IA, Erreur et Droit - La responsabilité des acteurs de l'IA.

Colloque : Should we Delete the Error? The value of Errors in the age of Robotics & AI UCLille , Sept. 7

1. Parler en droit d'erreur dans la conception ou le développement d'un produit en l'occurrence un système d'IA, c'est inévitablement s'interroger sur la responsabilité des acteurs de l'IA. A. Christie écrivait : « Je sais qu'il y a un proverbe qui dit que l'erreur est humaine mais une erreur humaine n'est rien à côté de ce qu'un ordinateur peut faire s'il essaye » … et j'ai envie d'ajouter bien plus encore s'il s'agit d'IA. Nous reviendrons sur l'amplification des risques nés du fonctionnement des systèmes d'IA par rapport au fonctionnement de nos systèmes informatiques classiques mais, auparavant, arrêtons-nous, si vous le voulez bien, à la notion d'intelligence artificielle.

2. L'intelligence artificielle désigne, selon une récente résolution du Parlement européen, *les systèmes logiciels qui, entre autres choses, collectent, traitent et interprètent des données structurées ou non structurées, identifient et établissent des modèles afin de tirer des conclusions ou prennent des mesures dans une dimension physique ou virtuelle sur la base de ces conclusions*; ». Il s'agit donc de tout système de *machine learning* supervisé ou non, travaillant de manière profonde ou non sur des bases de données structurées ou non, capable, suivant un modèle non point causaliste mais statistique de prendre des décisions en particulier relatives à une personne physique, morale ou non. La notion englobe également les outils robotiques, machines capables d'exécuter des tâches traditionnellement accomplies par des êtres humains, ces outils se multiplient : véhicules intelligents, enceintes connectées, drônes, robots aide-soignant, etc. Ces outils, qu'il s'agit de robots ou de systèmes IA, d'une part, se nourrissent de données

captées à travers de multiples sources (ainsi, les systèmes de santé *self quantified* ou un GPS, qui envoie en temps réel les données de localisation) et, d'autre part, échangent des données avec des terminaux plus ou moins intelligents (ainsi, le système de reconnaissance faciale, qui équipe un casque de policiers et qui permet dans une foule de reconnaître des personnes déjà suspectes). C'est à la fois en aval et en amont que l'infrastructure de la plupart des systèmes d'IA intègre ce qu'il est convenu d'appeler l'internet des objets.

3. Notre propos, dans un premier temps, identifie les textes légaux européens à prendre en considération. On notera l'élargissement des risques pris en compte par ces documents au fur et à mesure du temps et sur cette base, nous chercherons à dresser une typologie des risques liés à l'utilisation des systèmes d'IA et montrerons comment la première réponse légale s'est centrée sur les risques classiques visés par les textes de la responsabilité civile, la deuxième, encore que non totalement adéquate, sur les risques liés à la protection des données et ce avec le RGPD, la troisième, avec la nouvelle Commission et le nouveau Parlement européen, tente une approche globale et élargie dite éthique des risques provoqués par l'IA.

4. Le premier texte souvent cité date du début 2017. Il s'agit d'une Résolution du Parlement européen relative aux règles de droit civil applicables aux robots, singulièrement celles de la responsabilité du fait des produits et des services. Un rapport plus récent (novembre 2019) d'un groupe d'experts européens sur la responsabilité civile de l'IA : **LIABILITY FOR ARTIFICIAL INTELLIGENCE AND OTHER EMERGING DIGITAL TECHNOLOGIES** , devait fortement nuancer les propos de la résolution du Parlement. Le deuxième est sans contexte le RGPD dont l'application est sensée museler les dangers que représentent l'IA vis-à-vis de notre vie privée. Le troisième s'inscrit dans une perspective plus globale avec les **ETHICS GUIDELINES FOR TRUSTWORTHY AI,** publiées par le HLGE de la Commission en avril 2019, repris par la Commission en avril 2019. Depuis, le 'White Paper' de la Commission : On Artificial Intelligence - A European approach to excellence and trust publié en février de cette année et surtout les travaux du Parlement **"PROPOSAL FOR A RESOLUTION FROM THE EUROPEAN PARLIAMENT containing recommendations to the Commission concerning a framework of ethical aspects in artificial intelligence, robotics and related technologies »** marquent la volonté des instances européennes d'aborder de manière plus globale les risques liés

à l'iA et souligne l'importance d'une approche éthico-légale d'une gouvernance de tels risques.

