7

Les systèmes judiciaires européens à l'épreuve du développement de l'intelligence artificielle [1]

Yannick MENECEUR,

magistrat détaché,
conseiller en politiques du Conseil de l'Europe,
chercheur associé à l'Institut des Hautes Études sur la Justice (IHEJ) [2]

La forte médiatisation des derniers développements de l'intelligence artificielle (IA) n'est pas le fruit d'une révolution en recherche fondamentale mais provient de l'amélioration des performances des algorithmes d'apprentissage automatique, qui cherchent à représenter par des techniques statistiques un environnement donné. L'application à la justice de cette technologie à des fins « prédictives » reste toutefois une entreprise extrêmement périlleuse pour laquelle il est urgent de produire un cadre éthique. Discriminations, restrictions de l'accès au juge, atteintes à la vie privée sont autant de risques à prévenir de la conception des modèles à leur utilisation.

1 - Lorsque l'on s'intéresse à l'application concrète des outils relevant de l'intelligence artificielle (IA) dans les différents systèmes judiciaires, on est d'emblée frappé de constater à quel point l'Europe paraît beaucoup plus méfiante à l'égard de ces nouvelles technologies que le continent Nord-américain [3]. Aux États-Unis, l'apprentissage automatique est d'ores et déjà largement mis à profit pour tenter de prédire la commission d'infractions, évaluer la dangerosité des individus ou encore profiler les juges... En Europe, le mouvement est plus qu'amorcé avec le fort développement des legaltechs mais, malgré le « buzz » entretenu par ces sociétés commerciales, l'utilisation concrète de ces outils reste encore très contrastée et le sens des résultats obtenus sujet à caution [4].

2 - Il faut dire que personne ne semble réellement savoir où tout cela peut nous conduire au vu du « décalage entre la rapidité du développement de la science et la sagesse collective » [5]. Ce constat est tellement flagrant que des acteurs majeurs du numérique (tels que Google, Facebook IBM ou Microsoft) en sont venus eux-mêmes à promouvoir des démarches dites « d'éthique dès la conception » (*ethic by design*) ou de « droits humains dès la conception » (*human rights by design*) afin de bâtir une IA responsable [6]. Or, de quelle éthique parle-t-on ? Toute prétention à l'autorégulation pose immédiatement des questions de légitimité, de pertinence, et d'effectivité. Le dialogue entre ces acteurs majeurs des technologies et les diverses institutions (États, ONG et organisations internationales) s'est récemment enrichi mais sans pour autant aboutir encore à un cadre concret, formalisé et partagé [7].

3 - Il ne s'agira pas de disserter ici sur les motivations profondes des industries numériques pour initier de telles démarches mais simplement de constater que, dans cette phase de transition majeure pour nos sociétés, une certaine forme d'urgence s'impose à tous pour remettre ces technologies à leur juste place, surtout quand il s'agit de processus de décisions susceptible de porter atteinte à nos libertés, comme en matière judiciaire. On peut également tenter de rendre objectif le discours éthique de ces entrepreneurs qui diffère parfois à bien des égards de leurs actions. Certains d'entre eux cèdent en effet bien volontiers à la tentation du « solutionnisme » au mépris de toute précaution méthodologique et des acquis les plus évidents des sciences sociales : ils pensent trouver dans l'apprentissage automatique, appliqué à l'activité humaine, une solution à tous les problèmes et érigent volontiers les données issues de ces outils en véritables normes à suivre au lieu de les considérer comme le reflet imparfait et orienté d'une réalité qu'il s'agirait avant tout de décrire et de comprendre [8]. Appliqué au droit, leur solution est l'avènement d'une justice « prédictive », « quantifiée » ou « actuarielle » pour résoudre les divers maux dont celle-ci souffrirait. Mais ne nous y trompons pas : l'imagination sémantique (et marketing) des auteurs de ces outils recouvre une même et unique réalité qui consiste à établir divers types de probabilités (issues d'un litige sur la base de faits déjà qualifiés juridiquement ou risque de commission d'infraction à partir de statistiques pénales par exemple) à destination de professionnels de la justice, de directions juridiques ou d'assureurs. Une telle ambition, au demeurant fort modeste, ne serait pas critiquable si elle ne nourrissait pas les plus grands des fantasmes de nombre de

