

*Vol. 30, n° 3*

## **Quels impacts de l'intelligence artificielle sur les métiers du droit et du journalisme ?**

**Céline Castets-Renard\***

RÉSUMÉ / ABSTRACT . . . . .	987
1. INTÉRÊTS DE L'IA POUR LES MÉTIERS DU DROIT ET DU JOURNALISME . . . . .	993
1.1 Réduction des coûts. . . . .	993
1.2 Augmentation de l'efficacité . . . . .	994
2. RISQUES DE L'IA POUR LES MÉTIERS DU DROIT ET DU JOURNALISME (ET AU-DELÀ) . . . . .	996
2.1 Biais, discrimination et opacité des algorithmes de décisions de justice et journalistiques . . . . .	996
2.2 Responsabilité des résultats produits par IA pour les métiers du droit et du journalisme. . . . .	1001

---

© Céline Castets-Renard, 2018.

\* Membre de l'Institut Universitaire de France (IUF) ; Professeur de droit, Université Toulouse Capitole ; Directrice du Master Droit du Numérique ; Directrice adjointe de l'IRDEIC – Centre d'Excellence Jean Monnet.

[Note : cet article a été soumis à une évaluation à double anonymat.]



## **RÉSUMÉ**

L'intelligence artificielle influence de plus en plus nos vies, tant personnelles que professionnelles. Si toutes les professions ont vocation à être concernées, les métiers du droit et du journalisme se prêtent tout particulièrement à l'usage de ces technologies. Sont alors révélés les avantages, mais aussi les risques, en particulier liés aux potentiels biais et discrimination des procédés algorithmiques, ainsi qu'aux questions de responsabilité.

## **ABSTRACT**

Artificial intelligence technologies increasingly influence our personal and professional lives. Every occupation could be concerned, but the lawyers and journalists are particularly impacted. Indeed, such technologies present some advantages but also some risks. We focus on biased and discriminate algorithmic processing, as well as liability questioning.

## **MOTS-CLÉS**

données personnelles ; droit comparé ; droit computationnel ; droits des utilisateurs ; intelligence artificielle ; journalisme ; métiers du droit ; responsabilité ; transparence algorithmique



**Contexte.** L'heure est à l'intelligence artificielle (IA). Ne pas s'y mettre reviendrait à se condamner à l'obsolescence, voire à la disparition ! Il faut dire que ses promesses sont belles : applications pour une agriculture plus soutenable, un système financier plus stable, une meilleure médecine, des transports plus sûrs... Sans même évoquer les enjeux économiques<sup>1</sup>. Le marché de l'intelligence artificielle avoisinerait les 664 millions \$ US et devrait atteindre 38,8 milliards \$ US en 2025, d'après l'Opinion délivrée par le Comité économique et social européen en mai 2017. La machine seconde l'humain désormais libéré des tâches ingrates, pour se concentrer sur les tâches cognitives les plus valorisantes et augmenter son bien-être. Les gouvernements ne sont pas en reste pour encourager un tel progrès. En France, le rapport du mathématicien et député Villani affiche une volonté politique forte<sup>2</sup>, alors que le Canada, sous l'impulsion de l'Université de Montréal, a lancé une « Déclaration de Montréal pour une IA responsable »<sup>3</sup>. Ces deux initiatives ont en commun de vouloir considérer l'impact social du déploiement de l'IA et privilégier une approche éthique respectueuse de l'humain.

**Enjeux.** Des principes éthiques méritent en effet d'être affirmés et surtout défendus, alors que les développements technologiques de l'IA, menés en particulier aux États-Unis, montrent un certain nombre de risques, associés, par exemple, aux robots létaux. Mais sans aller jusqu'au déploiement d'une IA sophistiquée dans des secteurs de pointe, on constate dès à présent l'utilisation de ces procédés, intégrés dans la vie quotidienne de nombreux citoyens.

- 
1. Catelijne MULLER, « Artificial intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion) », *EESC*, INT/806-EESC-2016-05369-00-00-AC-TRA, 31 mai 2017.
  2. Cédric VILLANI, Marc SCHOENAUER, Yann BONNET, Charly BERTHET, Anne-Charlotte CORNUT, François LEVIN et Bertrand RONDEPIERRE, « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », France, mission confiée par le premier ministre Édouard Philippe, en ligne : <[https://www.aiforhumanity.fr/pdfs/9782111457089\\_Rapport\\_Villani\\_accessible.pdf](https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf)>.
  3. « Déclaration de Montréal pour le développement responsable de l'IA », Université de Montréal, en ligne : <<https://www.declarationmontreal-iaresponsable.com>>.

