Le livre blanc de la Commission européenne sur l'intelligence artificielle : vers la confiance ?

Céline Castets-Renard
Professeure, Université d'Ottawa, Faculté de droit civil
Titulaire de la Chaire de recherche "Law, Accountability and Social Trust in AI", ANITI[1]
Chercheuse à l'OBVIA

La nouvelle Commission européenne d'Ursula Von der Leyen montre la volonté de poursuivre une politique forte en matière de numérique[2]. Elle a dévoilé son programme de travail en 2020[3] *pour une Europe adaptée à l'ère du numérique*[4], ainsi qu'une *Stratégie européenne pour les données*[5] et un livre blanc sur *l'intelligence artificielle : excellence et confiance*[6]. Le livre blanc fait suite à deux communications de la Commission sur l'intelligence artificielle (IA)[7] et aux *Lignes directrices en matière d'éthique pour une IA digne de confiance* du Groupe d'experts[8].

Forte du constat que l'IA se développe rapidement, générant des opportunités mais aussi des menaces, la Commission identifie deux catégories de risques qui tiennent, d'une part, à l'atteinte aux droits fondamentaux et, d'autre part, à la sécurité et au fonctionnement effectif du régime de responsabilité.
L'utilisation des systèmes d'IA peut porter atteinte à un nombre élevé de droits fondamentaux[9], spécialement les droits à la liberté d'expression, à la liberté d'association et de réunion, à la dignité humaine, la non-discrimination fondée sur le genre, les origines raciales ou ethniques, la religion ou la croyance, le handicap, l'âge ou l'orientation sexuelle, la protection des données personnelles et de la vie privée, le droit à un recours judiciaire effectif et à un procès équitable, ainsi que la protection des consommateurs. Ces violations surviendraient en cas de défauts de conception des systèmes d'IA dans leur ensemble et/ou de l'utilisation de données sans correction de potentiels biais. Les biais algorithmiques et discrimination sont en outre plus menaçants qu'en présence d'erreurs et biais humains puisqu'ils peuvent affecter de nombreuses personnes, en l'absence d'un mécanisme de contrôle social. L'IA augmente par ailleurs le risque de traquer et surveiller les habitudes quotidiennes des individus. Par l'analyse d'une énorme masse de données et en identifiant des liens entre elles, l'IA pourrait être aussi utilisée pour retracer et désanonymiser les données concernant les personnes, y compris pour les jeux de données qui ne contiennent pas, en tant que telles, des données personnelles. Alors que les systèmes sont de plus en plus amenés à prendre ou aider la prise de décision concernant directement les citoyens et entités légales, les autorités de contrôle et personnes affectées

[1] L'auteure remercie l'ANR-3IA et ANITI (*Artificial And Natural Intelligence Toulouse Institute*) : https://aniti.univ-toulouse.fr.
[2] Une Union plus ambitieuse : mon programme pour l'Europe : https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.
[3] Communication sur le programme de travail, COM(2020) 37 final : https://ec.europa.eu/info/publications/2020-commission-work-programme-key-documents_en.
[4] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_fr.
[5] COM(2020) 66 final : https://ec.europa.eu/info/files/communication-european-strategy-data_en.
[6] COM(2020) 65 final : https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.
[7] Communication du 25 avril 2018 de la Commission au Parlement européen, au Conseil européen, au Comité économique et social européen et au comité des régions, « *L'intelligence artificielle pour Europe* », COM(2018) 237 final ; Communication du 8 avril 2019 de la Commission au Parlement européen, au Conseil européen, au Comité économique et social européen et au comité des régions, « Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain », COM(2019) 168 final.
[8] Groupe d'experts de haut niveau sur l'IA constitué par la Commission européenne (Avril 2019).
[9] https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published.

risquent de manquer de moyens de vérifier comment une décision, impliquant un processus d'IA souvent opaque et complexe, a été prise et si les règles applicables ont été respectées. L'accès effectif à la justice pourrait ne plus être garanti non plus.