1. Nous tenons à remercier Nicolas Régis, magistrat, et François Paychère, magistrat à la Cour des comptes de Genève, pour leurs apports respectifs.
2. La présente étude reflète une analyse personnelle de l'auteur et n'engage pas le Conseil de l'Europe.
3. Nous emploierons plus généralement le terme d'apprentissage automatique (*machine learning*) dans les développements de cette étude, car il qualifie précisément la famille de technologie utilisée. Le terme IA ne désignera pas les travaux, encore totalement spéculatifs, sur des IA dites « fortes », c'est-à-dire comparables à la cognition humaine.
4. V. par ex. *J.-B. Duclercq, Les algorithmes en procès : RFDA 2018, p. 131. – V. Vigneau, Le passé ne manque pas d'avenir : Recueil Dalloz 2018, p. 1095.*
5. *Y. Bengio, Présentation des délibérations citoyennes de la déclaration de Montréal sur une IA responsable : Montréal, 14 juin 2018.*
6. V. par ex. le partenariat pour une IA responsable : *www.partnershiponai.org/* et l'entretien du Monde avec *E. Horvitz, Intelligence artificielle. – Microsoft ne veut pas fournir d'outils qui pourraient violer les droits de l'homme, 3 juill. 2018 : www.lemonde.fr/pixels/article/2018/07/03/eric-horvitz-microsoft-ne-veut-pas-fournir-d-outils-qui-pourraient-violer-les-droits-de-l-homme_5324975_4408996.html.*
7. V. par exemple la coopération renforcée du Conseil de l'Europe avec le secteur privé afin de promouvoir un « internet sûr et ouvert, dans lequel les droits de l'Homme, la démocratie et l'État de droit sont respectés dans l'environnement en ligne », concrétisée par un échange de lettres entre le Secrétaire général du Conseil et 8 entreprises technologies de premier plan le 8 novembre 2017 lors du Forum mondial de la démocratie.

décideurs publics et un discours idéologique qui amalgame des aspirations de physique sociale, des enjeux de *new public management* et des impératifs issus du néo-libéralisme. En prétendant contribuer au bien commun et sous le couvert d'une scientificité dont la substance ne peut être débattue que par un nombre extrêmement limité d'experts, ces outils risquent pourtant de créer bien plus d'injustices qu'ils ne prétendent en corriger (comme le renforcement des discriminations du fait des biais des données ou la survalorisation des comportements moyens au détriment d'une réelle interprétation du droit).

4 - C'est pourquoi une éthique doit être partagée en urgence : éclairés de la mécanique profonde de ces outils **(1)** et des risques engendrés **(2)**, des principes directeurs devraient être posés et avoir pour ambition de soutenir le développement de l'apprentissage automatique avec des décisions judiciaires en lui donnant conscience de la nécessité de respecter tant son objet que les droits fondamentaux **(3)**.

1. L'apprentissage automatique, de quoi s'agit-il exactement ?

5 - Avant d'envisager les raisons d'une entreprise de régulation, il semble opportun de revenir sur le contexte de la résurgence de l'IA depuis 2010. La discipline a connu en effet un nouvel essor grâce à l'amélioration des performances des algorithmes d'apprentissage automatique, notamment pour la reconnaissance d'images et de sons. Cette « révolution » ne vient toutefois pas d'une découverte en recherche fondamentale mais d'un changement complet de paradigme par rapport à la précédente génération d'IA, les systèmes experts. Avec l'apprentissage automatique, l'approche se veut inductive [9] : il n'est plus question ici de modéliser et de programmer à la main des règles en miroir de la réalité à reproduire, comme pour les systèmes experts, mais de laisser les ordinateurs les découvrir par corrélation et classification, sur la base d'une quantité massive de données. Cela revient, en d'autres termes, à représenter par des techniques statistiques un environnement donné, en identifiant des similitudes et en les catégorisant.

Systèmes experts et apprentissage automatique : deux conceptions différentes de l'IA

