

Strasbourg, 3 décembre 2018

CEPEJ(2018)14

COMMISSION EUROPEENNE POUR L'EFFICACITE DE LA JUSTICE
(CEPEJ)

**Charte éthique européenne d'utilisation de l'intelligence artificielle
dans les systèmes judiciaires et leur environnement**

adoptée par la CEPEJ lors de sa 31^{ème} réunion plénière
(Strasbourg, 3-4 décembre 2018)



Table des matières

Introduction	4
Les 5 principes de la Charte éthique d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement	5
1. <i>Principe de respect des droits fondamentaux : Assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux</i>	6
2. <i>Principe de non-discrimination : Prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus</i>	7
3. <i>Principe de qualité et sécurité : En ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi disciplinaire, dans un environnement technologique sécurisé</i>	8
4. <i>Principe de transparence, de neutralité et d'intégrité intellectuelle : Rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes</i>	9
5. <i>Principe de maîtrise par l'utilisateur : Bannir une approche prescriptive et permettre à l'utilisateur d'être un acteur éclairé et maître de ses choix</i>	10
Annexe I : Etude approfondie sur l'utilisation de l'IA dans les systèmes judiciaires, notamment les applications d'IA assurant le traitement des décisions et des données judiciaires	11
Introduction	12
1. Etat des lieux de l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires des Etats membres du Conseil de l'Europe	14
2. Etat des lieux des politiques publiques d'open data des décisions judiciaires dans les systèmes judiciaires des Etats membres du Conseil de l'Europe	16
2.1 Définition de l'open data des décisions judiciaires	16
2.2 Etat du développement de l'open data des décisions judiciaires dans les Etats-membres du Conseil de l'Europe et conséquences sur le développement de la jurisprudence	17
2.3 La protection des données à caractère personnel dans les politiques d'ouverture des décisions judiciaires	21
3. Caractéristiques de fonctionnement de l'intelligence artificielle (apprentissage machine) appliquée aux décisions judiciaires	24
3.1 Les fonctionnalités théoriques d'un logiciel de justice dite prédictive	24
3.2 Le fonctionnement concret des intelligences artificielles : des machines statistiques construisant des modèles sur la base du passé	25
4. Les intelligences artificielles peuvent-elles modéliser <i>a priori</i> un raisonnement juridique?	29
5. Les IA peuvent-elles expliquer <i>a posteriori</i> le comportement des juges ?	32
6. Quelles applications pour la justice civile, commerciale et administrative ?	34
6.1 Une nouvelle puissance de calcul pour la constitution de barèmes	35
6.2 Les résolutions de litiges en ligne	35

6.3 Les principales garanties à réaffirmer dans le cadre de la procédure civile, commerciale et administrative	37
7. Les enjeux spécifiques à la justice pénale : prévention des infractions, du risque de récidive et évaluation de la dangerosité	39
7.1 Les outils utilisés par les services d'enquête avant le procès pénal	39
7.2 Les outils durant le procès pénal	40
7.3 Les défis du « prédictif » en matière pénale	42
8. Questions spécifiques relatives à la protection des données personnelles	44
9. Les potentialités et les limites des outils de justice dite prédictive	45
10. La nécessité d'un débat public approfondi autour de ces outils antérieurement à la mise en place des politiques publiques de développement de ceux-ci. L'impérieuse nécessité d'une cyberéthique pour encadrer le développement des algorithmes d'intelligence artificielle dans le respect des droits fondamentaux	47
10.1 L'importance de débattre, tester et continuellement revoir l'application de ces outils antérieurement à la mise en place de politiques publiques	47
10.2 La mise en place d'un cadre éthique	47
Annexe II : Quelles utilisations pour l'IA dans les systèmes judiciaires européens ?	49
Utilisations à encourager	49
Utilisations à envisager avec de fortes précautions méthodologiques	50
Utilisations à envisager au terme de travaux scientifiques complémentaires	51
Utilisations à envisager avec les plus extrêmes réserves	51
Annexe III : Glossaire	52
Annexe IV : Checklist d'intégration des principes de la Charte dans vos traitements	58
Checklist d'évaluation de votre traitement	59

Introduction

Prenant acte de l'importance croissante de l'intelligence artificielle (IA) dans nos sociétés modernes, et des bénéfices escomptables lorsqu'elle sera utilisée pleinement au service de l'efficacité et de la qualité de la justice¹, la CEPEJ adopte solennellement les 5 principes fondamentaux suivants appelés « Charte éthique européenne d'utilisation de l'IA dans les systèmes judiciaires et leur environnement ».

Cette Charte s'adresse aux acteurs publics et privés en charge de la conception et du déploiement d'outils et de services d'intelligence artificielle s'appuyant notamment sur le traitement des décisions juridictionnelles et des données judiciaires (apprentissage machine ou toutes autres méthodes issues des sciences de données).

Elle concerne également les décideurs publics en charge de l'encadrement législatif ou réglementaire, du développement, de l'audit ou de l'utilisation de tels outils et services.

L'utilisation de tels outils et services dans les systèmes judiciaires a vocation à améliorer l'efficacité et la qualité de la justice et mérite d'être encouragée. Elle doit toutefois se faire de manière responsable, dans le respect des droits fondamentaux des individus énoncés par la Convention européenne des droits de l'homme et par la Convention pour la protection des données à caractère personnel, ainsi que des autres principes fondamentaux énoncés ci-après, qui devraient orienter la définition de politiques publiques de la justice en ce domaine.

Le traitement des décisions juridictionnelles par l'IA est susceptible, selon leurs concepteurs, de contribuer, en matière civile, commerciale et administrative, à améliorer la prévisibilité de l'application de la loi et la cohérence des décisions judiciaires, sous réserve du respect des principes énoncés ci-après. En matière pénale, leur utilisation doit être envisagée avec les plus extrêmes réserves, afin de prévenir des discriminations sur des données sensibles, en conformité avec les garanties du procès équitable.

Qu'ils soient conçus dans le but d'apporter un support à une consultation juridique, une aide à la rédaction ou à la décision ou une orientation des justiciables, il est essentiel que lesdits traitements soient effectués dans des conditions de transparence, de neutralité et de loyauté certifiées par une expertise extérieure à l'opérateur et indépendante.

Application de la Charte

Les principes de la Charte devraient pouvoir faire l'objet d'une application, d'un suivi et d'une évaluation régulière par les acteurs publics et privés, pour améliorer constamment les pratiques.

A cet égard, il est souhaitable qu'un bilan régulier de la mise en œuvre des principes de la Charte soit effectué par ces acteurs, expliquant le cas échéant les raisons de non-réalisation ou de réalisation partielle, accompagné d'un plan d'action pour introduire les mesures nécessaires.

Les autorités indépendantes mentionnées dans la Charte pourraient être chargées d'évaluer périodiquement le niveau d'adhésion aux principes de cette Charte de tous les acteurs et de proposer des améliorations pour l'adapter à l'évolution des technologies et des utilisations de ces technologies.

¹ Pour la définition d'intelligence artificielle, voir le Glossaire en annexe.

Les 5 principes de la Charte éthique d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement

1

PRINCIPE DE RESPECT DES DROITS FONDAMENTAUX : assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux.

2

PRINCIPE DE NON-DISCRIMINATION : prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus.

3

PRINCIPE DE QUALITE et SECURITE : en ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi-disciplinaire, dans un environnement technologique sécurisé.

4

PRINCIPE DE TRANSPARENCE, DE NEUTRALITE ET D'INTEGRITE INTELLECTUELLE : rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes.

5

PRINCIPE DE MAÎTRISE PAR L'UTILISATEUR : bannir une approche prescriptive et permettre à l'utilisateur d'être un acteur éclairé et maître de ses choix.

1

Principe de respect des droits fondamentaux :

assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux

Le traitement de décisions juridictionnelles et de données judiciaires doit répondre à des finalités claires, dans le respect des droits fondamentaux garantis par la Convention européenne des droits de l'homme (CEDH) et de la Convention pour la protection des données à caractère personnel (Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, STE n° 108 telle qu'amendée par le protocole d'amendement STCE n°223).

Lorsque les outils d'intelligence artificielle sont utilisés pour trancher un litige ou en tant qu'outils d'aide à la décision judiciaire ou d'orientation du justiciable, ils ne doivent pas porter atteinte aux garanties du droit d'accès au juge et du procès équitable (égalité des armes et respect du contradictoire).

Ils devraient également être utilisés dans le plein respect des principes de l'Etat de droit et de l'indépendance des juges dans leur processus de décision.

Les approches dites d'« éthique² dès la conception » (*ethical-by-design*) ou de « droits de l'homme dès la conception » (*Human rights by design*) seraient ainsi à privilégier. Elles consistent à intégrer dès la phase de conception et d'apprentissage des règles interdisant de porter atteinte directement ou indirectement aux valeurs fondamentales protégées par les Conventions

² Le choix éthique est réalisé en amont par les concepteurs du programme et n'est donc pas laissé à l'utilisateur.

2

Principe de non-discrimination :

prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus

Conscients de la capacité de ces traitements à révéler des discriminations existantes, par regroupement ou classification de données relatives à des individus ou des groupes d'individus, les acteurs publics et privés doivent s'assurer que ceux-ci ne reproduisent pas ou n'aggravent pas ces discriminations et qu'ils ne conduisent pas à des analyses ou usages déterministes.

Une vigilance particulière doit être accordée tant dans la phase d'élaboration que d'utilisation, d'autant plus lorsque ces traitements se fondent de manière directe ou indirecte sur des données « sensibles ». Peuvent être considérées comme telles l'origine prétendument raciale ou ethnique, les conditions socio-économiques, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale ainsi que les données génétiques, les données biométriques, les données concernant la santé ou concernant la vie sexuelle ou l'orientation sexuelle. Des interventions correctives visant à limiter, et si possible neutraliser ces risques, de même qu'une sensibilisation des acteurs doivent être envisagées lorsque ces discriminations sont constatées.

En revanche, l'utilisation de l'apprentissage automatique et la conduite d'analyses scientifiques pluridisciplinaires visant à lutter contre de telles discriminations devraient être encouragées.

3

Principe de qualité et sécurité :

en ce qui concerne le traitement des décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles avec des modèles conçus d'une manière multi-disciplinaire, dans un environnement technologique sécurisé

Les concepteurs de modèles à partir de l'apprentissage automatique doivent pouvoir recueillir de manière large l'expertise des professionnels de la justice concernés (juges, procureurs, avocats, etc.) et des chercheurs / enseignants en droit et en sciences sociales (économistes, sociologues, philosophes par exemple).

La constitution d'équipes de projet mixtes dans des cycles de conception courts pour produire des maquettes fonctionnelles est l'un des modes d'organisation permettant de tirer le meilleur parti de cette pluridisciplinarité.

Les mesures éthiques existantes devraient pouvoir être constamment partagées par ces équipes de projet et enrichies sur la base des retours d'expérience.

Les données dérivant des décisions juridictionnelles et intégrées dans un logiciel qui exécute un algorithme d'apprentissage-machine doivent provenir de sources certifiées et ne doivent pas pouvoir être altérées jusqu'à leur utilisation effective par le mécanisme d'apprentissage. L'ensemble du processus doit ainsi pouvoir être tracé pour s'assurer qu'aucune altération, de nature à modifier 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 et l'intangibilité du système.

4

Principe de transparence, de neutralité et d'intégrité intellectuelle :

rendre accessibles et compréhensibles les méthodologies de traitement des données, autoriser les audits externes

Un équilibre doit être trouvé³ entre la propriété intellectuelle de certaines méthodes de traitement et les exigences de transparence (accès à la démarche de conception), de neutralité (absence de biais⁴), de loyauté et d'intégrité intellectuelle (faire primer l'intérêt de la justice), en raison des effets juridiques ou des incidences que ces méthodes peuvent avoir sur les personnes. Il convient de préciser que toute la chaîne de conception et de fonctionnement est concernée par ces mesures puisque la méthodologie de sélection, la qualité et l'organisation des données influent directement sur la phase d'apprentissage.