Les méthodes d'intelligence artificielle, en particulier de *machine learning*, conduisent à une utilisation massive de données pour notamment fonder la prise de décisions individuelles ou collectives. Toute information peut être collectée et codée pour refléter l'opinion d'un individu ou fournir des informations sur lui, afin de décider, par exemple, d'octroyer ou au contraire refuser un droit, accès ou avantage. Ces outils peuvent ainsi être utilisés pour attribuer un emploi, une allocation sociale, une prime d'assurance ou encore un prêt, à partir d'informations données spontanément par les individus eux-mêmes, mais aussi collectées à leur insu, en particulier sur les réseaux sociaux. En résumé, le déploiement des dispositifs d'intelligence artificielle influence significativement la vie des individus par des décisions automatiques prises sur la base d'algorithmes, sans que ces derniers en aient toujours connaissance ni même conscience. La gouvernance par les nombres<sup>4</sup> n'est certes pas nouvelle et prend depuis longtemps la forme traditionnelle des méthodes statistiques ou probabilistes, appliquées en particulier au secteur financier. Elle révèle cependant ici une autre ampleur, compte tenu de l'émergence d'algorithmes nouveaux, de la multiplication des jeux de données et du décuplement des puissances de calcul<sup>5</sup>. Ces méthodes peuvent conduire à construire une société fondée sur le score généré par des procédés d'IA<sup>6</sup> se nourrissant de données massives, issues de sources variées, pas toujours documentées par des métadonnées ni vérifiées.

**Définitions des méthodes d'IA.** Les méthodes d'intelligence artificielle supposent l'utilisation d'algorithmes. Un algorithme peut se définir comme une suite finie et non ambiguë d'opérations ou instructions permettant de résoudre un problème ou d'obtenir un résultat. De façon générale, l'intelligence artificielle est un ensemble de théories et techniques mises en œuvre pour permettre aux machines de simuler l'intelligence. Il s'agit de « créer des processus cognitifs

4. A. SUPIOT, « La gouvernance par les nombres », Cours au Collège de France (2012-2014), Fayard, coll. Poids et mesures du monde, 2015.
5. D. PONTILLE et D. TORNÏ, « La manufacture de l'évaluation scientifique : algorithmes, jeux de données et outils bibliométriques » et A. ROUVROY et Th. BERNS, « Gouvernamentalité algorithmique et perspectives d'émancipation : le disparate comme condition d'individuation par la relation ? », dans *Politique des algorithmes : les métriques du web*, Réseaux, La Découverte, févr.-avr. 2013 ; A. ROUVROY, « La digitalisation de la vie même : enjeux épistémologiques et politiques de la mémoire digitale », Documentaliste – Sciences de l'information, 2010, vol. 47, n° 1, p. 63 ; C. CASTETS-RENARD, « Traitement algorithmique des activités humaines : le sempiternel face-à-face homme/machine », dans *Les Cahiers de droit, Sciences et Techniques*, 2016, n° 6, p. 239-255, en ligne : <<https://cdst.revues.org/509>>.
6. D. KEATS CITRON et F. A. PASQUALE, « The Scored Society: due Process for Automated Predictions », *Washington Law Review*, 2014, vol. 89, p. 1 ; F. A. PASQUALE, « The Black Box Society: the Secret Algorithms That Control Money and Information », *Harvard University Press*, 2015.

comparables à ceux de l'être humain »<sup>7</sup>, en particulier l'apprentissage et la résolution des problèmes. Le *machine learning* ou « apprentissage machine » implique la mise en œuvre de techniques par lesquelles l'algorithme « apprend » une multitude de scénarios à partir d'une masse de données. Ces techniques sont censées donner à l'ordinateur la capacité de progressivement améliorer ses performances sur une tâche spécifique par la fouille de données, rendue possible par l'agrégation de données massives (« *big data* ») et sans être explicitement programmé (apprentissage non supervisé). Les méthodes de *machine learning* sont basées sur les représentations des données d'apprentissage, par opposition aux algorithmes basés sur des tâches spécifiques. La machine « apprend » par elle-même, en suivant cependant un but défini par le programmeur. L'intervention humaine est donc principalement focalisée sur la définition des buts et le choix des sources des données utilisées par le programme. Par ailleurs, les outils sont aussi capables d'analyser des faits historiques et actuels permettant aux modèles de faire des prédictions (modèles prédictifs). Également, le *natural langage processing* (traitement automatique du langage naturel) est un procédé d'intelligence artificielle qui vise à créer des outils de traitement du langage naturel pour diverses applications. Il est mis en œuvre dans des applications informatiques nécessitant l'intégration du langage humain à la machine.