La deuxième catégorie de risques porte sur la sécurité et le fonctionnement du régime de responsabilité dans l'hypothèse où l'IA est intégrée dans des produits et services. Le résultat d'une défaut d'une technologie de reconnaissance d'objet peut par exemple provoquer une erreur d'identification sur la route par une voiture autonome et causer un accident, générateur de dommages matériels et corporels. De tels risques peuvent être provoqués par la défaillance de la conception des systèmes d'IA et/ou être reliés à la disponibilité et qualité des données. Par ailleurs, les autorités traditionnelles de contrôle, telles par exemple les agences de protection de la concurrence et de la consommation, n'ont ni le pouvoir d'agir ni les capacités techniques appropriées pour inspecter les systèmes. Il est aujourd'hui difficile de tracer les décisions problématiques prises avec des systèmes d'IA, si bien que les victimes n'ont pas accès à la preuve nécessaire pour demander réparation.

Afin de faire face à ces enjeux, la Commission entend défendre ses valeurs, tout en encourageant le développement des systèmes d'IA et empêchant la fragmentation du marché unique numérique. La Commission rappelle à juste titre qu'il convient avant tout de faire respecter la loi, à laquelle les développeurs et acteurs du déploiement de l'IA sont soumis. Il faut se réjouir que le droit soit remis au centre du processus normatif, tant la prolifération internationale des textes éthiques[10] a pu inquiéter sur l'émergence d'un *ethical washing*[11]. Contrairement aux *Lignes directrices* du groupe d'experts qui intègre le concept de "droits fondamentaux" à celui "d'objectif éthique", la Commission redonne toute sa place aux droits fondamentaux. Toutefois, elle reprend les sept points clés mis en avant par le groupe d'expert portant sur : la gouvernance et le contrôle humain ; la robustesse et la sécurité technique ; la vie privée et la gouvernance des données ; la transparence ; la diversité, non-discrimination et loyauté ; le bien-être social et environnemental, la redevabilité. Or, si ces valeurs sont importantes, elles sont de nature différentes. Certaines d'entre elles renvoient à un corpus législatif, telle la protection de la vie privée et la lutte contre la discrimination, dont le respect s'impose déjà. En revanche, d'autres valeurs paraissent trop floues pour être intégrées au droit, à l'instar du « bien-être » social ou environnemental dont on ignore s'il doit être relié aux règles du droit social ou environnemental. Cet objectif n'est en outre pas propre à l'IA, alors que d'autres paraissent au contraire en épouser parfaitement les caractéristiques, telles « la gouvernance et la contrôle humain », « la robustesse et sécurité technique », « la transparence » et enfin « la redevabilité »[12]. Le droit pourrait parachever ces objectifs. Nous nous concentrerons sur l'analyse des facteurs de confiance reposant sur l'acquis communautaire (1), à compléter par l'adoption d'un cadre règlementaire spécifique à l'IA (2).

1. L'adaptation de l'acquis communautaire dans le contexte de l'IA

L'acquis communautaire peut permettre d'appréhender certains risques des systèmes d'IA (A), sous réserve d'ajustements (B).

---

[10] https://www.coe.int/fr/web/artificial-intelligence/cahai. J. Fjeld, H. Hilligoss, N. Achten, M. Levy Daniel, J. Feldman, S. Kagay, Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches, Juil. 2019 : https://ai-hr.cyber.harvard.edu/primp-viz.html.

[11] C. Castets-Renard, Comment construire une intelligence artificielle responsable et inclusive ?, Recueil Dalloz 2020, p. 225.

[12] De telles valeurs émergent dans le milieu de la recherche en IA, à l'image de la conférence internationale organisée en machine learning et sciences sociales ACM (Association for Computing Machinery) FAccT/ML *Fairness*, *Accountability* et *Transparency* : https://facctconference.org.

A. L'acquis communautaire pertinent

Les risques d'atteinte aux droits fondamentaux peuvent d'abord être couverts par les directives sur l'égalité et la non-discrimination visant à assurer l'égalité de traitement entre les races et ethnies[13], dans l'emploi et le travail[14] et l'égalité de traitement entre les hommes et les femmes dans l'emploi et l'accès aux biens et services[15]. Des règles de protection des consommateurs viennent les compléter[16]. L'inclusion des personnes handicapées dans l'accès aux produits et services devrait aussi être améliorée en 2025 par la directive (UE) 2019/882 du 17 avril 2019 relative aux exigences en matière d'accessibilité applicables aux produits et services.

