juin 2016, *Visant à lutter contre la discrimination à raison de la précarité sociale*.

[214] L. n° 2016-1547, 18 nov. 2016, *De modernisation de la justice du XXI e siècle*.

225-2 du Code pénal ou encore l'article L. 122-45 du Code du travail prohibent ainsi toute forme de discrimination.

L'article L. 1132-1 du Code du travail interdit d'écarter un salarié ou un candidat d'une procédure de recrutement ou de l'accès à un stage ou à une période de formation en entreprise, de le sanctionner ou de le licencier ou de lui infliger toute autre mesure discriminatoire, directe ou indirecte, en se fondant sur des critères ou des préconçus discriminatoires. Or ce sont là justement des domaines dans lesquels on est de plus en plus tenté de faire appel à des outils algorithmiques. Si le fruit de la décision résultant de l'algorithme revient statistiquement à occasionner un désavantage pour une catégorie de salariés en raison de leur âge, de leur état de santé, ou de tout autre motif discriminatoire légal, alors la discrimination pourra être retenue sans que l'employeur ne puisse tirer argument de l'objectivité des données traitées par l'algorithme de GRH[215]. Pour s'exonérer de sa responsabilité, il faudrait alors pouvoir prouver que l'on a poursuivi un but légitime au moyen d'instruments proportionnés. Quant au salarié ou candidat lésé, il devrait pouvoir engager une action en nullité de la mesure discriminatoire et demander le versement de dommages-intérêts afin de réparer son préjudice. Au niveau européen, le Comité économique et social européen juge dans le même sens que le recours à des IA ne saurait aller sans « la protection des droits et libertés concernant le traitement des données des travailleurs [et] le respect des principes de non-discrimination »[216]. Face au risque que des algorithmes provoquent, perpétuent ou renforcent des biais sociaux, le CESE préconise de former à l'éthique les ingénieurs, informaticiens et autres concepteurs d'IA, mais aussi d'instituer des entités de contrôle[217].

Par ailleurs, le droit français, aux côtés des discriminations absolument prohibées (celles qui se rapportent à la race, aux origines ou aux croyances), consacre des discriminations relativement prohibées. Cela signifie que certains critères de distinction peuvent être utilisés par la loi, par exemple à des fins de « discrimination positive », mais pas par des acteurs privés. On comprend alors que les outils algorithmiques, sauf s'ils sont déployés par des administrations publiques, ne sauraient dans aucun cas procéder à des discriminations à l'occasion de relations privées. Ce sont donc bien toutes les formes de discrimination qui se trouvent défendues aux IA privées, qu'elles touchent à l'âge, l'apparence physique, le sexe, l'identité de genre, le handicap, l'activité syndicale, le lieu de résidence etc.

Cela, étant donné le mode de fonctionnement naturel de ces techniques, est une grande limite aux possibilités de leur développement. La lutte contre les discriminations donne ainsi lieu à de redoutables conflits entre liberté et égalité — même si dans de nombreux cas les IA malmènent l'une et l'autre. Plus la liste des critères de distinction prohibés s'allonge, plus la libre détermination, par les individus, des personnes avec lesquelles ils vont contracter, interagir, cohabiter, travailler etc. est réduite. Dans l'idéal, le droit devrait prendre un tour plus fin et précis afin de distinguer les situations dans lesquelles des discriminations sont trop peu graves pour devoir être interdites et celles où il doit intervenir afin de les prohiber, en raison de leur impact sur la vie des personnes. La liberté même signifie pouvoir choisir en fonction des critères que l'on estime utiles ou pertinents. Finalement, la liberté est notamment une liberté de discriminer. Mais le principe d'égalité triomphe de cette liberté là — sans pour autant que cela soit indiscutable, comme en a témoigné la façon avec

[215] L. Malfettes, « Gestion du personnel par algorithmes et droits du salarié », *Droit social* 2019, p. 591.

[216] CESE, « L'intelligence artificielle : anticiper ses impacts sur le travail pour assurer une transition équitable », avis, 19 sept. 2018.

[217] *Ibid*.

laquelle la Cour de justice de l'Union européenne a jugé que la liberté d'entreprendre comprenait le choix fait par l'employeur d'une neutralité (religieuse, philosophique, politique) de l'image de l'entreprise et ne contrevenait pas, ce faisant, au droit de la non-discrimination : ici, la Cour a préféré protéger la liberté des acteurs sociaux, en reconnaissant aux employeurs le pouvoir d'imposer la neutralité[218].

3. Le droit des données personnelles comme garde-fous

L'IA oblige à prendre garde, plus encore que par le passé, aux risques de discrimination accompagnant les procédés de personnalisation. Dans le secteur des assurances, par exemple, la lutte contre les discriminations justifie l'interdiction de s'appuyer sur des données de santé, ce qui limite nettement le travail des algorithmes de personnalisation concernant certaines polices d'assurance[219]. A titre d'exemple, la Cour de justice de l'Union européenne interdit l'utilisation du genre comme variable au sein d'un modèle statistique en assurance au nom du principe d'égalité de traitement entre les hommes et les femmes[220]. Or une telle interdiction pourrait être malmenée par des IA capables d'inférer par elles-mêmes les critères permettant d'identifier le genre.

D'un côté, la combinaison des données et des algorithmes pourrait conduire à une plus grande diversité au nom d'une société dite « inclusive », cela au terme de procédés dits de « discrimination positive » — des pratiques qui, si elles visent à supprimer une discrimination subie par un groupe de personnes en les faisant bénéficier d'un traitement préférentiel, sont toutefois précisément encadrées sur le plan juridique[221]. Mais il convient avant tout de prévenir et atténuer les possibles « discriminations négatives » de certains groupes qui présentent un risque accru de voir leurs droits affectés par l'IA de manière disproportionnée. Sont concernés les femmes, les enfants, les personnes âgées, les personnes économiquement défavorisées, les membres de la communauté LGBTI, les personnes handicapées et les groupes « raciaux », ethniques ou religieux. Cela devrait justifier d'interdire le recours à des IA pour décider ou pour aider à décider dans les domaines les plus problématiques, y compris dans les services publics. Des mesures de contrôle et d'audit devraient être imposées afin d'identifier toute dérive vers des effets discriminatoires lorsque le recours à l'IA ne serait pas défendu.

En matière commerciale, si une entreprise peut se procurer les informations nécessaires concernant ses clients de manière à leur recommander des biens et services personnalisés, on pourrait s'inspirer du *Customer Privacy Bill of Rights* de 2012, proposition américaine visant à protéger les données des consommateurs et, par suite, à les préserver des personnalisations, surtout lorsqu'elles tournent à la discrimination[222]. Pour l'heure, le profilage est possible, malgré le RGPD qui se borne à l'encadrer. Or le profilage, par définition, est une forme de discrimination reposant sur des critères plus ou moins justes et pertinents. Le consentement au profilage devrait être non seulement obligatoire mais aussi effectif, éclairé. Or les législateurs et les juges sont susceptibles de mal réaliser quels sont les enjeux et les menaces. Un tel décalage pourrait donner un regrettable chèque en blanc aux technologies discriminantes. L'arrêt rendu le 12 avril 2018 par la chambre sociale de

---

[218] CJUE, C-157/15, 14 mars 2017, *Achbita c. G4S Secure Solutions*.

[219] L. n° 2016-41, 26 janv. 2016, *De modernisation de notre système de santé*.

[220] CJUE, aff. C-236/09, 1er mars 2011, *Test-Achats*.

[221] I. Desbarats, « Le recrutement à l'ère de l'IA : l'éthique au secours du droit ? », *RLDA* 2019, n° 153.

[222] B. Ancel, « La vie privée dans un monde digitalement connecté : la démocratie en danger ? », *RLDI* 2019, n° 159, p. 34.

la Cour de cassation témoigne de la perception décalée de certaines juridictions par rapport aux conséquences du recours aux IA sur les conditions de travail des salariés. Ici, les magistrats ont estimé que la consultation du comité d'hygiène, de santé et des conditions de travail (CHSCT) n'était pas justifiée dans la mesure où l'utilisation de l'IA Watson « se traduit en termes de conséquences mineures dans les conditions de travail directes des salariés dont les tâches vont se trouver facilitées »[223]. Pourtant, l'impact de l'IA sur les conditions de travail dépasse de beaucoup ce simple constat. Une telle décision est un mauvais signal à une époque où les entreprises s'apprêtent à utiliser massivement l'intelligence artificielle[224]. Les représentants élus du personnel doivent être pleinement impliqués dans les décisions qui vont changer profondément et durablement l'environnement de travail des salariés.

Face à la possible gestion algorithmique des personnels, le droit des données personnelles mérite d'être rigoureusement respecté. Les salariés vont voir de plus en plus de leurs données être collectées et traitées afin de servir les nouveaux outils de gestion des carrières, recrutements, formations, paies, horaires, annuaires, organigrammes, actions sociales etc. Si cela peut leur être profitable, le droit doit veiller à ce que les aspects pernicieux ne prospèrent pas, à l'exemple du suivi et du contrôle continu grâce à des objets connectés que chacun devrait porter en permanence. Les données de géolocalisation, comme les autres, ne sauraient être exploitées au mépris des droits des intéressés. Or, dans le Code du travail, les articles L. 1222-3 et L. 1222-4 posent déjà le principe selon lequel aucune information concernant personnellement un salarié ne peut être collectée par un dispositif qui n'a pas été porté préalablement à sa connaissance, de même qu'il doit être expressément informé, préalablement à leur mise en œuvre, des méthodes et techniques d'évaluation professionnelle utilisées à son égard. De plus, le comité social et économique (CSE) doit être informé, préalablement à leur utilisation, concernant les méthodes ou techniques d'aide au recrutement des candidats à un emploi ainsi que sur toute modification de celles-ci, sur les traitements automatisés de gestion du personnel et sur toute modification de ceux-ci préalablement à leur mise en œuvre. Le même CSE doit être informé et consulté en cas d'introduction de nouvelles technologies modifiant les conditions de santé et de sécurité, ainsi qu'au sujet des moyens ou techniques permettant un contrôle de l'activité des salariés, cela en vertu de l'article L. 2312-38 du Code du travail. Le non-respect de cette obligation expose l'employeur au délit d'entrave. Du côté de la Cour européenne des droits de l'homme, il a été jugé que le suivi des paramètres de communication des salariés (contenus, fréquences, destinataires) opéré via les outils de l'entreprise constitue une atteinte aux droits fondamentaux des travailleurs, qui doit être évaluée par conséquent en fonction du but légitime ou non qu'elle poursuit et de la proportionnalité de l'atteinte qu'elle engendre[225].

S'agissant des procédures de recrutement, l'article L. 1221-8 du Code du travail implique que les données collectées soient strictement nécessaires à l'évaluation des capacités du candidat à occuper le poste proposé. Dès lors, les formulaires de candidature ne peuvent imposer, par exemple, d'indiquer sa situation matrimoniale ou sa paternité[226]. Un employeur doit tout spécialement veiller au respect des exigences tenant à la licéité des traitements de données de ses collaborateurs et candidats à l'embauche, y compris — et surtout — si ces opérations sont automatisées au moyen d'algorithmes. En premier lieu, il importe de se trouver dans l'une des bases légales limitativement

[223] Cass. soc., n° 16-27.866, 12 avr. 2018.

[224] G. Loiseau, « Intelligence artificielle et conditions de travail des salariés : un impact à prendre au sérieux », *Dalloz IP/IT* 2018, p. 437.

[225] CEDH, n° 588/13, 22 févr. 2018, *Libert c. France*.

[226] CNIL, « Le recrutement et la gestion du personnel », 2018.

énumérées par le RGPD, suivant le mécanisme inauguré par la loi « Informatique et libertés » française : le « consentement de la personne concernée, dans les conditions mentionnées au 11 de l'article 4 et à l'article 7 du règlement [européen] », la nécessité de « l'exécution d'un contrat auquel la personne concernée est partie ou à l'exécution de mesures précontractuelles prises à la demande de celle-ci », le « respect d'une obligation légale à laquelle le responsable du traitement est soumis », la « sauvegarde des intérêts vitaux de la personne concernée ou d'une autre personne physique », l' « exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement » ou encore les « fins des intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel, notamment lorsque la personne concernée est un enfant ». Respecter cette disposition, c'est déjà prévenir en partie les risques de discrimination. Cela d'autant plus que le consentement du salarié au traitement de ses données, comme l'analyse depuis longtemps la CNIL, mérite d'être exclu en raison du lien de subordination qui le lie à son employeur et qui amène à douter du caractère libre de son consentement. Aujourd'hui, le RGPD confère une importance décisive à la liberté de choix de l'intéressé — le consentement donné doit être libre, donné spécifiquement pour le traitement pour lequel il est sollicité, éclairé, recueilli sans ambiguïté. Il en résulte que la plupart des traitements de données ne peuvent pas reposer sur le consentement du salarié. On prévient aussi les discriminations à la base lorsqu'on opère le contrôle de finalité : la collecte et le traitement des données par l'employeur doivent répondre à des objectifs légitimes eu égard à l'activité de l'entreprise. Et les données doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Respecter le droit des données personnelles est la première des exigences afin de limiter les discriminations[227]. Tel est encore le cas, par exemple, des données sensibles telles que l'appartenance syndicale, qui ne peuvent être collectées et qui, si elles l'étaient, menaceraient de nourrir des discriminations. Dans le cadre d'un recrutement, la collecte de données n'est concevable qu'à condition que ces données ne servent qu'à évaluer la capacité du candidat à occuper l'emploi (qualification, expérience).

4. Égalité mais pas égalitarisme : la nécessité de certains biais

Dans le secteur privé, Accenture, comme d'autres, a développé un instrument d'équité (« *fairness tool* ») censé lui permettre d'identifier et de corriger les discriminations potentielles au sein de l'entreprise. Des algorithmes visent ainsi à quantifier les biais tels que des écarts de traitement entre les hommes et les femmes ou la prise en compte de l'origine au cours des procédures de recrutement. Il s'agit là d'une exigence de justice sociale : il faut lutter contre la tendance naturelle des hommes à perpétuer le passé, y compris dans ses travers, et à préférer ce qui ressemble à soi par rapport à ce qui est différent. Les IA peuvent ainsi parfaitement servir l'équité et la diversité, tout dépend des usages qu'on en fait et des objectifs qu'on lui assigne. En matière de recrutement, elle peut être utile en permettant de dépasser le confort immédiat du conservatisme et du mimétisme pour s'appuyer sur les critères les plus objectifs de distinction entre les candidats. Une telle ambition est d'ailleurs assez logique en matière d'IA : promouvoir une forme d'égalitarisme utilitariste dans laquelle chacun se caractérise par son utilité réelle plutôt que par son origine ou son milieu. Une société sans biais serait sans doute plus efficace et plus prospère qu'une société fonctionnant à base de copinage et de

[227] L. Malfettes, « Gestion du personnel par algorithmes et droits du salarié », *Droit social* 2019, p. 591.

localisme plutôt que de mérite. Cependant, une telle délégation du choix de ses collaborateurs à des IA parfaitement neutres n'est pas du goût de tout le monde et Gaspard Kœnig observe que « cette logique trahit un double renoncement au libre arbitre. D'une part, l'objet du choix est réputé non responsable de ses caractéristiques personnelles, qu'elles soient innées ou acquises. D'autre part, l'agent du choix délègue sa décision à la machine de peur de céder, consciemment ou non, à des préjugés profondément enfuis »[228]. Et le philosophe de se demander : « Quels critères resterait-il à l'intelligence artificielle, une fois l'individu réduit à l'épure de lui-même et devenu une simple abstraction de 1 et de 0 ? L'utilité réciproque. Ainsi l'intelligence artificielle, non contente d'anticiper et de révéler nos préférences spontanées, irait jusqu'à les supprimer au nom même de la satisfaction équitable et optimale de l'ensemble des individus »[229].

Nul doute qu'on ne saurait accepter un monde dans lequel les entreprises placeraient sur un pied d'égalité, pour éviter toute différenciation, ceux qui sont diplômés et ceux qui ne le sont pas, ceux qui ont une expérience et ceux qui n'en ont pas, ceux qui font la preuve de certaines formes d'intelligence et ceux qui ne le font pas, ceux qui donnent envie de travailler avec eux et ceux qui ne le donnent pas. L'égalitarisme n'est pas l'égalité et cette dernière ne saurait justifier qu'on préfère un candidat pénalisé par ses mauvaises études ou son QI trop faible afin de ne pas le discriminer. On ne saurait pour autant plaider pour les biais et les discriminations et, s'il faut sans doute se préserver d'une société totalement lissée, totalement automatisée, totalement dé-subjectivisée, c'est un fragile équilibre, comme souvent, qu'il faut poursuivre. C'est bien entendu un gouffre qui sépare les biais injustes et les distinctions justes. Trop d'égalité tue la liberté ; et trop de liberté tue l'égalité. Entre ces deux extrêmes, nous devons œuvrer à l'édification d'une société ménageant à la fois la liberté et l'égalité. Comme souvent, les positions par trop radicales ne sont pas bonnes conseillères. C'est pourquoi, comme la justice, il convient de se bander les yeux et de peser les qualités, les arguments ou les intérêts en présence au moyen d'une balance. En fonction des emplois, la force physique ou le niveau d'éducation peuvent être des critères justes de distinction entre des candidats. Tel ne saurait être le cas de la couleur de peau, de l'origine socioculturelle ou de l'identité sexuelle.

Surtout aux États-Unis, la lutte contre les discriminations est devenue une obsession de tous les instants dans les cercles progressistes. Dès qu'on jouit d'un début de puissance ou de privilège, il faut prouver à chaque occasion que l'on n'est pas dans le camp de l'oppression. Il faut prendre garde à ce que l'objectif de lutter contre les discriminations dans le monde des IA ne se retourne pas contre lui-même, n'aboutisse pas à tout aplanir, ce qui serait une autre forme de déshumanisation liberticide. L'objectif ultime des antiracistes ou des féministes devrait être de rendre la couleur de peau ou le sexe indifférent, non de construire des communautés d'intérêts prêtes à entrer en conflit. Dans les entreprises américaines, aujourd'hui, les employés sont paralysés par la peur de laisser échapper un propos que d'aucuns qualifieraient de « discriminatoire » ou de se comporter d'une manière qui serait jugée « inappropriée », si bien que les relations interpersonnelles authentiques deviennent impossibles. Finalement, on peut plaider pour une société des IA moins biaisée, mais pas sans biais aucun car alors cela signifierait en même temps sans humain aucun, sans affects, sans goûts, sans passions. Le droit n'a pas à promouvoir des IA qui seraient devenues parfaitement neutres et objectives. Il doit se contenter de lutter contre les injustices tout en préservant la part d'humanité qui permet de choisir arbitrairement ses amours, ses amis, ses collaborateurs. Selon le bon vieux préjugé, « qui se ressemble s'assemble ». Les biais permettent la continuité, l'appariement

[228] G. Kœnig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de l'observatoire, coll. De facto, 2019, p. 224.
[229] *Ibid.*, p. 225.

culturel est ce qui permet de poursuivre une histoire commune. En définitive, il existe aussi bien des raisons de redouter que les IA reproduisent et renforcent des biais injustes que des raisons de craindre qu'elles altèrent les biais justes et nécessaires à toute société. Car il n'est pas de société sans *affectio societatis*.

Une première voie équilibrée semble se trouver dans l'obligation de réaliser une étude d'impact relative aux discriminations pour tout projet d'IA visant à prendre ou à aider à prendre des décisions individuelles, touchant directement des personnes. Dans le cadre juridique européen actuel, il est déjà nécessaire, avant de mettre en œuvre des traitements de données personnelles, d'effectuer préalablement une étude quant à l'impact potentiel de ces activités sur les droits et les intérêts des personnes concernées. Cette étude d'impact sur la vie privée (« *privacy impact assessment* ») peut conduire, en cas de risque identifié, à apporter préventivement des correctifs. En ce sens, les lignes directrices adoptées par le groupement européen des autorités de protection des données (G29) soulignent qu'un PIA est nécessaire chaque fois qu'un risque de discrimination ou d'exclusion émerge avec un traitement de données et qu'il s'agit là d'une condition forte de l'acceptabilité sociétale de l'IA. Les concepteurs d'IA, qu'ils en soient à l'initiative ou à la réalisation, devraient s'interroger systématiquement quant aux conséquences sociales de leurs projets et travaux. On pourrait aller plus loin encore en imaginant un système d'autorisation préalable, ce qui obligerait à répondre aux complexes questions de son organisation, notamment s'agissant de l'organisme qui serait chargé d'opérer les contrôles et délivrer les autorisations. Mais cela irait contre la logique de responsabilisation des acteurs pour l'heure dominante, notamment afin de laisser suffisamment de marge d'innovation aux industriels, pour qui le droit doit être un soutien et non un frein — sachant que l'innovation peut être une innovation en faveur de l'égalité numérique.