5. Avant d'aborder l'analyse des réponses juridiques tentons de répondre à deux questions. Quelles caractéristiques propres ou en tout cas intensifiées par l'intelligence artificielle justifient la prise en compte par le Droit de risques nouveaux ou amplifiés liés au développement de cette technologie. La **sécurité des traitements** d'IA représente un défi dans la mesure où ces systèmes reposent souvent sur la collecte de données via des réseaux vulnérables. On cite par ailleurs les risques de *hacking* et d'intrusion : la modification de quelques pixels ou l'ajout de quelques données peuvent fausser totalement les résultats affichés par le système.

   La **complexité du montage** des systèmes d'IA et le recours à de nombreux acteurs dans la conception, le déploiement et l'utilisation du produit ou du service incorporant un système d'IA représentent une autre difficulté lorsqu'en cas d'accident, on s'interrogera sur la question de savoir qui sera responsable dans cette chaîne d'acteurs ?. On évoque le ou les fournisseurs de données, le fournisseur de l'algorithme, l'intégrateur aux besoins particuliers de celui qui désire opérer le système. Ce dernier lui-même nourrira avec sa propre pratique et ses propres données le système, on ajoutera l'utilisateur final qui peut-être un particulier, acheteur d'un drône ou utilisateur d'un robot aide-soignant.

   L'opacité en particulier, les systèmes dits de '*deep learning*' opèrent des connexions entre réseaux neuronaux de manière opaque pour celui qui subit les décisions de la machine voire pour celui qui délègue à la machine le pouvoir de prendre de telles décisions. On ajoute les risques supplémentaires d'opacité due à **l'autonomie des systèmes**, qui évoluent au hasard des nouvelles données reçues et des interconnexions qu'elles suscitent, sans contrôle direct humain.

   Ensuite, les résultats sont basés sur des corrélations et non une logique humaine et peuvent être **biaisés** par la qualité ou la partialité des données sur lesquelles le système travaille ou par les préjugés conscients ou inconscients de leurs concepteurs. Le **conservatisme des résultats** s'explique par le fait que les systèmes d'IA travaillent sur des données du passé et s'avèrent donc peu aptes à proposer des solutions innovantes. Enfin, on dénonce la **puissance prédictive** de l'instrument, capable de prédire le comportement futur d'une personne, selon le slogan du patron

d'Amazon '*Before you adress your order, we have already packaged your command* ». L'instrument permet de nous profiler voire de nous manipuler ou de nous stigmatiser et cela , d'autant plus que les 'vérités' sorties de l'ordinateur sont considérées comme d'autant plus fiables qu'elles sont basées sur un principe : « Si les personnes mentent, leurs données, elles, ne mentent pas. »

6. A tous ses facteurs, correspondent des risques de natures diverses, illustrés dans les exemples qui font suivre par des cas bien réels. Traditionnellement, le droit n'envisageait que les risques physiques liés à l'utilisation, en particulier des robots. L'aide-soignant qui pousse la personne âgée dans les escaliers suite à une erreur de programmation, la personne écrasée par une voiture connectée qui a confondu la victime avec une simple nappe de brouillard, etc. Les risques financiers liés à des résultats erronés comme ceux d'un mauvais calcul d'une prime d'assurance apparaissent bien vite également comme à prendre en considération. A tout cela, on peut ajouter les dommages moraux que subit une personne, par exemple, une personne noire qui serait victime d'une suspicion suite à une reconnaissance faciale insuffisamment précise parce qu'entraînée sur des sujets en large majorité blancs. Dans tous les cas cités, il s'agit de **risques individuels** amplifiés dans le contexte des applications d'IA.