La Commission ne cite pas le règlement 2016/679/UE général de protection des données personnelles (RGPD), alors que l'article 22§1 du RGPD accorde un droit à la personne concernée 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. Trois catégories d'exception sont toutefois prévues en présence d'un contrat, d'un consentement explicite de la personne concernée ou si le droit de l'Union ou le droit de l'État membre l'autorise (art. 22§2). Ces exceptions doivent s'accompagner de sauvegarde des droits de la personne, au moins le droit d'obtenir une intervention humaine, d'exprimer son point de vue et de contester la décision (art. 22§3). Les décisions prises ne doivent en principe pas être fondées sur des données sensibles (art. 22§4). Ces dispositions créent une approche par les droits au profit des personnes concernées et pourraient servir de modèle pour encadrer l'IA dans son ensemble. La Commission devrait donc les intégrer à sa revue de l'acquis communautaire.

Quant à la sécurité et la responsabilité des produits, un corps de règles européennes peut être mobilisé, au travers de la directive générale sur la sécurité des produits[17], ainsi que les règles sectorielles couvrant différentes catégories de produits.

B. L'acquis communautaire à ajuster

Les règles de responsabilité posées par la directive relative aux produits défectueux[18] sont aussi susceptibles de s'appliquer, sous réserve toutefois d'ajustements eu égard à l'opacité, l'autonomie et l'évolutivité des systèmes d'IA. La notion de « produit » s'applique en effet mal à un système d'IA qui peut prendre la forme immatérielle d'un algorithme décisionnel. Ces systèmes s'apparentent davantage à des services plutôt qu'à des produits, alors que les règles de sécurité concernent principalement les produits. En outre, les systèmes d'IA peuvent être intégrés dès l'origine dans un produit ou intégrés *a posteriori*, ce qui pose la question du cycle

[13] Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en œuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique.

[14] Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail.

[15] Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès à des biens et services et la fourniture de biens et services ; directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte).

[16] Par ex. : directive 2005/29/CE sur les pratiques commerciales déloyales ; directive 2011/83/CE sur les droits de consommateurs ; directive 2019/2161/UE dite « omnibus » en droit de la consommation du 27 novembre 2019 modifiant la directive 93/13/CEE du Conseil et les directives 98/6/CE, 2005/29/CE et 2011/83/UE du Parlement européen et du Conseil en ce qui concerne une meilleure application et une modernisation des règles de l'Union en matière de protection des consommateurs.

[17] Directive 2001/95/CE du Parlement européen et du Conseil du 3 décembre 2001 relative à la sécurité générale des produits.

[18] Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux.

de vie des produits et de la robustesse de l'algorithme, tout comme celle du produit lui-même. La « défectuosité » sera aussi difficile à prouver et obligera à définir le défaut de sécurité « à laquelle on peut légitimement s'attendre » en matière d'IA. On peut aussi se demander si cette catégorie ne devrait pas être élargie, afin d'intégrer des faiblesses liées aux risques de cyberattaques ou de perte de connectivité. Enfin, la preuve du lien de causalité entre le dommage et le système d'IA sera difficile à rapporter. Par ailleurs, les relations homme-machine doivent être pensées, si on veut éviter les risques de manipulation mentale et intrusion dans l'intimité de la personne, spécialement par des robots humanoïdes, susceptibles de constituer des dommages moraux. La mise en œuvre de ces règles génère donc tout autant des problèmes conceptuels que probatoires, de nature à laisser les victimes sans droit effectif de recours en réparation.

Par ailleurs, les systèmes d'IA sont souvent complexes et nécessitent l'intervention de plusieurs catégories de producteurs et opérateurs, de nature à former une chaîne de responsabilité. À l'heure actuelle, les règles d'attribution de responsabilité ne sont pas encore définies ni à l'égard du consommateur ni des autorités de contrôle.