Par ailleurs, une autre et dernière voie pourrait être de penser les droits collectifs sur les données, comme y a invité le « rapport Villani »[230]. La législation actuelle, focalisée sur la protection de l'individu, se trouve en décalage par rapport aux logiques autour desquelles les systèmes d'IA fonctionnent — soit l'examen de masses d'informations afin de repérer des tendances et des comportements cachés. Elle ne prend pas suffisamment en compte les effets de ces technologies sur des groupes d'individus. C'est pour combler ce décalage qu'il a été proposé de créer des droits collectifs sur les données, de consacrer une « macro-protection de la vie privée »[231]. Et la loi *De modernisation de la justice du XXIe siècle* du 18 novembre 2016 a ouvert une action de groupe « données personnelles » (figurant à l'article 43 ter de la loi du 6 janvier 1978) permettant aux associations de défense des consommateurs et de protection des données personnelles d'agir en cas d'infractions à la législation sur ces données[232]. Cette action de groupe est toutefois très encadrée et elle ne permet que la cessation de l'infraction concernée, non la réparation du préjudice causé. Reste que les droits collectifs sur les données auront sans doute un rôle utile à jouer dans ce contexte où il convient d'adapter la protection des droits et des libertés à la nouvelle donne de l'IA. Dans ce contexte, alors que la technique saisit les données des individus mais aussi des groupes d'individus, et que les systèmes rétroagissent sur les individus mais aussi sur les groupes d'individus, le trouble jeu des données menace de causer des discriminations tant individuelles que collectives, appelant

[230] C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 140.

[231] G. Loiseau, « Intelligence artificielle et droit des personnes », in A. Bensamoun, G. Loiseau, dir., *Droit de l'intelligence artificielle*, LGDJ, 2019, p. 59.

[232] L. n° 2016-1547, 18 nov. 2016, *De modernisation de la justice du XXIe siècle*.

des réponses tant individuelles que collectives — les discriminations collectives étant peut-être plus graves encore que les discriminations individuelles.

B. Le droit à la neutralité numérique

1. Les plateformes : café du commerce ou place du village, censeurs ou intermédiaires ?

L'égalité appelle un autre principe très célèbre en « droit de l'internet » : la neutralité du net (l'expression « *net neutrality* » a été inventée par le juriste américain Tim Wu dès 2003), destinée à garantir l'égalité de traitement en ligne, c'est-à-dire l'obligation pour les opérateurs concernés de traiter identiquement tous les contenus auxquels ils ont affaire. Ainsi se placent-ils à l'abri de tout risque d'opérer des discriminations. La neutralité garantit un accès égalitaire et sans discrimination à tous les contenus, quel que soit l'opérateur. Aussi peut-on considérer que le principe de neutralité serait « le préalable indispensable à l'exercice effectif des libertés à l'âge numérique »[233]. Ce principe s'oppose notamment à ce que des lois imposent des obligations de surveillance, par exemple s'agissant des contenus violents ou choquants, aux plateformes du web. Il serait regrettable que demain les réseaux sociaux en viennent à s'autocensurer et à supprimer tout propos un tant soit peu tendancieux, car cela porterait une profonde atteinte à la liberté d'expression. La neutralité protège donc contre les discriminations potentielles, mais aussi contre des atteintes à d'autres droits fondamentaux : le respect de la vie privée, l'accès à l'information, le libre choix des consommateurs. Alors que, le 14 décembre 2017, la Commission fédérale des communications des États-Unis a abrogé le principe de neutralité du net, jugé obsolète, le droit de l'Union européenne garantit un internet ouvert et interdit aux acteurs de discriminer l'utilisateur sur la base de l'origine ou de la destination des données et garantit ainsi les droits de l'abonné en termes de vitesse et de qualité du service[234]. En France, la loi Pour une République numérique consacre le principe de la neutralité de l'internet[235], donc l'accès égal au réseau de tous les utilisateurs, quelles que soient leurs ressources, mais aussi de toutes les données, sans discrimination. Et il revient à l'ARCEP (Autorité de régulation des communications électroniques et des postes) de veiller au respect de ce principe par les opérateurs.

Cependant, la neutralité numérique n'est pas un principe relatif au fond des contenus et encore moins un principe politique. Elle consiste techniquement à traiter de manière indifférenciée les paquets IP, donc les données acheminées sur le réseau mondial, quelle que soit leur provenance. Par exemple, la neutralité interdit de donner la priorité à un serveur mail par rapport à un serveur de flux vidéos. Elle oblige à traiter identiquement les différents services du web. Mais elle ne porte pas sur le sort réservé aux divers contenus à l'intérieur d'un service donné. Défendre la neutralité du net revient à défendre l' « internet pour tous », l'égalité d'accès au réseau, donc l'interdiction de fournir des accès

---

[233] Commission de réflexion et de propositions sur le droit et les libertés à l'âge numérique de l'Assemblée nationale française et Commission sur les droits et devoirs sur internet de la Chambre des députés italienne, déclaration commune, 28 sept. 2015.

[234] Règl. (UE) 2015/2120 du Parlement européen et du Conseil, 25 nov. 2015, *Établissant des mesures relatives à l'accès à un internet ouvert et modifiant la directive 2002/22/CE concernant le service universel et les droits des utilisateurs au regard des réseaux et services de communications électroniques et le règlement (UE) n° 531/2012 concernant l'itinérance sur les réseaux publics de communications mobiles à l'intérieur de l'Union*.

[235] L. n° 2016-1321, 7 oct. 2016, *Pour une République numérique*.

aux réseaux différents en fonction des ressources financières des utilisateurs. Cela a bien sûr un lien direct avec la liberté d'expression, l'accès à l'information et la démocratie, mais ce lien ne doit pas être confondu avec celui qui relie la liberté d'expression, l'accès à l'information et la démocratie à la lutte contre les fausses informations en ligne. La neutralité du net se rapporte à des flux de données qui, en soi, n'ont ni sens informationnel ni sens idéologique. Les éventuelles obligations de contrôle imposées aux grands services du web se rapporteraient, pour leur part, à la signification des données. Que les algorithmes s'efforcent au maximum de respecter cette neutralité est donc un grand enjeu, même si elle vise classiquement les réseaux.

Derrière la responsabilité des plateformes et notamment des réseaux sociaux, c'est toute la question de la neutralité du net qui se pose. Pour beaucoup, il serait impossible d'obliger ceux-ci à intervenir à l'égard des contenus qu'ils hébergent sans porter excessivement atteinte à ladite neutralité, ainsi qu'à la liberté d'expression. Les devoirs de surveillance des grandes plateformes ont toutefois déjà été renforcés en matière de lutte contre les contenus racistes, antisémites, négationnistes et pédopornographiques. L'article 6-I-7 de la loi *Pour la confiance dans l'économie numérique* du 21 juin 2004 oblige les FAI et hébergeurs à mettre en place un dispositif facilement accessible et visible permettant à toute personne de leur signaler des contenus qu'ils doivent retirer s'il s'agit d'apologie de crimes contre l'humanité, d'incitation à la haine raciale, de pornographie enfantine, d'incitation à la violence, d'atteinte à la dignité humaine et de promotion de jeux d'argent illégaux. S'il est possible pour les grands hébergeurs d'identifier des contenus illicites quand ceux-ci sont objectivement contraires à la loi, on ne saurait étendre un tel dispositif à d'autres contenus dont la traque suppose d'analyser des preuves. Par définition, l'hébergeur n'a pas accès à ces preuves et ce n'est pas son rôle, si bien qu'il ne peut que s'en remettre au juge dans la majorité des cas, faute d'illicéité manifeste des contenus signalés[236]. La Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique s'oppose également à tout « renforcement par la loi des obligations de surveillance des intermédiaires privés dans la lutte contre les contenus illégaux. S'agissant du rôle actif que joueraient certains hébergeurs par leurs systèmes de référencement ou de classement, la Commission rappelle que la responsabilité limitée de ces acteurs résulte en réalité de ce qu'ils n'ont pas connaissance des contenus, contrairement aux éditeurs. Le fait que certains hébergeurs aient recours à des outils de classement des contenus mis en ligne par des tiers peut justifier d'imposer des règles en ce qui concerne ce classement mais ne justifie pas que la loi leur impose des obligations de surveillance et de censure supplémentaires sur ces contenus »[237]. Si la justice publique était en mesure de répondre aux besoins, il n'y aurait aucune difficulté à faire appel à elle. Mais, cette justice étant en proie à de grandes difficultés et appelée à opérer de profondes réformes, on propose de donner un rôle cadre aux intermédiaires du web, là où la loi *Pour la confiance dans l'économie numérique* du 21 juin 2004, transposant la directive communautaire 2000/31/CE du 8 juin 2000 sur le commerce électronique, avait préféré consacrer la quasi-irresponsabilité des intermédiaires techniques et se reposer sur des prérogatives fortes accordées aux juges, conformément à la tradition de protection d'une liberté de communication triomphante.

Après que Donald Trump, contestant les résultats de l'élection présidentielle de 2020, a appelé ses partisans à marcher sur la Capitole, à Washington, le 6 janvier 2021, et que ceux-ci ont envahi les assemblées, convaincus qu'il était légitime de reprendre par la violence ce qui avait été indûment perdu par le vote — simili coup d'État qui était inimaginable dans l'une des plus belles démocraties

---

[236] Ch. Bigot, « Légiférer sur les fausses informations en ligne, un projet inutile et dangereux », *D.* 2018, p. 344.

[237] Ch. Féral-Schuhl, Ch. Paul, Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique, *Numérique et libertés : un nouvel âge démocratique*, Assemblée Nationale, rapport n° 3119, p. 88.

au monde avant l'ère des réseaux sociaux —, Facebook, Twitter, Instagram et Snapchat ont « pris leurs responsabilités » : ils ont bloqué les comptes de celui qui terminait son mandat (pour une durée indéterminée concernant Facebook, Instagram et Snapchat et de façon permanente pour Twitter, qui a aussi fermé des centaines de comptes en lien avec la mouvance QAnon)[238]. S'il est sans doute bienvenu que la liberté d'expression soit limitée en cas d'incitations à la violence, que la normativité d'origine privée relaie ainsi le droit, cela au niveau constitutionnel qui est celui de la liberté d'expression — surtout aux États-Unis où elle est conçue tel un quasi-absolu —, est ô combien significatif de la faiblesse actuelle du droit. Et d'aucuns peuvent évidemment s'offusquer devant une telle censure privée du plus haut personnage de l'État et estimer que cela interroge l'état des libertés publiques en Occident ainsi que les risques que font peser ces multinationales sur ces libertés, quand d'autres se réjouissent de cette heureuse réaction, bien que trop tardive, tant Donald Trump a fait du mal à son pays et au monde à travers ses manœuvres de désinformation permanente, surtout sur Twitter — soutenir le droit de désinformer, de diffuser des fausses informations et des théories du complot, au nom de la liberté d'expression, d'information ou d'opinion est parfaitement illogique et inconséquent.

Peut-être Twitter est-il plus proche du bar du commerce que de la place du village et doit-il être régulé comme tel (le patron peut mettre dehors le client ivre qui crie trop fort et nuit à la paisibilité des autres clients). Mais sans doute y a-t-il aussi de bonnes raisons de ne pas accepter de s'en remettre aux puissances privées en matière de régimes juridiques des droits et libertés fondamentaux. S'agissant de la liberté d'expression, on remarquera que Facebook et Google deviennent, selon un rapport d'Amnesty International du 1er décembre 2020, des « zones de non-droits de l'homme » au Vietnam, où les géants de la technologie collaboreraient avec les pouvoirs publics afin de censurer l'opposition pacifique et la liberté politique dans le pays, cela afin de pouvoir continuer à y proposer certains de leurs services. Amnesty International a ainsi prévenu que, « bien qu'elles ont été autrefois le grand espoir pour l'essor de la liberté d'expression dans le pays, les plateformes des réseaux sociaux sont en train de devenir rapidement des zones de non-droits de l'homme ». Facebook et Google se conformeraient sans mot dire aux demandes de retrait « des mauvaises informations, de la propagande contre le Parti et l'État » (en 2020, Facebook aurait répondu favorablement à 95 % des demandes du gouvernement et YouTube à 90 % d'entre elles). Facebook a d'ailleurs admis plus tôt dans l'année qu'il bloque les contenus jugés illégaux par les autorités.

2. Vers un nouveau statut entre hébergeurs et éditeurs ?

La Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique se dit défavorable à la création d'un statut intermédiaire entre l'éditeur et l'hébergeur, proposition formulée notamment dans le *Rapport visant à renforcer la lutte contre le racisme et l'antisémitisme sur internet* remis le 20 septembre 2018 au Premier ministre par la députée Laetitia Avia, l'écrivain Karim Amellal et le vice-président du CRIF Gil Taieb. Selon ces derniers, il conviendrait de créer un nouveau statut, celui de « propulseur de contenus », assorti d'un régime de responsabilité renforcée. Pour la Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique, « l'extension de mécanismes de censure privée via la loi contreviendrait au droit à un

[238] AFP, « Twitter, Facebook, Instagram... Les réseaux sociaux s'organisent contre Donald Trump », lexpress.fr, 9 janv. 2021.

procès équitable et méconnaîtrait les principes qui sous-tendent l'État de droit. Face aux propositions et évolutions tendant à renforcer la responsabilité des hébergeurs à l'égard des contenus illégaux, la Commission estime que le statut de l'hébergeur, qui constitue une grande conquête et une garantie importante des libertés (libertés d'expression et liberté d'innovation), doit être réaffirmé. […] Elle est attachée à la réaffirmation des obligations limitées des hébergeurs à l'égard des contenus illégaux »[239].

Pour sa part, le Conseil d'État, dans son étude annuelle de 2014, estime que « la *summa divisio* issue de la directive sur le commerce électronique est aujourd'hui sujette à de fortes incertitudes quant à la démarcation entre les deux catégories d'éditeur et d'intermédiaire technique » et qu' « il est probable que, dans les prochaines années, des décisions juridictionnelles écarteront la qualification d'hébergeur pour les principales catégories de plateformes […] : après les places de marché et les moteurs de recherche, suivront les réseaux sociaux, les plateformes de partage et les magasins d'applications. Tous ces acteurs perdront alors le régime de responsabilité limitée qui favorise leur activité »[240]. Le Conseil d'État propose dès lors de consacrer une nouvelle catégorie, celle des plateformes, distincte de celle des hébergeurs mais qui, à l'égard des contenus mis en ligne par les tiers, se verrait appliquer le régime de responsabilité civile et pénale des hébergeurs. Seraient qualifiés de plateformes les services de référencement ou de classement de contenus, biens ou services édités ou fournis par des tiers et partagés sur le site de la plateforme (moteurs de recherche, réseaux sociaux, sites de partage de contenus, places de marché, magasins d'applications, agrégateurs de contenus et comparateurs de prix). Par rapport aux hébergeurs, les plateformes se distingueraient donc par l'existence d'un service de classement ou de référencement.

Selon le Conseil d'État, la distinction des plateformes et des hébergeurs permettrait de protéger la liberté d'expression en allant contre la tendance de la jurisprudence à remettre en cause le statut d'hébergeur des plateformes en raison de leur rôle actif à l'égard des contenus. En même temps, cela permettrait de mieux réguler les activités des plateformes en les soumettant à des obligations nouvelles non applicables aux hébergeurs. Mais ces obligations ne porteraient pas sur l'identification des contenus illégaux. Elles concerneraient leurs méthodes de classement et leurs rapports avec les utilisateurs et les tiers qui mettent en ligne des contenus. Pareilles propositions, à nouveau, sont protectrices d'une liberté d'expression et de communication écrasante, au détriment, par exemple, de la protection de la démocratie contre les manœuvres de désinformation en période électorale.

Par ailleurs, plutôt qu'aux plateformes privées, ne faudrait-il pas confier à une ou des autorité(s) administrative(s) (indépendante(s)) le soin d'intervenir afin d'exiger la suppression de certaines informations ou certains contenus, tirant les conséquences des lenteurs et inadaptations de la justice ? Cela existe déjà en droit positif avec le blocage sur décision administrative des contenus incitant au terrorisme autorisé par l'article 12 de la loi du 13 novembre 2014 *Renforçant les*

---

[239] Ch. Féral-Schuhl, Ch. Paul, Commission de réflexion et de propositions sur le droit et les libertés à l'âge du numérique, *Numérique et libertés : un nouvel âge démocratique*, Assemblée Nationale, rapport n° 3119, p. 77. La LCEN a défini deux grandes catégories d'acteurs :

* les éditeurs de services, pleinement responsables des contenus qu'ils mettent en ligne, sont soumis à un régime de responsabilité calqué sur celui de la presse ;

* les hébergeurs bénéficient d'un régime de responsabilité civile et pénale limitée à l'égard des contenus illégaux. Ce régime de responsabilité se justifie par leur rôle purement passif à l'égard des contenus de tiers qu'ils rendent accessibles sans en prendre connaissance. En application des articles 6-I-2 et 6-I-3 de la LCEN, la responsabilité civile ou pénale des hébergeurs ne peut pas être engagée « s'ils n'avaient pas effectivement connaissance » du caractère illicite des contenus stockés ou « si, dès le moment où ils en ont eu connaissance, ils ont agi promptement pour retirer ces données ou en rendre l'accès impossible ».

[240] Conseil d'État, *Le numérique et les droits fondamentaux*, La documentation française, 2014, p. 221.

dispositions relatives à la lutte contre le terrorisme. Sur ce point, de telles autorités administratives ne présentent guère plus que les acteurs privés de garanties de respect des droits et libertés des internautes. Il ne s'agit pas moins d'une manière de contourner l'autorité judiciaire et les garanties qui s'y attachent. Tous les pouvoirs extra-judiciaires de régulation des contenus portent atteinte à la conception classique de la liberté d'expression dans un État de droit. Les opposants à de telles prérogatives confiées à des autorités administratives rappellent que « le préalable d'une décision judiciaire apparaît comme un principe essentiel, de nature à respecter l'ensemble des intérêts en présence, lorsqu'est envisagé le blocage de l'accès à des contenus illicites sur des réseaux numériques. Ce préalable constitue une garantie forte de la liberté d'expression et de communication et de la neutralité des réseaux. […] L'intervention d'une autorité judiciaire est nécessaire à chaque fois qu'est en cause une liberté individuelle afin de s'assurer que la mesure prise ne présente pas de caractère arbitraire, qu'elle est nécessaire et proportionnée à l'objectif poursuivi et respecte les droits de la personne »[241]. À l'aune de tels principes, il est évident que ni la régulation des contenus par les plateformes elles-mêmes ni la régulation des contenus par des autorités administratives (indépendantes) ne sont possibles sans y porter gravement atteinte.

3. Une obligation de loyauté

Dans l'ensemble, et en allant plus loin que l'obligation de neutralité, c'est peut-être aussi une obligation de loyauté qui semble devoir s'imposer aux plateformes numériques. Si la « neutralité » des algorithmes, au-delà de celle des réseaux, est souvent utopique et impossible à vérifier, on attend forcément des plateformes du web qu'elles respectent leurs utilisateurs en faisant réellement faire à leurs calculateurs ce qu'elles disent et prétendent leur faire faire. En ce sens, le Conseil national du numérique et le Conseil d'État ont, dès 2014, revendiqué une nouvelle obligation : la loyauté envers les utilisateurs[242]. Une telle exigence de loyauté signifie avant tout apporter une information honnête et suffisante, donc ne pas masquer ses intentions réelles et encore moins ses pratiques réelles derrière quelques déclarations d'intentions et vagues explications générales. Il ne s'agit pas d'imposer une veine objectivité dans le travail des IA, mais d'éviter la manipulation des utilisateurs à travers des écarts injustifiables entre les services promis en théorie et les services apportés en pratique. Les algorithmes trient les informations, c'est là toute leur utilité. Encore faut-il que les services indiquent quelles priorités et quels objectifs président à ces tris. Il faudrait plus encore pouvoir s'assurer qu'aucun intérêt caché, aucune déformation cachée ou aucun favoritisme ne porte atteinte à la sincérité des résultats obtenus.