7. Plus récemment, la littérature a mis en évidence des **risques collectifs** subis par des groupes ou par la société elle-même. Ainsi, il a été montré que le profilage de l'IA permet l'exclusion de certaines catégories de personnes en ce qui concerne par exemple l'accès au logement ou la majoration du prix de certains produits offerts sur le web, pour certains consommateurs. Il s'agit de discrimination de groupes de personnes unies parfois selon des critères classiques comme l'appartenance ethnique, l'adhésion à des courants religieux ou philosophiques, etc., mais souvent selon des critères déterminés par la machine elle-même et donc imprévisibles par ceux qui subissent ces discriminations, ce qui rend difficiles leur défense commune. Au-delà, on note combien l'IA permet à certaines entreprises, en particulier les plateformes, de renforcer leur position sur le marché tant vis à vis des consommateurs que de leurs concurrents. On note, ensuite, combien l'IA peut également introduire insidieusement une forme de régulation nouvelle à distance des règles légales. Ainsi, l'assurance auto '*one-to–one* », calculée en fonction du risque représenté par chaque conducteur et mesurée grâce à mille

capteurs, s'écarte du principe de mutualisation chère au droit des assurances ; l'utilisation de systèmes de détection par l'IA des risques de délinquance induit une possibilité d'intervention avant même l'infraction et s'écarte du principe suivant lequel seule l'infraction peut déclencher l'action policière. Enfin, l'affaire Cambridge Analytica et la diffusion des *Fake news* montrent combien l'IA peut mettre nos démocraties en danger. Notre thèse est que le Droit européen s'est progressivement élargi à la prise en considération de tels risques sociétaux mais avant d'analyser comment il appréhende ces risques dans les textes les plus récents, voyons la façon dont il aborde, par le droit de la responsabilité civile et par le droit de la protection des données, les risques individuels plus classiques et les difficultés rencontrées face à la nouveauté radicale que représente l'IA.

1ère réponse : le droit de la responsabilité civile

8. Dans le cadre de l'exposé oral, nous nous limiterons à exposer les conclusions toutes récentes du *HLGE on liability,* qui témoigne à suffisance des difficultés d'appliquer à l'IA soit le régime général de responsabilité fondée sur la faute (oui mais la faute de qui ? comment la prouver ? quel critère retenir : référence à la personne ou au produit (*reasonable AI system*) OK mais comment se référer au produit normal lorsque le plus souvent il s'agit d'un produit innovant ?). Quoiqu'il en soit, exiger la faute tendrait à une exonération le plus souvent de celui qui a développé un système d'IA. A l'inverse, la responsabilité du fait des produits fait peser une responsabilité stricte sur le dos du « producteur », certes à condition qu'un système d'IA peut être appeler 'produit', que la notion de défaut puisse être étendue alors que le produit évolue et fonctionne de manière opaque et que celle de producteur puisse être éclaircie face aux nombreux acteurs intervenant dans la chaîne de production. La rigueur de ce régime apparaît contreproductive par rapport au souci d'encourager l'innovation. On conçoit dès lors que la proposition faite récemment par le groupe d'experts représente un compromis intéressant entre ces deux régimes de responsabilité extrêmes.

9. Résumons la de la manière suivante : la première règle est de créer légalement des **obligations préventives** du danger, c'est-à-dire pour ceux qui conçoivent des systèmes d'IA ou les mettent sur le marché, de veiller, avant même toute 'mise en circulation' mais également par la suite, à la qualité du produit, à son absence d'erreur et à son bon fonctionnement :

l'absence de suivi de ces devoirs entraîne une responsabilité automatique de celui qui s'en est abstenu ; la deuxième règle est d'affirmer que la personne qui utilise un système IA ou un robot est responsable de la même manière que si le dommage avait été causé par un humain sous ses ordres. La troisième règle fait supporter la responsabilité première à celui qui a le meilleur contrôle sur le système. Ainsi, bien souvent, ce ne sera pas l'utilisateur final qui sera responsable des dégâts causés par son robot mais bien le producteur ou celui qui a commercialisé le robot. La quatrième règle touche la question des preuves : le principe est que c'est à l'utilisateur professionnel d'un système IA de mettre à disposition les logs qui permettront de comprendre les causes de l'accident et que c'est en principe lui qui supportera la charge de la preuve de sa non responsabilité. La cinquième exige en cas d'utilisation d'un système d'IA à haut risque de prendre une assurance. Ces approches préventive et cette notion de système à haut risque se retrouvera dans les deux autres réponses et apparaît bien comme une constante de la réponse du Droit aux enjeux de l'IA. Enfin, le groupe d'experts exclut de manière claire l'attribution, pourtant suggérée par le Parlement, aux agents autonomes d'une personnalité juridique, qui obscurcirait encore l'imputation de la responsabilité.