Sur la forme, la Commission envisage à la fois la réforme de la directive sur la responsabilité des produits défectueux et l'harmonisation ciblée de règles nationales de responsabilité, en vue notamment d'assouplir les règles nationales sur la charge de la preuve en cas de dommages causés par des systèmes d'IA, ce qui nécessitera l'accord des Etats membres.

2. Au-delà de l'acquis communautaire : réflexion sur l'adoption d'un cadre légal spécifique à l'IA

Si on peut s'accorder pour dire que certaines caractéristiques de l'IA justifient l'adoption d'une règlementation spécifique, son champ d'application et son contenu restent à déterminer. Alors qu'il n'y a pas *une* mais *des* IA, il est d'autant plus important de définir le champ d'application matériel. À ce jour, la Commission n'a pas donné de définition de l'IA ni établi le périmètre d'une future norme. On peut communément admettre que l'IA désigne « les systèmes qui font preuve d'un comportement intelligent en analysant leur environnement et en prenant des mesures, avec un certain degré d'autonomie, pour atteindre des objectifs spécifiques »[19]. Une telle définition est particulièrement large et vise différents objets, allant des systèmes automatiques de prise de décision à la robotique et peuvent donc être basés sur des logiciels agissant dans un monde virtuel ou intégrés dans le matériel. Cette lacune définitionnelle est d'autant plus regrettable que la Commission a précisé davantage son objet dans la *Communication* de 2018. Le Groupe d'experts constitué par la Commission européenne a aussi élaboré un document[20], sur lequel la Commission pourrait s'appuyer. Quant au champ d'application territorial, la Commission pose un principe d'extraterritorialité, comme elle a pris l'habitude de le faire dans sa stratégie du marché unique numérique[21]. Elle prévoit ainsi que les exigences de la future législation devraient s'appliquer à tous les fournisseurs proposant des produits ou services de l'UE basés sur l'IA, peu important qu'ils soient établis ou non dans l'UE, faute de quoi les objectifs de l'intervention législative ne seraient pas pleinement atteints. Quant au contenu d'une future norme, la Commission européenne semble s'orienter vers une législation sectorielle de l'IA fondée sur les risques, afin de s'assurer que l'intervention législative soit proportionnée et ciblée (A). Seuls les systèmes d'IA présentant des risques élevés devraient faire l'objet de règles spéciales (B).

[19] *Artificial Intelligence: A European Perspective*, Joint Research Centre, EUR 29425 EN, 2018.

[20] *A Definition of AI: Main capabilities and Disciplines*, Independent High-Level Expert Group on Artificial Intelligence, Commission européenne, avril 2019.

[21] Par ex. : article 3 du RGPD et article 1§2 du règlement (UE) 2019/1150 du Parlement européen et du Conseil promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne.

A. Les critères de détermination des risques des systèmes d'IA

La Commission distingue les applications à « haut risque » de celles qui ne le sont pas. Alors que les nouvelles obligations légales ne s'appliqueraient qu'aux premières, les secondes devront néanmoins respecter le droit de l'Union. Plutôt qu'une approche binaire, il serait préférable que la Commission propose une vision plus nuancée et graduée des risques, à l'instar de la directive sur la prise de décision automatisée et de l'outil d'évaluation de l'incidence algorithmique adoptés par le gouvernement fédéral canadien qui incluent quatre niveaux[22]. Cette approche par les risques n'est pas nouvelle et est déjà consacrée à l'article 35 du RGPD qui dispose que « lorsqu'un type de traitement, en particulier par le recours à de nouvelles technologies, et compte tenu de la nature, de la portée, du contexte et des finalités du traitement, est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement effectue, avant le traitement, une analyse de l'impact des opérations de traitement envisagées sur la protection des données à caractère personnel ». Les lignes directrices du G29 adoptées le 4 avril 2017[23] soulignent d'ailleurs que l'appréciation du « risque élevé pour les droits et libertés des personnes physiques » vise en premier lieu la protection des données personnelles et de la vie privée mais concerne aussi les risques d'atteinte aux autres droits fondamentaux, telle que la liberté d'expression, la liberté de pensée, la liberté de mouvement, l'interdiction des discriminations, le droit à la liberté de conscience et de religion. L'analyse d'impact sur les risques d'atteinte à la protection des données personnelles va donc bien au-delà et peut permettre de fonder une étude d'impact algorithmique sur les droits fondamentaux des systèmes d'IA[24].