La totale transparence technique (par exemple *open source* du code et de la documentation) est une première possibilité, limitée parfois par la protection des secrets industriels. Le système pourrait être également explicable dans un langage clair et vulgarisé afin de décrire la manière dont il produit ses résultats, en communiquant par exemple sur la nature des prestations proposées, les outils développés, les performances et les risques d'erreur. Des autorités ou des experts indépendants pourraient être en charge de la certification et de l'audit des traitements ou fournir, en amont, du conseil. Des labels renouvelés régulièrement pourraient être délivrés par les autorités publiques.

³ La proposition faite dans l'étude du MSI-NET du Conseil de l'Europe intitulé « Algorithmes et droits humains », page 46, est intéressante : « La solution consistant à demander la divulgation publique d'algorithmes entiers ou de leur code source est utopique dans ce contexte, les entreprises privées considérant leurs algorithmes comme des logiciels propriétaires stratégiques, et les protégeant donc en conséquence. Il paraît en revanche envisageable d'exiger la publication d'informations partielles mais néanmoins importantes, comme les variables utilisées, les objectifs visés par l'optimisation des algorithmes, les jeux de données d'apprentissage, les valeurs moyennes et les écarts types des résultats obtenus, ou la quantité et le type de données traitées par l'algorithme ». Ou encore les suggestions du rapport intitulé « l'IA pour l'humanité » rédigé par le député Cédric Villani dans le cadre d'une mission confiée par le Premier Ministre de la République française, à page 144: « les auditeurs pourraient tester l'équité et la loyauté d'un programme (faire ce qui est censé faire), par exemple à travers la soumission de multiples fausses données d'entrée, créer des nombreux profils d'utilisateur suivant des caractéristiques précises »...Mêmes constats dans le rapport de la House of Lords, « AI dans le Royaume Uni : prêts, disposés et capables ? sur l' IA : paras 92, 96-99).

⁴ Les solutions envisagées de manière générale pour garantir la neutralité des algorithmes dans le rapport susmentionné de la House of Lords (§114, 115, 116, 119, 120) sont intéressantes: plus de données diversifiées, plus de diversité et de transdisciplinarité, plus d'audit sur des aspects comme le traitement des données ou la manière dans laquelle la machine est construite.

5

Principe de maîtrise par l'utilisateur :

bannir une approche prescriptive et permettre à l'utilisateur d'être un acteur éclairé et maître de ses choix

L'autonomie de l'utilisateur doit être renforcée et ne pas être restreinte par l'utilisation d'outils et de services d'intelligence artificielle.

Le professionnel de la justice devrait à tout moment pouvoir revenir aux décisions et données judiciaires ayant été utilisées pour produire un résultat et continuer à avoir la possibilité de s'en écarter au vu des spécificités de l'affaire concrète.

Le justiciable devrait être informé dans un langage clair et compréhensible du caractère contraignant ou non des solutions proposées par les outils d'intelligence artificielle, des différentes options possibles ainsi que de son droit à une assistance juridique et à saisir un tribunal. Le justiciable devrait également être informé de manière claire de tout recours à un traitement préalable d'une affaire par de l'intelligence artificielle avant ou pendant un processus judiciaire et être en mesure de s'y opposer pour que son affaire soit entendue directement par un tribunal au sens de l'Article 6 de la CEDH.

D'une manière générale, des actions d'alphabétisation numérique des usagers et des débats impliquant les professionnels de la justice devraient être conduits à l'occasion de la mise en œuvre de tout système d'information s'appuyant sur l'intelligence artificielle.

Annexe I

Etude approfondie sur l'utilisation de l'IA dans les systèmes judiciaires, notamment les applications d'IA assurant le traitement des décisions et des données judiciaires

*préparée sur la base des contributions de M. Xavier Ronsin, Premier Président de la Cour
d'Appel de Rennes, expert scientifique (France)*

et

*M. Vasileios Lampos, Chercheur en chef au Service de sciences de l'informatique au
University College de Londres, expert scientifique (Royaume-Uni)*

avec la contribution de

*Mme Agnès Maîtreperre, magistrat, membre du Comité consultatif de la Convention sur la
protection des données à l'égard du traitement automatisé des données à caractère personnel
du Conseil de l'Europe (France)*

Les experts suivants ont contribué au perfectionnement de l'Etude :

*M. Francesco Contini, Chercheur en chef à l'Institut de recherche sur les systèmes judiciaires –
Centre National de Recherche (IRSIG-CNR), Bologne (Italie)*

*M. Francesco De Santis, Professeur de procédures de tutelle des droits de l'homme,
Université de Naples (Italie)*

*M. Jean Lassègue, Philosophe et épistémologue, chargé de recherche au Centre National de
Recherche Scientifique (CNRS) et chercheur associé à l'Institut des Hautes Etudes sur la Justice
(IHEJ) (France)*

*Mme Dory Reiling, Juge principal honoraire, experte indépendante en technologies de l'information et
en réformes judiciaires (Pays-Bas)*

*M. Aleš Završnik, Chercheur en chef à l'Institut de criminologie, professeur associé à la Faculté de
droit de l'Université de Ljubljana (Slovénie) et chercheur EURIAS 2017-18 au Collegium Helveticum
à Zürich (Suisse)*

Introduction

1. La vague de transformation numérique de nos sociétés touche de manière encore inégale les systèmes judiciaires des Etats membres du Conseil de l'Europe. De nombreux pays d'Europe semblent déjà avoir développé une réflexion extrêmement avancée avec des applications concrètes - tant sur le plan technologique que sur le plan de l'accompagnement juridique - alors que pour d'autres le sujet apparaît encore comme émergent et se limiterait simplement à pouvoir mettre en œuvre une informatique de gestion efficace.
2. Parmi les technologies à l'œuvre dans cette grande transformation numérique, l'intelligence artificielle (IA) apparaît à la fois comme la plus spectaculaire et la plus redoutable. Aux Etats-Unis, des « robots avocats » sont déjà à l'œuvre et semblent converser en langage naturel avec les humains. Les *legaltechs*, start-ups spécialisées dans la conception de nouveaux services juridiques, proposent également des nouvelles applications aux professions juridiques, principalement les avocats, services juridiques et assureurs, permettant d'avoir un accès approfondi à l'information juridique et à la jurisprudence. Ces compagnies privées ambitionnent même de prévoir les décisions des juges avec des outils, improprement qualifiés nous le verrons¹, de « justice prédictive ».
3. Un premier examen de ce phénomène conduit toutefois à distinguer ce qui relève des discours commerciaux d'une réalité d'utilisation et de déploiement. L'utilisation concrète et quotidienne de logiciels prédictifs par des juges paraît pour l'instant inexistante dans les Etats membres du Conseil de l'Europe. Des expérimentations locales² ainsi que des travaux universitaires³ visant à explorer les potentialités de ces applications ont pu être menés, mais sans aucun développement généralisé. En effet, l'initiative du développement de ces outils provient très largement du secteur privé, qui a su pour l'instant trouver un début de clientèle auprès des compagnies d'assurance, des avocats et des services juridiques soucieux de réduire l'« aléa judiciaire », c'est-à-dire l'imprévisibilité de la décision judiciaire. Les décideurs publics commencent néanmoins à être de plus en plus sollicités par un secteur privé souhaitant voir ces outils – qui sont parfois des versions « bêta », c'est-à-dire appelés à évoluer ultérieurement - intégrés dans les politiques publiques.
4. En continuité avec la réflexion entamée dans ses *Lignes Directrices sur la conduite du changement sur la cyberjustice*⁴, la CEPEJ se propose d'offrir aux décideurs publics et aux professionnels de la justice des clés pour une meilleure compréhension du phénomène de la « justice prédictive ».
5. Il s'agira premièrement de clarifier les questions relatives à la nature intrinsèque de ces systèmes de traitement de masse de données de jurisprudence, ainsi qu'à leurs limites techniques et théoriques : celles-ci, très peu présentes dans le débat qui s'est installé dans le monde judiciaire, sont pourtant très connues et discutées par les spécialistes de ces technologies (mathématiciens, statisticiens, informaticiens) et méritent d'être relevées.
6. Deuxièmement, le document analysera les avantages et les risques de ces outils. Si leurs promoteurs mettent en relief leurs atouts en matière de transparence, de prévisibilité et d'homogénéisation de la jurisprudence, leurs détracteurs pointent du doigt non seulement les limites mais aussi les biais de raisonnement des logiciels actuellement sur le marché. Les risques inhérents à ces technologies peuvent même aller au-delà de l'acte de juger et toucher des éléments essentiels de fonctionnement de l'État de droit et des systèmes judiciaires, auxquels le Conseil de l'Europe est particulièrement attaché.
7. Il s'agit notamment de principes comme la primauté de la loi : ces outils, dont les effets ne seraient pas seulement incitatifs mais quasiment prescriptifs, dégageraient en effet une nouvelle

¹ Voir en particulier l'encadré du Chapitre 9.

² A titre d'exemple les deux cours d'appel de Douai et de Rennes en France ont expérimenté pendant trois mois en 2017 le bénéfice d'un logiciel présenté comme prédictif auprès d'un panel de juges.

³ Travaux sur un échantillon de 584 décisions de la Cour européenne des droits de l'Homme : Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiu-Pietro, Vasileios Lampos, « Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective », publié le 24 octobre 2016, [En ligne], <https://peerj.com/articles/cs-93/>

⁴ Voir notamment le paragraphe 51 du document CEPEJ(2016)13, Lignes directrices sur la conduite du changement vers la Cyberjustice.

forme de normativité, qui tendrait à compléter la loi en encadrant le pouvoir souverain d'appréciation du juge et pourrait aboutir à long terme à une uniformisation des décisions judiciaires sur la base non plus d'un raisonnement individualisé des tribunaux, mais d'un pur calcul statistique par rapport à la moyenne des indemnités allouées antérieurement par d'autres juridictions.

8. Par ailleurs, la compatibilité de ces solutions avec des droits individuels consacrés par la Convention européenne des droits de l'homme (CEDH) est aussi une source de questionnement. Il s'agit notamment du droit à un procès équitable (plus particulièrement du droit au juge naturel établi par la loi, du droit à un tribunal indépendant et impartial, de l'égalité des armes dans les procédures judiciaires) et du droit au respect de la vie privée et familiale, notamment en cas d'insuffisant respect de la protection des données communiquées en *open data*.
9. Tout en étant sensible à ces questions, le document relève l'immense potentiel de l'IA en tant que support au travail des professionnels de la justice. Nul ne doute que certaines applications d'IA en cours de développement ou d'expérimentation, comme par exemple celles visant à améliorer la recherche juridique, peuvent être d'une très grande utilité et rendre le travail judiciaire à la fois plus rapide et plus efficace. Le document met en avant ces exemples positifs et prône de manière générale une utilisation de l'IA qui soit, d'une part, au service des professionnels de la justice et en phase avec leurs besoins et d'autre part, respectueuse des droits individuels garantis par la CEDH et des normes du Conseil de l'Europe, notamment en matière pénale. Loin d'être un simple instrument d'amélioration de l'efficacité des systèmes judiciaires, l'IA devrait renforcer et ne pas diminuer les garanties de l'État de droit tout comme la qualité du service public de la justice.
10. Enfin, le document donne des pistes pour une « gouvernance » de ce phénomène, sous la forme d'une charte éthique qui dégage un certain nombre de principes qui devraient guider le déploiement de l'IA dans les systèmes judiciaires. Il met par ailleurs l'accent sur la nécessité d'une approche prudente s'agissant de l'intégration des outils d'IA dans les politiques publiques. Il est en effet indispensable qu'un débat public implique tous les acteurs – qu'ils soient professionnels de la justice, *legaltachs* ou scientifiques - afin de leur permettre d'une part de partager leur analyse sur la portée et le possible impact de l'introduction des applications d'IA dans les systèmes judiciaires et d'autre part de développer au profit de celles-ci un indispensable cadre éthique. Dans un second temps, ce débat pourrait sortir d'un cadre purement « métier », impliquer les citoyens eux-mêmes et participer ainsi à une certaine forme d'alphabétisation numérique, comme cela a été réalisé par exemple au Canada⁵.