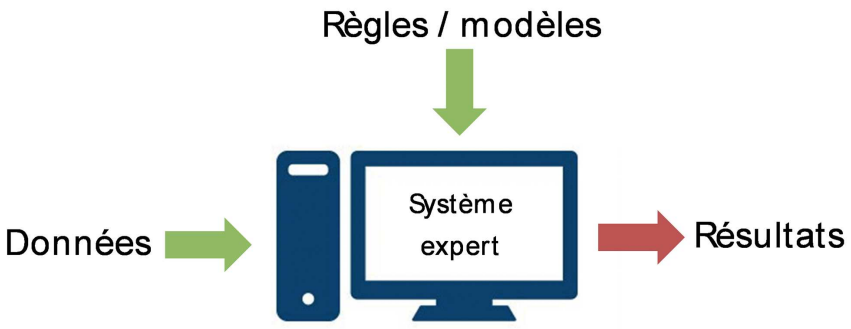

Fig.1. Systèmes experts

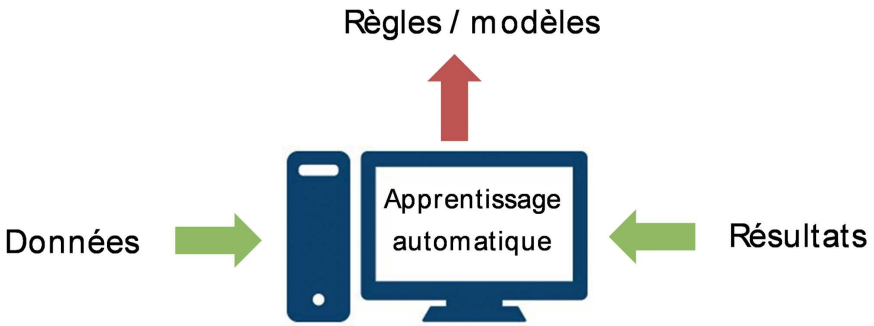

Fig.2. Apprentissage automatique

8. *E. Morozov, Pour tout résoudre, cliquez ici : Fyp éd., 2014, p. 22.*
9. Il ne sera pas développé ici la critique logique de l'induction. Il pourra toutefois être cité l'exemple de Karl Popper qui illustrait cette critique par l'exemple des cygnes blancs. Si l'on observe une série de cygnes qui sont tous blancs, l'on aura tendance à conclure que tous les cygnes sont blancs. Or, cette généralisation est fausse car elle a omis l'existence de cygnes noirs, qui n'ont simplement pas été observés.

6 - Les fondements, que l'on peut faire remonter selon les techniques employées au 18e siècle pour l'inférence bayésienne [10] ou aux années 1940 pour les neurones formels [11], sont relativement simples à comprendre : dans sa phase d'apprentissage, la machine va rechercher les liens entre des données préalablement sélectionnées (par exemple, la température de l'air et le nombre de crèmes glacées vendues par les glaciers d'une ville). Ce modèle peut ensuite être utilisé pour résoudre des questions du type : s'il fait 25°, combien de glaces puis-je espérer vendre dans tel lieu ? L'intervention humaine demeure encore essentielle, qu'il s'agisse de choisir les données d'apprentissage, d'identifier leurs éventuels biais ou alors, quand cela est possible, de distinguer parmi les corrélations celles pouvant être réellement la cause d'un phénomène (si l'on vend beaucoup plus de glaces pour un lieu donné, est-ce à cause de la température ou de la présence d'un très bon glacier ?).

7 - Appliqué à des décisions judiciaires, la technique est la même [12]. À partir, par exemple, d'un grand nombre de décisions de divorce attribuant des prestations compensatoires, la machine sera nourrie de données d'entrée (la durée du mariage, la situation professionnelle, la disparité de situation patrimoniale, l'âge et l'état de santé des parties) et de résultats (le montant de la prestation prononcée en fonction de ces critères). Elle recherchera ensuite la possible influence de ces différents ensembles sur la décision finale et construira un modèle qui, appliqué à de nouvelles données d'entrée, pourra suggérer des probabilités de réussite (prononcé ou non d'une prestation) et une distribution des différentes fourchettes de montants susceptibles d'être alloués. Ce qui sera intéressant pour le technicien manipulant ces données sera d'explorer, grâce aux possibilités de recherche de corrélations de l'apprentissage machine, si d'autres variables (présence ou non d'un adultère, montant excessif ou trop faible des prétentions par exemple) sont susceptibles d'avoir eu une certaine prépondérance sur la décision du juge. L'application de cette démarche au contenu d'un jugement exige toutefois une extrême rigueur : s'il est relativement aisé d'étudier le lien de causalité entre des variables mesurables physiquement (la température de l'air ou le nombre de glaces vendues par exemple), les corrélations d'ordre linguistique découvertes dans les décisions sont loin de pouvoir être toutes interprétées de manière irréfutable comme des rapports de cause à effet [13].

2. Le défi de l'application de l'apprentissage automatique en matière judiciaire

8 - Il ne sera pas développé ici les raisons pour lesquelles ces corrélations d'ordre linguistique ne peuvent prétendre à modéliser la prise de décision judiciaire ou à expliciter le comportement d'un juge [14], ni les limites techniques de ce type de modélisation (comme l'hyperspécialisation des modèles d'apprentissages, difficilement généralisables à d'autres tâches, ou la possibilité de produire des résultats tronqués en manipulant des données d'entrée [15]). Il s'agira de démontrer en quoi des apprentissages n'ayant pas conscience de la complexité de la matière traitée (l'application de la loi et les phénomènes sociaux) risquent de créer plus de problèmes qu'ils n'apportent de solution.