2ème réponse : le RGPD et les risques de l'IA.

10. Notre propos n'est pas d'envisager l'ensemble des mérites de la législation européenne de protection des données qui, certes, face aux applications de l'IA peut servir de protection contre les risques liés à l'IA. Les opérations de profilage et les applications de reconnaissance faciale sont ainsi sévèrement encadrées par le RGPD. Notre propos est plutôt, dans un premier temps, de montrer les limites de cette protection offerte et les difficiles questions d'interprétation que le RGPD confronté aux applications de l'IA soulève. A cet égard, nous relevons

- le fait que les dispositions du RGPD ne s'intéressent qu'aux seules données à caractère personnel. Or, les données utilisées dans le cadre de beaucoup d'applications mêlent tant des données à caractère personnel que des données anonymes. Par ailleurs, les possibilités incroyables de couplage des données au sein des systèmes IA conduisent à considérer que les technologies de pseudonymisation voire d'anonymisation constituent des garanties parfois insuffisantes pour assurer la protection des données. Le calcul du risque

d'individualisation des données doit faire l'objet d'une attention particulière et des engagements excluant toute tentative de ré-identification ou ré-individualisation devraient, dans certains cas, être recommandés voire imposés ;

- le statut de nombre d'acteurs, impliqués dans la constitution ou l'implémentation des applications IA, est peu clair au regard des classifications du RGPD. Par ailleurs, les obligations qui naissent des dispositions du RGPD sont limitées aux seuls acteurs : responsable de traitement et sous-traitant, alors qu'il apparaît important d'imposer certains devoirs à tous les acteurs impliqués dans la chaîne depuis la fourniture des données jusqu'au déploiement d'une solution IA ;

- l'application des principes de finalité et de compatibilité nécessite une attention particulière dans la mesure où les applications IA permettent très facilement, au hasard des interconnexions découvertes, de rencontrer de nouvelles finalités et que dès lors les responsables sont tentés d'élargir les potentialités des traitements ;

- les principes de minimisation et de proportionnalité peuvent difficilement être appliqués dans la mesure où leur mise en œuvre irait à l'encontre même du fonctionnement des systèmes de *machine learning.* A l'inverse des systèmes experts fondés sur une logique décrite *a priori,* le fonctionnement des systèmes d'IA repose sur des corrélations par définition non prévisibles mais significatives entre les données les plus diverses et les plus riches possible. Ceci dit, on sera attentif à brider les algorithmes, c'est-à-dire à leur imposer certaines contraintes de manière à ne pas utiliser n'importe quelles données sans rapport avec le but recherché et hors des prévisions des personnes concernée (en quoi, ma couleur de peau ou mon image peut=elle servir à deviner mes goûts musicaux ?) et à tester sur des données d'entrainement les premiers résultats de manière à éviter des corrélations insensées;

11. Ceci dit, nombre de points positifs existent du fait de l'application du RGPD afin de circonscrire les risques dus à l'IA :

    - le principe d'accountability consacré par le RGPD entraîne un renversement de la charge de la preuve : c'est au responsable du

traitement à démontrer qu'il a respecté les principes de la protection des données ;

- les exigences de la sécurité des traitements IA emportent le besoin d'une attention particulière et d'une évaluation, dès la conception du système et tout au long de la vie de l'application, des risques de biais, d'attaques malfaisantes ou non, d'erreurs de programmation et d'absence de qualité des données utilisées (manque de mises à jour, non pertinence des données ou non couverture des hypothèses à rencontrer (exemple : surreprésentation d'une race ou d'un genre, …) ;

- la loyauté des traitements d'IA exige une information renforcée (article 13 et 14) et aisément accessible et compréhensible par les personnes concernées : ainsi, sur l'existence (en particulier, en matière de robots), les catégories de données utilisées, le modèle retenu pour le traitement et l'impact du traitement sur les personnes concernées.