⁵ Déclaration de Montréal, disponible au lien : <https://www.declarationmontreal-iaresponsable.com/demarche>, visité le 16 juillet 2018.

1. Etat des lieux de l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires des Etats membres du Conseil de l'Europe

En 2018, l'utilisation des algorithmes d'intelligence artificielle dans les systèmes judiciaires européens demeure pour l'essentiel à l'initiative du secteur privé afin de développer une offre commerciale à destination des assurances, des directions juridiques, des avocats ou des particuliers.

11. L'utilisation de l'IA dans le champ judiciaire paraît assez largement vulgarisée aux Etats-Unis qui ont investi ces outils de manière assez décomplexée, tant en matière civile que pénale⁶.
12. Le recensement des initiatives en matière d'algorithmes d'IA dans les systèmes judiciaires des Etats membres du Conseil de l'Europe est une tâche moins aisée, car la plupart des initiatives proviennent du secteur privé et sont encore peu intégrées dans les politiques publiques.
13. La question de l'utilisation de l'IA dans les systèmes judiciaires a été traitée dans une enquête ligne spécifique, ouverte en avril 2018 de manière large tant aux représentants des Etats membres auprès de la CEPEJ que de la société civile. Le taux de réponse est apparu comme relativement faible et n'a pas permis de dessiner de tendances claires. Certains opérateurs privés n'ont pas semblé sensibles à ce recensement et les membres de la CEPEJ, qui appartiennent pour la plupart à des ministères de la justice ou des conseils supérieurs de justice, ne pas sont en mesure de citer d'autres outils que ceux effectivement utilisés par la sphère publique.
14. En conséquence, l'état des lieux dressé ci-dessous n'est que partiel et ne s'appuie que sur les recherches menées par les experts et le Secrétariat dans la littérature disponible publiquement⁷.
15. Une classification peut être opérée en fonction du service offert. L'intervention de l'IA peut être très variable en fonction des applications. Les grandes familles pourraient être organisées comme suit, à titre illustratif :
 - Moteur de recherche de jurisprudence avancé,
 - Résolution de litiges en ligne,
 - Aide à la rédaction d'actes,
 - Analyse (prédictif, barèmes).
 - Catégorisation des contrats selon différents critères et détection de clauses contractuelles divergentes ou incompatibles,
 - « Chatbots » de renseignement du justiciable ou de support de celui-ci dans sa démarche litigieuse.
16. La **Lettonie** a déclaré explorer les possibilités de l'apprentissage machine (*machine learning*) pour des besoins d'administration de la justice. Il s'agit essentiellement de traiter les statistiques des juridictions pour dresser de manière prévisionnelle des estimations d'évolution des moyens (humains, financiers) à allouer.
17. D'autres activités menées par des legaltechs n'ont pas été intégrées dans cette classification car ne relevant peu ou pas du tout d'un traitement par de l'intelligence artificielle : certains sites proposent ainsi de l'accès à de l'information juridique, des solutions « cloud », de la signature électronique, ...

⁶ Voir les algorithmes COMPAS ou les outils tels que RAVEL LAW ou l'agent conversationnel ROSS

⁷ Voir la bibliographie sommaire en Annexe IV – des apports substantiels proviennent de Benoît Charpentier (https://media.wix.com/ugd/c21db1_14b04c49ba7f46bf9a5d88581cbda172.pdf) et de Emmanuel Barthe (<http://www.precisement.org/blog/Intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html#nb14>), ainsi que de Giuseppe Contissa et Giovanni Sartor (<https://rm.coe.int/giuseppe-contissa-presentation-study-session-artificial-intelligence-2/16808b7586>)

18. Un aperçu non exhaustif de services juridiques utilisant de l'IA de manière opérationnelle apparaît ci-dessous.

Logiciel	Etat	Type
Doctrine.fr	France	Moteur de recherche
Prédictice	France	Analyse (hors pénal)
Case Law Analytics	France	Analyse (hors pénal)
JurisData Analytics (LexisNexis)	France	Moteur de recherche, Analyse (hors pénal)
Luminance	Royaume-Uni	Analyse
Watson/Ross (IBM)	Etats-Unis	Analyse
HART	Royaume-Uni	Analyse (pénal, risque de récidive)
Lex Machina (LexisNexis)	Etats-Unis	Analyse

2. Etat des lieux des politiques publiques d'*open data* des décisions judiciaires dans les systèmes judiciaires des Etats membres du Conseil de l'Europe

La disponibilité de données est une condition essentielle au développement de l'IA, permettant à cette dernière d'effectuer certaines tâches qui étaient autrefois menées par l'homme de manière non automatisée. Plus les données sont disponibles, plus l'IA est à même d'affiner des modèles de prédiction. L'*open data* des décisions judiciaires est donc le préalable à l'activité des legaltechs spécialisées dans les moteurs de recherche ou l'analyse de tendances (justice dite prédictive).

Le traitement de ces données pose un certain nombre d'enjeux, tels que la transformation de la formation de la jurisprudence ou la protection des données personnelles (notamment le nom des professionnels).

19. Les données collectées par l'informatique seraient actuellement le « pétrole » du XXIème siècle, puisque leur exploitation et leur croisement produiraient une toute nouvelle richesse. Même si cette assertion est contestée par certains acteurs ou auteurs, les succès de l'industrie numérique à l'échelle planétaire durant les dernières décennies confirment l'énorme potentiel de croissance de ce domaine d'activité.
20. La quantification des activités humaines, désormais à une échelle planétaire, ne pouvait laisser échapper les données produites par le secteur public. C'est dans ce cadre que le mouvement d'ouverture des données publiques s'est opéré, en s'appuyant sur des impératifs beaucoup plus anciens, fondateurs de nos Etats de droit.
21. Le développement majeur de ces dernières années provient de l'ouverture de données publiques téléchargeables (*open data*), notamment dans le cadre du « Partenariat pour un gouvernement ouvert » (« PGO » ou *Open Government Partnership* «OGP»). Le PGO est une organisation non gouvernementale réunissant près de 70 Etats membres (dont de nombreux Etats membres du Conseil de l'Europe) avec des représentants de la société civile et des géants du numérique. Cette ouverture vise à améliorer la transparence de l'action publique, à inciter les citoyens à élaborer et évaluer les politiques publiques et à garantir l'intégrité de l'action publique, ainsi que celle de ses agents, par le traitement de masses considérables d'informations structurées en bases de données (*big data*).

2.1 Définition de l'*open data* des décisions judiciaires

22. Redéfinissons tout d'abord la notion d'*open data* avant de traiter de la question de l'impact de l'ouverture des données sur l'activité judiciaire. **Tout d'abord, il existe une fréquente confusion entre accès à l'information et accès aux données (plus précisément, l'accès à l'information en base de données⁸).**
23. En effet, un certain nombre d'informations publiques, nécessitant une large publicité, bénéficient déjà des technologies de l'information pour assurer leur diffusion : ainsi en France, un site gouvernemental *Légifrance.fr* est le premier canal dématérialisé d'informations publiques certifiées, où l'on trouve non seulement des textes législatifs ou réglementaires, mais aussi de la jurisprudence et des nominations à des emplois publics. Cette information centralisée, bien que disponible sur internet, est cependant totalement distincte de l'accès direct à des données organisées et constituées en base, pouvant être téléchargées et traitées par un ordinateur.

⁸ Les données sont des lettres et des chiffres sans signification. L'information est une donnée incluse dans un contexte. C'est le contexte qui donne un sens aux données. Nous pouvons deviner que 2005 est une année, mais sans contexte, nous ne savons pas. Mais dans « en 2005, nous avons réalisé 3 projets » le contexte donne un sens au nombre. Par conséquent, les « données ouvertes » ne sont pas des données au sens de la définition, mais des informations. De même, des grandes données sont aussi de grandes quantités d'informations, et non des données.

24. **L'open data ne concerne donc que la diffusion de données « brutes » en bases informatiques structurées** ; ces données, agglomérées en tout ou partie avec d'autres sources structurées, constituent ce que l'on qualifie des mégadonnées (*big data*). Le Comité Consultatif de la Convention 108 du Conseil de l'Europe définit le *big data* comme « *la capacité technologique croissante de collecter, traiter et extraire très rapidement des connaissances nouvelles et prédictives à partir d'un gros volume, d'une grande variété de données et à une vitesse considérable. Sous l'angle de la protection des données, les principaux problèmes ne viennent pas uniquement du volume, de la variété des données traitées et de la vitesse du processus, mais également de l'analyse de ces données au moyen d'un logiciel dans le but d'extraire des connaissances prédictives de nature à orienter un processus décisionnel à l'égard de personnes ou de groupes. Aux fins des présentes lignes directrices, la définition des mégadonnées englobe donc à la fois les données elles-mêmes et le procédé analytique* »⁹.
25. **Comme le montre cette définition, on ne doit pas confondre l'ouverture des données avec leurs moyens de traitement.** Nombre de discours sur cette ouverture concernent en réalité les traitements opérés par diverses méthodes avancées que l'on qualifie globalement de *data science*. Justice dite prédictive à l'aide d'intelligences artificielles, moteurs de recherches avancés avec des critères extrêmement précis, robots juristes sont autant d'applications algorithmiques qui sont nourries de données mais qui n'ont rien à voir avec la politique d'ouverture elle-même.
26. Cette politique d'ouverture doit toutefois être interrogée au regard des possibilités qu'elle offre pour des traitements ultérieurs, quelle que soit leur nature. Car si l'on filtre en amont certaines données, par exemple au titre de la confidentialité et du respect de la vie privée, les risques ultérieurs de détournement de finalité semblent pouvoir être réduits.

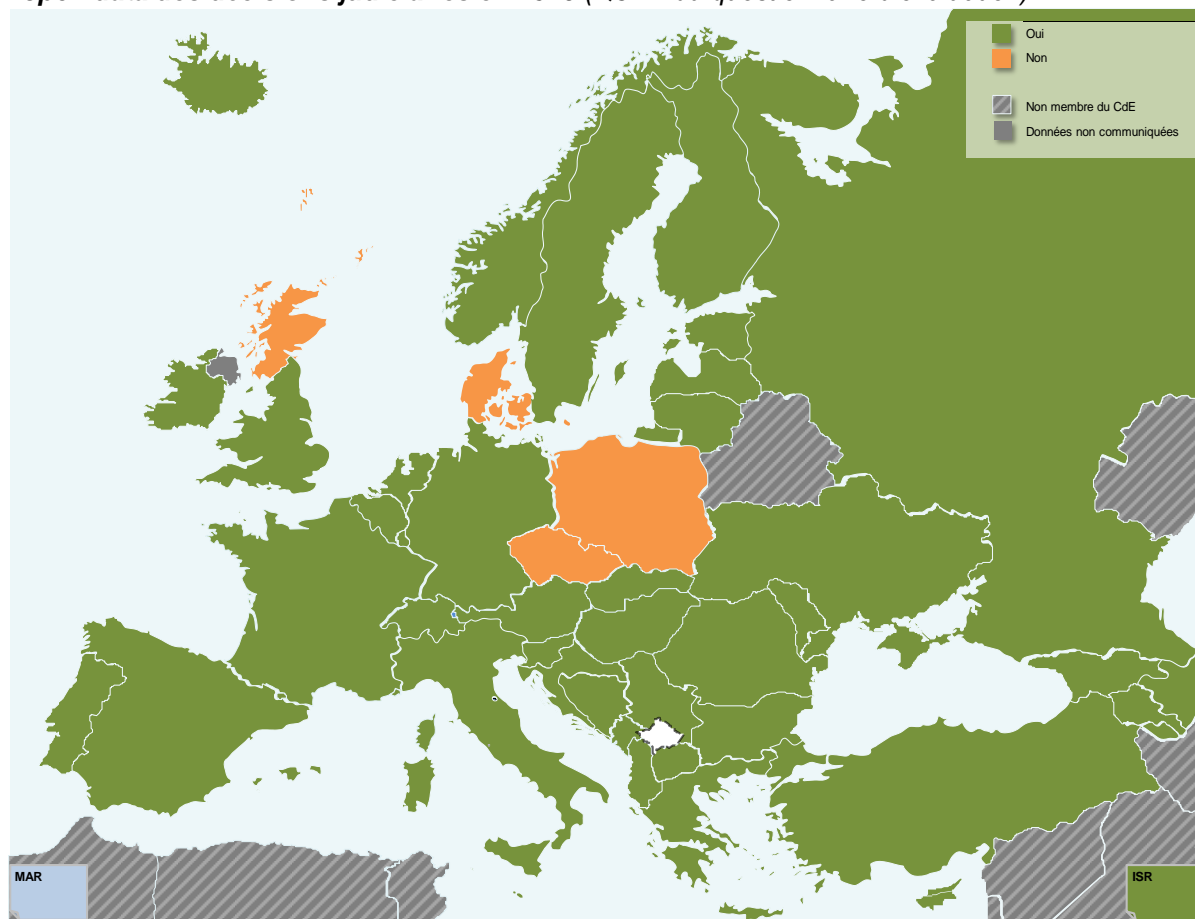
2.2 Etat du développement de l'open data des décisions judiciaires dans les Etats membres du Conseil de l'Europe et conséquences sur le développement de la jurisprudence