L'approche par les risques suppose d'établir des critères précis d'évaluation. La Commission pose deux critères cumulatifs, liés au secteur d'activité et à l'usage envisagé, impliquant des risques significatifs, en particulier du point de vue de la protection de la sécurité, des droits des consommateurs et des droits fondamentaux. Tel sera le cas si les usages de l'IA produisent des effets sur les droits d'une personne ou d'une société ; s'il y a un risque de blessure, de mort, d'un dommage matériel ou immatériel significatif ; si le système produit des effets qui ne peuvent raisonnablement pas être évités par les individus ou entités légales. Quant aux secteurs d'activité concernés, ils devraient être listés de façon exhaustive et pourraient concerner par exemple la santé, le transport, l'énergie et certaines activités du secteur public. Il est dès lors fort probable que l'on s'oriente vers une législation sectorielle en matière d'IA, ce qui paraît être la meilleure réponse possible, eu égard à la diversité des objectifs, méthodes et impacts sociaux des systèmes amenés à se déployer dans des contextes variés et cadres législatifs préexistants différents. Le déploiement de la voiture autonome s'inscrit par exemple dans les règles du code de la route, de la responsabilité et des assurances, ce qui n'a rien à voir avec le cadre légal d'une autre application comme la police prédictive, rattachée à l'action de forces de l'ordre. Certaines situations peuvent être considérées systématiquement à haut risque, comme par exemple l'utilisation de systèmes automatiques d'IA dans le processus de recrutement ou encore dans les situations impactant les travailleurs, eu égard au principe européen d'égalité de l'emploi et du travail. Il en est de même des systèmes biométriques d'identification et autres technologies, utilisés à des fins de surveillance.

[22] https://www.tbs-sct.gc.ca/pol/doc-fra.aspx?id=32592.

[23] Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purpose of Regulation 2016/679, WP 248.

[24] En ce sens, voir Margot Kaminski et Gianclaudio Malgieri, Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations (September 18, 2019). U of Colorado Law Legal Studies Research Paper No. 19-28. Disponible sur SSRN: https://ssrn.com/abstract=3456224.

Même si la Commission européenne n'en fait pas mention, il serait pertinent de se référer à l'article 35§3 du RGPD qui prévoit l'analyse d'impact relative à la protection des données personnelles. Ce type d'analyse (*privacy impact assessment*) ne permet pas la prise en compte de tous les risques algorithmiques qui peuvent reposer sur des discriminations de groupe et non pas uniquement sur l'utilisation de données personnelles. Cependant, la méthode et les critères retenus sont utiles pour déterminer les exigences en matière d'étude d'impact algorithmique. Ainsi, une telle étude est requise, précisément dans l'hypothèse où des traitements automatisés produisent des décisions ayant des effets juridiques à l'égard d'une personne physique ou l'affectant de manière significative de façon similaire, c'est-à-dire en cas de prise de décision algorithmique négative pour les individus. L'étude d'impact est également requise en cas de traitement à grande échelle de données sensibles ou de données relatives à des condamnations pénales ou encore pour la surveillance systématique à grande échelle d'une zone accessible au public. L'article 35§4 du RGPD prévoit en outre que l'autorité nationale de contrôle doit publier une liste des types d'opérations de traitement pour lesquelles une analyse d'impact est requise ou non. La CNIL a publié deux listes en ce sens[25].
Dans l'hypothèse où le traitement envisagé n'entre dans aucune des deux listes, il convient alors de se référer aux neuf critères publiés dans les lignes directrices du G29 : évaluation ou *scoring*, y compris profilage et prédiction ; décision automatique avec effet légal ou similaire ; surveillance systématique ; collecte de données sensibles ou de données à caractère hautement personnel ; collecte de données à large échelle ; croisement ou combinaison de jeux de données ; données concernant les populations vulnérables ; usage innovant, application de nouvelles technologies ou solutions organisationnelles ; exclusion du bénéfice d'un droit, d'un service ou d'un contrat. La CNIL estime que si deux des critères au moins sont remplies, le traitement présente des risques élevés et requiert une étude préalable d'impact. Sinon, le responsable de traitement devra justifier et documenter les raisons pour lesquelles il n'a pas réalisé cette analyse. Si le traitement présente des risques résiduels élevés, le responsable doit consulter l'autorité nationale de protection des données.
Même si la Commission de mentionne pas cette méthodologie issue du RGPD, l'étude d'impact algorithmique pourrait se baser sur une approche comparable. Les autorités nationales de protection des données personnelles n'ont cependant pas compétence pour la prendre en charge, aussi faudrait-il créer des autorités de contrôle algorithmique.