9 - Parmi ces problèmes, il sera plus spécifiquement étudié la manière dont les modèles d'apprentissage peuvent potentiellement reproduire et aggraver les discriminations s'ils ne sont pas nativement construits pour les neutraliser. Les différentes techniques de l'apprentissage automatique paraissent en effet en elles-mêmes neutres en termes de valeurs sociales : que l'apprentissage soit supervisé ou non, avec ou sans renforcement, s'appuyant sur des machines à support de vecteur ou des réseaux de neurones profonds, les sciences fondamentales qui les animent sont avant tout un formalisme. En revanche, l'utilisation de ce formalisme avec une méthode et des données biaisées entraînera systématiquement des résultats biaisés.

10 - Prenons l'exemple de l'algorithme COMPAS [16] qui est utilisé de manière effective dans certains États américains afin d'évaluer la dangerosité des individus en vue de leur éventuel placement en détention provisoire ou lors du prononcé d'une condamnation pénale. Cet algorithme s'appuie sur des études académiques en criminologie et en sociologie, sur différents modèles statistiques et le traitement d'un questionnaire de 137 entrées, relatif à la personne concernée et à son passé judiciaire sans aucune référence à son origine ethnique [17]. Le système fournit ensuite au juge différents « scores » à un horizon de 2 années : risque de récidive, risque de comportement violent et risque de non-comparution pour les situations de placement en détention provisoire. La démarche apparaît *a priori* pluridisciplinaire et fondée scientifiquement.

11 - Toutefois, en mai 2016, les journalistes de l'ONG ProPublica ont analysé l'efficacité des « prédictions » de COMPAS sur une population de près de 10 000 individus arrêtés dans le comté de Broward (Floride) entre 2013 et 2014 [18]. Cette étude a révélé non seulement un taux relativement faible de « prédictions » justes (61 %) mais, en procédant à l'analyse approfondie des « faux positifs », elle a par ailleurs établi que les populations afro-américaines étaient pondérées d'un plus fort risque de récidive que les populations blanches. Inversement, les populations blanches ayant effectivement récidivées avaient été deux fois plus classifiées comme étant en risque faible que les populations afro-américaines. En d'autres termes, sans inclure l'éthnie des individus ou avoir été spécifiquement conçu pour traiter cette caractéristique, le croisement des données (dont le lieu de résidence) a indirectement surpondéré cet aspect au détriment d'autres facteurs sociaux individuels (éducation, emploi, parcours familial) et a conduit à influencer les juges avec des indicateurs proprement discriminatoires.

10. Le théorème de Bayes (1763) a servi de fondement aux développements ultérieurs pour calculer la distribution de probabilité de survenance d'un phénomène (pouvant être apparemment aléatoire) à partir de l'étude d'événements similaires passés.
11. Le premier modèle mathématique et informatique du neurone biologique (neurone formel) a été mis au point par Warren McCulloch et Walter Pitts en 1943. Il est en réalité aussi semblable à un neurone biologique que l'aile d'un avion est semblable à celle d'un oiseau.
12. Pour être précis, il convient de rappeler que le traitement de langage naturel (*natural language processing*) est une classe d'algorithme pouvant être mobilisée avant la phase de traitement par de l'apprentissage automatique, afin de rendre possible la recherche des différentes variables contenues dans les décisions judiciaires.
13. Sur la confusion entre corrélation et causalité, V. not. *D. Cardon, À quoi servent les algorithmes. Nos vies à l'heure des big data : Seuil, La République des idées, 2015*.
14. Pour un exposé de ces raisons, *V. Y. Meneceur, Quel avenir pour une justice prédictive : JCP G 2018, doctr.190*.
15. *V. K. Quach, How we fooled Google's AI into thinking a 3D-printed turtle was a gun : The Register, 6 nov. 2017*.
16. Correctional Offender Management Profiling for Alternative Sanctions (Profilage des délinquants correctionnels pour des sanctions alternatives) est un algorithme développé par la société privée Equivant (ex-Northpointe) : *www.equivant.com/solutions/inmate-classification*.
17. *Practitioner's Guide to COMPAS Core, Northpointe, 2015 : www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf*.
18. L'étude et sa méthodologie est accessible en ligne : *www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing*.