Pour la CNIL, les « deux principes fondateurs pour le développement des algorithmes et de l'intelligence artificielle » sont la loyauté et la vigilance[243]. Tous deux se retrouvent au cœur du RGPD qui impose aux entreprises d'élaborer des principes éthiques en respectant un principe de loyauté (conçu comme fait de faire primer les intérêts des utilisateurs) et un principe de vigilance. Concernant la loyauté, étant donné l'opacité des systèmes d'IA qui accompagne des menaces d'ordres différents, elle est à rapprocher de l'exigence de transparence : toute plateforme devrait

---

[241] *Ibid.*, p. 92.

[242] Conseil d'État, *Rapport sur la neutralité des plateformes*, La Documentation française, 2014 ; Conseil national du numérique, « Ambitions numériques », rapport remis au Premier ministre, juin 2015.

[243] CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 48.

expliquer ce qu'elle fait et pourquoi, être transparente quant aux collectes et aux traitements de données. Mais elle doit aussi être associée à l'objectivité, à la bonne foi et, finalement, à la neutralité. On ne saurait se satisfaire de services manipulant les utilisateurs tout en leur expliquant qu'ils les manipulent — sans doute en termes édulcorés et relativement abscons afin de faire passer cette manipulation pour un service rendu.

Dans son étude annuelle de 2014, le Conseil d'État a formulé trois recommandations pour « repenser les principes fondant la protection des droits fondamentaux ». Parmi ceux-ci, on trouvait le principe de loyauté, appliqué non pas à toutes les IA mais seulement aux « plateformes ». Pour les auteurs, il convenait de soumettre celles-ci à une obligation de loyauté envers leurs utilisateurs parce qu'elles sont des acteurs classant les contenus de tiers. Et de définir la loyauté comme consistant à « assurer de bonne foi le service de classement ou de référencement, sans chercher à l'altérer ou à le détourner à des fins étrangères à l'intérêt des utilisateurs »[244]. Ainsi, parmi les obligations des plateformes envers leurs utilisateurs, il y aurait la loyauté et, découlant de cette loyauté, d'une part, la pertinence des critères de classement et de référencement mis en œuvre au regard de l'objectif de meilleur service rendu à l'utilisateur et, d'autre part, l'information sur les critères de classement et de référencement utilisés[245]. La loyauté limiterait donc la liberté de choix des critères de l'algorithme en même temps qu'elle obligerait à des actes positifs d'information quant au mode de fonctionnement de l'IA. L'intérêt du principe de loyauté tel qu'il est préconisé par le Conseil d'État se trouve dans la notion d' « intérêt des utilisateurs ».

Quant au Conseil national du numérique, il a initié dans son rapport « Ambition numérique » de 2015 un projet audacieux : créer une « agence de notation de la loyauté des algorithmes » appuyée sur un réseau ouvert de contributeurs. L'objectif est à la fois de rendre accessible en un lieu unique diverses informations déjà rassemblées par les différents acteurs et de proposer un espace de signalement de pratiques problématiques ou de dysfonctionnements. *In fine*, cette loyauté a vocation à encourager les bonnes pratiques et à favoriser une prise de conscience citoyenne et une meilleure connaissance des problématiques qui entourent ces techniques. Mais la loyauté des algorithmes — en réalité la loyauté des concepteurs et entraîneurs d'algorithmes — se heurte à cet obstacle technique redondant : si des problèmes peuvent se poser pour les droits des personnes, y compris à l'insu de leurs concepteurs, la portée de la notion de loyauté se trouve largement diminuée en raison des difficultés à comprendre comment fonctionne l'IA, comment elle aboutit à certains résultats à partir de certaines entrées. Dès lors, il devient difficile de donner corps à cette loyauté. Il n'en demeure pas moins que tous ceux qui contribuent au fonctionnement d'une IA doivent se garder de toute ambition manipulatrice ou autrement dolosive et expliquer clairement, sans rien cacher, quels objectifs ils poursuivent à travers leurs travaux.

4. Le droit à des décisions humaines

Le droit à des décisions humaines pourrait être un dernier gage d'égalité en ce que le regard humain pourrait permettre de déceler et éviter certains biais dans les décisions automatiques. Il est vrai, cependant, que ce regard humain pourrait aussi aboutir à biaiser une décision automatique *a priori* non biaisée. C'est d'ailleurs peut-être plus symboliquement qu'autre chose que le droit à une décision humaine serait important : on accepte plus facilement le choix d'un humain, qui peut l'expliquer si

[244] Conseil d'État, *Le numérique et les droits fondamentaux*, La Documentation française, 2014, p. 278.
[245] *Ibid.*

l'on lui pose des questions, que le choix issu de traitements automatisés de données dont on saisit mal ou même pas du tout les mécanismes. Reste que le droit de réclamer la transparence des résultats des systèmes informatiques pourrait comprendre le droit de faire appel des décisions prises par l'IA et de demander leur validation ou révision par un être humain.

Allant plus loin, la loi *Informatique et libertés* s'oppose à ce qu'une machine puisse prendre seule, sans intervention humaine, des décisions emportant des conséquences cruciales pour les personnes. Elle dispose qu' « aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage »[246]. Cela ne concerne toutefois que l'administration et les institutions publiques — l'approche moderne-étatiste de la juridicité faisant que l'adjectif « juridique » est ici réservé aux seules sources publiques de normes. Or les mastodontes de l'économie numérique prennent aussi des décisions produisant des effets considérables à l'égard d'une personne sur le seul fondement d'un traitement automatisé de données destiné à définir son profil ou à évaluer certains aspects de sa personnalité. Et ces décisions, à portée juridique ou non, n'impactent pas moins de façon redoutable les vies des personnes qu'elles visent. La loi *Informatique et libertés* ajoute, depuis sa modification par la loi du 20 juin 2018[247], qu' « aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne. Aucune autre décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données destiné à prévoir ou à évaluer certains aspects personnels relatifs à la personne concernée »[248]. Une telle disposition est une précaution importante contre les dérives du profilage, qui peuvent porter atteinte à l'égalité entre les hommes. Mais sa portée est là encore par trop réduite en raison du fait qu'elle touche les pouvoirs publics et non les pouvoirs privés. Si cette méfiance du législateur a été avant-gardiste, puisqu'une norme semblable existe depuis 1978, son effectivité a toujours posé question et pose de plus en plus question. Par conséquent, une telle disposition appliquée de façon rigoureuse et étendue à toutes les décisions produisant des effets à l'égard de certaines personnes serait une avancée pour l'égalité, mais aussi pour la liberté[249].

Il faut se souvenir qu'en 1978, la loi *Informatique et libertés* posait qu' « aucune décision administrative *ou privée* impliquant une appréciation sur un comportement humain ne peut avoir pour seul fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé »[250]. Et elle ajoutait que « toute personne a le droit de connaître et de contester les informations et les raisonnements utilisés dans les traitements automatisés dont les résultats lui sont opposés ». Ces dispositions sont aujourd'hui reprises, en substance, par l'article 22 du Règlement général sur la protection des données : « La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire ». Mais cet article prévoit des limites importantes : « Le paragraphe 1 ne s'applique pas lorsque la décision : a) est nécessaire à la conclusion ou à l'exécution d'un contrat entre la

[246] L. n° 78-17, 6 janv. 1978, *Relative à l'informatique, aux fichiers et aux libertés*, art. 47, al. 2.

[247] L. n° 2018-493, 20 juin 2018, *Relative à la protection des données personnelles*.

[248] L. n° 78-17, 6 janv. 1978, *Relative à l'informatique, aux fichiers et aux libertés*, art. 95.

[249] S. Merabet, *Vers un droit de l'intelligence artificielle*, th., Université d'Aix-Marseille, 2018, p. 244.

[250] Non souligné dans la forme originale.

personne concernée et un responsable du traitement ; b) est autorisée par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée ; ou c) est fondée sur le consentement explicite de la personne concernée. Dans les cas visés au paragraphe 2, points a) et c), le responsable du traitement met en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision »[251]. Le principe se trouve donc largement vidé de sa substance par cette adjonction d'exceptions très larges. De la même manière, en France, la loi du 20 juin 2018 a étendu le spectre des « cas dans lesquels, par exception, une décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel ». Une évolution de la loi *Informatique et libertés* datant de 2004 avait déjà facilité de fait la prise de décision automatisée, par exemple dans le secteur bancaire : si l'intervention humaine était encore requise, il s'agissait désormais d'un droit à faire valoir *a posteriori*, permettant de demander à ce que, en cas de décision défavorable, celle-ci soit réexaminée par une personne. La portée du droit à une intervention humaine a donc été progressivement diminuée.

Le RGPD, inspiré par la loi française, interdit donc qu'une machine puisse prendre seule, sans supervision humaine, des décisions emportant des conséquences graves pour les personnes, en matière d'octroi de crédit par exemple. Cela permet de conserver une salvatrice responsabilité humaine derrière le travail des IA. Finalement, les traitements automatisés de données à caractère personnel ne devraient toujours être que des aides à la décision humaine. À tout le moins, toute personne devrait pouvoir exiger une intervention humaine dans une prise de décision automatisée et pouvoir faire valoir ses arguments avant que la décision soit prise.

[251] Règlement (UE) 2016/679 du Parlement européen et du Conseil, 27 avr. 2016, *Relatifà 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*, art. 22.

IV. Contrôle

Question 1. Des mécanismes de contrôle humain du travail des IA existent-ils ?

Question 2. Pour chaque opération réalisée par une IA, existe-t-il un humain qui en est responsable et qui peut répondre du résultat ?

Question 3. Les résultats produits par les IA sont-ils explicables ?

Question 4. Existe-t-il des procédures d'évaluation de la fiabilité des résultats et d'évaluation des risques ?

Question 5. En cas d'apprentissage automatique des IA, des procédures de contrôle et de supervision ont-elles été mises en place ?

Question 6. Les utilisateurs sont-ils informés du fait qu'ils interagissent avec une IA et non avec un humain ?

Question 7. Des mécanismes permettent-ils d'informer les utilisateurs des raisons et critères expliquant les résultats fournis par l'IA ?

Question 8. Des mécanismes permettant un recours en cas de préjudice ou d'effet néfaste ont-ils été prévus ?

Question 9. Des procédures sont-elles prévues afin de faciliter l'auditabilité du système d'IA par des acteurs internes et/ou indépendants (notamment en assurant la traçabilité et la journalisation du fonctionnement de l'IA) ?

Question 10. Est-ce que des méthodes de garantie de la fiabilité de l'IA existent ?

Question 11. Est-ce qu'une étude des possibles attaques que le système d'IA pourrait subir a été menée ?

Question 12. Des mesures sont-elles prévues afin de prévenir ces attaques et préserver l'intégrité de l'IA ?

Question 13. L'IA a-t-elle été conçue ou entraînée pour continuer à fournir des résultats pertinents en cas de changement brutal de circonstances ?

Question 14. Les potentiels dommages et préjudices causés par l'IA à des utilisateurs ou à des tiers ont-ils été envisagés, notamment du point de vue de leur gravité et de leur probabilité ?

Question 15. En cas de risque qu'une IA cause des dommages, est-ce que des règles de responsabilité et de protection des consommateurs ont été prévues ?

Question 16. Des contrôles et des mesures visant à réduire l'impact environnemental de la création et de l'utilisation de l'IA ont-ils été prévus ?

A. Le droit à des décisions humaines

1. L'homme, un dernier recours vital

Peut-être le premier de tous les droits de l'homme est-il le droit à l'homme. L'humanisme juridique se trouve tout entier incarné en lui. Sans doute est-il indispensable si l'on veut proposer des intelligences artificielles au service de l'homme et du bien commun. En matière de droits et libertés fondamentaux de l'homme numérique, on devrait pouvoir en toutes circonstances exiger qu'une décision qui nous impacte directement soit prise par une intelligence humaine plutôt que par une intelligence artificielle en acceptant les éventuelles sous-optimalités, lenteurs ou erreurs dans ces décisions humaines. Bien sûr, un tel droit semble délicat à respecter dans de nombreuses situations où des entreprises ou des administrations sont conduites à prendre d'innombrables décisions individuelles, tâche que des humains ne pourraient matériellement pas exercer. Dès lors, il faudrait repenser des systèmes organisationnels entiers, éventuellement pour en revenir à ce qui existait avant les progrès de l'informatique. On pourrait aussi établir la liste des actes qui impactent fortement les vies des personnes, à l'exemple, typiquement, d'une procédure de recrutement ou de l'affectation d'un bachelier dans un établissement d'enseignement supérieur, et n'appliquer le droit à l'homme qu'à ces situations. Mais c'est aussi et surtout l'effet symbolique du droit à l'homme qui importe : en le mettant en avant, même s'il n'a d'existence que dans la pensée juridique et ne trouve pas de traduction juridique positive parce que ce serait trop compliqué de lui donner vie concrètement, on contribue à informer la population des effets normatifs des IA et de leurs lourdes conséquences, souvent insoupçonnées, sur nos existences.

Le recours désormais massif à l'IA dans certains domaines sensibles comme la police, la banque, l'assurance, la justice ou l'armée (avec la question des armes autonomes) pose la question des limites à lui imposer et du minimum de responsabilité humaine en dessous duquel on ne devrait jamais passer. La tension entre sophistication technologique et dépossession humaine s'illustre de façon remarquable dans le trading algorithmique : ici, des robots ordonnent « librement » l'achat ou la vente de titres, engageant au quotidien les transferts de sommes les plus importants de la planète, tandis que l'humain n'est même plus présent à titre de superviseur dans la boucle des opérations. Plus généralement, la place de l'automatisation dans les décisions humaines doit être interrogée : y a-t-il des cas, plus ou moins nombreux, dans lesquels le jugement humain, aussi faillible soit-il, ne devrait pas être remplacé par le calcul informatique ? Il importe de s'efforcer à une méfiance permanente à l'égard de l'aura et de la puissance des calculateurs, lesquelles favorisent le dessaisissement de la décision humaine. C'est ainsi que la crise financière de 2008 a, en partie, été provoquée par des calculs de risques bancaires abandonnés à des modèles mathématiques qui négligeaient l'amplitude des fluctuations potentielles. Au moment même où l'évaluation algorithmique prenait son essor, elle montrait ses limites, mais la leçon n'a pas été retenue.

On se souviendra que, durant la Guerre froide, à deux reprises la décision souveraine d'un humain a permis d'éviter une conflagration nucléaire par le refus d'exécuter des instructions émises par des machines. Le 28 octobre 1962, le capitaine américain William Bassett reçut en effet un signal correspondant au code de lancement des missiles. Il fit alors le choix, libre et éclairé, de contrevenir à la procédure prévue, cela malgré la rigueur et le sens de la hiérarchie des militaires. Il s'avéra que sa radio avait en fait commis une erreur en décryptant le véritable message[252]. À l'identique, le 26

[252] A. Gratchev, *Un nouvel avant-guerre ? Des hyperpuissances à l'hyperpoker*, Alma, 2017.

septembre 1983, le lieutenant-colonel soviétique Stanislav Petrov reçut un message d'alerte indiquant une attaque de missiles américains. Il préféra ne pas transmettre cette information à l'état-major, imaginant quels effets en cascade cela aurait généré. Il découvrit ensuite qu'il s'agissait d'une erreur de signalement des satellites soviétiques[253]. Avec moins de lucidité humaine, moins de libre-arbitre et plus d'autorité du code informatique, la loi des robots aurait provoqué une guerre nucléaire et le cours de l'histoire moderne aurait été radicalement changé. Aujourd'hui, ne vit-on pas dans un monde où l'autorité du code informatique s'est sensiblement renforcée, tandis que la lucidité humaine et le libre arbitre ont sensiblement régressé ? N'assiste-t-on pas à l'évaporation progressive de la capacité souvent vitale et salvatrice des hommes à se déterminer librement, à agir et à penser loin du déterminisme et du fatalisme, qui affirment que la volonté serait déterminée dans chacun de ses actes par des forces extérieures telles que le nudge des IA ?

En janvier 2018, un autre exemple d'erreur informatique aurait pu donner lieu à un scénario catastrophe : l'agence de gestion des urgences d'Hawaï a envoyé un message d'alerte téléphonique annonçant une attaque de missiles immédiate et ordonnant aux résidants de se mettre à l'abri. Cette fausse alerte issue d'une erreur technique a causé une vague de panique qui, sans intervention humaine, aurait pu contaminer les relations déjà tendues avec la Corée du Nord, notamment. Plaider pour un droit à l'homme, c'est aussi soutenir la recherche de tout moyen afin que les hommes qui se trouvent derrière les machines demeurent responsables, restent conscients de leur responsabilité pleine et entière quand les IA ne sont que des outils parfois efficaces et parfois maladroits. Sans capacité humaine de déroger aux commandes de la machine, grâce à un regard critique maintenu en permanence sur les résultats qu'elle produit, bien des désastres individuels et collectifs risquent d'advenir. La valeur ajoutée de l'humain ne doit jamais s'effacer, quelles que soient les capacités des supercalculateurs. Elle doit demeurer au cœur de nos écosystèmes.

Les États devraient garantir aux individus le droit de n'être jamais soumis à une décision les affectant de manière significative lorsque cette décision est fondée sur un processus décisionnel automatisé sans réelle intervention humaine. A minima faudrait-il permettre à cette personne d'exiger qu'un humain révise ou valide la décision de l'IA après avoir entendu ses arguments. Tel est déjà le régime juridique retenu en Europe par le RGPD, à son article 22 : toute personne « a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire ». Par exemple, une candidature ne saurait être exclue sur le seul fondement de méthodes et techniques automatisées d'aide au recrutement et doit faire l'objet d'une appréciation humaine. Mais cela est largement illusoire puisque bien souvent les hommes valideront les options des systèmes informatiques sans procéder à aucun contrôle, les vérifications attendues étant matériellement impossibles à réaliser face à la vitesse des calculs et du traitement de l'information des ordinateurs. On peut s'interroger quant à la part de liberté qui subsiste dans la prise de décision face au poids d'une réponse présentée comme « scientifique » ou « mathématique ». Cela conduit souvent à la simple apposition de signatures aux côtés des résultats produits par les algorithmes. L'intervention humaine est alors purement factice. L'existence d'une aide informatisée à la décision, y compris lorsqu'elle n'est supposément qu'un simple outil d'aide à la décision, est susceptible de conduire les professionnels à s'aligner sur le diagnostic de l'IA pour ne pas « prendre de risque ». Cette dilution de la responsabilité humaine accompagnant des comportements moutonniers n'est évidemment pas satisfaisante.

[253] *Ibid.*

Reste que, logiquement, beaucoup d'institutions plaident pour le « droit à l'homme » en matière d'IA, à l'image de l'Agence des droits fondamentaux de l'Union européenne qui, dans son rapport « Bien préparer l'avenir : l'IA et les droits fondamentaux » de décembre 2020, estime que chaque IA devrait être soumise à un système de supervision et laisser la possibilité à n'importe qui de contester ses décisions, permettant d'avoir droit à une appréciation humaine en cas de désaccord[254]. Les députés européens, en octobre 2020, avaient déjà adopté un texte tendant à imposer que les hommes conservent toujours le dernier mot par rapport aux machines douées d'intelligence artificielle. Si cela semble constituer une décision logique et nécessaire, elle ne va pas de soi pour tout le monde et 44 députés ont voté contre ce texte, avançant qu'il risquerait d'accroître le retard technologique et économique de l'Europe par rapport aux États-Unis et à la Chine, lesquels préfèrent, à des fins de puissance, ne mettre aucun bâton juridique dans les roues des robots et leur accorder des pouvoirs illimités.