- le principe de l'article 22 du RGPD interdit que les personnes concernées soient soumises à des décisions prises exclusivement sur base d'un traitement automatisé. Il mérite toute notre attention et, sans doute, une réglementation particulière, dans la mesure où l'IA est souvent un instrument direct ou indirect de décision en matière d'IA. En toute hypothèse, la transparence ou du moins l' « explicabilité » (*explainability*) et l'auditabilité des traitements d'IA en particulier de *deep learning* doit être rendue obligatoire lorsqu'il s'agit de systèmes utilisés par l'autorité publique ou de systèmes à hauts risques. Le droit à une réelle intervention humaine, d'une personne disposant des compétences à la fois techniques et décisionnelles, doit être consacré, d'autant plus que mettre en doute la 'vérité' sortie de l'ordinateur est difficile ;

- enfin, la procédure prévue par l'article 35 du RGPD relative à l'obligation de *Privacy Impact assessment*, affirme la nécessité d'un contrôle humain (*human oversight)* à tous les moments de la vie des systèmes AI lorsque ces derniers représentent un haut risque. Ce contrôle doit exister depuis la conception du système, le choix des données, des algorithmes, la mise en place, le *testing*, l'exploitation du système et à différents moments de son fonctionnement. Il

implique une attention particulière non seulement à la sécurité du système (confidentialité, intégrité, disponibilité) mais, au-delà, aux risques que le système comporte à la fois vis à vis des libertés individuelles. Cette nécessité devrait se traduire par l'obligation à charge des concepteurs et des développeurs de mettre en place des procédures internes d'évaluation multidisciplinaire et inclusive, dont les résultats devraient être transparents.

12. Pour conclure, le RGPD constitue une réponse effective même si limitée. Deux objections nous paraissent devoir lui être adressées : La première est le rôle trop, important laissé au consentement par le RGPD. Il est clair qu'au regard du déséquilibre informationnel entre celui qui opère un système IA et la personne concernée et vu la nécessité et l'urgence (sans doute relatives mais ressenties par la personne concernée) de pouvoir bénéficier du service offert (… et souvent apparemment gratuitement) par cet opérateur (e.g. l'accès à un moteur de recherche ou à une information sur le net), le consentement individuel, sans possibilité de choix entre diverses alternatives de traitement est un leurre. Seul, un contrôle par l'autorité ou par la collectivité des utilisateurs permettrait de légitimer les traitements d'IA opérés. La seconde constate que le RGPD, s'il permet, sans doute de manière non optimale, de contrer les risques individuels se révèle, sauf à étendre sa portée au-delà de sa '*ratio legis',* inapte à prendre en charge, en tant que tels, les risques collectifs et sociétaires. C'est cette volonté de considérer l'ensemble des risques tant individuels que sociétaires qui justifient les récentes avancées de la Commission et des discussions au sein du Parlement européen : « *Le Parlement conclut, à la suite des réflexions précédentes concernant les aspects liés à la dimension éthique de l'intelligence artificielle, de la robotique et des technologies connexes, que cette* ***dimension éthique doit être définie comme une série de principes aboutissant à un cadre juridique*** *à l'échelle de l'Union, supervisés par les organismes nationaux compétents, coordonnés et renforcés par une agence européenne chargée de l'intelligence artificielle et dûment respectés et certifiés au sein du marché intérieur*. » Cette déclaration souligne le caractère global de l'approche proposée par le Parlement et loin d'opposer éthique et droit, considère le Droit comme la garantie de l'effectivité de la mise en œuvre des valeurs éthique.

3ème réponse : Vers une approche ethico-légale et globale des risques liés aux systèmes d'IA ?