12 - Sans pour autant rendre leur méthodologie transparente en réponse à la critique, les concepteurs de COMPAS (la société Northpointe) ont contesté les résultats de cette enquête en affirmant que les journalistes avaient mal interprété les données. Ils ont en effet démontré de leur côté que le risque de récidive pour les profils à haut risque est le même, quelle que soit l'ethnie de la personne concernée [19]. Au final, à qui donner raison ? Pour Krishna Gummadi, chercheur à l'Institut Max Planck de Sarrebruck, les deux conclusions ne se contrediraient pas car Northpointe se serait focalisé sur l'absence de discrimination des valeurs prédictives positives et ProPublica sur les discriminations présentes dans les « faux positifs » [20]. Il estime au final qu'il s'agit avant tout d'une perception différente de l'équité, le calibrage de l'algorithme pour augmenter l'identification de personnes à haut risque (« vrai positif ») amplifiant mécaniquement le nombre de personnes injustement classées comme récidivistes probables (« faux positif »). Inversement, un calibrage pour réduire le nombre de ces « faux positifs » entraînerait un accroissement des récidivistes passant dans les mailles du filet (« faux négatif »). Est-ce à des concepteurs privés de décider de cette équité ? Qui doit se prononcer sur la pertinence d'appliquer des prédictions basées sur des statistiques générales et des algorithmes opaques pour une situation individuelle ? Certainement la loi et les tribunaux, même si la Cour suprême des États-Unis a refusé de connaître le 26 juin 2017 de l'appel de la décision opposant Eric Loomis contre la Cour suprême du Wisconsin, qui avait estimé que le juge continuait de disposer de son entier pouvoir d'appréciation face à COMPAS [21].

13 - En reprenant du champ, on pourrait considérer que la problématique ainsi posée pour la matière pénale est singulière. Elle reste en réalité la même avec des affaires civiles, administratives ou commerciales : la nature de la matière contentieuse est en effet étrangère à la présence de biais dans la méthode et les données. Il serait intéressant par exemple d'examiner si, pour une série de prestations compensatoires dans lesquelles la situation maritale et économique est objectivement équivalente, les montants « prédits » par de tels algorithmes apparaissent pondérés différemment selon le lieu de résidence des parties. Dans la positive, quels critères, explicites ou sous-jacents, ont pu avoir une influence ? Sans explication ni transparence sur cet état de fait, cela revient à laisser une « boîte noire » influer de manière tout à fait discrétionnaire sur l'issue d'un contentieux et à reproduire des inégalités.

14 - Réduire l'interprétation du droit au chiffre, c'est donc aller d'une société fondée sur la représentation démocratique et l'État de droit vers une société enfermée dans les équations conçues par les informaticiens [22]. Une forme de coup d'État numérique... un « coup data » comme le nomme les avocats Adrien Basdevant et Jean-Pierre Mignard [23]. Il ne s'agit naturellement pas de nier les apports d'une analyse de la jurisprudence avec les moyens de la *data science* mais de ne pas penser pouvoir réduire la matière judiciaire en des formules prédictives à des fins prescriptives [24]. Pablo Jensen, directeur de recherche au CNRS, rappelle bien les différences fondamentales existant entre les sciences naturelles, où il est bien plus aisé d'identifier « des faits têtus qu'on ne peut éliminer d'un revers de main », et les sciences sociales, dans lesquelles « il est bien plus difficile de retrouver des relations stables ». Il insiste sur l'utilité de la formalisation mathématique pour appréhender la réalité (elle rend visible par exemple la mortalité due à la pollution, indétectable dans des cas individuels) mais il prévient de toute simplification et insiste sur les bénéfices d'un travail avec « une communauté scientifique pluraliste, capable d'objecter, de mettre en doute les suppositions et les calculs » [25]. En somme, ne pas réduire la manière de concevoir la société à une simple forme logique.

3. Une éthique pour soutenir un développement de l'apprentissage automatique avec des décisions judiciaires dans le respect des droits fondamentaux

15 - Lentement, le numérique est donc susceptible de nous conduire vers un monde qui n'a pas été choisi en pleine conscience démocratique. Un monde qui ne semble d'ailleurs plus pouvoir se résumer en la dichotomie du siècle passé État nation/marché mais semble se recomposer en État de droit/gouvernance par la donnée [26]. Dans le contexte d'un tout nouveau type de choc de civilisation, il pourrait être objecté que la *soft law* n'a aucun caractère contraignant *per se* et ne serait en fait que l'expression d'une forme de faillite d'un « État garant » [27]. Seul le renforcement des moyens les plus forts d'encadrement (conventions et protocoles internationaux ou constitutions et lois nationales par exemple) serait à même de garantir une certaine efficacité [28]. De telles constructions vont vraisemblablement se mettre en œuvre [29] mais des formes moins contraignantes peuvent s'avérer être des laboratoires bien utiles permettant de distinguer progressivement l'essentiel de l'accessoire. Il s'agit même d'une opportunité pour alimenter un débat public responsabilisant tous les acteurs et pour contribuer à un processus législatif de qualité, au lieu de céder à une certaine précipitation voulue par notre temps politique, qui ne conduit