**IA et professions.** Ces méthodes d'intelligence artificielle paraissent particulièrement pertinentes pour aider, voire remplacer l'humain dans l'accomplissement de certaines tâches professionnelles. On pense par exemple à la médecine, secondée aujourd'hui par des outils d'aide au diagnostic, à la prise de décision médicale ou encore à la prédiction de maladies, notamment sur la base d'analyses génétiques. Les métiers du droit et du journalisme<sup>8</sup> se prêtent aussi parfaitement à l'utilisation d'outils d'intelligence artificielle gérant une masse de textes ou d'informations. C'est ainsi par exemple que le *machine learning* peut permettre de compiler des textes et produire des nouvelles, que l'on espère de qualité et exhaustives, à un moindre coût. L'outil *Automated Insights* propose de tels services. Sont également utiles aux journalistes les outils de traduction automatique et de révision linguistique ou encore les *natural langage technologies* comme *Narrative Science*<sup>9</sup> pour interagir facilement avec les demandes exprimées par les journalistes.

---

7. Rapport Villani, *supra*, note 2.

8. P.-J. OMBELET, A. KUCZERAWY et P. VALCKE, « Supervising Automated Journalists in the Newsroom: Liability for Algorithmically Produced News Stories », *Revue du droit des technologies de l'information*, 2016.

9. En ligne : <<https://narrativescience.com>>.

Des outils présentant des performances similaires peuvent aussi s'envisager pour les métiers du droit. Les algorithmes de compilation et traitement des décisions de justice permettent de prédire les chances de réussite d'un procès, ainsi que le montant probable des dommages et intérêts alloués, à l'instar de ce que propose le logiciel COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*)<sup>10</sup>. Ces capacités peuvent aussi être utilisées pour aider le juge dans sa prise de décision (justice prédictive). Outre ces fonctions prédictives, l'intelligence artificielle sert l'exécution automatique du droit, au travers des *smart contracts* ou contrats intelligents<sup>11</sup>. Par ailleurs, le traitement automatique du langage humain par la machine donne la possibilité à l'avocat d'interroger le système et de répondre à des questions légales de manière succincte et précise, en fouillant dans les textes juridiques. Telle est la principale mission de *Ross*, système *Watson* d'IBM. Les fonctions de mémorisation de l'IA évitent aux professionnels de fastidieuses séances de recherches. La rédaction d'un contrat et le conseil juridique sont en revanche des tâches cognitives plus complexes.

Au-delà de la communication vantant les prouesses de l'IA, force est de constater que de nombreux dispositifs utilisés aujourd'hui sont encore de simples aides à la décision, laquelle continue d'être prise par des êtres humains.

Il est fréquemment dit qu'une technique est neutre et que ses effets, positifs ou négatifs, dépendent de l'usage qui en est fait. La réalité est sans doute plus subtile, dans la mesure où toute technologie d'IA peut présenter en soi des avantages et inconvénients. Afin d'illustrer cette proposition, nous présenterons les intérêts (1) et risques d'utilisation des technologies de l'IA (2) dans les métiers du droit et du journalisme. Ces deux métiers ont en commun d'exiger notamment l'usage d'une documentation complexe et en grand nombre, pour le maniement de laquelle les technologies de l'IA peuvent s'avérer utiles.

---

10. B. BARRAUD, « Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ? », *Les Cahiers de la justice*, 2017, n° 1, p. 121-139, en ligne : <[http://www.dalloz-revues.fr/revues/Les\\_cahiers\\_de\\_la\\_justice-47.htm](http://www.dalloz-revues.fr/revues/Les_cahiers_de_la_justice-47.htm)>.

11. Les contrats intelligents sont des protocoles informatiques qui facilitent, vérifient et exécutent la négociation ou l'exécution d'un contrat. Les contrats intelligents ont généralement une interface utilisateur et émulent la logique des clauses contractuelles. Ils visent à assurer une sécurité supérieure à la mise en application de la loi sur les contrats et à réduire les coûts de transaction associés à leur conclusion.



## 1. INTÉRÊTS DE L'IA POUR LES MÉTIERS DU DROIT ET DU JOURNALISME

Les deux principaux intérêts que l'on prête fréquemment aux systèmes d'IA appliqués aux métiers du droit et du journalisme tiennent à la réduction des coûts (1.1) et à l'augmentation de l'efficacité des procédures (1.2).