En France, la loi Pour une République numérique impose aux plateformes une obligation d'information des consommateurs dans l'utilisation de procédés algorithmiques et à l'administration une obligation de transparence[255]. Le législateur tend ainsi à imposer aux acteurs de l'IA, tant lors de la phase de conception que lors de la phase d'utilisation, des exigences en termes de loyauté, de transparence et d'explicabilité des opérations réalisées par des procédés automatiques. Plus les IA se dotent d'un pouvoir performatif sur nos vies, plus devient nécessaire un droit à l'explication et à la justification de toute décision assistée par machine, cette justification, devant être compréhensible par tous, pouvant notamment servir de base pour régler un différend légal en cas de besoin.

Dans son Traité de droit et d'éthique de la robotique civile, allant au-delà des simples décisions individuelles, Nathalie Nevejans consacre de longues pages au « principe roboéthique du droit pour la personne à refuser d'être prise en charge par un robot »[256]. Face aux relations de plus en plus intimes et de plus en plus nombreuses entre les hommes et les robots dits « sociaux », notamment avec la multiplication des cas de prise en charge de personnes vulnérables telles que les personnes âgées ou handicapées, elle insiste sur l'importance de donner à chacun la possibilité de s'opposer à toute dépendance à l'égard d'un système informatique, en bref le droit d'avoir toujours le choix entre une alternative humaine et une alternative automatique. Il s'agit d'un enjeu fort du point de vue du respect de la dignité de la personne humaine[257]. Cette dignité, en tant que fondement des droits de l'homme, implique qu'une intervention et une participation humaines significatives doivent être possibles dans toutes les interactions entre la technologie et les êtres humains.

Après avoir frénétiquement automatisé tout un tas de services, remplaçant les humains par des machines, on constate désormais un besoin de plus en plus impérieux de réhumaniser lesdits services, de réintroduire de l'humain dans des processus devenus trop froids et trop rigides. C'est un équilibre entre l'efficacité des systèmes informatiques et la souplesse de la relation humaine qu'il faudrait trouver. Dans l'éducation, par exemple, on verra peut-être les mêmes qui ont voulu augmenter la place des terminaux interactifs et intelligents en remplacement des professeurs se rendre compte de l'inanité de telles politiques qui robotisent les hommes dès leur plus jeune âge. Ils plaideront alors pour que l'on redonne à l'humain toute sa place dans l'école. Cela fait déjà quarante

---

254 Agence des droits fondamentaux de l'Union européenne, « Bien préparer l'avenir : l'IA et les droits fondamentaux », 13 déc. 2020.

255 L. n° 2016-1321, 7 oct. 2016, *Pour une République numérique*.

256 N. Nevejans, *Traité de droit et d'éthique de la robotique civile*, LEH édition, coll. Science, éthique et société, 2017, p. 811 s.

257 *Ibid*.

ans que des chercheurs en sciences sociales tirent la sonnette d'alarme face aux risques de déshumanisation de la société au prétexte que la plupart des tâches pourraient être exercées par des robots — sauf que la très grande majorité d'entre elles, mis à part lorsqu'elles font appel à l'intelligence logico-mathématique, ne sauraient l'être que moins bien que par des hommes. Pourtant Dominique Wolton peut constater combien « nous tirons la sonnette d'alarme, mais personne ne veut entendre parce que cela ne fait pas "moderne" »[258]. De nombreux travaux mettent en garde contre l'usage immodéré des corrélations au détriment du raisonnement. Et il semble toutefois que de plus en plus nombreux soient ceux qui entendent cet avertissement.

2. Signaler toute interaction avec une IA

Toute personne confrontée à une IA dont l'action ou la décision la concerne ou l'affecte devrait être informée de la présence de cette IA et du rôle qu'elle joue. La « Déclaration de Montréal » retient en ce sens que « le développement des services d'IA doit éviter de créer des dépendances par les techniques de captation de l'attention et par l'imitation de l'apparence humaine qui induit une confusion entre les services d'IA et les humains »[259]. Elle ajoute que « tout utilisateur d'un service qui recourt à des agents conversationnels doit pouvoir identifier facilement s'il interagit avec un service d'IA ou une personne »[260]. Le droit à l'homme implique en premier lieu d'informer l'utilisateur très explicitement du fait qu'il interagit avec une intelligence artificielle et non avec un humain à chaque fois que tel est le cas. Il faudrait donc, dès la conception d'un système informatique, prendre toutes les précautions pour que ses futurs utilisateurs ne risquent jamais d'hésiter sur le point de savoir si, derrière l'écran ou la voix, se cache un homme ou un robot. En cas de doute, on devrait pouvoir facilement trouver cette information.

Alors que le « test de Turing » n'est plus un défi insurmontable pour les IA et que celles-ci se comportent de façon de plus en plus confondante comme le ferait un humain, les voix synthétiques étant remplacées par des voix naturelles et les réponses maladroites et stéréotypées laissant la place à des réponses précises et personnalisées, il devient indispensable de s'assurer que les intelligences artificielles ne puissent jamais être prises pour des intelligences humaines. Cela doit permettre de conserver cette « summa divisio » des hommes et des machines, une séparation nette entre les principes, comportements et valeurs de ces deux catégories. C'est là une exigence morale et éthique à laquelle il ne serait guère difficile d'accorder une pleine force juridique. Alors que des androïdes hyper-réels s'immiscent dans la société, le droit doit protéger l'humanité de l'homme en faisant en sorte que, dans un maximum d'instants du quotidien, il soit rappelé à chacun les différences essentielles entre un être humain et un outil informatique. Alors que l'IA a pour effet de modifier notre perception des humains et de l'humanité, l'humanisme juridique ne saurait accepter pareille évolution qui va forcément dans le sens non d'une surhumanisation des hommes mais dans celui de leur soushumanisation. Quand on confond le vivant et la machine, on accorde peut-être un souffle de vie à la machine, mais on retire surtout un peu de ce souffle au vivant. L'homme, pendant qu'il s'attache à un robot, oublie de s'attacher à ses semblables. Il se robotise et il robotise autrui. Les droits de l'homme numérique limitent forcément les possibilités de développement des robots

[258] D. Wolton, Internet, et après ? Une théorie critique des nouveaux médias, Flammarion, coll. Champs essais, 2010, p. 207.
[259] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.
[260] *Ibid*.

androïdes et humanoïdes. Le jour où l'on ne fera plus la différence entre une personne et une chose, tout s'effondrera : l'humanisme, le droit et l'humanisme juridique.

Plus généralement, les informations rendues publiques devraient permettre une véritable évaluation de l'IA. Et, même si une telle situation est très peu plausible, il faudrait idéalement qu'aucun système informatique n'atteigne un degré de sophistication tel qu'il en devienne impossible à surveiller et à contrôler par des humains. Les systèmes qui ne peuvent être soumis à des normes de transparence et de responsabilité appropriées ne devraient pas être utilisés. Les IA — en fait ceux qui les conçoivent et ceux qui les utilisent — doivent être soumises à une obligation de rendre des comptes. Il s'agit de garantir leur vérifiabilité, éviter le scénario de la « boîte noire » dans lequel on ne sait pas expliquer comment la machine passe des données en entrée aux données en sortie. Tout au contraire, il convient d'assurer la traçabilité des actions majeures et mineures du robot suivant la logique des boîtes noires du secteur aéronautique. Cela doit notamment permettre de repérer les éventuelles injustices imputables aux calculs d'une IA et de permettre leur réparation. Et, si des accidents surviennent, l'IA doit être transparente et rendre des comptes à l'enquêteur afin que le processus interne qui a posé problème puisse être cerné. Comme l'explique le Conseil constitutionnel, le responsable de traitement est notamment tenu de s'assurer de « la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard »[261].

Paradoxalement, et à condition de consentir les efforts nécessaires, l'IA pourrait finalement nous permettre de redécouvrir le proprement humain, réaffirmer la singularité humaine. Se comparant au robot non afin de saisir leur ressemblance mais afin de mieux comprendre leurs profondes différences, l'homme se redéfinirait en tant qu'homme. En découlerait un nouvel appétit d'humanité. Cela toucherait en particulier cette propriété centrale de l'espèce humaine : son ultra-sociabilité. Renouant avec lui-même, l'homme trouverait de nouvelles raisons et de nouveaux moyens de « faire société ». On serait donc en présence d'une technologie qui permettrait aux hommes de devenir encore plus hommes, mais cela à condition de s'en donner les moyens, notamment en termes de droits de l'homme numérique. Pour l'heure, on ne peut que constater que l'IA robotise les hommes davantage qu'elle les humanise. La réhumanisation par la réhabilitation de notre ultra-sociabilité suppose la revalorisation de la place des émotions dans l'intellect. Or, pour l'heure, les IA renforcent moins les émotions des individus qu'elles les plongent dans des comportements routiniers et stéréotypés dans lesquels « on fait sans penser ». En replaçant l'humain au centre du jeu, spécialement avec le méta-principe de droit à l'homme, on souhaite éviter ces dérives délétères de l'immixtion de la technique informatique dans nos quotidiens et dans nos intimités.

3. Consacrer partout le statut d'outil de l'IA

Un défi majeur lié à l'essor de l'IA, peut-être le plus essentiel de tous, est de « faire en sorte que ces nouveaux outils soient à la main humaine, à son service, dans un rapport de transparence et de responsabilité »[262]. Il faudrait en permanence rappeler le statut d'outil de l'informatique, y compris dans ses développements les plus perfectionnés. D'un point de vue humaniste, conserver une

[261] Cons. const., déc. n° 2018-765 DC, 12 juin 2018, *Loi relative à la protection des données personnelles*.

[262] I. Falque-Pierrotin, « Préface », in CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 3.

conscience claire et nette de la différence entre l'homme et la machine est un enjeu fort. Or le droit, par exemple en octroyant une personnalité juridique aux robots, en les rendant responsables ou en leur attribuant de très paradoxaux « droits de l'homme », peut parfaitement contribuer à l'humanisation des robots. Mais il peut aussi renforcer la séparation entre les êtres vivants et les choses-objets, affirmer avec force une hiérarchie dans laquelle l'être humain reste juridiquement très supérieur à l'intelligence artificielle, cela se traduisant par la reconnaissance de droits à l'égard des entreprises et des administrations proposant des services fonctionnant grâce à l'IA. Aussi des principes de loyauté, de transparence et de « rendre compte » sont-ils posés.

En avril 2018, plus de deux cents experts, dont de nombreux juristes, ont signé une lettre ouverte dénonçant la résolution du Parlement européen qui préconisait de créer une personnalité juridique robotique. Ils soulignaient tout simplement la contradiction qu'il y avait à attribuer des droits à des machines tout en défendant les droits de l'homme tels qu'ils sont consacrés dans différents textes de droit européen. Puisque les droits naturels ont été conçus par et pour l'homme, on ne saurait les étendre à d'autres formes d'existence. Si les droits et les obligations subjectifs jaillissent de décisions individuelles, on ne peut que refuser par principe qu'un robot puisse prendre une décision, qu'un choix, un acte ou, pire, un contrat puisse échapper à l'humain. L'introduction d'une personnalité morale au profit des systèmes informatiques bouleverserait — en mal — les structures fondamentales de nos systèmes juridiques dans lesquels c'est l'individu et son libre arbitre qui font naître les droits et les obligations, qui sont responsables en cas de transgression de la loi. Cela se trouve au fondement du droit européen, mais aussi du droit américain et, selon la Cour suprême, « l'intention personnelle est universelle dans la plupart des systèmes de droit matures qui croient dans la responsabilité de l'homme normal choisissant librement entre le bien et le mal »[263]. Une IA, elle, n'a pas d'intention personnelle puisque des calculs et des statistiques ne sont pas des pensées ni des raisonnements.

En outre, cette lettre ouverte souligne, de façon un peu provocatrice, que des robots personnifiés jouiraient de droits fondamentaux et devraient notamment être rémunérés et se voir proposer des conditions de travail décentes, ce qui n'aurait aucun sens. Plus sérieusement, ils soulignent combien il serait dangereux de rendre les robots responsables de leurs actes : ainsi les fabricants d'armes létales autonomes pourraient-ils s'absoudre de toute responsabilité, comme les fabricants de voitures autonomes. L'accident mortel provoqué par une voiture autonome d'Uber, en mai 2018, a livré un avant goût des casse-têtes juridiques à venir : les juges ont estimé que la responsabilité de l'entreprise n'était pas engagée dans ce drame. Alors que les querelles entre assureurs promettent d'être nombreuses, le droit serait bien avisé de clarifier les choses en prenant des positions fermes et précises.

Les machines doivent donc conserver le statut juridique d'outils et des personnes physiques ou morales doivent en conserver le contrôle à tout moment. Cela implique aussi que les IA soient conçues et exploitées de façon à se conformer au droit existant. Alors que l'on peut craindre une perte de contrôle de l'humain tant l'IA permet de déléguer un nombre croissant de tâches à des systèmes automatiques et peut être amenée à prendre des décisions sensibles comme des diagnostics et prescriptions médicaux, des actes judiciaires ou le lancement d'une bombe dans un conflit armé, il paraît nécessaire de rappeler partout et tout le temps qu'un ordinateur n'est qu'un outil. Au moins dans ces situations sensibles, la décision devrait rester humaine — formellement mais aussi et surtout matériellement.

[263] Cour suprême des États-Unis, 342 U.S. 246, 7 janv. 1952, Morissette c. United States.

Or, sitôt qu'on recourt à un algorithme comme aide à la décision, celui-ci risque d'exercer son pouvoir performatif, empêchant l'humain, qui s'incline devant ses incommensurables capacités de calcul, de le contredire. L'IA est si puissante qu'elle dilue les figures d'autorité traditionnelles. Dans ces conditions, on pourrait définir tout un tas de cas sensibles dans lesquels on devrait tout bonnement s'interdire de s'appuyer sur une IA pour décider. En matière de justice, par exemple, les magistrats, au moment de juger, se fondent sur la loi et la jurisprudence, mais aussi sur l'éthique et autres formes de droit non officiel qui permettent d'adapter les décisions aux particularités de chaque cas. En témoigne l'exemple des « délits altruistes », ces infractions commises dans l'intérêt général et non afin d'en retirer un profit personnel. Parfois, il est bon que la morale, forme de droit souple, tempère le droit dur. Il est beaucoup plus délicat d'apprendre à des algorithmes le droit souple que le droit dur. C'est pourquoi le juge-humain ne saurait laisser sa place à un juge-robot. Par exemple, ne faudrait-il pas redouter le « panurgisme judiciaire » qui risquerait d'accompagner le déploiement de la justice assistée par ordinateur ? Les juges, sachant dans quel sens les tribunaux ont précédemment tranché les situations similaires, pourraient préférer suivre la tendance juridictionnelle afin de ne pas renvoyer l'image d'une justice balbutiante. Les algorithmes possèderaient ainsi un effet performatif, génèreraient des prophéties auto-réalisatrices. Or, si l'allégorie de la justice a les yeux bandés, n'est-ce pas afin de se prémunir contre l'influence des bases de données ? Certes, la loi prévoit déjà que le diagnostic du médecin ou la décision du juge ne peuvent pas faire l'objet d'une automatisation. Mais, face aux limites incertaines entre délégation et aide à la décision, on pourrait rappeler solennellement ces principes.

Par ailleurs, s'agissant des robots ayant une existence physique et même des IA embarquées dans des systèmes et s'exprimant à travers des écrans, il n'est pas infondé de plaider pour l'obligation de les doter d'un bouton physique ou immatériel (à cliquer) permettant à tout moment de les arrêter et d'en reprendre le contrôle. On devrait toujours pouvoir recourir à un tel mécanisme d'arrêt d'urgence sous la forme d'un bouton d'arrêt, d'une combinaison clavier ou d'une commande vocale. Ainsi l'humain serait-il supposé conserver sans cesse le contrôle du robot. L'autonomie du robot, dans sa mobilité ou sa capacité à décider, commande d'imposer un tel mécanisme qui permettrait d'éviter des accidents.

Depuis 2020, Spot, le « super chien-robot » de Boston Dynamics qui a fait la une des journaux lorsqu'il a été aperçu en train de faire respecter la distanciation sociale dans un parc de Singapour durant l'épidémie de covid-19, est disponible à la vente pour un prix de 75 000 dollars. Ce nouveau meilleur ami de l'homme quadrupède semble devoir toujours rester un outil. Mais cela ne saurait suffire, car il existe de bons et de mauvais outils. Spot, grâce à son agilité, son niveau de charge utile et ses capacités de détection, pourrait remplacer les humains dans certains environnements à haut risque — à l'image de ces chiens policiers déjà envoyés en éclaireurs dans certains terrains dangereux, souvent au prix de leurs vies. Dans le Massachusetts, la police utilise déjà Spot pour intervenir dans des situations périlleuses. Mais le chien de garde automatisé pourrait aussi devenir l'instrument d'un État policier robotisé malmenant grandement les libertés individuelles.

4. Déresponsabiliser les robots

À mesure que le fonctionnement et les actes des IA se complexifient, il devient difficile d'en tenir responsable les hommes qui les ont conçues afin qu'elles soient les plus fiables et les plus éthiques. Pourtant, dès lors qu'on refuse de rendre ces IA elles-mêmes responsables, on ne peut que conserver la responsabilité humaine face aux risques juridiques accrus et aux contraintes d'indemnisation qui

en découlent. Si cela peut conduire à réduire les possibilités de déploiement concret de l'IA, il ne s'agit pas moins, peut-être, de la meilleure solution, du moins du point de vue de l'humanisme juridique. Que l'on songe aux problèmes de responsabilité que vont fatalement provoquer l'utilisation des voitures autonomes, des drones ou encore des robots chirurgicaux. Jusqu'à présent, parce que l'IA restait accessoire dans nos vies, on considérait que le régime de responsabilité du robot était nécessairement un régime de responsabilité sans faute. Mais les régimes de responsabilité classiques sont désormais mis en doute par certains, quand d'autres estiment qu'il faudrait coûte que coûte conserver le système dans lequel la responsabilité est fondée sur la maîtrise du bien. L'Office parlementaire d'évaluation des choix scientifiques et technologiques juge, pour sa part, que quatre régimes de responsabilité pourraient trouver à s'appliquer : celui du fait des produits défectueux, celui du fait des animaux, celui du fait d'autrui ou encore celui du fait des choses. Ils permettraient ainsi de retenir la responsabilité, selon les cas, du concepteur du système d'analyse, du fabricant du robot, de son propriétaire ou encore de son utilisateur[264]. Un régime spécial de responsabilité du fait du robot, calqué sur celui de la responsabilité du fait des animaux prévu par l'article 1243 du Code civil, pourrait être opportun. Mais en aucun cas ne saurait-on retenir le robot comme seul responsable, à l'exclusion de tout être humain. On doit déresponsabiliser le robot pour responsabiliser l'humain, ce qui est un gage d'humanité.

L'une des grandes questions juridiques posées par l'IA est celle de la responsabilité civile. Le recours massif à des systèmes informatiques aidant plus ou moins fortement à prendre des décisions fait craindre que les responsabilités en cas d'erreurs soient compliquées à établir, par exemple dans le secteur médical où les règles relatives à la responsabilité ne pourront être appliquées comme elles le sont actuellement. Sans doute faudra-t-il réformer le cadre relatif à la répartition de la responsabilité entre l'utilisateur (le médecin ou praticien), le producteur de la solution technologique et l'établissement de soins proposant le traitement. On peut en effet craindre que « l'utilisateur soit amené à céder systématiquement à la solution de diagnostic ou au traitement proposé par l'instrument technologique par crainte de subir des mesures de responsabilité civile lorsque le jugement professionnel et éclairé l'amène à tirer des conclusions pouvant être en partie divergentes »[265]. On risque bel et bien, à l'avenir, de devoir répondre à la question de savoir si le fait pour un professionnel de ne pas suivre les recommandations ou préconisations d'une IA intervenant dans son processus décisionnel est constitutif ou non d'une faute. Dès lors qu'un robot médical établirait un diagnostic et/ou proposerait un traitement adapté, le médecin qui choisirait de s'en affranchir pour suivre son intuition serait-il responsable en cas d'erreur ? Sans doute serait-il regrettable qu'on institue un régime de faute, éventuellement professionnelle, en cas de « non-respect » des indications fournies par les systèmes de prévision automatique. Ces outils dits « d'aide à la décision » deviendraient en pratique des outils de prise de décision. Mais on peut aussi renverser la question : dès lors qu'un robot médical établirait un diagnostic et/ou proposerait un traitement adapté et que le médecin suivrait ses préconisations, ce robot serait-il responsable en cas d'erreur ? Le droit prévoit que le critère de la responsabilité est la personnalité juridique, ce qui exclut de fait les IA dans cette responsabilité et dans l'indemnisation qui s'ensuit. Revient alors cette solution facile mais insatisfaisante à de nombreux égards : accorder aux systèmes informatiques, qui jouent un rôle

[264] Office parlementaire d'évaluation des choix scientifiques et technologiques, « Pour une intelligence artificielle maîtrisée, utile et démystifiée », 15 mars 2017, p. 153 s.