13. L'objectif de la proposition, discutée en avril de cette année, au Parlement est large : « *Any artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, shall be developed, deployed and used in a* ***human-centric manner (****Based on Man (Human in the loop) and developed by Man (Human on the loop)****) with the aim of contributing to the existence of a democratic, pluralistic and equitable society by safeguarding human autonomy and decision-making and ensuring human agency*.". Pour être plus clair encore, **«** *l'intelligence artificielle, la robotique et les technologies connexes socialement responsables doivent préserver et promouvoir les valeurs fondamentales de notre société telles que la démocratie, des médias diversifiés et indépendants et des informations objectives et librement accessibles, la santé et la prospérité économique, l'égalité des chances, les droits des travailleurs et les droits sociaux*, *une éducation de qualité, une diversité culturelle et linguistique, l'équilibre hommes-femmes, l'habileté numérique, l'innovation et la créativité. »*

14. Cet objectif vaste met en évidence, à la fois, l'importance de combattre les risques de discrimination sous toutes ses formes et, à la fois, la **responsabilité sociétale de tous les acteurs** (les concepteurs, développeurs, opérateurs et utilisateurs professionnels ou non), qui participent à la production et à la mise en circulation des technologies de l'IA que de celles corrélées : l'Internet des objets et la robotique. La proposition parlementaire, comme déjà le White Paper de la Commission, reprend deux idées déjà présentes dans le RGPD mais en élargit les contours. Ainsi, la notion de systèmes à haut risque, déjà présente dans le cadre des travaux en matière de responsabilité, est reprise du RGPD mais désormais prend en compte l'ensemble des conséquences potentielles non seulement physiquement dommageables ou sur la protection de nos données à caractère personnel mais au-delà les risques physiques et économiques susceptibles d'être causés et au-delà les risques sociétaires et environnementaux : « *L'intelligence artificielle, la robotique et les technologies connexes, y compris les logiciels, les données et les algorithmes utilisés ou produits par ces technologies, qui comportent un risque élevé risque élevé d'enfreindre les principes en matière de sécurité, de transparence, de responsabilité, d'absence de biais ou de discrimination, de responsabilité sociale et d'équilibre hommes-femmes,*

de respect de l'environnement et de durabilité, de vie privée et de gouvernance doivent être considérées comme étant à haut risque relativement au respect des principes éthiques lors de la conclusion d'une évaluation des risques impartiale, réglementée et externe par un organisme national de surveillance»**.** Cette qualification réclame, voir la dernière phrase de la citation, pour de tels systèmes un contrôle (*risk assessment*), sur le modèle du RGPD certes, mais cette fois externe et indépendant.

15. C'est précisément, sur le plan de ce contrôle, de cette gouvernance des risques que la proposition de résolution se montre la plus innovante. Elle propose la création tant au niveau national qu'au niveau européen d'organes « multistakeholders » (industrie, partenaires sociaux, recherche, consommateurs, associations de libertés civiles et APD), multidisciplinaires (ingénieurs, philosophes, juristes, économistes, sociologues) et largement ouverts aux débats publics chargés en outre d'évaluer ou de faire évaluer via des organsismes accrédités les systèmes d'IA . Cette volonté s'inspire de la création dans des pays comme le DK, l'Allemagne, le RU de Data Ethics Commissions chargées de missions semblables.

CONCLUSIONS

16. Quelques mots pour conclure ce tour d'horizon du droit des risques de l'IA. Notre propos tend à montrer qu'une approche globale des risques tant collectifs et sociétaires qu'individuels est nécessaire, dans la mesure où l'AI est un phénomène ubiquitaire et conduit à une transformation forte de notre vivre ensemble. Cette approche, désormais souhaitée par les instances européennes, est tant éthique que juridique : les principes éthiques devant trouver dans le droit leur traduction non seulement formelle (les droits de l'homme consacrent ces principes éthiques) mais surtout effective par la création d'une gouvernance et l'affirmation d'une responsabilité sociétale des acteurs de l'AI. Cette gouvernance aura soin d'instaurer un débat public de '*Technology Assessement*' et d'exiger une coopération, ou du moins une coordination, entre les divers organismes qui s'occupent à l'heure actuelle chacun séparément d'une partie des risques : les autorités de protection des données, les autorités de concurrence, les associations de droits de l'homme, les centres pour l'égalité des chances, etc. Elle veillera à mettre en place des mécanismes de contrôle externe et préventif de façon à veiller à maintenir la maîtrise

de l’homme sur les produits artificiels de son intelligence et à minimiser ses dangers.

Yves Poullet
Moxhe, le 27 août