### 1.1 Réduction des coûts

L'IA peut permettre de réduire les coûts de nombreuses activités professionnelles. À titre d'exemple, le service *Narrative Science* comprend et produit du langage naturel à partir d'une multitude de données. Il est aussi capable d'écrire des articles journalistiques se limitant à des textes factuels assez simples, tels des comptes rendus d'épreuves sportives et de brèves analyses boursières. Cette technologie interprète les données, les transforme en informations significatives, les rédige en langage humain, à grande échelle et rapidement. Si toutes les promesses ainsi faites peuvent être tenues, nul doute que les rédactions, souvent en grande difficulté financière, seront tentées d'y recourir. L'économie ainsi réalisée n'implique pas nécessairement la suppression du métier de journaliste, mais les tâches de base, consistant à rechercher des histoires ou des faits, avant l'analyse du sujet et la rédaction elle-même, pourraient être de plus en plus confiées à une machine. Les tâches cognitives d'analyse et de rédaction ont aussi vocation à être réalisées par IA, ce qui devient plus directement menaçant pour les journalistes. Le développement de l'IA pose en effet immanquablement la question de la place de l'humain au travail : que va devenir son rôle en présence d'une machine puissante dont certaines capacités dépassent les siennes et sans que les contraintes liées au droit du travail ne s'appliquent ? Si ces questions préoccupent depuis longtemps les chercheurs et institutions<sup>12</sup>, il est difficile d'anticiper avec précision leur ampleur. En effet, si les coûts du travail sont élevés, ceux liés à la conception d'une IA ne sont pas non plus négligeables. On s'accorde souvent à dire que les métiers vont évoluer et qu'une nouvelle répartition des tâches se mettra en place entre l'humain et la machine. S'agissant des métiers du journalisme, les tâches les plus élémentaires, souvent confiées à des stagiaires ou journalistes peu expérimentés, risquent de disparaître, réduisant d'autant le nombre de journalistes dans les rédactions. Notons que le rapport Villani accorde une place importante à cette

12. Parmi les nombreux rapports réalisés en France, voir celui du Conseil d'orientation pour l'emploi (COE) intitulé « Automatisation, numérisation et emploi » publié le 12 janvier 2017.

question de l'évolution des métiers et encourage la formation initiale et continue des professionnels, afin de leur permettre d'accomplir de nouvelles tâches en remplacement de celles qui seront assurées par IA<sup>13</sup>. Les obligations de formation mises à la charge des employeurs pourraient d'ores et déjà inclure au moins une sensibilisation aux transformations prévisibles à brève échéance, *a minima* dans les secteurs d'activité susceptibles d'être les plus affectés. Ce conseil risque cependant de rester lettre morte pour certains acteurs. Compte tenu de la crise financière des médias, les entreprises de presse (qui déjà investissent peu dans la formation continue) ne sont probablement pas en mesure d'assumer une telle responsabilité.

On le voit, les tâches de recherche, à la base de toute activité journalistique, risquent d'être assumées à l'avenir par l'IA et non plus par des jeunes journalistes en début de carrière. Ceux-ci devront donc apprendre à se servir de ces outils. Cette anticipation doit toutefois être nuancée par le fait que de nombreux faits sont collectés par interview, dont la mise en œuvre est plus difficilement substituable par la machine. L'humain ne va sans doute pas encore être rayé d'un trait de plume, mais verra très certainement ses tâches évoluer pour se concentrer sur celles que la machine réalise moins bien.

Un constat identique peut être fait pour les métiers du droit. Il a déjà été relevé<sup>14</sup> que le *machine learning* appliqué au droit peut être utile pour établir, à partir des données, des relations entre individus ou pour faire ressortir des informations non visibles autrement. Également, les procédures de type *discovery process*<sup>15</sup> supposent de gérer une très grande quantité de documents entre lesquels il faut trier, en fonction de la preuve que l'on veut apporter et de la stratégie à adopter. Ce travail délicat de tri requiert une grande vigilance pour ne pas manquer un document essentiel. Dans ces hypothèses de classification documentaire, l'IA révèle tout son intérêt. Au demeurant, l'espoir est ici de faire baisser les coûts de procédure qui constituent une barrière à l'entrée élevée pour le justiciable et régulièrement dénoncée<sup>16</sup>.

## 1.2 Augmentation de l'efficacité

À l'actif des méthodes algorithmiques, on peut aussi relever qu'une décision automatique peut être plus efficace qu'une déci-

13. Partie 3 du Rapport Villani, *supra*, note 2, p. 114.

14. H. SURDEN, « Machine Learning and Law », (2014) 89 *Wash. L. Rev.* 87.

15. R. SFADJ, « L'IA juridique (entretien) », *Expertises*, janvier 2017.

16. A. GARAPON, « Les enjeux de la justice préventive », *La Semaine juridique*, janvier 2017, en ligne : <[https://www.cnll.fr/sites/default/files/atoms/files/justice\\_predictive.pdf](https://www.cnll.fr/sites/default/files/atoms/files/justice_predictive.pdf)>.

sion humaine. Dans des cas simples qui ne suscitent pas d'enjeux importants, les décisions fondées sur un algorithme semblent plus performantes. Tel est par exemple le cas lorsqu'il s'agit d'appliquer un barème tarifaire ou encore d'exécuter automatiquement un contrat en l'absence de toute contestation. Au demeurant, les risques d'erreurs humaines sont souvent plus élevés que ceux générés par les algorithmes.