[265] Parlement européen, « Résolution sur une politique industrielle européenne globale sur l'intelligence artificielle et sur la robotique », 2018/2088 (INI), 12 févr. 2019, § 77.

décisif dans beaucoup d'actes et de faits, la personnalité juridique, leur permettant de devenir sujets d'obligations mais aussi titulaires de droits.

La personnalité juridique « n'est pas seulement l'aptitude à recueillir des droits subjectifs devenir propriétaire, créancier… mais, beaucoup plus largement, la vocation à être pris en compte dans les diverses situations définies et régies par le droit objectif : se marier ; divorcer ; payer des impôts ; voter aux élections politiques ; conclure des contrats… »[266]. On imagine mal un robot se marier ou voter — mais, certes, on imagine tout aussi mal déjeuner avec une personne morale. La personnalité juridique du robot est surtout une folie en raison des effets symboliques que cela provoquerait en l'homme et dans la société. On doit s'interdire tout acte et toute pensée tendant à produire une confusion entre la sphère robotique et le monde humain. Alors que déjà trop d'hommes, après avoir été tentés par la bestialité, sont portés à délaisser l'humanité au profit de la machinité, il importe d'encourager l'humanité des hommes par tous les moyens. Un de ces moyens est de ne pas les obliger à partager la personnalité juridique avec des outils informatiques. Cette personnalité est un attribut essentiel d'une personne. Jean Carbonnier le notait : les personnes « sont les êtres capables de jouir de droits ; ce sont, d'une expression équivalente, les sujets de droit »[267]. Une « personne-robot » serait une bien étrange figure qui brouillerait beaucoup de repères chez des hommes déjà en manque de repères. La reconnaissance de la personnalité juridique est un droit fondamental, devant bénéficier à tous. Ainsi l'article 6 de la Déclaration universelle des droits de l'homme et l'article 16 du Pacte international relatif aux droits civils et politiques proclament-ils que « chacun a le droit à la reconnaissance en tous lieux de sa personnalité juridique ». Attribuer la personnalité juridique à des IA, ce serait donc ouvrir la boîte de Pandore de ces droits de l'homme des robots qu'on ne saurait accepter sans ruiner des siècles de progrès de l'humanisme juridique. Si l'on peut craindre que le transhumanisme conduise au transjuridisme, il serait encore plus douloureux de voir le transjuridisme favoriser le transhumanisme.

L'attribution d'une personnalité, même « singulière », aux robots serait source de nombreuses incohérences, en plus de signer l'acte de migration de notre société humaniste vers une néo-société robotiste. On ne s'étonnera pas que de nombreux civilistes s'opposent fermement à toute personnalité morale des robots suivant l'exemple de celle des sociétés et autres collectivités publiques , allant jusqu'à évoquer une « monstruosité juridique » qui « nous entraînerait dans une spirale schizophrénique »[268], qui serait une façon de « revisiter Frankenstein »[269]. Ils dénoncent ce qui aboutirait à obscurcir la distinction, pourtant essentielle, entre les personnes et les choses. La catégorie des personnes serait alors constituée des personnes physiques, des personnes morales et des personnes robots. Les personnes non humaines concurrenceraient les personnes humaines, tandis que nombre d'êtres vivants continueraient à se voir refuser tout droit : les animaux, mais aussi les embryons humains. Et il en resterait de même des entités humaines sans vie, appréhendées souvent comme des choses. De plus, la personne robot deviendrait un sujet de droit sans cesser d'être un objet de droit car les droits de propriété sur les instruments informatiques ne disparaîtraient pas. Aussi peut-on déplorer qu'une telle perspective « détraque le droit » car « seule une chose peut être objet de droit. La personnification des robots dérèglerait le construit juridique en donnant vie à une

[266] J.-L. Aubert, *Introduction au droit et thèmes fondamentaux du droit civil*, 7e éd., Armand Colin, 1998, p. 188.

[267] J. Carbonnier, *Droit civil – I. Les personnes*, 21e éd., Puf, 2000, p. 1.

[268] G. Loiseau, « La personnalité juridique des robots : une monstruosité juridique », *JCP G* 2018, p. 597.

[269] F. Rouvière, « Le robot-personne ou Frankenstein revisité », RTD civ. 2018, p. 778.

chimère, mi-personne mi-chose, qui pervertirait la summa divisio des personnes et des choses et l'ordre de valeur qui lui correspond »[270].

Alors que des juristes comme Alain Bensoussan peuvent plaider que, « dans les dix prochaines années, les humains et les robots devraient évoluer de manière tangentielle : les robots s'humaniseront et les humains se robotiseront »[271], d'autres peuvent défendre l'humanisme, l'humanisme juridique et donc tout ce qui singularise l'homme dans le monde. Ainsi ne saurait-on, par exemple, aller dans le sens d'un droit à la dignité des robots qui ne ferait qu'ajouter de la confusion à la confusion et de l'inanité à l'inanité : la dignité humaine est peut-être le plus essentiel de tous les droits humains : le partager avec des êtres non humains ou, pire, avec des êtres non vivants serait un coup fatal porté à l'humanité du droit. Ce dernier doit, au contraire, rappeler sans cesse qu'un robot est un instrument construit par et pour l'homme. S'il sort de ce cadre, alors il perd toute raison d'être. La dignité signifie qu'on ne doit pas traiter un homme comme une chose, par exemple en le louant ou en le vendant, en louant ou en vendant ses données personnelles, non qu'on devrait traiter des choses comme si elles étaient des hommes. Comme l'a écrit Emmanuel Kant, l'être humain doit être toujours traité comme une fin et jamais comme un moyen. Un objet ou un outil, au contraire, doit toujours être considéré comme un moyen et jamais comme une fin. En incitant à voir en l'homme un moyen et dans le robot une fin, l'IA a tendance à avilir l'humanité, ce contre quoi le droit doit agir, rappelant que l'humain est la valeur suprême et son guide fondamental. Instrumentaliser et réifier les personnes tout en personnifiant les choses, ce serait une source de malheurs immenses. Mais les droits de l'homme numérique sont là pour rappeler que toute avancée technologique qui sortirait du cadre triangulaire formé par la dignité, la liberté et l'égalité doit être refusée.

Pour être clair, disons que les animaux devront toujours posséder plus de droits que les robots. Les chats, les écureuils, les cochons, les poules, les alligators, les araignées, les punaises de lit et même les chiens sont plus dignes de voir protégées leur dignité, leur liberté et leur égalité que les androïdes et les algorithmes des réseaux sociaux. Ceux qui soutiennent l'attribution d'une personnalité juridique et même de droits fondamentaux aux robots mettent en avant leur autonomie qui serait un critère suffisant. Or il ne l'est absolument pas et les animaux sont des êtres autonomes. Surtout, contrairement aux robots, ils sont vivants. Mais bon nombre d'entités autonomes, animées et vivantes ne se voient pas reconnaître d'aptitude à jouir de droits. Malgré leur immense intelligence attestée scientifiquement, les dauphins, les perruches et les poulpes, qui apprennent et évoluent naturellement, ne sont pas dotés pour autant d'un patrimoine juridique. Lorsqu'en 1978 on a rédigé une « Déclaration des droits des animaux », certains y ont vu un risque de déshumanisation. Il ne s'agirait pas de reconnaître des droits au profit des animaux mais seulement d'imposer aux hommes des devoirs à l'égard des autres êtres vivants. Aujourd'hui, dans le Code civil, les atteintes aux animaux restent des atteintes aux biens. Il demeure une distinction essentielle entre l'humain et le non humain. Même les extraterrestres, s'ils sont doués de raison et de conscience, devraient se voir accorder bien plus de droits que les IA.

Par ailleurs, s'il est vrai qu'on peut attendre une protection des IA par le droit, par exemple contre des utilisations abusives, des destructions par des tiers ou des copies sans autorisation, il convient de ne pas confondre personnalité juridique et protection juridique, la seconde pouvant parfaitement aller sans la première. Les œuvres d'art sont rigoureusement protégées juridiquement sans avoir été

[270] G. Loiseau, « La personnalité juridique des robots : une monstruosité juridique », JCP G 2018, p. 597.

[271] A. Bensoussan, J. Bensoussan, *Droit des robots*, Larcier, 2015.

pour autant personnifiées. Il n'est pas utile d'attribuer des droits à une chose pour la préserver d'éventuels abus. C'est à l'humain propriétaire de cette chose que les droits doivent aller. Quant aux animaux, si on leur a récemment reconnu la qualité d'êtres vivants doués de sensibilité, ils n'en demeurent pas moins classés dans la catégorie des biens, meubles ou immeubles lorsqu'ils sont attachés à l'exploitation d'un fonds. Et pourtant le droit les protège assez fortement : l'article 515-14 du Code civil dispose ainsi que, « sous réserve des lois qui les protègent, les animaux sont soumis au régime des biens ». Or les lois en question interdisent les sévices et les actes de cruauté. Quant aux robots, il est difficile de défendre à leurs propriétaires de les démembrer, les modifier ou les détruire si bon leur semble : ils ne sont pas, eux, des êtres vivants doués de sensibilité. L'environnement, pour sa part, est l'objet de régimes juridiques de plus en plus substantiels visant à le préserver des pollutions et autres atteintes humaines. Et pourtant il n'a jamais été question de lui accorder une personnalité juridique.

Pour ce qui est de la protection des personnes physiques, l'attribution d'une personnalité juridique est a fortiori incongrue. L'identification et la traçabilité des robots, nécessaires, peuvent parfaitement se passer de cette personnalité, comme cela est déjà le cas pour les biens dont le suivi est assuré grâce à un numéro d'immatriculation, permettant de définir les responsables en cas d'incident impliquant, par exemple, une voiture ou un lot de jambon sous vide. On ajoutera que les personnes physiques sans autonomie conservent leur personnalité juridique bien qu'elles soient incapables, dans l'impossibilité d'exprimer une volonté libre et éclairée. Les concernant, la protection attachée aux personnes est encore plus indispensable. On voit que la personnalité juridique est inhérente à la nature humaine et que, en tant que droit de l'homme, le droit ne fait que la reconnaître — en théorie puisque l'histoire a connu l'esclavage ou la mort civile.

L'interdiction de l'esclavage est fondée sur la nature commune des êtres humains et donc sur l'absence de hiérarchie entre eux. À l'inverse, il faut autoriser l'esclavage des robots, cela au nom des natures différentes entre eux et les hommes et de l'existence d'une hiérarchie faisant que les hommes sont très supérieurs, surtout en droits, aux robots. Il y a donc bien, d'une part, les hommes et les personnes, qui ne font qu'un, et, d'autre part, les IA et les choses, qui ne font qu'un. Accorder une légitimité équivalente aux IA et aux humains mettrait en péril la primauté de ces derniers et risquerait d'aboutir à une guerre de souverainetés juridiques. Mais sans doute cela n'arrivera-t-il jamais, car on ne saurait croire que les hommes, qui ont seuls le pouvoir de faire et défaire le droit, pourraient perdre à ce point toute lucidité qu'ils en viendraient à forger eux-mêmes les armes de leur autodestruction.

On peut dès lors conclure que « le souhait de l'attribution d'une personnalité aux robots semble plus répondre à l'expression d'une empathie déplacée qu'à une nécessité juridique »[272]. S'il s'agit de pouvoir dédommager les victimes d'un dommage causé par une IA en puisant dans le capital dont celle-ci serait titulaire, il faut bien déterminer à qui il revient d'approvisionner ce capital. Dès lors, autant faire de ce dernier le responsable direct plutôt qu'un responsable indirect. Une responsabilité propre aux robots serait contre-productive en faisant écran à celle des fabricants ou des utilisateurs.

[272] M. Bouteille-Brigant, « Intelligence artificielle et droit : entre tentation d'une personne juridique du troisième type et avènement d'un “transjuridisme” », *LPA* 27 mars 2018, p. 7.

5. Une IA responsable est une IA dont des hommes sont responsables

En 2020, le collectif Impact AI a publié un livre blanc « IA digne de confiance » dans lequel il explique qu'une telle IA inspirant la confiance serait déjà une IA qui ne serait jamais abandonnée aux services informatiques des entreprises ou des administrations mais qui serait l'objet d'une véritable gouvernance, d'une véritable politique concertée et ordonnée, permettant d'en conserver une parfaite maîtrise. « L'utilisation de l'IA doit impliquer toute l'entreprise, quelle que soit sa taille », explique Impact AI[273]. Prolongeant cette approche, un principe directeur gouvernant le monde des IA semble devoir être le principe de responsabilité en vertu duquel les hommes et les organisations qui déploient et utilisent ces systèmes sont pleinement responsables devant la loi des éventuels dommages causés par ceux-ci. À aucun moment le déploiement des services informatiques ne saurait contribuer à une déresponsabilisation des êtres humains lorsque des décisions sont prises. « Seuls des êtres humains peuvent être tenus responsables de décisions issues de recommandations faites par des services d'IA et des actions qui en découlent »[274], affirme la « Déclaration de Montréal », avant d'ajouter que, « dans tous les domaines où une décision qui affecte la vie, la qualité de la vie ou la réputation d'une personne doit être prise, la décision finale devrait revenir à un être humain et cette décision devrait être libre et éclairée »[275], avant de rappeler que, logiquement, « les personnes qui autorisent des services d'IA à commettre un crime ou un délit, ou qui font preuve de négligence en les laissant en commettre, sont responsables de ce crime ou de ce délit »[276]. Seuls les hommes, doués de raison et de conscience selon l'article 1er de la Déclaration universelle des droits de l'homme, sont responsables. On comprend dès lors qu'une IA responsable ne signifie pas une IA dotée de la personnalité juridique et sommée de répondre de ses actes. Il s'agit d'une IA fabriquée de façon responsable et dont les producteurs puis les utilisateurs demeurent à tout moment responsables, loin de pouvoir se décharger sur un mode « ce n'est pas ma faute mais celle du robot ». L'imprévisibilité, le caractère évolutif et potentiellement surprenant des algorithmes et de leurs effets ne sont pas des raisons de déresponsabiliser les hommes qui leur ont donné naissance mais plutôt des raisons de n'avancer que par petits pas en matière d'IA, de ne lancer des produits sur le marché qu'une fois que l'on est certain de leur parfaite sécurité et de leur entier respect des droits de l'homme numérique.

S'il est bien sûr délicat d'encadrer et de contrôler un objet instable, dont les effets changent à mesure de son apprentissage, la CNIL propose de poser une opportune « obligation de vigilance et de réflexivité » en vertu de laquelle les concepteurs et ceux qui déploient l'intelligence artificielle devraient prendre en compte cette caractéristique[277]. Ce principe, sorte de « principe de précaution à l'envers » faisant que chaque acteur de la chaîne de fabrication et d'utilisation est coauteur des résultats produits par l'IA et se doit d'être attentif[278], doit permettre de répondre dans le temps au défi constitué par le caractère instable et imprévisible des IA capables d'apprendre automatiquement.

---

[273] Cité par E. Bembaron, « Intelligence artificielle : les entreprises en quête d'éthique », lefigaro.fr, 10 déc. 2020.

[274] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

[275] *Ibid.*

[276] *Ibid.*

[277] CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 50.

[278] A. Gattolin, C. Kern, C. Pellevat, P. Ouzoulias, « Intelligence artificielle : l'urgence d'une ambition européenne », Rapport d'information au Sénat, Commission des affaires européennes, 31 janv. 2019.

Il constitue aussi une réponse aux formes d'indifférence, de négligence et de dilution de responsabilité que peut générer le caractère très compartimenté et segmenté des systèmes algorithmiques[279]. De plus, l'informatique, depuis longtemps, suscite une confiance exagérée dans le fait qu'elle serait généralement infaillible quand l'action et le jugement humains sont souvent biaisés. Mais un outil de géolocalisation peut vous inviter à traverser à pieds une autoroute, ce que ne ferait jamais un humain. Les IA font moins d'erreurs que les humains mais elles font des erreurs que jamais un humain ne ferait et qu'un minimum de vigilance permet de détecter. Le principe de vigilance et de réflexivité a vocation à prendre en compte et contrebalancer ce biais cognitif qui amène humain à accorder une confiance excessive aux décrets des algorithmes. Des procédures et mesures concrètes doivent favoriser le questionnement permanent à l'égard de ces objets techniques. L'IA actuelle peut dysfonctionner plus ou moins gravement et il est important que des hommes puissent à tout moment reprendre les commandes, l'empêchant d'engendrer une spirale infernale. Alors que l'essor des systèmes informatiques va de pair avec une érosion des vigilances individuelles, il faut encourager, au contraire, l'accompagnement humain des IA. En ce sens, le Comité économique et social européen prône une approche dite « *human-in-command* », soit des développements scientifiques et industriels de l'IA qui restent toujours responsables, sûrs et utiles, dans le cadre desquels « les machines restent les machines, sous le contrôle permanent des humains »[280]. Selon le CESE, « il n'est pas éthiquement acceptable qu'un être humain soit contraint par l'IA ou qu'il soit considéré comme un exécutant de la machine qui lui dicterait les tâches à accomplir »[281]. Le Parlement européen a repris à son compte l'approche « *human-in-command* », insistant sur la nécessité de préserver le principe selon lequel « l'humain contrôle la machine »[282], tandis que, pour l'Organisation internationale du travail, l'IA doit rester « sous contrôle humain », les décisions finales touchant au travail doivent être prises par des êtres humains[283]. La Commission européenne, de son côté, propose de retenir une obligation de contrôle humain qui « contribue à éviter qu'un système d'IA ne mette en péril l'autonomie humaine ou ne provoque d'autres effets néfastes » et dont l'intensité devrait dépendre de l'utilisation prévue de l'IA et de ses incidences potentielles[284]. Une personne physique devrait donc systématiquement être responsable en dernier ressort du travail des algorithmes. Le Conseil constitutionnel a ainsi pu retenir que « ne peuvent être utilisés, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement »[285].

Ensuite, il n'est guère aisé, étant donné la pluralité d'intervenants autour d'une IA, de déterminer celui ou ceux, parmi eux, qui doit/doivent être tenu(s) responsable(s). Une intelligence artificielle est d'abord une machine complexe qui a des concepteurs et des fabricants multiples. Ce sont eux qui doivent être responsables en cas de difficulté. Mais ils ne sauraient l'être tous ensemble lorsque le

---

279 CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 6.

280 Comité économique et social européen (CESE), « Les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société », avis, 31 mai 2017.

281 Comité économique et social européen (CESE), rapport public, 5 oct. 2018.

282 Parlement européen, « Résolution relative à la politique industrielle sur l'IA et la robotique », 12 févr. 2019.

283 Organisation internationale du travail, « Travailler pour bâtir un avenir meilleur », rapport de la commission mondiale sur l'avenir du travail de l'OIT, 22 janv. 2019.

284 Commission européenne, « Intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance », livre blanc COM(2020) 65 final, 19 févr. 2020.