B. Les obligations inhérentes aux systèmes d'IA à haut risque

La Commission européenne tient compte des propositions du Groupe d'expert pour déterminer les obligations applicables aux systèmes d'IA présentant des risques élevés. Ces obligations concernent les données d'entraînement, la conservation des enregistrements et données, les informations à fournir, la robustesse et la justesse et le contrôle humain. Enfin, des dispositions spécifiques pourraient concerner certaines applications d'IA, relatives à l'identification biométrique.
Ainsi, les données d'entraînement devraient être suffisamment larges et couvrir tous les scenarii pertinents pour éviter les situations dangereuses, mais aussi représentatives pour s'assurer que les dimensions de genre, ethnies et autres terrains de discrimination soient correctement reflétées dans les jeux de données. L'utilisation des données d'entraînement devrait naturellement respecter les règles de protection des données personnelles.
La conservation des enregistrements et données d'entraînement devrait permettre d'assurer une traçabilité et vérification. Pourraient ainsi être conservées les données exactes des jeux de données utilisés pour entraîner et tester les systèmes d'IA, y compris une description des caractéristiques principales et la façon dont les jeux de données ont été sélectionnés. Dans

[25] https://www.cnil.fr/fr/ce-quil-faut-savoir-sur-lanalyse-dimpact-relative-la-protection-des-donnees-aipd.

certains cas, il pourrait aussi être exigé que le jeu de données lui-même soit conservé. La documentation sur la programmation et les méthodologies d'entraînement, les processus et technologies utilisées pour construire, tester et valider les systèmes d'IA, y compris dans le respect de la sécurité et en évitant les biais qui pourraient entraîner une discrimination interdite. Ces informations devraient être disponibles à la demande, en particulier pour permettre les tests et inspections des autorités de contrôle. Des arrangements permettraient d'assurer la confidentialité de l'information et le respect des secrets d'affaire.
La transparence exige par ailleurs la fourniture d'informations aux usagers concernant les capacités et limites du système, ainsi que le fait d'interagir avec un système d'IA et non avec un humain. Cette exigence ne devrait pas être respectée s'il est évident pour le consommateur qu'il interagit avec une machine. L'information doit être objective, concise et compréhensible.
Les systèmes d'IA devraient aussi être robustes, exacts et gérer les erreurs et inconsistances tout au long de leur cycle de vie. La robustesse implique aussi la résilience des systèmes qui doivent résister aux cyberattaques, à la manipulation des données ou algorithmes. Des mesures de minimisation des risques doivent alors être prises.
Le degré du contrôle humain doit dépendre des risques et peut prendre plusieurs formes. Par exemple, les résultats d'un système d'IA pourraient ne pas devenir effectifs sans vérification humaine ou encore un contrôle humain pourrait être exigé après coup. Egalement, le contrôle d'un système d'IA peut aussi se faire pendant son utilisation et le contrôleur aurait alors la capacité d'intervenir en temps réel. Enfin, des opérations contraignants les systèmes d'IA sont susceptibles de s'imposer dans la phase de conception.
Enfin, des exigences spécifiques concerneraient certaines applications d'IA, comme les procédés d'identification biométrique. Prenant acte des risques particulièrement élevés en la matière, liés par exemple à l'usage de la reconnaissance faciale dans les espaces publics, la Commission rappelle la nécessité de respecter la règlementation sur la protection des données personnelles et la Charte des droits fondamentaux de l'UE. Alors que l'on s'attendait à une position politique forte de l'Union européenne vers un moratoire ou la prohibition de la reconnaissance faciale, le livre blanc ne prône pas une telle mesure radicale et, bien au contraire, prévoit peu de dispositions spécifiques. Il se contente de rappeler l'acquis communautaire sans envisager des règles plus sévères. Cette disposition paraît bien timorée et en retrait par rapport aux débats actuels[26]. La reconnaissance faciale est une méthode encore en cours d'évolution qui présente des risques élevés d'erreur et discrimination. Les finalités et le contexte de leur utilisation doivent être pris en compte, en différenciant leur usage à des fins de sécurité nationale par les forces de police *versus* un usage commercial, étant entendu que les principes de finalité, nécessité, proportionnalité, limitation dans le temps et l'espace sont à respecter en tout état de cause.