Appliquée au droit et poussée à l'extrême, l'IA permettrait une plus grande efficacité des métiers du droit, par une analyse automatisée des règles. Le droit computationnel<sup>17</sup> est une branche de l'analyse informatique du droit qui vise l'automatisation de l'analyse juridique. L'objectif est d'élaborer des systèmes informatiques capables de faire des calculs juridiques sur des situations variées, tel que le contrôle de conformité ou l'analyse de la réglementation. L'analyse juridique est alors faite par des machines plutôt que par des êtres humains. Turbotax, système aidant à la déclaration fiscale aux États-Unis, est l'exemple d'un système rudimentaire de droit computationnel. Basé sur les données fournies par les utilisateurs, il compile automatiquement les obligations fiscales des usagers et les met dans la forme fiscale appropriée. En cas de demande, il peut fournir les explications de ces résultats en les reliant aux parties pertinentes du code fiscal. De telles technologies ont le potentiel de changer les professions du droit, en augmentant la qualité et l'efficacité des services juridiques et en perturbant le modèle d'affaire des cabinets d'avocats. Cet effet négatif ne doit pas dissimuler un impact positif consistant à rendre le droit plus accessible à tous et non plus seulement aux professionnels du droit. Un tel effet va améliorer l'accès à la justice et le système juridique dans son ensemble. La prévisibilité du droit peut être améliorée également si un système informatique rationalise l'ensemble.

Cependant, force est bien de constater que la règle de droit n'est pas toujours réductible en commandes claires et uniques. Un tel procédé d'encodage du droit ne peut être utile en présence de concepts à texture ouverte, tels les notions floues ou standards juridiques. En résumé, le droit computationnel trouve son pouvoir dans le raisonnement déductif, aussi ne peut-il s'appliquer au raisonnement analogique ou inductif, pourtant fréquent en droit. Son apport est donc limité, mais on peut se consoler en considérant que le nombre de cas impliquant un raisonnement déductif est suffisamment important pour justifier le recours à ce type d'outil encore en cours d'améliora-

---

17. M. GENESERETH, « Computational Law: the Cop in the Backseat », *CodeX: The Center for Legal Informatics, Stanford University*, en ligne : <<http://logic.stanford.edu/complaw/complaw.html>>.

tion. En outre, pour certaines tâches et dans certains contextes légaux, un résultat approximatif peut être acceptable<sup>18</sup>.

Même si l'utilisation de techniques d'IA dans les métiers du droit et du journalisme peut améliorer les processus et réduire les coûts, les risques associés ne doivent cependant pas être négligés.

## **2. RISQUES DE L'IA POUR LES MÉTIERS DU DROIT ET DU JOURNALISME (ET AU-DELÀ)**

Les risques et limites de l'IA peuvent d'abord être propres aux métiers du journalisme, eu égard à la production textuelle qui leur est associée. Est ici posée la question de savoir quelles règles de propriété intellectuelle s'appliquent à une œuvre générée par IA, eu égard aux conditions d'originalité et titularité des droits. Ces questions ne seront pas envisagées ici, car l'objectif est de se concentrer sur des risques et limites en général, détachés d'un secteur d'activité. Ces risques témoignent des difficultés auxquelles l'IA doit encore faire face aujourd'hui, tant pour lutter contre les biais et opacité des algorithmes de décision (2.1), que pour définir des règles de responsabilité (2.2).

### **2.1 Biais, discrimination et opacité des algorithmes de décisions de justice et journalistiques**

Le développement de l'IA fait émerger de nombreuses questions éthiques : quel impact les systèmes autonomes auto-apprenants ont-ils sur l'intégrité personnelle, l'autonomie, la dignité, l'indépendance, l'égalité, la sécurité et la liberté de choix ? Comment nous assurons-nous que nos normes fondamentales, valeurs et droits humains demeurent respectés et préservés<sup>19</sup> ? Ces questions sont particulièrement cruciales pour l'exercice des métiers du droit et de la justice. La procédure, marquée notamment par les règles du procès équitable et du contradictoire, traduit en principe ces exigences qui doivent pouvoir être garanties, même dans l'hypothèse où la prise de décision est assurée en tout ou partie par un procédé d'IA.

Par ailleurs, l'algorithme est souvent considéré comme non discriminant et plus objectif que le jugement humain. Cependant, un algorithme n'est pas neutre par nature et peut perpétuer des stéréotypes préexistants, de nature à renforcer des ségrégations sociales ou raciales<sup>20</sup>. Un biais survient quand un système informa-