27. Où en sont les Etats membres du Conseil de l'Europe en matière d'*open data* des décisions judiciaires ? Le cycle d'évaluation 2016-2018 de la CEPEJ s'est intéressé pour la première fois à la question de l'*open data* des décisions de justice, source de certains traitements basés sur l'IA. La question de l'anonymisation ou de la pseudonymisation des données, au regard du cadre normatif européen de protection des données personnelles¹⁰, a fait l'objet d'une question spécifique afin d'identifier les mesures particulières mises en œuvre par les Etats membres et les observateurs dans ce domaine particulièrement sensible.

⁹ T-PD(2017)1, Lignes directrices sur la protection des données à caractère personnel à l'ère des mégadonnées.

¹⁰ Règlement général de protection des données (RGPD, Règlement de l'Union européenne 2016/679, et Convention n°108 du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données personnelles.

Graphique 1. Tableau illustratif des Etats membres du Conseil de l'Europe ayant introduit l'open data des décisions judiciaires en 2016 (Q62-4 du questionnaire d'évaluation)



28. Sur l'ensemble des Etats et observateurs évalués, seuls 5 ont déclaré ne pas avoir mis en œuvre de politique d'open data des décisions judiciaires en 2016. Si ce taux de réponse est à relativiser au regard de la confusion de certaines réponses assimilant l'accès public à une base de décisions à de l'open data (**Arménie, Belgique, Bosnie-Herzégovine, Espagne, Luxembourg, Fédération de Russie, Turquie**), il révèle d'une part, un souhait de transparence des institutions judiciaires européennes, et, d'autre part, une volonté de beaucoup de pays de rendre publiques les décisions des tribunaux et de rendre ainsi possible ultérieurement leur traitement par des outils d'IA. Cela demande par ailleurs des efforts par les institutions concernées, car un certain nombre de mesures techniques doivent être mises en place à cet effet. En **France** spécifiquement, une certaine partie de la jurisprudence administrative est déjà disponible en téléchargement sur le site data.gouv.fr (voir infra).

29. S'agissant de la protection des données personnelles, 23 pays déclarent procéder à de la pseudonymisation¹¹ au moins de certains contentieux (état des personnes, famille par exemple) en occultant les données rendant identifiables les parties ou des témoins (noms, adresses, numéros de téléphone, numéros d'identité, numéros de compte en banque, numéros fiscaux, état de santé, etc). Ce travail apparaît être placé sous la responsabilité de personnels judiciaires (par exemple **Israël, République de Moldova**) ou d'opérateurs publics (par exemple **Bosnie-Herzégovine, Espagne**). La **Bosnie-Herzégovine** et la **Hongrie** ont en revanche déclaré explicitement publier les noms des professionnels.

¹¹ Selon la définition opérée par le T-PD dans ses « Lignes directrices sur la protection des données à caractère personnel à l'ère des mégadonnées » T-PD(2017)1, la pseudonymisation qualifie un traitement de données à caractère personnel opéré « de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable. »

30. Il existe toutefois une réelle difficulté de mesure de l'impact de l'*open data* sur l'efficacité et la qualité de la justice. Comme indiqué précédemment, l'initiative du réemploi de ces données est en effet essentiellement privée, avec comme cible une clientèle professionnelle (avocats, directions juridiques) et il paraît difficile d'identifier exhaustivement ces réalisations dans le cadre d'une activité seulement intergouvernementale.
31. L'expérience française est représentative des questionnements posés par cette approche et révèle un certain nombre d'enjeux. Il paraît tout d'abord important de souligner que la France a légiféré en 2016 pour imposer un cadre obligatoire de diffusion en *open data* des décisions de justice à ses juridictions.
32. En effet, les articles 20 et 21 de la loi pour une République numérique¹² ont rompu avec la précédente logique de sélection qui disposait¹³ un principe de publicité des « décisions présentant un intérêt particulier rendues par les autres juridictions de l'ordre judiciaire ». Désormais, la loi française pose le principe inverse que tout est publiable, sauf dans des cas précis identifiés par la loi (pour les seules décisions judiciaires) et dans le respect de la vie privée des personnes concernées. Il est toutefois prévu que cette mise à disposition du public des décisions - tant de de l'ordre judiciaire qu'administratif - soit précédée d'une analyse du risque de réidentification des personnes.
33. Les gains d'une large diffusion, promus notamment par la Cour de cassation lors de deux colloques en octobre 2016 et en février 2018, sont nombreux : meilleure connaissance de l'activité judiciaire et des tendances jurisprudentielles, gain de qualité pour une justice qui se sait ainsi observée et création d'une toute nouvelle base de référence de faits.
34. Cette volonté théorique de « discipliner » la jurisprudence par le levier numérique soulève toutefois un certain nombre de questions d'ordre général applicables à l'ensemble des pays qui s'engageraient dans la même démarche. Il convient tout d'abord de rappeler certains principes posés par la Cour européenne des droits de l'homme dans des affaires mettant en cause des divergences de jurisprudence internes : la Cour souligne clairement l'exigence de mise en balance entre l'objectif de sécurité juridique (qui implique la prévisibilité de la décision) et l'intérêt de la « vitalité » dans l'interprétation judiciaire¹⁴.
35. Plusieurs éléments doivent être examinés à cet égard, en premier lieu des éléments techniques :
- a) **La collecte de l'intégralité des décisions judiciaires éligibles à une publication n'est pas nécessairement bien organisée entre tous les degrés de juridictions** : certains des « applicatifs métiers » dans les tribunaux européens n'ont pas été conçus pour cela, notamment s'agissant des décisions de première instance, et cette collecte, si elle se veut exhaustive, imposera de créer dans certains pays des nouveaux processus de collation des jugements ;
 - b) Même si des développements prometteurs sont en cours, **il n'existe pas encore de mécanisme automatisé d'anonymisation a posteriori totalement efficace et à même de prévenir tout risque d'identification ou de réidentification** ;
- Et ensuite, également des éléments de fond :
- c) **La transformation de la logique même de production de la jurisprudence est à considérer** ; Quelle est la valeur de la « norme » issue du nombre des décisions rendues sur une matière déterminée ? Cette « norme » ajoute-t-elle à la loi ? Dans ce cas, s'agit-il d'une nouvelle source de droit ?

¹² Cette loi a été adoptée dans le but de mettre le droit français en conformité avec la directive 2013/37/UE du Parlement européen et du Conseil européen du 26 juin 2013, qui modifiait à son tour la directive du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (dite « directive PSI »).

¹³ Article R433-3 du Code l'organisation judiciaire.

¹⁴ Paroisse gréco-catholique Lupeni et autres c. Roumanie [GC]. n. 76943/11, 29/11/2016, § 116.

- d) **Est-il approprié de demander aux magistrats des efforts supplémentaires de motivation pour expliquer les écarts à cette « moyenne »¹⁵ ?** Cette question n'est pas anodine et ne trouve pas de réponse uniforme dans l'ensemble des systèmes judiciaires. Pour s'en tenir au droit civil français¹⁶, « *le juge tranche le litige conformément aux règles de droit qui lui sont applicables* ». Dans ce contexte particulier français, lui demander de justifier sa décision, en intégrant dans la motivation de cette décision, les raisons pour lesquelles il s'écarte de la prétendue tendance majoritaire de la jurisprudence dans la solution du litige (tout en se conformant aux règles de droit qui lui sont applicables), ne reviendrait-il pas à le faire sortir de son office ? Le juge ne trancherait pas le litige conformément aux seules règles de droit, mais aussi conformément à des tendances jurisprudentielles issues des statistiques élaborées par un outil numérique (qui pourrait par ailleurs être biaisé ou élaboré sans contrôle extérieur par un acteur privé, voir infra 6.1 et 9). La question qui se pose ici va bien au-delà d'une spécificité nationale et renvoie d'une manière plus large à la place des outils d'IA dans le processus décisionnel du juge. Elle n'a pas vocation à remettre en cause la place de la jurisprudence dans les sources du droit ou l'autorité du précédent jurisprudentiel, qui sont des principes acquis au niveau européen¹⁷.
- e) **De plus, la " norme " issue du nombre ne tendrait-elle pas à niveler les décisions de justice, sans les ordonner en fonction de la hiérarchie des juridictions dont elles émanent, ce qui méconnaîtrait la portée des décisions des juridictions suprêmes qui dans beaucoup d'Etats européennes sont les garantes de l'interprétation uniforme du droit ?** Quel serait le rapport de " la norme issue du nombre " avec la jurisprudence de celles-ci ? Devenirait-elle, en étant " issue du nombre " un critère déterminant pour ces cours dans la détermination de leur propre jurisprudence, qu'elles devraient à leur tour justifier lorsqu'elle s'écartent de la " moyenne " ?
- f) **Enfin, n'y a-t-il pas un risque de cristallisation de la jurisprudence, selon une logique reproductive ?** Alors que les décisions judiciaires ont vocation à évoluer en fonction de l'évolution du cadre normatif (national, européen ou international), de la jurisprudence de référence (par exemple des juridictions suprêmes, des cours européennes) ou du contexte socio-économique, la « norme » issue du nombre ne deviendrait-elle pas un standard auquel les magistrats seraient incités à se référer sans questionnement, avec un effet induit d'uniformisation excessive des décisions judiciaires ?
36. **Par ailleurs, on peut s'interroger sur les conséquences pour l'utilisateur, qui est censé être le premier bénéficiaire de la transparence de l'action :** est-ce que celui-ci va tirer réellement un avantage du fait de la publication de la totalité des décisions judiciaires sur internet ou ne risque-t-il pas plutôt de se trouver « noyé » par la masse des décisions, sans en tirer forcément des bénéfices sur le plan qualitatif, faute de disposer des connaissances juridiques et de l'appareil critique pour les analyser ?
37. Paradoxalement, on peut se demander si l'*open data* est en mesure de délivrer une information signifiante puisqu'il permettra certes, par un simple clic sur un lien de télécharger un ensemble considérable de données brutes, mais dont le sens restera totalement abscons pour la majorité des citoyens.
38. Prêter des vertus de transparence et d'information à l'*open data*, c'est donc d'abord espérer que des tiers puissent s'en emparer pour les analyser puis les valoriser. Un modèle économique dans lequel les données publiques de jurisprudence, faisant l'objet d'un travail soutenu par les autorités judiciaires afin de les rendre techniquement « lisibles » par l'IA et conformes aux exigences légales de protection des données personnelles, seraient remises gratuitement au secteur privé et revendues ensuite par celui-ci aux juridictions, aux professionnels et aux citoyens, mérite donc d'être questionné, voire de faire l'objet de critiques.