285 Cons. const., déc. n° 2018-765 DC, 12 juin 2018, *Loi relative à la protection des données personnelles*.

problème est le fait de l'un d'entre eux. Pour la Commission européenne, « chaque obligation devrait s'appliquer à l'acteur ou aux acteurs qui est/sont le/les mieux placé(s) pour éliminer tout risque potentiel »[286]. Par exemple, le développeur est mieux en mesure d'éliminer les risques liés à la phase de développement et le déployeur est celui qui doit maîtriser les risques au cours de la phase d'utilisation. Pour sa part, le Parlement européen, dans sa résolution du 16 février 2017, suggère, s'agissant de la chaîne de responsabilité, que la responsabilité soit proportionnelle au niveau d'instructions données. Ainsi, si le problème vient de la période de formation par l'homme, ce dernier serait davantage responsable que si les difficultés sont apparues durant l'auto-apprentissage de l'IA. On ne saurait en tout cas trop démultiplier les responsables potentiels car la définition d'une chaîne de responsabilité allant du concepteur d'IA aux utilisateurs finaux est une ambition difficile à mettre en œuvre en pratique.

Enfin, il serait logique de traiter différemment les IA en fonction des secteurs dans lesquels elles interviennent et des risques potentiels qu'elles font courir aux humains. Il ne semble guère pertinent d'appliquer le même régime de responsabilité à une voiture autonome et à un service de prise de rendez-vous à distance. La Commission européenne va en ce sens en proposant de distinguer les applications d'intelligence artificielle « à haut risque » et les autres[287]. Et de définir ce « haut risque » à partir de deux critères cumulatifs : d'une part, le secteur (par exemple, « les soins de santé, les transports, l'énergie et certains pans du secteur public ») et, d'autre part, l'utilisation en elle-même, puisqu'un secteur à haut risque peut recourir à des applications ne portant pas de risque en soi[288]. Des usages non compris dans ces catégories pourraient cependant aussi être considérés à risque et donc être soumis à la même réglementation. La Commission prend les exemples de « l'utilisation d'applications d'IA dans les procédures de recrutement et dans des situations ayant une incidence sur les droits des travailleurs » et de « l'utilisation d'applications d'IA à des fins d'identification biométrique à distance et pour d'autres technologies de surveillance intrusive »[289].

B. Le droit à la transparence des IA

1. « Nul n'est censé ignorer la loi des IA »

Certaines IA aboutissent à des résultats très pertinents, bien supérieurs à ceux que pourraient produire des humains. Le problème est que la façon dont ils y parviennent demeure très opaque. C'est là le fameux problème des « boîtes noires » : on voit ce qui y rentre, on récolte ce qui en sort, mais on n'a qu'une vague idée de ce qui s'y passe, des mécanismes entre les entrées et les sorties. Cet enjeu est de taille car pouvoir expliquer comment l'IA fonctionne est crucial pour de multiples usages, qu'il s'agisse de l'attribution d'un crédit, du choix du meilleur candidat pour un emploi, de l'affectation d'un bachelier dans un établissement d'enseignement supérieur, d'un diagnostic médical ou de la conduite autonome, surtout lorsqu'elle engendre des accidents. Dans son rapport du printemps 2019, le groupe d'experts de haut niveau de la Commission européenne notait ainsi que, « dès qu'un système d'IA a une incidence importante sur la vie des personnes, il devrait être possible

[286] Commission européenne, « Intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance », livre blanc COM(2020) 65 final, 19 févr. 2020, p. 26.
[287] *Ibid.*, p. 20.
[288] *Ibid.*, p. 21.
[289] *Ibid.*, p. 21.

d'exiger une explication appropriée du processus de décision ». Plus généralement, on devrait pouvoir s'assurer à tout instant qu'une IA fait ce pourquoi elle a été programmée et ce pourquoi elle est utilisée. Toute IA devrait pouvoir être évaluée en temps réel et *a posteriori*. Nul mécanisme algorithmique ne devrait atteindre un degré de complexité tel qu'il ne puisse plus être surveillé ni contrôlé par des êtres humains. Dans la « Déclaration de Montréal », il est retenu ainsi que « les décisions des services d'IA devraient toujours être justifiables dans un langage compréhensible aux personnes qui les utilisent ou qui subissent les conséquences de leur utilisation. La justification consiste à exposer les facteurs et les paramètres les plus importants de la décision et doit être semblable aux justifications qu'on exigerait d'un être humain prenant le même type de décision »[290].

Pour contrôler le travail des IA, notamment afin de prévenir les discriminations, il est impératif de pouvoir « soulever le capot et plonger les mains dans le moteur ». Dès lors qu'une décision individuelle est prise sur la base des résultats d'un algorithme, il faudrait avoir le droit de se voir expliquer comment ceux-ci ont été obtenus, à travers quels mécanismes, à partir de quelles données, suivant quels critères. Or cela est en réalité assez utopique. Poser un tel principe reviendrait largement à interdire l'usage de ces technologies, d'une part car il est souvent impossible de comprendre exactement comment un système d'IA aboutit à certains résultats plutôt que d'autres et, d'autre part, parce que même lorsque cela est possible pour ses concepteurs, traduire le mode de fonctionnement d'une IA afin de le rendre accessible au grand public sans néanmoins trop le simplifier est une gageure. Il est bien possible de préconiser, comme la CNIL, de « constituer une plateforme nationale d'audit des algorithmes », mais, en pratique, la réalisation d'un tel audit achopperait sur de nombreux obstacles techniques. Et encore faudrait-il convaincre les autorités compétentes de ne pas limiter les contrôles au seul respect de la vie privée et d'inclure aussi les atteintes aux principes de liberté et d'égalité. On ne saurait pour autant renoncer au principe selon lequel « nul n'est censé ignorer la loi des algorithmes », car cette loi emporte des conséquences trop importantes sur nos existences individuelles, collectives, privées, professionnelles etc. Le fonctionnement des services d'intelligence artificielle devrait être intelligible de tous ou, au moins, de leurs concepteurs. Quiconque s'estime victime d'une violation des droits de l'homme causée par une IA devrait pouvoir accéder aux données personnelles le concernant qui ont été utilisées, mais aussi aux informations d'apprentissage et de test, aux informations sur la façon par laquelle l'IA fonctionne et a été utilisée, aux informations utiles et compréhensibles sur la manière avec laquelle l'IA a produit la recommandation, la prévision ou la décision. Enfin, cette personne devrait pouvoir comprendre comment les données issues de l'IA ont été interprétées et quelles suites leur ont été données et pourquoi. Il s'agirait donc d'enregistrer et documenter le travail des systèmes : description de la collecte et de l'étiquetage des données, description de l'algorithme utilisé, explication du processus de prise de décision algorithmique.

2. Le secret des algorithmes

La transparence des algorithmes suscite logiquement de nombreuses réserves et est souvent jugée insatisfaisante ou même impraticable. Une transparence se contentant de la publication pure et simple d'un code source laisserait la quasi-totalité du public dans l'incompréhension et correspondrait en pratique à une absence de transparence. Il s'agit pourtant là de la façon la plus

[290] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

évidente de donner un accès au mode de fonctionnement de l'IA. En outre, une telle transparence risque de porter atteinte aux droits de propriété intellectuelle et aux secrets industriels et des affaires, qui ne sont pas de peu de poids dans la balance des intérêts. Dévoiler un algorithme peut aboutir à menacer un modèle économique. Concernant le secret des affaires, la directive européenne relative à la « protection des secrets d'affaires contre l'obtention, l'utilisation et la divulgation illicites » entérine l'impossibilité d'accéder à des demandes d'accès à une modélisation, aux objectifs d'un dispositif ou aux annotations d'un jeu de données[291]. Quant au considérant 63 du RGPD, il précise que le droit d'accès aux informations personnelles qui ont justifié une décision automatisée « ne devrait pas porter atteinte aux droits ou libertés d'autrui, y compris au secret des affaires ou à la propriété intellectuelle ». Avec la protection des données personnelles, le droit des brevets, le secret entourant certaines fonctions régaliennes afin de protéger la sécurité et l'ordre public ou encore l'article L. 341-1 du Code de la propriété intellectuelle qui interdit « toute extraction, par transfert permanent ou temporaire, de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu d'une base de données sur un autre support, par tout moyen et sous toute forme que ce soit », ce sont divers obstacles juridiques qui s'opposent à la possibilité d'obtenir des comptes quant aux décisions automatisées[292]. Au-delà, on peut aussi refuser de renseigner le monde entier quant au fonctionnement de son algorithme afin de le protéger des tentatives de manipulation. En effet, quiconque connaît les critères décisifs du travail d'une IA peut chercher à en tirer parti. C'est pourquoi Google, par exemple, veut éviter que les résultats fournis par l'algorithme PageRank de son moteur de recherche ne soient faussés par des acteurs qui auraient compris comment en exploiter la logique à leur profit, faisant remonter certains résultats favorables en tête et faisant disparaître d'autres.

Reste que s'il est difficile d'entrer dans les rouages hyper-complexes et d'accéder aux variables versatiles des algorithmes, on peut en revanche attendre de ceux qui les commandent et les fabriquent qu'ils rendent publics les objectifs qu'ils leur attribuent. Ceux-ci sont normalement à la fois plus explicites, plus compréhensibles et plus importants, mais à condition bien sûr qu'ils soient exprimés de bonne foi et sérieusement. On peut aussi demander à accéder aux données qui nourrissent l'algorithme, surtout s'il s'agit de ses propres données personnelles. Mais cela restera le plus souvent très insuffisant et ne permettra pas de faire le lien direct avec les résultats. On pourrait réclamer des explications quant aux pondérations affectées aux différentes hypothèses en fonction du profil et du contexte d'utilisation. Cependant, dès lors que des méthodes non paramétriques sont appliquées, les conditions des calculs sont révisées en permanence selon les actions des utilisateurs. Il est par conséquent vain de réclamer la levée d'un quelconque secret des algorithmes qui n'existe pas en tant que tel — ou qui existe sans que personne ne le connaisse précisément. On peut alors s'en remettre aux flux de données utilisés dans le calcul.

Par rapport aux algorithmes qu'on programme et dont on saisit donc plutôt bien le mode de fonctionnement, les réseaux de neurones artificiels permettant l'apprentissage profond peuvent difficilement sortir de leur condition de « boîtes noires ». Ces IA sont opaques par nature, à tel point que l'obligation de transparence et d'explicabilité semble devoir en interdire l'usage, ce qui est une conséquence radicale mais inéluctable. Les concernant, il est impossible de saisir quel raisonnement artificiel permet de passer des données initiales au résultat final. Il faut distinguer l' « apprentissage

[291] Dir. 2016/943/UE, 8 juin 2016, *Relative à la protection des secrets d'affaires contre l'obtention, l'utilisation et la divulgation illicites.*

[292] J.-M. Deltorn, « Quelle(s) protection(s) pour les modèles d'inférences », *Cahiers Droit, Science et Technologies* 2017, n° 7, p. 127 s.

supervisé », qui permet en théorie d'accompagner pas à pas l'apprentissage du système et donc d'en contrôler la logique, et l' « apprentissage non supervisé », dans lequel la machine jouit d'une entière autonomie et on se borne à contrôler la qualité a priori des résultats qu'elle produit. Cependant, une science naissante, forme de « rétro-ingénierie », tâche de développer des outils permettant d'un peu mieux saisir le fonctionnement des réseaux de neurones. L'explicabilité des algorithmes d'apprentissage automatique est un sujet si pressant qu'il constitue aujourd'hui un champ de recherche spécifique, qui doit être soutenu par la puissance publique. Trois axes en particulier paraissent devoir être approfondis spécifiquement : la production de modèles plus explicables, mais aussi la production d'interfaces utilisateurs mieux intelligibles et la compréhension des mécanismes cognitifs à l'œuvre pour produire une explication satisfaisante.

C'est pourquoi de nombreux spécialistes se réfèrent non à la transparence, trop illusoire que l'on soit en présence d'un algorithme programmé ou, surtout, d'un système d'apprentissage automatique, mais à l'intelligibilité ou explicabilité. Ce n'est pas d'avoir accès à un code source et à des paramètres incompréhensibles et parfois incertains qui importerait mais de pouvoir saisir la logique générale de fonctionnement de l'algorithme. Cette logique devrait pouvoir être comprise par tous et donc être énoncée textuellement, non à travers des lignes de code informatique. En ce sens, pour la CNIL, « rendre transparent un calculateur, cela doit avant tout être un travail pédagogique, pour essayer de faire comprendre ce qu'il fait. Ce qui est essentiel, ce n'est pas que le code soit transparent, c'est que l'on comprenne ce qui rentre et ce qui sort de l'algorithme ainsi que son objectif. C'est cela qui doit être transparent »[293].

Alors que la DARPA (Defense Advanced Research Projects Agency), l'agence chargée de la recherche au sein du département de la défense américain, mène un projet sur l' « XAI » (« *Explainable Artificial Intelligence* »), les principaux éditeurs de programmes d'IA (IBM, Google, Microsoft) jouent plus ou moins honnêtement le jeu de la transparence en proposant des suites logicielles (AI OpenScale chez IBM, Google AI Explanations) qui permettent à leurs utilisateurs de comprendre comment ces IA aboutissent à tel ou tel résultat plutôt qu'un autre. Ces instruments d'analyse demeurent toutefois limités : ils permettent seulement d'identifier les données ayant joué un rôle déterminant pour l'algorithme, sans dévoiler son mécanisme exact ni quelques relations explicatives. Ils permettent simplement aux ingénieurs non spécialistes de faire une grossière analyse de ces outils informatiques[294].

3. L'explicabilité : comprendre la « logique » de fonctionnement d'un algorithme

Quelles que soient les limites des opérations de contrôle qu'il serait possible d'envisager, des exigences de loyauté et de transparence à l'égard des individus dont les IA affectent la qualité de vie, les choix de vie les plus essentiels ou encore la réputation (numérique) doivent être posées. Tel est déjà le cas dans le RGPD, à la suite de la loi *Informatique et libertés*. Le texte européen, par son article 22, prévoit au bénéfice des personnes concernées le « droit de ne pas faire l'objet d'une décision fondée sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire ». Dans le même sens, le considérant 71 souligne que, « en tout état de cause, un traitement de ce type devrait être assorti de

[293] CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 51.

[294] J. Henno, « L'intelligence artificielle à l'heure de la transparence algorithmique », lesechos.fr, 9 mars 2020.

garanties appropriées, qui devraient comprendre une information spécifique de la personne concernée ainsi que le droit d'obtenir une intervention humaine, d'exprimer son point de vue, d'obtenir une explication quant à la décision prise à l'issue de ce type d'évaluation et de contester la décision ». Le RGPD insiste plus généralement sur le besoin de transparence. Les personnes touchées par des traitements de données doivent en être averties et doivent être alertées en cas de collecte de leurs données personnelles au plus tard au moment de celle-ci. L'information doit, selon l'article 12, être réalisée de façon « concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples », ce qui fait peser sur l'auteur du traitement de données une exigence renforcée tant sur la forme que sur le fond de l'information à transmettre. Au-delà, la satisfaction de l'exigence de transparence devrait dépendre de la lisibilité des algorithmes mobilisés, de la capacité à expliquer clairement et simplement, sans ambiguïté, comment ils traitent les données personnelles. Par exemple s'agissant d'un salarié ou d'un candidat dont la mutation ou le recrutement dépend des indications fournies par un outil informatique, il est impératif qu'il puisse saisir comment ce dernier opère et contrôler la validité des résultats qu'il génère. Mais, encore une fois, si le principe est aisé à poser, sa mise en œuvre effective est assez chimérique. Au moins faut-il pouvoir vérifier les informations détenues sur la personne et ayant été utilisées, ainsi que les corriger si nécessaire ou demander leur suppression, droits que le RGPD consacre.

L'obligation d'information est déjà un devoir crucial en robotique, pour les fabricants, les vendeurs et les prestataires de services[295]. Cela semble devoir être vrai dans tout le secteur de l'IA en général. Pour prévenir et encadrer les traitements algorithmiques, le droit français a renforcé l'exigence de transparence à l'occasion de la loi « Pour une République numérique »[296] et de ses décrets d'application[297]. Les entreprises sont ainsi tenues à une obligation de transparence concernant leurs algorithmes de classement, de référencement et de déréférencement. Les plateformes en ligne, en vertu de l'article L. 111-7 du Code de la consommation, sont tenues de « délivrer au consommateur une information loyale, claire et transparente sur […] les modalités de référencement, de classement et de déréférencement des contenus, des biens ou des services auxquels ce service permet d'accéder ». Cela est important tant les GAFAM et autres services de réseautage social devraient expliquer clairement au public ce que font leurs IA, quels présupposés elles suivent et en quoi elles influencent les comportements. Mais, alors que les algorithmes ont été, à grand renfort de savante communication, élevés au rang de mythes, jamais ces entreprises ne sauraient s'engager dans cette voie de leur propre initiative. C'est de façon très utopique que la CNIL suggère aux acteurs privés de contribuer à un cercle vertueux de transparence en ajoutant dans les espaces personnels des individus possédant un compte un onglet à cet effet, permettant notamment de consulter la liste des informations personnelles détenues par l'entreprise, de les corriger et de les supprimer, ainsi que des explications concernant la logique de l'algorithme[298]. Déjà dans la loi « Informatique et libertés » l'article 47 évoque la communication des règles du traitement et de ses principales caractéristiques par celui qui souhaite se prévaloir des exceptions au principe de prohibition des décisions prises sur le seul fondement d'un traitement automatisé de données à caractère personnel. Quant au RGPD, certains expliquent que la lecture conjointe des articles 22, 13 et 15 (relatifs au droit d'accès aux

---

[295] N. Nevejans, *Traité de droit et d'éthique de la robotique civile*, LEH édition, coll. Science, éthique et société, 2017.

[296] L. n° 2016-1321, 7 oct. 2016, *Pour une République numérique*.

[297] D. n° 2017-1434, 29 sept. 2017, *Relatif aux obligations d'information des opérateurs de plateformes numériques*.

[298] CNIL, *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une République numérique, déc. 2017, p. 56.

données collectées et au droit de connaître la finalité du traitement) induirait nécessairement, bien que tacitement, un « droit d'explication », donc le droit, pour tout individu, de demander des explications concernant une décision algorithmique prise à son sujet[299]. L'Union européenne souhaite en tout cas renforcer ce droit[300]. Et, plus clairement, les articles 12 à 15 du RGPD consacrent le droit de l'utilisateur d'exiger une explication sur le traitement des données dont il est la source. Ainsi le responsable du traitement doit-il informer la personne de « l'existence d'une prise de décision automatisée, y compris un profilage, [et] des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour [elle] ». L'article 12§1 précise que cette information doit être communiquée « d'une façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples ».

Des IA opaques sont la porte ouverte à tous les biais et toutes les discriminations dès lors que ceux-ci s'avèrent difficilement perceptibles. C'est pourquoi, malgré toutes les limites pratiques d'un tel principe, il faut plaider pour la transparence des algorithmes. Une telle transparence, par exemple lorsque le code source est rendu public ou communiqué aux individus concernés, est un premier garde-fou, bien que ces individus ne soient que rarement en mesure de comprendre le sens de ce code source. La « Déclaration de Montréal » pose en ce sens que « le code des algorithmes, publics ou privés, doit toujours être accessible aux autorités publiques compétentes et aux parties prenantes concernées à des fins de vérification et de contrôle »[301]. Mais un code source ouvert n'est ni nécessaire ni suffisant pour assurer la transparence — le risque est de voir la technicité extrême étouffer la clarté. En cas de prise de décision automatisée et même en cas de prise de décision assistée par informatique, il doit être nécessaire de garantir, dès la conception, la transparence et l'explicabilité des algorithmes. Dans l'idéal, il faudrait fournir un moyen simple de comprendre ce que fait l'IA, comment et pourquoi. Les informaticiens, comme tous ceux qui contribuent à la fabrication des IA, doivent garder en permanence à l'esprit l'objectif de faciliter leur vérifiabilité et définir des méthodes en vue de leur explication. Une IA digne de confiance suppose également d'assurer la « traçabilité » des systèmes d'IA[302]. Plus encore, il pourrait s'agir d' « intégrer des mécanismes de sécurité par conception et de sûreté dans le système d'IA permettant d'en vérifier l'innocuité à chaque stade »[303]. Et la Commission européenne de proposer de retenir une obligation de conservation des dossiers et des données, plus précisément des « dossiers de programmation de l'algorithme et des données utilisées pour entraîner les systèmes d'IA à haut risque, et, dans certains cas, […] des données elles-mêmes » en vue de pouvoir vérifier la conformité des systèmes d'intelligence artificielle[304].