18. H. SURDEN, *supra*, note 14.

19. Rapport Villani, *supra*, note 2.

20. C. ANUPAM, « The Racist Algorithm? », (2017) 115 *Michigan Law Review*.

tique discrimine systématiquement et de façon déloyale un groupe ou des individus sur le fondement de critères sociaux ou ethniques par exemple<sup>21</sup>. Des biais introduits par un procédé de *machine learning* peuvent l'être à des étapes variables, y compris au cours de l'élaboration de l'algorithme. La nature et la qualité des données utilisées à l'entrée du système d'apprentissage (*inputs*) constituent aussi un enjeu essentiel, de nature à impacter la fiabilité des résultats en sortie (*outputs*). Les systèmes d'intelligence artificielle sont entraînés avec des données en principe collectées à partir des activités et décisions humaines et qui peuvent aussi refléter des biais cognitifs implicites de ces dernières<sup>22</sup>. Conséquemment, de nombreux systèmes d'intelligence artificielle ont appris à classer les données à partir de jeux de données d'entraînement reflétant des biais socio-culturels. Rien de surprenant alors de constater que les résultats produits par ces technologies répliquent les inégalités si le système a « appris » à partir d'informations biaisées<sup>23</sup>. D'un autre côté, une sous-représentation d'un groupe minoritaire parmi les données historiques renforce la discrimination contre ce groupe pour les futures demandes fondées sur un traitement automatisé. Par ailleurs, les algorithmes sont basés sur un code source qui reçoit des instructions des développeurs. Ces commandes peuvent aussi contenir des biais implicites contre certaines catégories de personnes ou groupes sous-représentés, lorsque le cerveau réalise des associations mentales et opère sans conscience, intention ou contrôle. Le plus souvent, aucune information n'est fournie sur les instructions données, ce qui peut être facilement une source de biais, d'erreurs ou de discrimination. Le risque est accru par le fait que les capacités analytiques de l'intelligence artificielle et du *machine learning* ont rendu plus facile la création de profils et la prise automatique de décisions dont l'impact potentiel sur les droits et libertés des individus est aussi devenu plus significatif<sup>24</sup>. Leur application aux activités de police ou à la justice prédictive en est une illustration éclatante.

- 
21. B. FRIEDMAN et H. NISSENBAUM, « Bias in Computer Systems », *ACM Transactions on Information Systems*, juillet 1996, vol. 14, n° 3, p. 330-347, en ligne : <[http://vsdesign.org/publications/pdf/64\\_friedman.pdf](http://vsdesign.org/publications/pdf/64_friedman.pdf)>.
  22. S. BAROCAS et A. D. SELBST, « Big Data's Disparate Impact », (2016) 104 *California Law Review* 671.
  23. Par exemple, le « White Guy problem » par lequel les biais algorithmiques sont difficiles à déceler : comment une femme peut-elle répondre à une offre d'emploi si elle n'est pas ciblée pour recevoir cette information ? Comment une communauté minoritaire peut-elle comprendre qu'elle est traitée différemment par des algorithmes ? ; K. CRAWFORD, « Artificial Intelligence's White Guy Problem », 25 juin 2016, dans *The New York Times*.
  24. C. O'NEIL, « Weapons of Maths Destruction: how big data increases inequality and threatens democracy », *Broadway Books*, 2016.

La qualité de la donnée est donc un enjeu majeur, ce qui suppose qu'elle soit juste, variée, suffisamment détaillée et non biaisée. La difficulté aujourd'hui est de mesurer ces biais et de mettre en œuvre des dispositifs pour « débiaiser » les systèmes, tout en améliorant l'efficacité des résultats. Un exemple célèbre de ces risques de biais a été révélé par des chercheurs américains<sup>25</sup> sur le site *ProPublica*, concernant un dispositif de justice prédictive COMPAS<sup>26</sup>, utilisé dans plusieurs États américains pour évaluer les risques de récidive. Si ce système a été contesté dans plusieurs affaires, il n'a pour l'instant pas été remis en question à notre connaissance. Dans l'affaire *Loomis*<sup>27</sup>, le requérant contesta l'utilisation de cet outil en invoquant notamment la violation du *due process*, eu égard à l'impossibilité de comprendre les décisions prises et donc à les contester, en violation du principe du contradictoire. La Cour suprême du Wisconsin rejeta la requête en estimant que COMPAS avait seulement servi d'aide à la décision et non à la prendre.

L'IA dans le journalisme révèle par ailleurs d'autres risques en cas d'usage ciblé de ces technologies. Elle peut permettre de réaliser de nouvelles agrégations de données, par exemple sur les réseaux sociaux, et de détourner les données de leur usage, comme l'a révélé le scandale *Cambridge Analytica*. Elle peut aussi constituer des « bulles informationnelles » au sein desquelles les individus reçoivent uniquement l'information qui les intéresse, basée sur leur profilage, indépendamment des exigences d'objectivité et d'exhaustivité du journalisme professionnel. Plus encore, le phénomène des « *fake news* » sur Twitter ou Facebook préoccupe de nombreux législateurs<sup>28</sup>, spécialement à l'approche d'élections nationales. Des systèmes d'IA (*bots*) ont pu être utilisés pour faire circuler des contenus propagandistes, en vue d'influencer les votes de la population<sup>29</sup>. Les résultats