¹⁵ Eloi Buat-Menard et Paolo Giambiasi, « La mémoire numérique des décisions judiciaires », *Recueil Dalloz*, 2017, p. 1483.

¹⁶ Article 12 du Code de procédure civile français.

¹⁷ A titre d'exemple, la CJUE a affirmé, à propos des conditions d'engagement de la responsabilité d'un Etat membre du fait du contenu d'une décision d'une juridiction nationale suprême, qu'une « violation du droit communautaire est suffisamment caractérisée », « lorsque la décision concernée est intervenue en méconnaissance manifeste de la jurisprudence de la Cour en la matière » (ECJ, affaire C-224/01, Köbler, § 56).

2.3 La protection des données à caractère personnel dans les politiques d'ouverture des décisions judiciaires

2.3.1. *Le nom des parties et des témoins*

39. Afin de tendre à l'ère numérique vers un juste équilibre entre la nécessaire publicité des décisions de justice et le respect des droits fondamentaux des personnes concernées, qu'elles soient parties ou témoins, leurs noms et leurs adresses ne doivent pas apparaître dans les décisions publiées, notamment au regard du risque de détournement et de réutilisation de ces informations personnelles, ainsi que de la sensibilité particulière des données susceptibles d'être contenues dans certaines décisions¹⁸. L'occultation de ces informations peut se faire de manière systématique au moyen de la mise en œuvre de processus automatisés.
40. D'autres informations « identifiantes » peuvent aussi être occultées (par exemple, numéros de téléphone, adresses électroniques, dates de naissance, prénoms des enfants, prénoms rares, surnoms, lieu-dit). Cette occultation revient, au regard des principes de protection des données personnelles, à une simple pseudonymisation des données, et non à une complète anonymisation. En effet, le volume et la variété d'informations qui sont contenues dans les décisions de justice, alliés à la facilité grandissante de procéder à des croisements avec d'autres bases de données, rendent en pratique impossible de garantir que la personne concernée ne soit pas ré-identifiée. En l'absence d'une telle garantie, ces données ne peuvent être qualifiées d'anonymes et relèvent donc du régime de protection des données personnelles.
41. Le caractère sensible de certaines données personnelles mérite une attention particulière, comme cela est prévu à l'article 6 de la Convention 108. Il en va ainsi des données qui révèlent l'origine ethnique ou raciale, les opinions politiques, l'appartenance syndicale, les convictions religieuses ou autres convictions, ainsi que la santé (physique ou mentale) ou la vie sexuelle, ce qui relève de la sphère la plus intime de la personne concernée.
42. Les décisions de justice peuvent contenir d'autres types de données personnelles, très variées, qui entrent dans cette catégorie des données sensibles. Une autre catégorie de données sensibles intéresse particulièrement l'activité pénale des juridictions. Il s'agit des données concernant les procédures et les condamnations pénales. L'ensemble de ces données sensibles mérite donc une vigilance particulière. Leur diffusion en masse présenterait de sérieux risques de discrimination, de profilage¹⁹ et d'atteinte à la dignité humaine.

2.3.2. *Le nom des professionnels, notamment celui des juges*

43. De manière évidente, la connaissance de la formation de jugement a toujours été un élément d'anticipation essentiel pour les avocats, pour qui connaître son juge est parfois presque aussi important que de connaître la loi. Ceux-ci s'efforcent depuis longtemps de réaliser des rapprochements et des comparaisons entre les formations de jugement, de manière plus ou moins empirique, afin de mieux conseiller leurs clients ayant affaire à tel ou tel juge ou formation de jugement.
44. Cette méthode était suffisante lorsqu'un avocat n'intervenait que devant un nombre limité de tribunaux, mais la levée progressive dans de nombreux pays des monopoles locaux du barreau, de même que la liberté d'installation au sein de l'Union européenne légitiment que tout avocat national et même européen puisse désirer connaître finement la jurisprudence de chaque juridiction nationale ou européenne, devant laquelle il est susceptible de plaider.

¹⁸ Voir CEDH, *Z. c. Finlandi* no. 22009/93, §§ 95 et suivants ainsi que le livre vert sur l'information émanant du secteur public dans la société de l'information de la Commission européenne COM(1995)585 (« *Si des précautions particulières ne sont pas prises, les bases de données jurisprudentielles, qui sont des instruments de documentation juridique, peuvent devenir des fichiers d'information sur les individus si ces bases de données sont consultées pour obtenir une liste des décisions de justice sur un individu spécifique plutôt que pour s'informer sur la jurisprudence* »)

¹⁹ Le profilage est défini à l'article 4 du RGPD. Il s'agit d'un traitement utilisant les données personnelles d'un individu en vue d'analyser et de prédire son comportement ou sa situation, comme par exemple déterminer ses performances au travail, sa situation financière, sa santé, ses préférences, ses habitudes de vie, etc.

45. Il n'est donc pas exclu, dans l'avenir, que des applications à forte valeur ajoutée, et donc forcément onéreuses, de *machine learning* puissent avoir une efficacité bien supérieure à l'expérience et au « bon sens » contentieux des avocats travaillant leurs dossiers de manière traditionnelle. L'utilisation de ces applications pourrait ainsi encore plus accentuer la distorsion de concurrence et une inégalité des armes entre des cabinets d'avocats ayant eu ou non recours de manière onéreuse à des tels logiciels d'analyse « prédictive » des jurisprudences.
46. Au prix d'un tel avantage compétitif, les risques sont réels de détournement du principe du juge naturel établi par la loi. En effet, les possibilités de « profilage du juge » dérivant du croisement des données publiques et privées pourraient permettre à des entreprises privées, et à leurs avocats, de se livrer à des pratiques encore plus accrues de *forum shopping*. Ce phénomène est déjà observé depuis longtemps aux Etats Unis, mais aussi en France pour le contentieux des délits de presse et d'atteintes à la vie privée par voie de presse, à l'occasion duquel le demandeur choisit déjà le tribunal qui lui paraît accorder les montants les plus élevés de dommages et d'intérêts.
47. Par ailleurs, un grand nombre de discours confondent l'*open data* avec les impératifs de publicité pesant sur un certain nombre d'informations publiques : il est ainsi parfois avancé que le nom des professionnels devraient obligatoirement figurer dans les données ouvertes au regard des impératifs de publicité et de transparence.
48. Or la mise à disposition de données informatiques de jurisprudence est totalement autonome des principes de publicité des originaux ou les copies certifiées des décisions. L'objectif de l'*open data* est de permettre un traitement automatisé de la jurisprudence au titre d'une licence d'exploitation non-onéreuse. Comme indiqué précédemment, cette mise à disposition s'effectue sous la forme d'une base de données informatique complète, peu claire et non compréhensible directement par le citoyen.
49. Cette mise à disposition ne permet évidemment pas de répondre à l'impératif de publicité du nom des professionnels ayant concouru à une décision spécifique. Rappelons qu'en droit, la mention du nom des juges dans les décisions de justice est une obligation commune aux Etats membres, en lien avec le principe de publicité de l'article 6§1 de la Convention européenne des droits de l'homme, qui vise à assurer de l'impartialité objective de juges (lesquels doivent être identifiables, régulièrement nommés et affectés aux fonctions qu'ils exercent) et du respect des règles de forme (publicité, collégialité par exemple)²⁰.
50. La réponse à la question de la légitimité de la publication du nom des professionnels²¹ en *open data* n'est donc pas à rechercher sur le terrain de l'obligation de la publication du nom des professionnels dans les décisions. Il semble plutôt que l'enjeu se situe dans la conciliation d'impératifs souvent opposés : d'une part, la transparence de l'action publique, en permettant aux citoyens de connaître et « d'évaluer » leurs juges, et, d'autre part, la protection de la vie privée des professionnels (dont la fonction ne saurait diminuer leurs garanties fondamentales dans ce domaine). Des défis se posent ainsi sur le terrain des garanties d'impartialité du juge, voire de l'institution judiciaire dans son ensemble, que les politiques d'*open data* viseraient à renforcer²² : comment préserver celles-ci *in concreto* des possibles tentatives de déstabilisation qui dériveraient du croisement des données nominatives des juges présentes dans la base de données avec d'autres sources (réseaux sociaux, sites marchands), et ceci dans le but de faire émerger d'hypothétiques biais politiques, religieux, etc ?
51. Ces questions ne se posent pas dans les mêmes termes partout en Europe et dépendent des spécificités du système judiciaire concerné (et de la nature de l'organe de gestion des carrières des juges), du caractère collégial ou non du jugement, et du niveau d'instance concerné. En Suisse par exemple où les juges sont élus, la publication est un gage de transparence et de

²⁰ CEDH, Vernes c. France, n. 30183/06 en ce qui concerne l'identification des juges ayant rendu la décision, Pretto c. Italie, n°. 7984/77, Grande Chambre, pour la publicité du jugement, Kontalexis c. Grèce, n. 59000/08, § 38, DMD GROUP, a.s., v. Slovakia, n°. 19334/03, § 66, Miracle Europe KFT v. Hungary, n° 57774/13, § 58 pour le droit à un tribunal établi par la loi en conjonction avec le droit à un juge impartial.

²¹ Ces questions peuvent aussi s'appliquer aux Rechtspfleger qui rendent des décisions judiciaires et aux greffiers, assistants du juge mentionnés dans la composition de la formation de jugement (quoique en moindre mesure).

²² Voir CEDH Previti c. Italie, n° 45291/06, §§ 249 et suivants qui rappelle les principes en matière d'impartialité objective du juge.

responsabilité sociale des juges vis-à-vis des citoyens ainsi que des groupes politiques. Ces informations sont d'ailleurs déjà disponibles dans les moteurs de recherche en ligne (qui ne sont pas à proprement parler de *open data*)²³.

52. Ces questions ne se posent également pas dans les mêmes termes selon les degrés de juridiction. L'intérêt de caractériser la jurisprudence de juges d'instances inférieures n'est peut être pas le même que pour des juges de Cours suprêmes ou de cours internationales. La Cour européenne des droits de l'homme autorise par exemple la recherche d'arrêts par le nom des juges de la formation de jugement, mais sans pour autant permettre le calcul de des statistiques concernant un juge en particulier²⁴. En revanche, dans les pays où les organes juridictionnels ne connaissent pas la pratique (existante dans cette juridiction internationale) de l'opinion dissidente, il peut paraître illégitime d'imputer à un juge, peut être mis en minorité lors du délibéré d'une formation collégiale, la responsabilité personnelle d'une telle décision.
53. Ces débats ont été bien circonscrits par une mission d'étude conduite en France par le professeur Loïc Cadiet. Les conclusions de la mission restent mesurées puisqu'elle ne recommande pas d'en prohiber le principe mais de réserver cette publication à certains types de contentieux et de l'exclure pour d'autres (pour les matières pénales spécialisées par exemple). La possibilité de publier seulement le nom des magistrats de la Cour suprême a été avancée, tout en soulignant la possibilité « d'aller sans retour » qu'une telle politique constituerait.
54. Un simple principe de précaution pourrait donc être appliqué en l'état afin d'évaluer, par type de contentieux et degré de juridiction, l'intérêt de la publication du nom des professionnels en base téléchargeable. De même, la possibilité d'exploiter cette information hors *open data*, par les institutions judiciaires même ou des tiers habilités, à des fins de connaissance de la jurisprudence n'est pas à omettre.
55. Une telle publication resterait en toute hypothèse à examiner au regard des règles européennes applicables en matière de protection des données personnelles, telles que celles de la Convention n°108 ou de l'Union européenne (RGPD et Directive 680/2016).

²³ Voir l'exemple du tribunal fédéral suisse, dont la jurisprudence est téléchargeable (<https://www.bger.ch/fr/index/juridiction/jurisdiction-inherit-template/jurisdiction-recht.htm>) ou des cantons (canton de Genève par exemple <http://ge.ch/justice/dans-la-jurisprudence>).