La Commission soulève par ailleurs la question de l'imputabilité de la responsabilité et considère que le futur cadre règlementaire devrait prévoir que chaque obligation serait assumée par l'acteur le mieux placé pour prendre en charge le risque potentiel. C'est ainsi que les développeurs devraient satisfaire les obligations relatives à la phase de conception, alors que les acteurs du déploiement prendraient en charge les obligations finales, comme le contrôle humain. De telles règles sont sans préjudice des droits des utilisateurs finaux qui doivent pouvoir agir contre les producteurs responsables du fait des produits défectueux, à charge ensuite pour les acteurs de la chaîne de responsabilité de se retourner contre le responsable par

[26] Par exemple la collecte massive d'images faciales sur les réseaux sociaux par la société Clearview AI pour revendre aux autorités de police une technologie de reconnaissance faciale. Une enquête a ainsi été ouverte par les autorités de protection des renseignements personnels du Canada, du Québec, de la Colombie-Britannique et de l'Alberta : https://www.cai.gouv.qc.ca/commissaires-lancent-enquete-conjointe-clearview-ai-dans-contexte-preoccupations-croissantes-quant-utilisation-technologie-reconnaissance-faciale.

l'exercice d'une action récursoire. Une telle approche peut toutefois paraître un peu naïve, tant il n'est pas simple de distinguer dans les faits le rôle joué par chacun. Il faudrait donc exiger des acteurs eux-mêmes une clarification de leur rôle à toutes les étapes du processus, de la conception au déploiement.

Enfin, le respect de la loi suppose d'établir des règles obligatoires et de donner compétence à des autorités nationales et européennes de contrôle pour en garantir le respect. Une conformité préalable pourrait en outre être assurée par des procédures de *testing*, inspection ou certification. Cela pourrait inclure des vérifications des jeux de données et algorithmes pendant la phase de développement. Il pourrait être pris modèle sur les mécanismes de conformité déjà existants. Dans la mesure où certains systèmes évoluent et apprennent de l'expérience, cela suppose aussi de réaliser des études pendant la durée de vie des systèmes d'IA. Un droit de recours judiciaire effectif aux parties affectées négativement doit être garanti. La Commission envisage aussi une labellisation des systèmes ne présentant pas de risques élevés. La participation serait volontaire mais le respect des exigences obligatoire. Il conviendrait ici d'ajouter des règles strictes de contrôle externe et indépendant, pour ne pas permettre aux acteurs de s'auto-règlementer. De façon générale, des organes de contrôle algorithmique suffisamment compétents et dotés de moyens doivent être créés. Etant donné la nécessité des contrôles fréquents *in situ*, le niveau national paraît plus pertinent que le niveau européen. D'ailleurs, la Commission souhaite créer une gouvernance de l'IA par la coopération entre les autorités nationales compétentes.

La démarche de la Commission européenne va dans le bon sens. Il convient à présent d'aménager l'acquis communautaire et de réfléchir plus précisément encore au contenu d'une législation propre à l'IA, afin de donner corps aux vœux d'une IA responsable et inclusive.