19. *W. Dieterich, C. Mendoza, T. Brennan, COMPAS risk scales : Demonstrating accuracy equity and predictive parity : Rapp. technique, Northpointe Inc., 2016.*
20. *K. P. Gummadi, Discrimination in machine decision making : www.european-big-data-value-forum.eu/wp-content/uploads/2017/12/Krishna-Gummadi-Max-Planck-Institute-Discrimination-in-Machine-Decision-Making-EBDVF17.pdf.*
21. La procédure du recours est accessible sur le site suivant : *www.scotusblog.com/case-files/cases/loomis-v-wisconsin/.*
22. *A. Garapon, J. Lassègue, Justice digitale : PUF, 2018, p. 37. – V. Entretien avec J. Lassègue, dans ce numéro : RPPI 2018, entretien 3.*
23. Formule développée par *A. Basdevant, J.-P. Mignard, L'empire des Données : Don Quichotte, 2018.*
24. Pour une analyse de la manière dont tirer parti des mathématiques et des outils d'apprentissage automatique à des fins autres que prescriptives, *V. J. Dupré, J. Lévy Véhel, L'intelligence artificielle au service de la valorisation du patrimoine jurisprudentiel : Dalloz IP/IT n° 10 2017, oct. 2017.*
25. *P. Jensen, Pourquoi la société ne se laisse pas mettre en équation : Éd. Seuil, 2018.*
26. *A. Basdevant, J.-P. Mignard, préc., note n° 23, p. 249.*
27. *P. Legendre, Sur la question dogmatique en Occident : Fayard, Paris, 1999, p. 172 cité par D. Forest, La régulation des algorithmes, entre éthique et droit : RLDI, 31 mai 2017 ; www.actualitesdudroit.fr/browse/tech-droit/intelligence-artificielle/7176/la-regulation-des-algorithmes-entre-ethique-et-droit.*
28. En ce qui concerne la nécessité de réguler le numérique par la voie constitutionnelle, V. par ex. *S. Soriano, Le numérique a toute sa place dans la Constitution : Le Nouveau Magazine Littéraire, 4 juill. 2018 ; www-nouveau--magazine--litteraire-com.cdn.ampproject.org/c/s/www.nouveau-magazine-litteraire.com/idees/le-numérique-a-toute-sa-place-dans-la-constitution.* – Ou encore *D. Forest, La régulation des algorithmes, entre éthique et droit : RLDI, 31 mai 2017 ; www.actualitesdudroit.fr/browse/tech-droit/intelligence-artificielle/7176/la-regulation-des-algorithmes-entre-ethique-et-droit.*
29. Les principes et la logique de la convention d'Oviedo relative au domaine biomédical sont déjà cités comme un exemple de démarche. V. l'entretien avec *J. Kleijssen, directeur au Conseil de l'Europe, The case for human-based regulation – www.youtube.com/watch ?v=i2q3MvxzNjM.*

au final qu'à consacrer des textes relatifs au numérique n'ayant qu'une « fonction subsidiaire » [30].

16 - Relevons également que prétendre édifier une éthique générale de l'IA revient à édulcorer les enjeux particuliers. Eric Horvitz, directeur du laboratoire de recherche Microsoft, faisait remarquer lors d'une conférence la grande difficulté à établir des règles générales pour un champ aussi large, en rappelant qu'il n'existe pas d'éthique générale de l'ingénierie par exemple [31]. Il semble donc opportun de composer des principes s'imposant aux concepteurs de modèles avec de l'apprentissage automatique en mêlant les fondements communs à toute application de cette technologie avec des règles plus spécifiques applicables au champ judiciaire.

A. - Les principes directeurs relatifs aux droits fondamentaux des individus

17 - Les finalités d'un traitement de données juridictionnelles avec de l'apprentissage automatique devraient être compatibles avec les droits fondamentaux et les principes de protection des données personnelles. Il pourrait être esquissé trois premiers principes, notamment à la lumière de la Convention EDH, du règlement général sur la protection des données et des premières chartes et de déclarations portant sur la régulation de l'IA [32] :

1) Prévention des discriminations [33]. – Les concepteurs, conscients de la capacité de l'apprentissage automatique à représenter les biais existants dans la société (selon des catégories ethniques, politiques, religieuses, philosophiques ou de genre), devraient tenter de neutraliser nativement le poids et l'effet de ces biais dans les calculs, en documentant la démarche dans un langage clair et accessible [34], ou les signaler à l'utilisateur final du traitement ;

2) Garantie des droits à un procès équitable et à un recours effectif [35]. – Le juge doit rester responsable de ses décisions juridictionnelles, qui ne sauraient être motivées sur le seul fondement de la proposition d'un système automatisé. De même, les avocats doivent pouvoir rester libres du choix de leur stratégie de défense. La responsabilité des professionnels de la justice ne pourrait être engagée sur le seul motif d'un écart avec la proposition de tels systèmes. Utilisés en amont d'une procédure judiciaire, ces systèmes ne devraient pas conduire à restreindre, directement ou indirectement, l'accès à un tribunal indépendant et impartial si une partie souhaite que sa cause soit entendue et débattue contradictoirement devant un juge [36] ;

3) Prévention du profilage des individus, protection de la dignité et de la vie privée et familiale [37]. – La finalité d'un traitement de données juridictionnelles avec de l'apprentissage automatique ne devrait pas conduire, seul ou en conjonction avec d'autres traitements, à contribuer ou établir un profil des individus (notamment les professionnels, parties, témoins, tiers) de nature à compromettre la protection de leur vie privée et familiale, à exercer une pression à leur encontre ou à porter atteinte à leur dignité ou à leur sécurité [38].