²⁴ Les décisions de la Cour européenne des droits de l'homme sont collégiales. Les publications comportent les éventuels avis dissidents.

3. Caractéristiques de fonctionnement de l'intelligence artificielle (apprentissage machine) appliquée aux décisions judiciaires

Le traitement du langage naturel et l'apprentissage machine sont les deux techniques au cœur du traitement des décisions judiciaires à l'aide de l'intelligence artificielle.

Dans la plupart des cas, l'objectif de ces systèmes n'est pas de reproduire un raisonnement juridique mais d'identifier les corrélations entre les différents paramètres d'une décision (par exemple dans un litige de séparation d'un couple : la durée de mariage, les revenus des époux, l'existence d'un adultère, le montant de la pension alimentaire prononcée, etc) et, en s'appuyant sur l'apprentissage machine, d'en déduire un ou plusieurs modèles, sur lesquels s'appuyer afin de « prédire » ou « prévoir » la future décision judiciaire.

Ce chapitre vise à apporter des éclaircissements sur la nature intrinsèque des logiciels qualifiés de « prédictifs » - parfois des versions « bêta », c'est-à-dire en phase de développement ultérieur - en évoquant tant leurs possibilités que leurs limites. Il visera à vulgariser les algorithmes dits d'apprentissage machine (*machine learning*) qui sont au cœur de l'analyse automatisée de la jurisprudence.

3.1 Les fonctionnalités théoriques d'un logiciel de justice dite prédictive

56. A titre d'introduction, il convient de revenir brièvement sur les fonctionnalités promises par ces logiciels dits prédictifs. Ils proposent d'établir des probabilités de succès (ou d'échec) d'une procédure devant un tribunal. Ces probabilités sont établies par la modélisation statistique des décisions antérieures via des méthodes provenant de deux domaines spécifiques de l'informatique : traitement de langage naturel (*natural language processing*) et apprentissage machine. Ces approches de modélisation statistique sont souvent qualifiées d'intelligence artificielle (IA) ; en réalité, il s'agit d'IA « faibles » (voire le glossaire)
57. Il doit être immédiatement précisé que le terme d'intelligence artificielle (IA) est débattu par les experts car il conduit à de nombreuses ambiguïtés. Le terme IA est désormais entré dans notre registre usuel de langage pour désigner un assemblage de sciences et de technologies assez diverses et qui permettent à des ordinateurs de battre les meilleurs champions du jeu de Go²⁵, de conduire une voiture, de converser avec les humains. Les chercheurs préfèrent pour leur part désigner les différentes applications par les technologies exactes qui les sous-tendent, dont l'apprentissage machine, et qualifie parfois l'ensemble de ces moyens d'IA très spécialisés comme des IA « faibles » (ou « modérées »). Ce qui les distingue d'un but ultime, encore totalement théorique, qui serait de créer une IA « forte », c'est-à-dire une machine auto-apprenante en capacité d'appréhender automatiquement le monde en général, dans l'ensemble de sa complexité.
58. Spécifiquement à la justice, les systèmes de justice dite prédictive ont pour ambition d'être utilisés par des directions juridiques, des assureurs (tant pour leurs besoins internes et que pour leurs assurés), ainsi que des avocats afin d'anticiper l'issue d'un contentieux. Ils auraient aussi vocation à assister les magistrats dans leurs prises de décisions.
59. Ils fournissent sous forme de représentation graphique des probabilités de succès pour l'issue d'un contentieux, en fonction de critères saisis par l'utilisateur (spécifiques pour chaque type de contentieux). Ces systèmes prétendent être en capacité de calculer les montants probables d'indemnisation par les tribunaux.

²⁵ <https://www.nature.com/articles/nature16961>

3.2 Le fonctionnement concret des intelligences artificielles : des machines statistiques construisant des modèles sur la base du passé

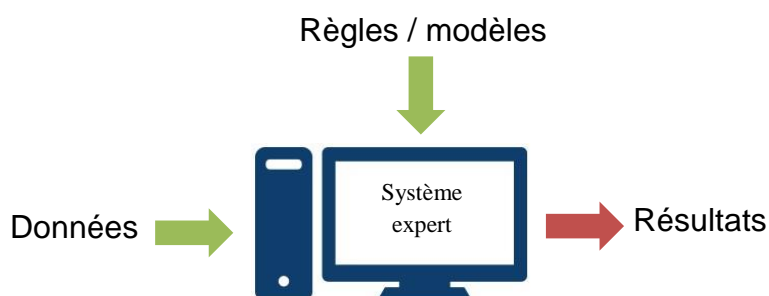
60. Une distinction paraît d'emblée devoir être opérée entre ce qui relève de la « *prédiction* » et ce qui relève de la « *prévision* ». La prédiction est l'acte d'annoncer (*prae*, avant - *dictare*, dire) par avance des événements futurs (par inspiration surnaturelle, par voyance ou prémonition). La prévision, quant à elle, résulte de l'observation (*visere*, voir) d'un ensemble de données afin d'envisager une situation future. Cet abus de langage et sa diffusion paraissent pouvoir être expliqués par un transfert du terme en provenance de l'univers des sciences « dures », où le terme fait référence à une variété de techniques de « *data science* » issues des mathématiques, des statistiques et de la théorie des jeux qui analysent des faits présents et passés pour réaliser des hypothèses sur le contenu d'événements futurs.
61. Il convient également de noter que la logique du fonctionnement d'un logiciel de justice prédictive repose essentiellement sur des méthodes génératives (communément appelées bayésiennes) ou discriminatoires qui tentent d'estimer la fourchette actuelle ou future des valeurs d'une variable (par exemple, le résultat d'un procès) à partir de l'analyse des exemples passés.
62. D'une manière générale, il convient également de garder à l'esprit « l'idée anthropomorphe que les machines calculatoires seraient intelligentes et que leurs concepteurs seraient parvenus à glisser un esprit à l'intérieur de leurs mécanismes »²⁶. Cette idée imprègne malheureusement encore de nombreuses analyses sur la justice dite prédictive qui prêtent à ces dispositifs des capacités immédiates ou futures de quasi réplication de l'intelligence humaine. Ce contexte, nourri chaque jour d'un lot de progrès révolutionnaire des IA²⁷, nous porte ainsi tous à aborder ces outils « prédictifs » avec une certaine part, consciente ou non, de mysticisme, en énonçant parfois que ce qui n'est pas tout à fait possible aujourd'hui le sera nécessairement demain.
63. En effet, les IA dites « fortes » de la littérature de science-fiction n'existent pas. Ce type d'IA, qui seraient douées non seulement d'intelligence mais aussi de conscience, demeurent purement fictionnel. Les systèmes d'apprentissage machine actuellement en cours de développement sont décrits comme des IA « faibles » et sont capables d'extraire des tendances complexes et d'apprendre à partir de grandes quantités de données de manière efficace et souvent avec des niveaux élevés de précision prédictive.
64. Pour dépasser toute considération instinctive ou fantasmée, ces technologies de traitement et d'analyse de l'information sont à appréhender autour de trois concepts :
- **l'IA n'est pas un objet unique et homogène** : il s'agit en réalité d'un assemblage de sciences et techniques (mathématiques, statistiques, informatique) en capacité de traiter des données pour concevoir des tâches très complexes de traitement informatique ;
 - **le moteur de l'IA ne produit pas de l'intelligence en soi mais fonctionne par une approche inductive** : l'idée est d'associer de manière plus ou moins automatisée un ensemble d'observations (entrées) à un ensemble de résultats possibles (sorties) à l'aide de diverses propriétés préconfigurées ; spécifiquement pour la « justice prédictive » le moteur construit des liens entre les différents groupes lexicaux composant les décisions judiciaires. Ces groupes sont corrélés entre ceux identifiés en entrée (les faits et la motivation) et ceux en sortie (le dispositif de la décision) et classifiés ;
 - **la fiabilité du modèle (ou fonction) construit dépend fortement de la qualité des données utilisées et du choix de la technique d'apprentissage machine**
65. Il convient de revenir brièvement sur les ancêtres de ces systèmes, les systèmes experts (SE), qui s'appuyaient pour leur part sur des règles de traitement écrites par un informaticien. Les SE ont connu un rapide développement à la fin des années 80 et dans les années 90, surtout dans

²⁶ Dominique Cardon, *A quoi rêvent les algorithmes, nos vies à l'heure des big data*, La République des idées, Editions du Seuil, 2015, p. 58.

²⁷ « AlphaZero : l'IA de DeepMind apprend et maîtrise le jeu d'échecs en moins de 4 heures », *Génération NT*, article publié le 8 décembre 2017, [En ligne], <https://www.generation-nt.com/alphazero-ia-deepmind-maitrise-echecs-4-heures-actualite-1948919.html> (page consultée le 14 décembre 2017).

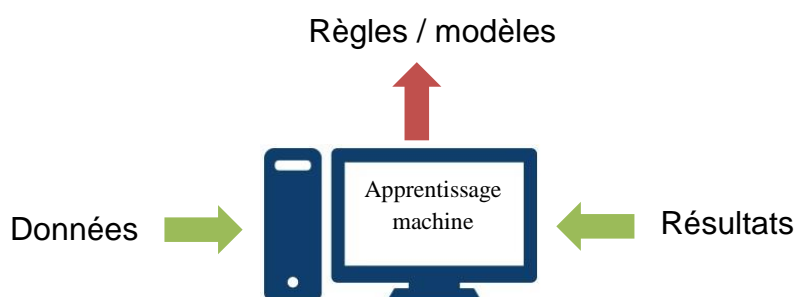
les domaines de la médecine et de la finance²⁸. Ces systèmes étaient capables de répondre à des questions spécialisées et effectuaient un raisonnement à partir de faits connus, en exécutant des règles d'encodage prédéfinies dans un « moteur ». Malgré le succès de SE comme *Deep Blue* face à Garry Kasparov lors d'une série de parties d'échecs en 1997, ces systèmes ont fini par échouer notamment parce qu'ils étaient dans l'incapacité d'interpréter « l'infinie variété des situations et des contextes »²⁹ et devenaient peu efficaces au-delà de 200 à 300 règles d'encodages, tant en terme de performance d'exécution que de maintenance (le raisonnement suivi par le système devenait quasiment impossible à appréhender pour ses concepteurs).

Graphique.2 : Les anciens systèmes experts étaient programmés avec des règles reproduisant la logique du raisonnement juridique



66. Aujourd'hui, l'idée n'est plus d'écrire des règles de raisonnement en miroir du raisonnement humain, comme pour les anciens systèmes experts³⁰, mais de laisser les systèmes d'apprentissage machine procéder eux-mêmes à l'identification de modèles statistiques existants dans les données et les mettre en correspondance avec des résultats spécifiques.

Graphique 3 : L'apprentissage machine produit seul des modèles en recherchant automatiquement des corrélations entre données et résultats.



67. Ce faisant, ces systèmes ne reproduisent pas et ne modélisent pas un raisonnement (par exemple juridique). A titre d'exemple les actuels traducteurs en ligne ne réalisent pas un raisonnement abstrait. Ils infèrent une estimation probable de la meilleure correspondance entre des groupes de lexicaux avec des traductions déjà réalisées. Les actions des utilisateurs

²⁸ Il peut être pris l'exemple du « High Frequency Trading (HFT) » ou transaction à haute fréquence, qui est un type transaction financière effectué à grande vitesse par des logiciels se basant sur des algorithmes. En 2013 et en Europe, 35 % des transactions étaient déjà réalisées avec le HFT. Aux USA, le HFT représente 70 % des volumes échangés sur le marché des actions. Au vu de l'automatisation des transactions, la chute est encore plus rapide en cas de krachs, comme ce fut le cas durant la crise financière de 2007.

²⁹ Hubert Dreyfus, *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, The MIT Press, 1992 cité par Dominique Cardon in "A quoi rêvent les algorithmes, nos vies à l'heure des Big Data", p.59.