25. J. LARSON, S. MATTU, L. KIRCHNER, J. ANGWIN, « How we Analyzed the compas Recidivism Algorithm », *ProPublica*, 23 mai 2016, en ligne : <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>>.
26. *Correctional Offender Management Profile for Alternative Sanction* créé par la société Northpointe Inc.
27. *State of Wisconsin v. Loomis*, 2016 WI 68, 13 juillet 2016, en ligne : <<https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690>>.
28. Voir les réflexions menées aux États-Unis autour du projet de loi *Honest Ads Act* présenté en octobre 2017 en réaction aux interférences russes lors de la campagne présidentielle de 2016. Voir aussi les règles éthiques présentées en réaction par Facebook et Twitter. En France, voir la proposition de loi relative à la lutte contre les fausses informations, n° 799, présentée en mars 2018 et en cours de discussion devant le Parlement.
29. N. W. JANKOWSKI, « Researching Fake News: A Selective Examination of Empirical Studies », *Journal of European Institute for Communication and Culture*, 2018, vol. 25, p. 248-255, en ligne : <<https://www.tandfonline.com/doi/full/10.1080/13183222.2018.1418964>>.

du référendum ayant conduit au Brexit au Royaume-Uni, ainsi que ceux des élections présidentielles américaines en 2016, sont ainsi devenus douteux et font l'objet d'une enquête du FBI s'agissant de ces dernières.

Outre le problème de biais, l'opacité des modèles doit également être soulignée<sup>30</sup>, eu égard à l'ignorance accompagnant le plus souvent la nature des données utilisées, ainsi que les instructions mises en œuvre par les procédés algorithmiques. Les *data scientists* eux-mêmes ne sont pas en mesure d'expliquer la plupart des processus algorithmiques et ne peuvent que constater une meilleure efficacité des résultats obtenus. Les modèles les plus utilisés aujourd'hui sont construits sur des corrélations et inférences plutôt que sur des causalités, ce qui rend impossible toute explication humaine claire et compréhensible et peut mener à des erreurs. Les données et systèmes utilisés dans le cadre des métiers du droit et du journalisme sont particulièrement sensibles et méritent plus qu'ailleurs de respecter des principes de transparence et d'imputabilité (ou de reddition de comptes).

Eu égard à l'émergence des risques associés aux procédés algorithmiques, s'est exprimé le besoin d'une gouvernance rigoureuse du *machine learning*<sup>31</sup>, autrement appelé « éthique » des algorithmes<sup>32</sup>, afin de garantir une meilleure transparence, loyauté<sup>33</sup>, ou encore « explicabilité » ou rendre compte (*accountability*)<sup>34</sup> des décisions prises sur le fondement de telles méthodes. En Europe et en France, la prise de conscience de ces enjeux s'est traduite par l'adoption de normes législatives. Le règlement n° 2016/679/UE 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

30. F. A. PASQUALE, *supra*, note 6.

31. J. BALKIN *et al.*, « Governing Machine Learning: Exploring the Intersection Between Machine Learning », *Law and Regulation: Information Society Project (ISP)*, Yale Law School.

32. Voir B. MITTELSTADT, P. ALLO, M. TADDEO, S. WACHTER et L. FLORIDI, « The Ethics of Algorithms: Mapping the Debate », *Big Data & Society*, 1<sup>er</sup> novembre 2016, vol. 3(2), DOI : 10.1177/2053951716679679. Ces auteurs identifient six types de problèmes éthiques avec les algorithmes : la preuve non conclusive, la preuve non vérifiable, la preuve faussée, les résultats déloyaux, les effets transformatifs, la traçabilité.

33. La difficulté de cette notion est qu'elle reçoit de nombreuses définitions selon les disciplines et différentes traductions. Elle est assimilée à la notion de « fairness ». Voir D. KEHL, P. GUO et S. KESSLER, « Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing », *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School*, 2017.

34. J. A. KROLL, J. HUEY, S. BAROCAS, E. W. FELTEN, J. R. REIDENBERG, D. G. ROBINSON, H. YU, « Accountable Algorithms », (2017) 165 *University of Pennsylvania Law Review* 633.

(RGPD) adopté le 27 avril 2016 pose à l'article 22§1 un droit à ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques concernant, ou affectant de manière significative de façon similaire, la personne concernée par le traitement. En cas d'exception à l'exercice de ce droit, le responsable du traitement doit au moins respecter le 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 (§3). En France, la loi n° 1321 *Pour une République Numérique* (LRN) du 7 octobre 2016 prévoit des mesures spécifiques codifiées à l'article L. 311-3-1 du Code des relations entre le public et l'administration, afin d'imposer aux administrations de rendre compte des décisions prises sur la base de procédés automatiques. L'entrée en application du RGPD implique une révision de la loi n° 78-17 du 6 janvier 1978 dite « informatique et liberté », laquelle vient modifier les dispositions de la LRN. L'adoption de ces textes, censés répondre aux exigences d'« éthique » algorithmique, pose la question de savoir si on aboutit véritablement à une transparence et « explicabilité » des décisions ainsi prises. On peut douter de l'efficacité réelle de ces dispositions, d'autant que l'explication ne garantit pas d'éviter les biais et discrimination. En conséquence, si le législateur a bien identifié les risques, il reste encore à poser des mesures plus efficaces pour les éviter.