Enfin, les administrations sont aussi soumises au principe de transparence. Sur ce point, l'article L. 311-3-1 du Code des relations entre le public et l'administration, introduit par l'article 4 de la loi « Pour une République numérique », prévoit que « toute décision individuelle prise sur le fondement

---

[299] B. Goodman, S. Flaxman, « European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation" », *AI Magazine* 2017, p. 50 s.

[300] Parlement européen, « Résolution sur une politique industrielle européenne globale sur l'intelligence artificielle et la robotique », 2018/2088(INI), n° 158, 12 févr. 2019.

[301] « Déclaration de Montréal pour un développement responsable de l'intelligence artificielle », Université de Montréal, 4 déc. 2018.

[302] Commission européenne, « Lignes directrices en matière d'éthique pour le développement et l'utilisation d'une IA », 8 avr. 2019.

[303] *Ibid*.

[304] Commission européenne, « Intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance », livre blanc, COM(2020) 65 final, 19 févr. 2020.

d'un traitement algorithmique comporte une mention explicite en informant l'intéressé. Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées par l'administration à l'intéressé s'il en fait la demande ». En pratique, ce sont surtout les décisions administratives assistées par informatique qui génèrent des contentieux autour du manque d'explication et de motivation — ce qui constitue un autre témoignage des peurs infondées du public à l'égard de l'État quand il fait une totale confiance aux multinationales privées qui, pourtant, à l'inverse des pouvoirs publics, défendent des intérêts privés et non l'intérêt général. Le Conseil constitutionnel a d'ailleurs ajouté que, « lorsque les principes de fonctionnement d'un algorithme ne peuvent être communiqués sans porter atteinte à l'un des secrets ou intérêts énoncés au 2° de l'article L. 311-5 du Code des relations entre le public et l'administration, aucune décision individuelle ne peut être prise sur le fondement exclusif de cet algorithme »[305]. La même règle devrait certainement s'appliquer aux relations entre personnes privées. Dans tous les cas, cependant, le principe de transparence et d'explicabilité ne saurait permettre seul de garantir l'effectivité des droits et libertés fondamentaux. Il n'en doit pas moins être considéré tel un principe essentiel. La transparence et l'explicabilité sont des préalables qui permettent de garantir que les autres principes seront observés.

4. Une condition de la confiance et de la responsabilité

Le résultat d'un algorithme doit pouvoir être compris et contesté. L'intelligence doit être augmentée, c'est-à-dire au service de l'humain. Ce qui se joue est le binôme homme-machine. Dans l'idéal, l'IA devrait être intuitivement compréhensible de l'humain. On ne peut pas avoir confiance dans ce que l'on ne comprend pas, ou bien alors il s'agit d'une confiance aveugle, soit une confiance précaire et insatisfaisante. La transparence et l'explicabilité sont importantes parce qu'elles sont une question de confiance pour les utilisateurs. Or on sait qu'une IA qui profite à tous, des industriels aux consommateurs en passant par les gouvernants et par les citoyens, est forcément une IA qui inspire la confiance[306]. Les enjeux économiques sont d'ailleurs élevés. George Akerlof, Prix Nobel d'économie en 2001, a montré que si un marché devient opaque pour certains de ses acteurs, il finit par s'effondrer[307]. Même si l'économie numérique repose aussi sur de fortes addictions et de grandes utilité et praticité, voilà ce qui pourrait lui arriver si un grand nombre de consommateurs perdaient confiance par la faute de manques de transparence.

Au-delà de l'aspect économique, l'explicabilité de ces technologies est une condition essentielle de leur acceptabilité sociale et, finalement, de leur viabilité à long terme. Il n'est ni admissible ni tolérable que, dans une société où règnent la démocratie et l'État de droit, certaines décisions importantes pour les citoyens et les individus puissent être prises sans explication et donc sans compréhension. Finalement, toute décision inexpliquée est forcément une décision injustifiée. Dans des domaines aussi décisifs pour la vie d'un individu que l'accès au crédit, à l'emploi, au logement, à la justice ou à la santé, on ne peut concevoir d'accepter l'injustifiable[308]. Plus généralement, on devrait toujours savoir suivant quel objectif un programme a été conçu et entraîné, notamment

---

[305] Cons. const., déc. n° 2018-765 DC, 12 juin 2018, *Loi relative à la protection des données personnelles*.

[306] A. Gattolin, C. Kern, C. Pellevat, P. Ouzoulias, « Intelligence artificielle : l'urgence d'une ambition européenne », rapport d'information, Sénat, 31 janv. 2019.

[307] J. Henno, « L'intelligence artificielle à l'heure de la transparence algorithmique », lesechos.fr, 9 mars 2020.

[308] C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 142.

lorsqu'il s'agit d'applications commerciales destinées essentiellement à manipuler le « temps de cerveau disponible » des consommateurs et, ce faisant, à maximiser les profits de l'entreprise, sans tenir compte ni de l'intérêt personnel de l'utilisateur ni de l'intérêt général. Il semble donc tout particulièrement nécessaire de faire preuve de transparence vis-à-vis de l'utilisateur ou du client en l'informant des motifs qui sous-tendent les propositions qui lui sont faites. Il conviendrait également de mieux informer les utilisateurs afin de ne pas les maintenir dans un hermétisme source d'une asymétrie d'informations préjudiciable. On devrait notamment leur permettre de mieux saisir la valeur réelle des informations qu'ils abandonnent, et qu'ils perçoivent comme insignifiantes, et leur détailler précisément les usages qui en sont faits, notamment lorsqu'il s'agit de ventes ou d'échanges.

Par ailleurs, consacrer un droit à l'explication et à la justification — compréhensible de tous — de toute décision assistée par machine doit pouvoir servir de base pour régler d'éventuels différends. Le fait de pouvoir expliquer comment une IA fonctionne est aussi un enjeu du point de vue de la responsabilité — même si l'on pourrait aussi poser le principe selon lequel quelqu'un qui prend une décision individuelle à l'aune d'une indication algorithmique en est toujours pleinement responsable. Il reste indispensable de mettre en place des mécanismes permettant d'assurer la responsabilité à l'égard des systèmes d'IA et de leurs résultats, et de les soumettre à une obligation de rendre des comptes. La responsabilité suppose donc la vérifiabilité. Pourtant, l'impossibilité pratique de suivre le trajet des données au sein des procédés décisionnels est un obstacle à l'établissement d'une chaîne de responsabilités des participants au processus de traitement des données. Cela contrarie aussi la possibilité de vérifier que différents principes, tels que celui de finalité du traitement, sont respectés. Si des accidents surviennent, l'IA doit être transparente et rendre des comptes à l'enquêteur pour que le processus interne qui a abouti à l'accident puisse être compris, que l'erreur ou le dysfonctionnement éventuel puisse être identifié. Suite à un accident, les parties, les juges, les avocats, les témoins et les experts doivent pouvoir s'appuyer sur toute information utile concernant le fonctionnement de l'algorithme. *In fine*, les systèmes qui ne peuvent être soumis à des normes de transparence et de responsabilité appropriées ne devraient pas être utilisés. Plus encore, un système d'IA honnête serait un système qui comporte des mécanismes d'identification des injustices et de réparation de celles-ci le cas échéant.

Enfin, c'est aussi par souci de transparence et afin de conforter la confiance des utilisateurs en même temps que la responsabilité des opérateurs que le droit, notamment à travers le RGPD, oblige à systématiquement alerter les autorités publiques compétentes et, éventuellement, les personnes affectées en cas de découverte d'erreurs de fonctionnement de l'intelligence artificielle, d'effets imprévus ou indésirables, de failles de sécurité ou de fuites de données.

5. Auditer ou certifier les algorithmes ?

Puisque les succès des IA dépendront de leur capacité à inspirer confiance — une confiance véritable, solide, pérenne, non une confiance aveugle comme aujourd'hui, susceptible de s'effondrer dès le premier scandale —, rendre les décisions automatisées ou semi-automatisées explicables et compréhensibles est un défi incontournable. Nul doute que bientôt l'acceptabilité de ces décisions sera étroitement liée à leur explicabilité. Cependant, étant donné la technicité du domaine, on pourrait préférer des explications fournies à des autorités compétentes et certificatrices plutôt qu'au grand public — mais encore faudrait-il que ce dernier fasse pleinement confiance à ces premières, ce qui est très incertain à une époque où l'on se méfie davantage de la puissance publique que des

puissances privées. Tandis que d'aucuns plaident pour la création de « chaînes d'audits »[309] permettant de compléter les contrôles *ex post* par des examens *ex ante*, Cédric Villani a proposé en ce sens d'accroître l'auditabilité des systèmes d'IA en constituant un groupe d'experts publics assermentés à qui il reviendrait de procéder à des audits d'algorithmes et de bases de données et de procéder à des tests[310]. Ces experts pourraient être saisis à l'occasion d'un contentieux judiciaire, dans le cadre d'une enquête diligentée par une autorité administrative indépendante ou suite à une demande du Défenseur des droits[311]. Cela permettrait d'imposer une stricte obligation de rendre des comptes (*accountability*) aux exploitants de techniques d'intelligence artificielle tout en préservant le secret des affaires, le modèle économique et les droits de propriété intellectuelle associés. Ce serait en quelque sorte une fonction tampon qui serait assurée entre les sphères du secret légitime et de l'information légitime.

De son côté, la Commission européenne préconise dans le même sens que les futurs « systèmes d'intelligence artificielle à haut risque » (concernant la santé par exemple) soient certifiés, testés et contrôlés, comme le sont les voitures, les cosmétiques ou les jouets[312]. L'Agence des droits fondamentaux de l'Union européenne, dans son rapport de décembre 2020 consacré à l'éthique de l'IA, préconise que chaque IA fasse l'objet d'une évaluation avant et pendant son utilisation[313]. Et certains ont proposé la création d'une agence des algorithmes, à l'image de la *Federal Drug Administration* aux États-Unis qui examine, teste et approuve des médicaments couverts par des brevets et des droits de propriété intellectuelle. Un tel organisme pourrait aussi exercer des fonctions de certification et de standardisation pour les objets fondés sur l'IA et interagissant dans l'environnement humain (voiture autonome, drone, robot etc.). Le public des utilisateurs des services devrait alors s'en remettre à la qualité du processus de certification et considérer que, si l'outil a été certifié, c'est qu'il répond à des standards qualitatifs suffisants. Le certificat serait la garantie que l'IA ne représente pas un risque pour les utilisateurs et les consommateurs. Il pourrait être délivré par un organisme indépendant sur le modèle des normes ISO. La CNIL envisage de placer cette activité sous le joug de la puissance publique, par le truchement d'une plateforme nationale constituée d'un comité d'experts. Elle interroge aussi l'opportunité de la labélisation privée, par des sociétés homologuées sur le modèle de la notation financière. On pourrait imaginer que les IA soient « notées » en fonction du niveau de confiance que l'on peut avoir en elles, à l'image du nutri-score ou de l'éco-score. Par ailleurs, toutes les IA pourraient ne pas être concernées de la même façon : les plus critiques, parce que des vies humaines en dépendraient, seraient très rigoureusement surveillées, tandis que les plus anecdotiques ne seraient pas contrôlées. On a ainsi imaginé une « échelle de risque » comprenant quatre catégories d'IA en fonction de leur impact sur l'intégrité d'une personne ou d'une communauté : si l' « explicabilité » serait le seul prérequis de l'entrée en fonction d'un algorithme de type 1, des publications scientifiques et des interventions humaines conditionneraient l'adoption d'un algorithme de type 4[314].

---

309 J. Charpenet, C. Lequesne Roth, « Discrimination et biais genrés », *D.* 2019, p. 1852.

310 C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 21.

311 *Ibid*.

312 Commission européenne, « Lignes directrices en matière d'éthique pour le développement et l'utilisation d'une IA », 8 avr. 2019.

313 Agence des droits fondamentaux de l'Union européenne, « Bien préparer l'avenir : l'IA et les droits fondamentaux », 13 déc. 2020.

314 J. Murawski, « Canada Prepares AI Standards for Government Agencies », *Wall Street Journal* 29 mai 2010.

Il s'agit donc ici d'une autre forme de transparence et d'explicabilité, peut-être davantage équilibrée, réaliste et faisable. Sans transparence, il serait impossible d'examiner le système en cause. Mais cette transparence est incomparable à celle qui consisterait, notamment, à publier le code source de ce système. De toute façon, le grand public, constitué de profanes, est généralement mal en mesure d'identifier, dans la conception ou la maintenance d'une IA, la présence de préjugés concernant le genre, l'origine, l'orientation sexuelle, l'âge etc. Par exemple, pour identifier une discrimination algorithmique, son existence et sa mécanique doivent être mises en évidence par un audit. Mais ce dernier risque de rencontrer des obstacles techniques, mais aussi juridiques. L'expérience du contentieux « admission post-bac » a témoigné des complexités technologiques que suppose un tel contentieux. En effet, si les discriminations alléguées devaient être établies à l'aune de la modélisation algorithmique et de son fonctionnement, grâce à une série de tests d'ingénierie, cela aboutit, en raison de l'incompétence et aussi de l'incompréhension des juges, à accorder un pouvoir immense aux ingénieurs, aux experts, conditionnant y compris l'introduction de l'action[315].

Certaines applications, notamment dans les systèmes critiques comme le transport aérien, par exemple, ne peuvent se passer de certification, donc d'un contrôle sérieux et rigoureux de leur fiabilité. L'informatique classique, fonctionnant avec des programmes et des règles, le permet. Les outils d'intelligence artificielle dont les résultats varient en fonction des données qui servent à les entraîner posent problème de ce point de vue. L'ambition d'auditer et certifier les IA achoppe sur l'incapacité par définition de garantir *a priori* le comportement d'un système qui apprend seul, de façon autonome. La certification formelle de l'apprentissage est ainsi en soi un nouveau champ de recherche. Pour que cette ambition puisse se réaliser, il faudra parvenir à développer des outils, des méthodes et des procédures suffisamment fiables et permettant en particulier de confronter les IA à leur cadre juridique et éthique. Pour l'heure, de telles capacités sont quasi inexistantes en raison de l'opacité des techniques d'apprentissage automatique et d'apprentissage profond. Des protocoles d'audit ou de certification ne peuvent donc qu'être balbutiants. Aujourd'hui, aucun pilote automatique d'avion ne repose sur des outils d'intelligence artificielle. Et les voitures autonomes ne sauraient être déployées à grande échelle sans aucune certification. Des initiatives existent toutefois déjà, à l'image de TransAlgo, plateforme d'évaluation portée par l'Institut national de recherche en informatique et en automatique (INRIA) et la CNIL et qui vise à mesurer la conformité aux règles juridiques et éthiques des services et produits issus des technosciences, donc à encourager la conception d'algorithmes responsables par construction. Certains, constatant qu'une tutelle humaine sur l'IA est impossible à mettre en œuvre, proposent de réguler les algorithmes par d'autres algorithmes[316]. Cette forme d'autorégulation se retrouve dans le projet du laboratoire allemand Algorithmic Accountability Lab ou dans la société Online Risk Consulting & Algorithmic Auditing fondée par Cathy O'Neil, dont l'objet est de déceler les biais algorithmiques des outils numériques. De tels algorithmes de contrôle d'autres algorithmes seraient évidemment fort utiles, par exemple dans les cas de contentieux autour de décisions algorithmiques.

Plus généralement, on pourrait investir significativement dans les techniques de rétro-ingénierie pour « tester » les IA, en particulier leur caractère non discriminatoire, l'absence de biais dans les jeux d'apprentissage. Il est même concevable de développer les explications en langage naturel par les outils informatiques basés sur l'apprentissage automatique. Tout acteur public ou privé souhaitant recourir à un service d'IA devrait spécialement se concentrer sur ces besoins. Il devrait,

[315] J. Charpenet, C. Lequesne Roth, « Discrimination et biais genrés », *D.* 2019, p. 1852.

[316] A. Sée, « La régulation des algorithmes : un nouveau modèle de globalisation ? », *RFDA* 2019, p. 830.

en outre, pouvoir prouver qu'il a procédé à des tests suffisants s'agissant de la répétabilité de son système, soit le fait que le même algorithme avec les mêmes données engendre les mêmes résultats, ainsi que de sa robustesse, c'est-à-dire son insensibilité aux légères modifications dans les données fournies en entrées.

On a aussi proposé de mettre en place pour la robotique et l'IA un système de test similaire à ceux des médicaments : les produits technologiques seraient évalués avant d'être commercialisés[317]. Pour obtenir la certification, il faudrait répondre à des exigences en matière de sécurité, d'efficacité et de non-discrimination. Enfin, on pourrait aller encore plus loin et imaginer des procédures d'audit ou de certification visant à garantir que les IA ne portent pas atteinte aux droits et libertés fondamentaux, mais aussi à d'autres formes de droits humains et sociaux. Tel pourrait être le cas, par exemple, des droits des travailleurs tels que consignés dans les chartes des droits de l'homme, les conventions de l'OIT et les conventions collectives, que le déploiement et la généralisation des IA ne devraient pas malmener. Un algorithme reflétant les conventions fondamentales de l'OIT et inclus directement dans le système pourrait précisément y veiller. En cas d'échec, le système s'arrêterait automatiquement[318]. On plaide ainsi de plus en plus pour la mise en place de mécanismes de certification des algorithmes, éventuellement avec des niveaux cumulatifs de certification[319].

Les derniers développements en la matière, très intéressants, se trouvent dans cette proposition de loi adoptée par le Sénat français en fin d'année 2020 et dont l'objectif est d'instaurer un « cyber-score » sur le modèle du nutri-score que l'on trouve désormais sur les emballages alimentaires, qui comprend cinq niveaux allant de A à E et du vert au rouge en fonction de la qualité nutritionnelle du produit. Il s'agirait ainsi d'informer le grand public de manière ludique et simple du niveau de sécurité qu'offrent les réseaux sociaux, les services cloud, les marketplaces et autres services fonctionnant à base d'IA. Si le Parlement dans son ensemble validait ce projet, serait créé un label pour informer les utilisateurs du niveau de protection des données personnelles lors de l'utilisation de certains services en ligne. Visible à chaque connexion, ce label reposerait sur des « critères objectifs et techniques » définis par l'Agence nationale de la sécurité des systèmes d'information (ANSSI). En outre, la proposition de loi prévoit que ce label ne devrait pas figurer dans les conditions générales d'utilisation, car il risquerait alors de se trouver noyé sous une masse d'informations illisibles pour l'internaute[320]. Au-delà, une telle certification devrait avoir pour effet de rendre les acteurs en question plus transparents et, plus généralement, davantage vertueux.

---

317 F. Pasquale, *New Laws of Robotics – Defending Human Expertise in the Age of AI*, Belknap Press, 2020.

318 UNI Global Union, « Les 10 grands principes pour une intelligence artificielle éthique », 2018.

319 Y. Meneceur, L'intelligence artificielle en procès – Plaidoyer pour une réglementation internationale et européenne, Bruylant, coll. Macrodroit microdroit, 2020, p. 357.

320 A. Vitard, « Données personnelles : bientôt un "cyberscore" pour afficher le niveau de protection des plateformes ? », usine-digitale.fr, 23 oct. 2020.

Conclusion

Emmanuel Kant a conduit l'humanisme jusqu'à son extrême aboutissement en érigeant l'humanité en fin de tout acte, de toute pensée ou encore de toute technique. Il a installé la personne humaine au-dessus de toute chose, hors de tout prix, inaliénable et imprescriptible, la déclarant insusceptible d'être réduite à un simple moyen permettant de subvenir aux besoins d'autrui. Le respect de la dignité que confère à l'homme la dimension suprasensible de sa nature et sa raison législatrice, source de la morale et de la loi, a donc vocation à s'imposer à toute action individuelle ou collective.