³⁰ Les règles de traitement écrites pouvaient s'appuyer sur du code spécifique, dans différents langages de programmation tels que le LISP ou sur des éditeurs permettant de modéliser des règles. Voir par exemple Radboud Winkels, « CLIME : Un projet de développement de serveurs juridiques intelligents », dans Danièle Bourcier, Patricia Hasset et Christophe Roquilly (dir.), *Droit et intelligence artificielle*, 2000, Romillat, p. 59.

contribuent évidemment à l'amélioration de la recherche de correspondance, mais l'algorithme d'apprentissage machine ne réalise pas à proprement parler une traduction en comprenant le sens des phrases traitées³¹.

68. Cet exemple montre la différence d'approche qui peut exister entre les juristes et les chercheurs : un juriste cherchera à comprendre l'effectivité au regard de règles préexistantes, ce que le chercheur ne pourra lui apporter qu'au travers du langage de sa science et de sa capacité à interpréter les milliers de paramètres issus d'un apprentissage machine.
69. En effet, l'apprentissage machine est un domaine des sciences informatiques dans lesquels les programmes des ordinateurs apprennent de l'expérience. Ces algorithmes font réaliser à une machine un processus d'entraînement, tel un enfant qui apprendrait dans son environnement. Sommairement, ces techniques d'apprentissage peuvent être supervisées ou non par un humain. La catégorie d'apprentissage machine la plus en vue est actuellement celle dite d'apprentissage non supervisé par renforcement : la machine récolte seule des « récompenses » virtuelles si l'action, dans un environnement donné, produit le résultat attendu. Les méthodes d'apprentissage machine qui incluent les réseaux de neurones (ou leur version plus complexe dite d'apprentissage profond - *deep learning*), ont été assez largement médiatisées au vu de leur autonomie et de leurs applications assez saisissantes comme celle parvenant à obtenir des scores élevés avec d'anciens jeux vidéo sous Atari 2600³² (sur la seule base du positionnement des pixels et des scores). D'autres applications concrètes de ces technologies affectent déjà notre quotidien et commencent à poindre dans le monde professionnel de la justice³³.
70. Toutefois, ces algorithmes restent très spécialisés pour une tâche et présentent des problèmes de discernement face à des situations chaotiques ou comportant un nombre insuffisant de données pour permettre une « prédiction » (comme la compréhension réelle du langage naturel). Dans le domaine des sciences sociales, auquel le droit et la justice appartiennent, l'échec apparaîtrait même inéluctable en absence de modèle convaincant de la cognition. Pour Pierre Lévy, « *l'intelligence artificielle se contente dans les faits de fournir une boîte à outils hétéroclite (règles logiques, syntaxes formelles, méthodes statistiques, simulations neuronales ou socio-biologiques...) qui n'offrent pas de solution générale au problème d'une modélisation mathématique de la cognition humaine* »³⁴. Ainsi, un réel apprentissage « prédictif » devrait en réalité s'appuyer sur une bonne représentation systémique du monde, qui se heurte, pour Yann LeCun, chercheur dans le domaine de l'intelligence artificielle, à une « *question fondamentale scientifique et mathématique, pas à une question de technologie* »³⁵.
71. Par ailleurs, l'originalité des systèmes actuels de traitement des *big data* n'est pas de prétendre reproduire notre modèle de cognition mais de réaliser une statistique des contextes sur une taille inédite de données, sans réelle garantie d'exclure les corrélations dites fallacieuses³⁶.
72. Alors qu'il est communément admis que les modèles statistiques et probabilistes sont améliorés par l'augmentation des données les alimentant, certains mathématiciens ont alerté sur les risques d'augmentation des corrélations dites fallacieuses, (c'est-à-dire des liens entre des facteurs n'ayant absolument aucun lien de causalité) dans le Big Data. Les mathématiciens Cristian Sorin Calude et Giuseppe Longo évoquent même le risque d'un « *déluge de corrélations fallacieuses dans le big data* » : « *Plus une base de données exploitée pour des corrélations est grande, plus grandes sont les chances de trouver des régularités récurrentes et plus grandes sont les chances*

³¹ Li Gong, « La traduction automatique statistique, comment ça marche ? », *Interstices.info*, publié le 29 octobre 2013, [En ligne], https://interstices.info/jcms/nn_72253/la-traduction-automatique-statistique-comment-ca-marche (page consultée le 14 décembre 2017).

³² <https://www.nature.com/articles/nature14236>

³³ Intelligence artificielle d'IBM Watson équipant le domaine médical ou servant de moteur à « Ross », présenté comme un avocat « virtuel » – Ross, « Do more than humanly possible », [En ligne], <http://rossintelligence.com> (page consultée le 14 décembre 2017).

³⁴ Pierre Lévy, « Intelligence artificielle et sciences humaines », *Pierre Levy's blog*, 8 octobre 2014, [En ligne], <https://pierrelevyblog.com/2014/10/08/intelligence-artificielle-et-sciences-humaines/> (page consultée le 30 décembre 2017).

³⁵ Yann LeCun, « Qu'est-ce que l'intelligence artificielle », recherche publiée sur le site du Collège de France, [En ligne], https://www.college-de-france.fr/media/yann-lecun/UPL4485925235409209505_Intelligence_Artificielle_Y._LeCun.pdf (page consultée le 14 décembre 2017).

³⁶ Dominique Cardon, *op. cit.*, p. 60.

de commettre des erreurs »³⁷. Ce qui peut apparaître en effet comme des régularités pour une IA (des liens récurrents entre différentes données, concepts, contextes ou groupes lexicaux) peut s'avérer en réalité aléatoire. Même si l'argument des deux mathématiciens ne doit pas être généralisé trop hâtivement, ceux-ci constatent que dans certains ensembles vastes de nombres, de points ou d'objets apparaissent des motifs réguliers aléatoires et il paraît impossible de les distinguer algorithmiquement des motifs révélant des causalités³⁸.

73. Enfin, les algorithmes, quelle que soit leur sophistication actuelle, demeurent nécessairement réduits aux interactions opérées dans la phase d'apprentissage et, partant, à leur passé. Le contenu et la qualité des flux de données entrant dans la composition des calculs est donc fondamentale pour avoir une compréhension des résultats obtenus et identifier des éventuels biais d'analyse. Les défis sont ici encore nombreux puisque, avec une approche exhaustive, l'analyse de la plus grande quantité possible de données relatives à une activité va produire des résultats dont le sens devra nécessairement être éclairé au regard de tous les facteurs ayant eu une influence. Avec une approche plus restrictive par échantillonnage de données entrantes, il existera aussi des risques au vu des biais propres aux arbitrages nécessaires pour sélectionner une donnée plutôt qu'une autre.

³⁷ Cristian Sorin Calude, Giuseppe Longo, « Le déluge des corrélations fallacieuses dans le big data », dans Bernard Stiegler (dir.), *La toile que nous voulons*, 2017, FYP éditions, p.156.

³⁸ Théorie dite de « Ramsey », Ronald L. Graham, Joel H. Spencer – « Ramsey Theory », *Scientific American*, vol.263, n°1, juillet 1990, p112-117 cité par Cristian Sorin Calude, Giuseppe Longo, « Le déluge des corrélations fallacieuses dans le big data », *ibid.*

4. Les intelligences artificielles peuvent-elles modéliser *a priori* un raisonnement juridique ?

Les intelligences artificielles contournent les difficultés rencontrées avec les anciens systèmes experts : elles ne tentent pas d'être un miroir d'un raisonnement juridique, dont la reproduction ne constitue pas pour elles un objectif. L'apprentissage machine conduit à effectuer des catégorisations entre les différents paramètres identifiés par les concepteurs ou ceux découverts par la machine.

74. Les systèmes juridiques continentaux sont très éloignés de l'idéal de rationalité que devait incarner par exemple le code civil de 1804 en France. Il existe une multitude de sources qui ne s'imbriquent pas de manière parfaitement cohérente et qui portent sur un ensemble de règles dont la signification demeure très indéterminée, ce que le théoricien du droit Herbert L. A. Hart qualifiait de « texture ouverte du droit »³⁹. Les systèmes de Common Law, même réputés plus efficaces « économiquement » car s'adaptant dynamiquement aux nouveaux besoins juridiques⁴⁰, sont également évolutifs et n'offrent pas plus de sécurité juridique. Les règles juridiques ne connaissent donc pas une évolution linéaire, à l'inverse des lois empiriques, celles des sciences dites exactes, où chaque nouvelle règle vient en générale compléter les précédentes et non invalider un ensemble complet de raisonnement.
75. Rappelons que face à cette complexité générale des systèmes juridiques, les anciens systèmes experts informatiques ont rapidement trouvé leurs limites dès lors que l'on atteignait une imbrication de 200 à 300 règles logiques. Le découpage du droit sous forme de « règles de production » n'a pas suffi à donner une représentation valable de l'ensemble des connaissances et méthodes qui guident un juriste.
76. Si la modélisation informatique des systèmes juridiques positifs se révèle en conséquence impossible, il en va de même du raisonnement des juges. Comme l'a mis en lumière la théorie du droit, le raisonnement judiciaire est surtout affaire d'appréciation et d'interprétation, celle des faits prouvés et pertinents du litige, des règles de droit applicables (textuelle ou des précédents), dont la signification demeure, on l'a dit, très indéterminée⁴¹, mais aussi de l'interprétation subjective par les juges de la notion d'équité, qui devrait connaître en Europe de nouveaux développements avec l'exigence encouragée par la Cour européenne des droits de l'homme de « contrôle de proportionnalité ».
77. Xavier Linant de Bellefonds soulignait que, « la complexité du droit tient à son caractère téléologique et contentieux : deux raisonnements cohérents peuvent conduire à des arbitrages différents en fonction de deux ordres d'impératifs⁴². »
78. C'est que le fameux syllogisme judiciaire est davantage un mode de présentation du raisonnement juridique que sa traduction formelle. Celui-ci ne rend pas compte de l'intégralité du raisonnement tenu par le juge, lequel est en réalité ponctué d'une multitude d'éléments décisionnels, de choix, non formalisables *a priori* et parfois relevant de son pouvoir d'appréciation : quels sont les faits pertinents ? Ces faits sont-ils prouvés ? Quelle règle leur est applicable ? Quelle est le sens de cette règle au regard du cas à trancher ? Quelle source faire prévaloir entre une pluralité de sources en conflit ? La cohérence d'ensemble des décisions de justice n'est jamais assurée et relève davantage d'une mise en récit *a posteriori*, celle qu'opère notamment la doctrine juridique ou le juge dans sa motivation, ayant davantage pour fonction de convaincre du bien-fondé de telle ou telle solution que de décrire strictement et objectivement l'intégralité des étapes ayant concouru à la prise des décisions.

³⁹ Herbert L. A. Hart, *Le concept de droit*, Bruxelles, Facultés universitaires Saint-Louis, 1976.