État de l'utilisation de l'intelligence artificielle avec des décisions judiciaires dans les systèmes judiciaires européens

L'évaluation du phénomène en Europe est délicate si elle s'opère du point de vue des pouvoirs publics. Les études de la CEPEJ par exemple, qui s'appuient sur un réseau intergouvernemental de représentants des ministères de la Justice, sont limitées aux outils effectivement utilisés par les tribunaux et ne recensent pas les logiciels utilisés par les autres professionnels du droit. Dans ce périmètre public, et si l'on s'en tient à la seule utilisation d'apprentissage automatique pour l'analyse de jurisprudence, peu ou pas d'applications concrètes ont été réalisées à ce jour (l'Autriche, la France, l'Italie, la Lettonie et certains pays scandinaves semblent avoir démarré des tests ou préfiguré des usages pour les tribunaux).

Si l'on s'intéresse maintenant au secteur privé et à une large clientèle de juristes, certaines offres (par exemple Lex Machina, Watson d'IBM, Luminance) sont commercialisées simultanément dans plusieurs pays européens alors que d'autres en sont implantées que localement. Sans prétendre à l'exhaustivité, il peut être cité l'Autriche (Watson), l'Espagne (ejusticia), l'Italie (Luminance) et la Grande-Bretagne (Luminance, Westlaw Edge). En France (JurisData Analytics, CaseLaw Analytics, Doctrine.fr et Predictice) semblent parmi les plus développées sur le marché. Citons également Jus Mundi (en développement) qui agrège différentes sources nationales et internationales.

Le logiciel HART, testé en Grande-Bretagne, n'utilise pas à proprement parler de la jurisprudence pour évaluer la dangerosité des individus mais utilise un algorithme à la philosophie comparable de celui ayant fait débat aux États-Unis (COMPAS). Toujours en Grande-Bretagne, Premonition dresse un classement des cabinets d'avocats basé sur leur taux de succès devant la chambre commerciale de la Haute Cour d'Angleterre. Enfin, aux Pays-Bas l'offre privée d'arbitrage en ligne e-Courts a fait polémique pour rendre plus complexe l'accès aux tribunaux et son manque de transparence.

B. - Les principes directeurs relatifs à la procédure de conception

18 - Les procédures de mise en conformité dès la conception (*security by design*, *privacy by design*) devraient être complétées d'une large approche de « droits de l'homme dès la conception » (*human rights by design*). Cette démarche consisterait non seulement à intégrer dès le début de l'élaboration des modèles des règles interdisant de porter atteinte directement ou indirectement à ces droits fondamentaux, mais aussi à garantir la qualité et la

30. *D. Forest, préc., note n° 27.*
31. *E. Horvitz, Fireside chat on AI, People and Society : Bruxelles, 18 juin 2018.*
32. Les principes développés s'appuient notamment sur les travaux des déclarations de Toronto et de Montréal, de la CNIL (Les enjeux éthiques des algorithmes et de l'intelligence artificielle), de Google (sept principes), d'*Open Law* (Charte éthique pour un marché du droit en ligne et ses acteurs) et d'AI-Ethics.com. Les principes développés ici n'ont naturellement pas de prétention exhaustive et ont une vocation exploratoire.
33. *Conv. EDH, art. 14.*
34. V. le principe 6 sur la question spécifique de la transparence.
35. *Conv. EDH, art. 6 et 13.* – S'agissant de décisions automatisées, V. de plus l'article 22 du règlement de l'Union européenne n° 2016/679 du 27 avril 2016 dit règlement général sur la protection des données (RGPD) qui ne les prohibe pas s'ils ont un effet juridique mais y ajoute des obligations spécifiques.
36. S'agissant de l'arbitrage par des dispositifs de résolution de litiges en ligne, la renonciation volontaire à saisir un tribunal suppose malgré tout un certain contrôle sur la procédure d'arbitrage par les tribunaux internes. Pour déterminer si ce contrôle a été exercé correctement, il y a lieu de tenir compte non seulement du compromis d'arbitrage intervenu entre les parties et de la nature de la procédure d'arbitrage privée, mais également du cadre législatif prévoyant une telle procédure *(Comm. des droits de l'homme, 27 nov. 1996, n° 28101/95, Nordström-Janzon et Nordström-Lehtinen c/ Pays-Bas, déc. et rapp. 87-A, p. 116).*
37. *Conv. EDH, art. 8.* – *PE et Cons. UE, règl. (UE) n° 2016/679, 27 avr. 2016, art. 4.*
38. Il ne sera pas développé ici la question du sens du traitement du nom des professionnels (magistrats et avocats notamment) par de l'apprentissage automatique mais il pourra utilement être fait référence aux travaux de la mission de préfiguration sur l'ouverture au public des décisions de justice du professeur Loïc Cadiet *(www.justice.gouv.fr/publication/open_data_rapport.pdf.* – *V. RPPI 2018, entretien 2)*. L'intérêt d'un tel traitement paraît variable en fonction du degré d'instance et pose, surtout, un certain nombre de difficultés de principes dont le sens de la statistique produite, *V. Y. Meneceur, Quel avenir pour une justice prédictive, préc., note n° 14.*