Dans son célèbre discours *Oratio de hominis dignitate* (1486), Pic de la Mirandole affirmait, de façon très originale pour l'époque, que la dignité spécifique à l'homme réside dans son libre arbitre et est ce qui lui permet de s'humaniser, contre tous les dogmatismes qui robotisent les êtres[321]. Nul doute que le libre arbitre, l'autonomie, la souveraineté individuelle et finalement la liberté se trouvent au cœur de la dignité des hommes. Toute personne est déjà une personne morale, un « être individuel, en tant qu'il possède les caractères qui lui permettent de participer à la société intellectuelle et morale des esprits : conscience de soi, raison, c'est-à-dire capacité de distinguer le vrai et le faux, le bien et le mal »[322]. Le nudge permanent diminue cette personne morale, qui tend alors à se réduire à l'aspect physique de la personne. La machinité des conduites ne va pas sans indignité. Et ceux qui succombent à des tendances animales sont plus dignes que ceux qui succombent à des tendances machinales. Mais la dignité dépend aussi du respect de la vie privée — le droit à la vie privée étant sans doute une liberté. Sans intimité, on n'est plus grand chose. Violer l'intimité, c'est violer la personne humaine. La protection des informations personnelles est, comme la défense du libre-arbitre, une question de dignité. Le personnalisme, cette doctrine morale et sociale fondée sur la valeur absolue de la personne[323], exposée dans le *Manifeste au service du personnalisme* d'Emmanuel Mounier en 1936, défendra la dignité de la personne tant dans sa liberté que dans sa vie privée. La dignité dépasse donc la question de la vie privée tout en étant profondément renouvelée par elle à l'ère des IA et de l'économie numérique.

Beaucoup voient dans la dignité le premier des droits de l'homme, celui qui aurait le plus naturellement vocation à l'universalité — même si l'on trouvera toujours des temps et des lieux réfractaires à toute raison d'être de la dignité humaine. Le respect de la personne humaine est un impératif moral indérogeable, sans rapport avec l'utilité ou la rationalité. Avec la dignité, ce qui fait le sacré de l'homme, ce qui fonde l'humanité, ce n'est pas l'inclusion dans une collectivité, le service de la communauté, mais l'inviolabilité de la personne humaine. L' « homme révolté » d'Albert Camus se révolte ainsi déjà contre sa condition, pour la défense de la dignité : « En assignant à l'oppression une limite en deçà de laquelle commence la dignité commune à tous les hommes, la révolte définit une première valeur »[324] — à la fin de ses *Réflexions sur la guillotine*, Albert Camus rappelle qu'on ne transige pas avec le respect de la personne humaine, ce qui suppose tout d'abord de refuser les camps de concentration et la peine de mort.

---

[321] M. Delmas-Marty, « Des humanismes à l'humanisme juridique : naissance et métamorphoses du mythe », cours au Collège de France, 5 janv. 2011.

[322] V° « Personne », in A. Lalande, *Vocabulaire technique et critique de la philosophie*, 10e éd., Puf, coll. Quadrige dicos poche, 2010, p. 758.

[323] V° « Personnalisme », in A. Lalande, *Vocabulaire technique et critique de la philosophie*, 10e éd., Puf, coll. Quadrige dicos poche, 2010, p. 756.

[324] A. Camus, « L'Homme révolté (1951) », in *Œuvres*, Gallimard, coll. Quarto, 2013, p. 1060.

Comme la personne humaine se divise en une personne physique et une personne morale, la dignité de cette personne comprend à la fois un pan physique et un pan moral. C'est essentiellement la dignité morale que les IA interrogent ; et elles le font tant en diminuant le libre-arbitre qu'en diminuant la vie privée. Protéger la dignité, c'est donc protéger la personnalité, soit la « personne qui réalise à un haut degré les qualités supérieures par lesquelles la personne se distingue du simple individu biologique »[325]. En droit, les droits de la personnalité sont des droits extrapatrimoniaux, donc des droits situés hors du patrimoine, sans valeur monétaire ou marchande, qui sont intransmissibles et insaisissables, donc qui ne peuvent être abandonnés ou vendus. Ils visent à protéger l'intégrité de la personne, qu'elle soit physique ou morale. Le droit au respect de la vie privée est au cœur de ces droits de la personnalité, notamment décliné en droit à l'image mais aussi sous la forme du droit à la protection des données personnelles ou encore du droit à l'honneur.

La dignité, notion centrale de la Déclaration universelle des droits de l'homme conçue tel un miroir inversé de la doctrine nazie, a pu sembler de plus en plus évidente à mesure que le XXe siècle se terminait. Au XXIe siècle, les bouleversements informatiques de l'homme redonnent une actualité aux combats menés en son nom. Les données personnelles sont l'ADN numérique de toute personne. Elles dévoilent ses valeurs culturelles et sa vie privée. Leur protection constitue dès lors un enjeu démocratique fort. Au moins quantitativement, c'est sur ce terrain que se déroule principalement le combat des droits de l'homme numérique. L'IA de confiance sera forcément une IA qui préserve l'intégrité de cette ADN numérique. Et l'IA qui inspire la méfiance est celle du capitalisme de surveillance, qui fonctionne grâce à des manipulations sans éthique des données personnelles qui ruinent la dignité morale des personnes et nient leurs personnalités. La difficulté, cependant, lorsqu'on pense les droits de l'homme numérique et spécialement la dignité numérique est l'absence de précédents historiques — qui ont nourri la DDHC ou la DUDH. Il faut avoir lu George Orwell pour mieux se faire une idée des enjeux en cause. La recherche de performances accrues des algorithmes et des réseaux de neurones appelle la collecte, le traitement et la conservation accrus de données à caractère personnel. Le développement de l'intelligence artificielle entre donc forcément en tension avec la dignité de la personne et ses normes dérivées, inscrites notamment dans le Règlement général sur la protection des données de l'Union européenne et la loi Informatique et libertés française. Le principe de finalité, en premier lieu, doit conduire à minimiser les collectes de données personnelles et à limiter la durée de leur conservation, alors qu'il est tentant de récupérer un maximum d'informations afin d'entraîner de futures IA.

La Commission européenne propose de créer un nouveau droit sur les données industrielles dans l'objectif de favoriser leur exploitation et encourager de nouveaux modèles économiques proposant des services innovants, socles du développement de l'industrie du futur[326]. L'insécurité et les asymétries résultant du cadre juridique relatif aux données générées par les machines entraverait le développement de l'économie européenne de la donnée. Mais l'urgence n'est-elle pas de plus et mieux protéger les données personnelles et la personnalité des citoyens européens, en termes de droit positif et surtout de droit effectif ? Le RGPD, entré en vigueur en 2018, va certainement dans ce sens en exigeant des collectes raisonnées et parcimonieuses des données, destinées à un usage précis. Il s'attaque ainsi directement au carburant de l'IA et proclame que la protection des personnes

[325] V° « Personnalité », in A. Lalande, *Vocabulaire technique et critique de la philosophie*, 10e éd., Puf, coll. Quadrige dicos poche, 2010, p. 757.

[326] Commission européenne, communication « Building a European Data economy », COM(2017) 9 final, 2017, p. 10 ; C. Zolynski, « Un nouveau droit de propriété intellectuelle pour valoriser les données : le miroir aux alouettes ? », *Dalloz IP/IT* 2018, p. 94.

physiques à l'égard du traitement de leurs données à caractère personnel est un droit fondamental protégeant le nom, les données de localisation ou encore les éléments spécifiques à l'identité physique, physiologique, génétique, économique ou sociale, ainsi que les informations biométriques et l'adresse IP. Peut-on s'en contenter ou bien de nouveaux progrès de la législation relative aux données personnelles peuvent-ils être espérés ? Le RGPD, « mélange de vieux et de neuf »[327], a été largement critiqué, notamment en raison de son trop faible impact concret au-delà des principes qu'il proclame. Et l'organisation Privacy Tech a par exemple publié un livre blanc intitulé « Une nouvelle gouvernance pour les données au XXIe siècle » esquissant de nouvelles pistes à suivre. Comme le souligne Isabelle Falque-Pierrotin, ancienne présidente de la CNIL, « la période du chèque en blanc sur les données est terminée »[328]. Le renouveau du régime juridique des données personnelles est nécessaire pour rééquilibrer le rapport des forces entre les individus et les multinationales du numérique. Pour l'heure, le troc « services contre données » n'est pas équitable.

Certains s'en satisfont pourtant, arguant qu'ils n'ont « rien à cacher ». Mais, comme l'expliquait l'ex-analyste-lanceur d'alerte de la NSA Edward Snowden, « lorsque vous dites "le droit à la vie privée ne me préoccupe pas, parce que je n'ai rien à cacher", cela ne fait aucune différence avec le fait de dire "je me moque du droit à la liberté d'expression parce que je n'ai rien à dire", ou "je me fiche de la liberté de la presse parce que je n'ai rien à écrire" »[329]. Une telle position est significative de l'égoïsme ambiant. Ce n'est pas parce que je n'ai pas d'intérêt à jouir personnellement d'un droit fondamental que je ne dois pas le défendre pour ceux qui en ont besoin. Les droits fondamentaux sont de toute façon des droits auxquels on ne peut pas renoncer, qui nous sont imposés pour notre bien même si l'on ne le conçoit pas ainsi — et Jean-Jacques Rousseau pouvait dire que « de lui-même le peuple veut toujours le bien, mais de lui-même il ne le voit pas toujours. La volonté générale est toujours droite, mais le jugement qui la guide n'est pas toujours éclairé. Il faut lui faire voir les objets tels qu'ils sont, quelquefois tels qu'ils doivent lui paraître, lui montrer le bon chemin qu'elle cherche, la garantir des séductions des volontés particulières »[330].

Face à l'automatisation du monde et de la gestion des vies humaines par des systèmes informatiques dont la prétendue neutralité dissimule la défense des intérêts économiques de quelques grandes firmes, il est un impératif moral et juridique : celui du respect de la dignité des hommes. On devrait pouvoir, par exemple, se suicider numériquement, pouvoir être un homme physique sans ombre numérique. Un droit de l'homme numérique pourrait être celui de ne pas être. L'invisibilité pourrait constituer un droit fondamental.

Il est toutefois aussi vrai que ce n'est que par l'équilibre, que par le « en même temps » qui est le seul mode de pensée juste, qu'on peut identifier une issue viable pour tout le monde. Protéger la vie privée, la dignité ou encore la liberté ne doit pas impliquer l'interdiction de l'IA qui est aussi la source de grands progrès. Si cette IA n'était plus utilisée que par des « passagers clandestins », profitant des services sans abandonner aucune donnée lui permettant de fonctionner, elle finirait par péricliter. On ne peut raisonnablement justifier d'utiliser gratuitement les services de Google ou Facebook sans rien donner ou payer en échange. Si l'on en venait à trop malmener le modèle économique de l'intelligence artificielle, on se priverait des nombreux progrès qu'elle permet ou, du

---

[327] V. Mayer-Schönberger, Y. Padova, « Regime Change? Enabling Big Data through Europe's New Data Protection Regulation », *Colum. Sci. & Tech. L. Rev.* 2016, n° 17, p. 315.

[328] I. Falque-Pierrotin, *Le Monde de l'économie* 28 mai 2017.

[329] Cité par P. Crochart, « Données personnelles : les ressources pour s'informer, les outils pour reprendre le contrôle », clubic.com, 21 févr. 2020.

[330] J.-J. Rousseau, *Du contrat social*, 1762, L. II.

moins, on devrait faire avec une baisse de la qualité des services en raison du tarissement des sources de données personnelles. Patrimonialiser les données à travers un droit de propriété produirait une perte d'efficacité des IA. Dès lors, la ville intelligente serait moins intelligente et la voiture autonome moins autonome. En laissant à chacun la possibilité de décider du sort des biens immatériels qu'il produit, cela nuirait à l'optimisation et à l'utilité recherchées par les IA, engendrerait des accidents et des pertes. Serait-ce le prix de la dignité et de la liberté ? C'est sur cette ligne de crête que se déroule la lutte pour les droits de l'homme numérique, entre le nudge permanent et l'autonomie contreproductive, entre la servitude volontaire et la liberté qui égard.

En matière d'IA, la prise en compte des droits de l'homme et autres valeurs éthiques doit être effectuée dans l'idéal au moment de la conception de l'outil, c'est-à-dire par ses créateurs eux-mêmes. C'est ce que l'on nomme l'éthique « by design » ou droit « by design ». Ce n'est alors plus le code qui fait le droit — « code is law » selon Lawrence Lessig, observant combien la régulation des comportements est largement déterminée par le fonctionnement des systèmes informatiques[331] —, mais le droit qui fait le code : law is code. Le code informatique est supposé reprendre à son compte et refléter des exigences juridiques ou éthiques. Il s'agit d'intégrer les principes éthiques et les exigences liées à une IA digne de confiance dans les produits et services utilisant l'IA sous la forme de précautions techniques dès le stade de leur conception.

On peut aujourd'hui encapsuler l'éthique dans des lignes de code, la formaliser afin de la programmer. Une façon radicale de répondre à la question de l'éthique de l'IA est donc de rendre la machine éthique en soi, respectant ces principes moraux automatiquement comme elle prend des décisions automatiquement. Cependant, des expressions comme « éthique des algorithmes » ou « IA éthiques » ne doivent pas être prises au pied de la lettre. Elles comprennent une part d'anthropomorphisme revenant à attribuer des capacités humaines à des machines. Si une IA se comporte de manière éthique, ce n'est que parce que des hommes l'ont voulu et l'ont conçue à ces fins. Les exigences ne concernent toujours que les hommes qui conçoivent, entraînent, déploient et utilisent les IA, tandis que ces dernières, si elles peuvent bien répondre d'une manière prédéfinie à une situation envisagée à l'avance, ne sauraient appliquer un raisonnement éthique, à base de conscience et de sensation du bien et du mal, à des cas nouveaux.

L'article 5 du RGPD pose le principe de « minimisation des données » : dans son fonctionnement même, l'IA doit prélever et exploiter uniquement les données personnelles qui lui sont strictement nécessaires en fonction de la finalité de l'application. On peut même imaginer la mise à disposition des utilisateurs de moyens techniques leur permettant de contrôler leurs données tels qu'un espace numérique sécurisé, sorte de portefeuille numérique. Seulement peut-on douter que la règle du by design soit aussi bien respectée lorsqu'elle est défendue par le législateur que lorsqu'elle fait suite à une initiative des acteurs privés eux-mêmes. Aussi observe-t-on que la règle selon laquelle les données collectées doivent être « adéquates, pertinentes et limitées » au regard de la finalité du traitement est peu respectée. Rares sont les opérateurs qui intègrent dans leurs algorithmes des mécanismes visant à limiter les prélèvements et réutilisations de données. On recourt beaucoup à l'idée de « privacy by design », des mécanismes de protection de la vie privée devant être intégrés nativement dans les systèmes d'IA. Mais, en pratique, on imagine bien que les multinationales du web y sont largement réfractaires. De la même manière, s'il est logique de plaider pour des « human rights by design » afin de protéger la liberté et prévenir les biais et les discriminations, on peut

[331] L. Lessig, *Code and Other Laws of Cyberspace*, Basic books, 1999.

difficilement s'attendre à ce que les acteurs concernés s'y conforment spontanément dès lors que cela nuirait à leurs modèles économiques et à leurs positions sur leurs marchés.

Il n'en faut pas moins défendre le respect des droits de l'homme numérique by design, l'intégration des droits et libertés fondamentaux dans le code, dès l'initiative d'un produit et au cours de sa conception. La commission Villani n'a pas manqué de le relever : « Les considérations éthiques doivent irriguer le développement même des algorithmes d'intelligence artificielle »[332]. Les entreprises pourraient d'ailleurs en tirer profit : en offrant à leurs utilisateurs un haut niveau de protection, elles ne seraient pas freinées dans l'innovation et le développement de services et elles pourraient même gagner certains avantages concurrentiels. Le très haut niveau d'expertise exigé pour la mise en œuvre de tels systèmes pourrait être garanti par la création d'une toute nouvelle profession réglementée, qui détiendrait à la fois des compétences techniques et juridiques.

D'autres exigences pourraient être imposées afin de favoriser une IA responsable. On pourrait notamment étendre les obligations de procéder à des études d'impact. Sur le modèle de l'étude d'impact sur les risques en matière de vie privée, rendu obligatoire pour certains traitements de données par le RGPD, il pourrait être institué une étude d'impact sur la liberté, sur l'autonomie ou sur le libre-arbitre, ainsi qu'une étude d'impact concernant les risques de discrimination. Il s'agirait ainsi d'amener les développeurs d'IA à ne pas se précipiter et à prendre le temps de se poser les bonnes questions aux bons moments.

Les études d'impact peuvent être un outil intéressant du point de vue de la « compliance », c'est-à-dire en matière de vérification de la conformité au droit, dans le cadre des processus destinés à assurer qu'une entreprise, ses dirigeants et ses salariés respectent les normes juridiques et éthiques qui leur sont applicables. Les « compliance officers », dans les entreprises recourant à l'IA et qui sont censées, depuis le RGPD, désigner un délégué à la protection des données chargé dans de veiller au respect du droit des données personnelles, ont forcément un grand rôle à jouer. Cela relève de leur intérêt car, en cas de mise en cause, les conséquences peuvent être financières et commerciales, mais aussi humaines : la société en question va forcément pâtir de la réputation qui lui sera faite. Or avoir procédé à une étude d'impact sérieuse et pouvoir en attester en cas d'attaque au sujet de la déontologie de l'entreprise est évidemment une force par rapport au concurrent qui s'est précipité sans réfléchir afin d'être le premier à investir un nouveau marché — un peu comme on traitera différemment le laboratoire qui met en vente un nouveau vaccin après de longues et minutieuses études et une multitude de contrôles et celui qui sort son produit sans garanties dans le seul but d'augmenter son chiffre d'affaires et de contenter ses actionnaires, une attitude qui risque fort de se retourner contre lui.

On pourrait souhaiter que des évaluations de l'impact sur les droits de l'homme à tous les stades de l'élaboration et de la mise en œuvre des systèmes d'intelligence artificielle soient menées. L'entreprise peut aussi s'efforcer à obtenir une certification, comme un label « IA responsable ». En droit européen, l'article 42 du RGPD prévoit l'éventuel recours à la certification pour démontrer le respect des exigences prévues à l'article 25, portant sur le principe de privacy by default. En contrepartie, l'entreprise doit bénéficier d'une confiance accrue, mais aussi de facilités administratives et juridiques. On pourrait même imaginer doter les systèmes d'IA d'une « boîte noire éthique ». La transparence de l'IA serait facilitée par la présence de ce dispositif enregistrant les informations pertinentes concernant son fonctionnement, permettant d'expliquer pourquoi et

[332] C. Villani, *Donner un sens à l'intelligence artificielle – Pour une stratégie nationale et européenne*, mission parlementaire, 2018, p. 22.

comment une décision a été prise, et contenant des précisions claires sur les considérations éthiques intégrées à ce système. Une lecture de la boîte noire éthique simple et rapide serait un gage de meilleures relations avec les utilisateurs.

Enfin, on ajoutera la « privacy by default » et les « human rights by default ». La première est définie par l'article 25 du RGPD qui impose l'application de « mesures techniques et organisationnelles appropriées, telles que la pseudonymisation », afin de protéger les données personnelles. Cette fois, il s'agirait de faire en sorte que les programmes informatiques soient par défaut protecteurs des droits de l'homme en général et de la vie privée en particulier, chaque utilisateur pouvant ensuite opter pour un service plus efficace mais plus intrusif. Cependant, on ne peut là aussi guère s'attendre à des miracles s'agissant de la coopération d'entreprises privées dont les niveaux de revenu sont à peu près proportionnels aux niveaux d'intrusion dans la vie privée de l'utilisateur et de placement dans un sillon de sa liberté.

Avec la compliance, les études d'impact et l'éthique by design ou by default, nous sommes pleinement dans une logique de responsabilisation (« accountability ») et de co-régulation. On attend des acteurs privés qu'ils fassent preuve de maturité, de sérieux et de responsabilité, en insistant sur le fait qu'il en va de leur intérêt car ce qu'ils ont de plus précieux est la confiance que leurs utilisateurs ou clients leur accordent. Le jour où celle-ci disparaîtra, ils péricliteront dans le même mouvement.

Table des matières