Par ailleurs, les systèmes, et parfois même les données, ne peuvent pas être facilement audités, non seulement en raison de la complexité technique, mais aussi en présence de droits de propriété intellectuelle et secrets d'affaire portant sur les systèmes ou sur les données elles-mêmes, qui font obstacle à la transparence. Dans l'affaire *Loomis* précitée, le requérant s'est également vu opposer les droits de propriété intellectuelle de la société Northpointe, développeuse de la solution COMPAS, pour motiver le refus d'accès au dispositif. En conséquence, ces obstacles légaux et techniques créent une asymétrie d'information entre, d'une part, les concepteurs et utilisateurs des systèmes algorithmiques et, d'autre part, les personnes affectées par les résultats générés. Pourtant, de tels obstacles ne sont pas juridiquement justifiés et de tels droits de secret et propriété ne devraient pas faire obstacle à la transparence<sup>35</sup>, *a fortiori* lorsqu'il s'agit d'auditer des systèmes ayant un fort impact social. L'articulation des normes reste ici à mieux préciser.

---

35. R. WEXLER, « Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System », (2018) 70 *Stan. L. Review*.



Par ailleurs, les règles de responsabilité ne semblent pas encore adaptées à l'IA.

## **2.2 Responsabilité des résultats produits par IA pour les métiers du droit et du journalisme**

En droit français de la presse, le journalisme répond à des règles particulières de responsabilité éditoriale en cascade, dans le souci de satisfaire un équilibre délicat entre liberté d'expression et lutte contre l'injure, la diffamation et autre contenu illicite, à l'instar des propos négationnistes ou racistes par exemple. Se pose alors la question de savoir qui sera responsable d'un contenu informationnel illicite généré par IA. Si on peut penser en premier lieu au concepteur, la solution est moins évidente dans l'hypothèse où le dommage serait dû à un système d'IA auto-apprenant qui continue d'apprendre après son entrée en application. Dans un tel contexte, l'utilisateur ou les utilisateurs jouent sans doute un rôle important dans le processus et l'évolution des résultats générés. Quant à l'idée de responsabilité du système lui-même, si le Parlement européen a pu paraître favorable à la reconnaissance d'une personnalité juridique du robot<sup>36</sup>, cette hypothèse, fortement critiquée, semble écartée aujourd'hui. Une telle approche aurait pour effet de déresponsabiliser les humains contribuant à la création et mise en œuvre de l'IA, ce qui, sans même évoquer les débats éthiques sur la place de l'humain et de la machine, ne semble ni justifié ni souhaitable pour des raisons de justice sociale, mais aussi de solvabilité du responsable du dommage. Il reste que, en France, les règles classiques de responsabilité délictuelle du fait des choses ne s'appliquent pas aux biens immatériels qui ne sont pas sous la « garde » d'un gardien (C.civ., art. 1242 al. 1<sup>er</sup>). De même, la responsabilité du fait des produits défectueux visant un « produit », c'est-à-dire « tout bien meuble » (C. civ., art. 1245-2), n'est pas plus adapté. Non seulement ces règles de responsabilité ne tiennent pas compte de l'émergence des biens immatériels, mais aussi de leur « puissance » et des risques inhérents.

Sans doute faudra-t-il par ailleurs envisager un système d'assurance pour garantir la réparation des dommages générés par IA, sans avoir à en rechercher les causes. L'obligation de s'assurer pourrait être mise à la charge des concepteurs, mais aussi des utilisateurs. La situation nouvelle à laquelle il faut faire face aujourd'hui n'est pas nécessairement très différente des risques générés pré-

---

36. Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103 (INL)).

cédemment par le progrès technique, en matière de transport par exemple. Les solutions en matière de responsabilité et d'assurance pour la couverture des dommages générés par l'IA restent encore à préciser, mais les modèles existent.

La réflexion sur la régulation de l'IA ne fait que débuter. S'il faut se garder de trop se précipiter, il convient d'être vigilant et de réagir dès la survenance des risques. La loi n'est pas forcément le meilleur outil normatif et il faudra certainement accorder une attention aux bonnes pratiques professionnelles. On constate une prise de conscience des techniciens, amenés à travailler avec les éthiciens, philosophes et juristes, ce dont on ne peut que se réjouir.