© LEXISNEXIS SA - REVUE PRATIQUE DE LA PROSPECTIVE ET DE L'INNOVATION - N° 2 - OCTOBRE 2018

sécurité des données et à rendre possible l'audit. Des institutions au niveau national ou européen, pourraient être chargées de fonctions de contrôle, d'audit et de conseil tels que des comités d'éthique *ad hoc* ou des commissariats aux algorithmes.

1) Évaluation des impacts sur les droits fondamentaux dès la conception. – Les concepteurs de modèles avec de l'apprentissage automatique devraient s'engager à conduire durant leurs travaux des ateliers pluridisciplinaires avec des professionnels du droit et de la justice ainsi que des experts en sciences sociales afin d'évaluer les éventuels impacts sur les droits fondamentaux. Ces impacts devraient pouvoir être constamment partagés de manière large entre différentes communautés de pratique (recherche, industrie, professionnels) et réévalués de manière continue au vu des différents retours d'expérience.

2) Garantie de la qualité des données juridictionnelles et de la sécurité des traitements. – Les données d'apprentissage basées sur des décisions juridictionnelles devraient provenir de sources certifiées et ne pas être altérées jusqu'à leur utilisation effective par le mécanisme d'apprentissage [39]. L'ensemble du processus devrait ainsi pouvoir être tracé à tout moment pour s'assurer qu'aucune modification, de nature à changer le contenu ou le sens de la décision traitée, n'est intervenue. Les modèles et les algorithmes créés devraient pouvoir être également stockés et exécutés dans des environnements sécurisés, garantissant l'intégrité du système ;

3) Transparence, neutralité et loyauté. – Les méthodologies d'apprentissage devraient être accessibles et rendues compréhensibles. Des audits externes par des autorités certifiées devraient pouvoir être réalisés régulièrement. Un équilibre semble devoir être trouvé entre les secrets industriels et des exigences de transparence (accès à la démarche de conception), de neutralité (absence de biais) et de loyauté (faire primer l'intérêt général). Dans l'impossibilité d'offrir une totale transparence technique [40], la logique sous-jacente du système devrait en toute hypothèse pouvoir être expliquée dans un langage clair et vulgarisé (décrire la manière de produire ses résultats ainsi que la quantité et la qualité des données utilisées pour l'apprentissage) et pouvoir être éventuellement critiquée lors d'un débat contradictoire devant un juge au vu des circonstances particulières de l'espèce qui lui est soumise. Les jeux de données spécifiquement utilisés pour l'apprentissage devraient être librement accessibles. La qualité de la prévision et le taux d'erreur associé à l'utilisation de l'algorithme d'apprentissage devraient aussi pouvoir être mesurés et communiqués [41]. ■

Mots-Clés : Prospective - Justice prédictive - Algorithmes - Systèmes judiciaires européens

Technologies de l'information - Intelligence artificielle - Justice prédictive - Éthique

39. En ce qui concerne la nécessité de travailler sur des originaux certifiés, *V. J.-P. Jean, Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence, colloque de la Cour de cassation « La jurisprudence dans le mouvement de l'open data », 14 oct. 2016 : www.courdecassation.fr/IMG///Open%20data,%20par%20Jean-Paul%20Jean.pdf.* – Supplément au *JCP G 2017, n° 9.*

40. Il ne semble parfois guère possible d'entièrement expliciter les modèles bâtis avec de l'apprentissage profond (*deep learning*). – *V. I. Daubechies, Machine Learning Works Great – Mathematicians Just Don't Know Why, Wired, 12 déc. 2015 – www.wired.com/2015/12/machine-learning-works-greatmathematicians-just-dont-know-why/.*

41. V. en ce sens *P. Besse, C. Castets-Renard, A. Garivier, préc.*