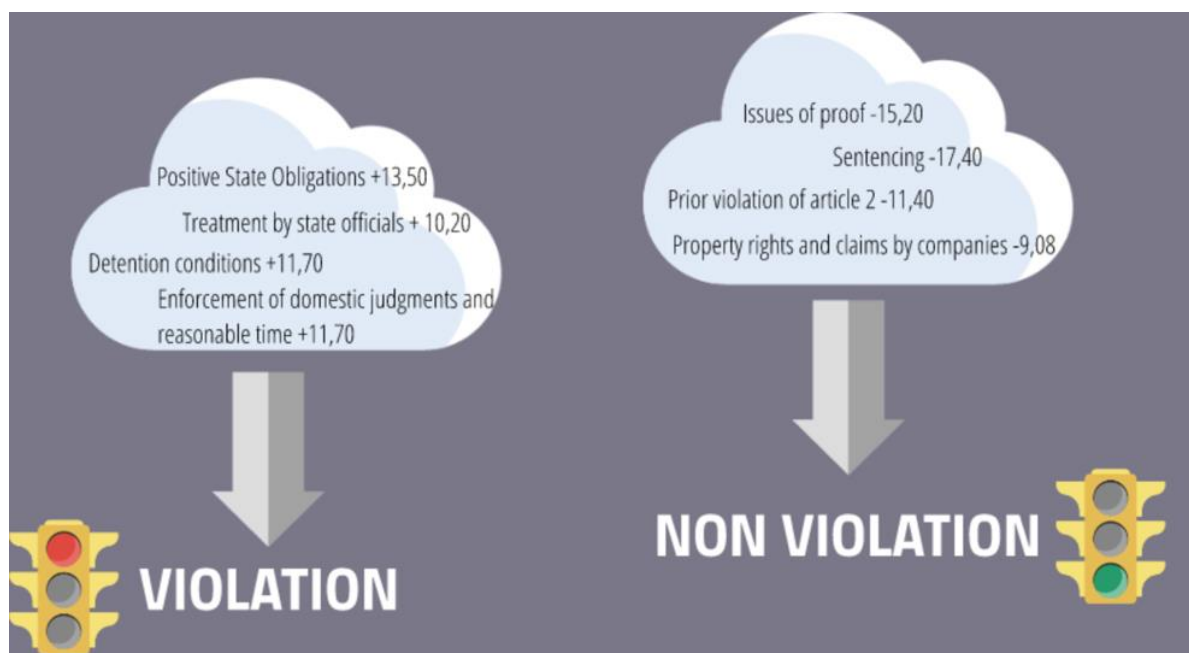
⁴⁰ <https://www.contrepoints.org/2014/08/15/177160-common-law-contre-droit-civil-lexperience-francaise-de-lancien-regime>

⁴¹ Sur ces questions, Michel Troper, *La théorie du droit, le droit, l'Etat*, Paris, PUF, 2001, spéc. p. 69-84.

⁴² Xavier Linant de Bellefonds, « L'utilisation des systèmes experts en droit comparé », *Revue internationale de droit comparé*, 1994, Vol. 46, n° 2, p. 703-718.

79. Or ce travail d'interprétation est précisément ce que ne réalisent pas – et ne veulent pas réaliser – aujourd'hui les techniques d'apprentissage machine, puisqu'elles procèdent, nous l'avons vu, à un traitement automatisé où la corrélation de masses d'information est censée remplacer la compréhension des causalités réelles d'une décision. Celles-ci n'ambitionnent nullement de formaliser le raisonnement juridique mais elles espèrent que les modèles qu'elles capturent puissent anticiper les probables décisions d'un juge face dans des situations similaires.
80. Ces résultats auxquels les IA parviennent sont en réalité étrangers à la question de la conformité, en droit, de telle ou telle solution et ne peuvent discriminer entre des raisonnements conformes ou non en droit.
81. Un examen des travaux de l'University College of London (UCL) sur la jurisprudence de la Cour européenne des droits de l'homme confirme bien ce diagnostic. L'étude de l'UCL partait du postulat qu'un simple modèle d'apprentissage automatique pouvait prédire l'issue d'une affaire avec une précision de 79 % pour ce tribunal particulier. Le modèle issu de l'apprentissage machine s'est avérée plus efficace sur la partie descriptive des faits des décisions étudiées que sur la partie de motivation relative à l'application de la Convention au cas d'espèce⁴³.
82. L'examen des termes corrélés au constat d'une « violation » (avec une pondération positive) et ceux corrélés à une non-violation de la Convention européenne des droits de l'Homme (avec une pondération négative) éclaire sans ambiguïté la mécanique à l'œuvre, qui n'est en rien comparable avec un raisonnement juridique.

Graphique 4 : Illustration des travaux de l'UCL – Poids théorique alloué à des mots ou termes en fonction de leur lien avec des constats de violation ou de non violation de la Convention européenne des droits de l'Homme.



83. Car, précisément, le choix des faits pertinents et leur qualification constituent l'un des éléments de la décision du juge. Autrement dit, l'étude de UCL n'a été en réalité capable de produire une probabilité qu'avec un matériel lexical en grande partie issu du raisonnement et de la motivation du juge et non pas avec celui du requérant, qu'elle a assemblé en fonction des seules fréquences d'apparition. Leur IA a donc établi une forte probabilité de correspondance entre des groupes de mots et une décision d'ores-et-déjà formalisée. Elle ne saurait en aucun cas reproduire seule le raisonnement des juges européens ni surtout prédire une issue, sur la base par exemple d'un

⁴³ Travaux sur un échantillon de 584 décisions de la Cour européenne des droits de l'Homme : Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotjuc-Pietro, Vasileios Lampos, « Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective », publié le 24 octobre 2016, [En ligne], <https://peerj.com/articles/cs-93/> (page consultée le 14 décembre 2017).

récit brut d'un futur requérant devant la Cour de Strasbourg, dont la requête sera soumise à un examen de recevabilité très strict (près de 70 356 requêtes ont été déclarées irrecevables ou rayées du rôle en 2017⁴⁴) reposant en grande partie sur l'application de standards d'appréciation (importance et sérieux du grief, etc.) ménageant un espace décisionnel important.

⁴⁴ Source : Analyse statistique 2017 de la CEDH, janvier 2018.

5. Les IA peuvent-elles expliquer *a posteriori* le comportement des juges ?

L'explication *a posteriori* d'un comportement d'un juge, notamment la révélation de biais nécessiterait d'identifier tous les facteurs potentiellement causatifs au travers d'un cadre interprétatif et de procéder à un travail d'analyse contextualisé. Le fait, statistiquement, que la garde des enfants est plus souvent confiée aux mères qu'aux pères ne démontre pas nécessairement un parti pris des juges mais conduit à devoir mobiliser différentes disciplines issues des sciences sociales pour éclairer ce phénomène.

84. D'un point de vue scientifique, expliquer un phénomène ou, pour ce qui nous concerne, un comportement humain, revient à déterminer les mécanismes de causalité ayant conduit à ce comportement en présence d'un certain nombre de données contextuelles.
85. Cela nécessite, de manière très schématique, la constitution préalable d'un cadre interprétatif, lui-même né de l'observation répétée de la réalisation de ce type d'évènements ou de comportements en présence de certains facteurs ou éléments. Ce cadre interprétatif est constitué par les hypothèses ou points de vue adoptées par les différentes disciplines des sciences sociales. Il s'agit d'une étape supplémentaire d'analyse dont peuvent être nourris les algorithmes, mais à laquelle ils ne procèdent pas seuls.
86. Certaines *legaltechs* sont allées relativement loin et ont pensé pouvoir identifier d'éventuels biais personnels des magistrats et nourrir des suspicions de partialité. La mise en *open data* du nom de certains présidents de formation de jugement des tribunaux administratifs et cours administratives d'appel en France a ainsi permis de construire un indicateur du taux de rejet de recours contre les obligations de quitter le territoire français prises par les autorités administratives. Certains commentateurs avaient alors soutenu avec beaucoup d'enthousiasme que la prétendue impartialité des juges était ainsi « mise à mal par l'intelligence artificielle ».
87. Mais peut-on réellement parvenir à de telles interprétations sur la base d'un traitement algorithmique des décisions de justice ? Caractériser un biais personnel dans la prise de décision d'un magistrat (en dehors naturellement des cas d'expression personnelle et publique de ce magistrat dans l'affaire concernée) consiste à fournir une explication d'un comportement, en l'occurrence d'une décision, par les traits de personnalité, les opinions ou encore la confession d'un juge. Or, on l'a dit, une telle explicitation causale ne peut être purement et simplement déduite du résultat probabiliste fourni par les algorithmes. Cela requiert au contraire un travail d'analyse supplémentaire devant permettre d'isoler, parmi les nombreux facteurs corrélés (dont l'identité de membres de la formation de jugement), ceux véritablement causatifs. Par exemple, le fait qu'un juge aux affaires familiales fixe statistiquement davantage la résidence habituelle d'un enfant chez la mère ne traduit pas nécessairement un parti-pris de ce magistrat en faveur des femmes mais davantage l'existence de facteurs psycho-sociaux, économiques voire culturels propres au ressort de la juridiction, tels que le temps de travail de chacun des parents, le montant de leurs revenus, l'offre locale de garde collective, la scolarisation ou non de l'enfant, la refondation ou non par un des parents d'un nouveau couple ou même tout simplement l'absence d'intérêt d'un des deux parents à s'occuper d'un jeune enfant.
88. De même, on ne peut pas comparer purement et simplement les décisions rendues en matière d'expulsion du territoire par une juridiction administrative située à proximité d'un grand centre de rétention avec celles d'une juridiction qui connaît ce contentieux de manière marginale.
89. D'ailleurs quel que soit le lieu de la juridiction, la question de la « jurisprudence » d'un juge unique de permanence qui ne traite qu'occasionnellement d'un type de contentieux, mais qui s'inspire (ou non) de la jurisprudence de ces collègues est particulièrement intéressante et elle peut poser légitimement la question de l'égalité des citoyens devant la production judiciaire, mais elle relève plus d'une culture judiciaire de la « collégialité » à réinventer ou à préserver que d'un « fichage » ou d'une « stigmatisation » via les outils de *machine learning*.

90. Que déduire, enfin, de la personnalité du président d'une formation de jugement collégiale, alors que son nom est la seule information nominative visible dans les décisions ouvertes de la justice administrative ?
91. Comment comparer également deux approches philosophiques et culturelles distinctes de la décision de justice, lorsque dans certains pays européens ou en France en matière de justice administrative règnent la culture du « précédent » et une connaissance fine par les juges des bases de données factuelles de toutes les décisions de 1^{ère} et de 2^{nde} instance (base Ariane), alors que dans d'autres pays ou systèmes sont privilégiés l'indépendance « intellectuelle » de chaque juridiction, de même que la volonté de traiter au « cas par cas » chaque situation ?
92. Fournir une véritable explication d'une décision de justice nécessite donc d'analyser beaucoup plus finement les données contingentes de chaque affaire et les règles de droit applicables, plutôt que de nourrir le vain espoir que de la masse des liens nait le sens.

6. Quelles applications pour la justice civile, commerciale et administrative ?

L'état de développement des techniques d'apprentissage machine ne permet pas à ce jour de parvenir à des résultats fiables s'agissant de la « prédiction » des décisions judiciaires. En revanche, leur application dans le champ de la justice civile, commerciale et administrative est à considérer pour la constitution de barèmes ou la résolution précontentieuse de litiges en ligne, dès lors qu'un recours ultérieur au juge demeure possible.

93. On le voit, la question que pose en premier lieu une telle utilisation des intelligences artificielles ne porte pas tant sur le fait de savoir si elles sont bénéfiques ou néfastes, souhaitables ou non, mais sur celui de savoir si les algorithmes que l'on nous propose peuvent parvenir ou non au type de résultat recherché. Les conclusions de l'expérimentation conduite dans les cours d'appel de Douai et de Rennes en France démontrent bien qu'en présence d'un discours séduisant de promotion d'un produit présenté comme né de l'IA peut cacher des défauts rédhibitoires de conception et des résultats d'analyse totalement erronés.

94. Indépendamment de la qualité du logiciel testé, l'anticipation de la décision du juge apparaît cependant, en matière civile, commerciale et administrative, comme une opportunité, pour des préoccupations parfois très différentes, tant pour les responsables de politiques publiques judiciaires que pour les professionnels privés du droit.

95. Quelle que soit la tradition juridique du pays, l'aléa judiciaire, c'est-à-dire le risque de voir sa demande en justice satisfaite ou rejetée, suscite l'envie de pouvoir être quantifié à l'aide de ces nouvelles applications technologiques.

96. Les avocats (ou les services juridiques d'une entreprise) y voient la possibilité d'apporter à leurs clients un conseil mieux éclairé en évaluant de manière empirique et systématique les chances de succès d'une procédure, mais aussi de favoriser la conclusion de transactions permettant, le cas échéant, d'échapper à un long et coûteux procès. Certains assureurs proposent déjà l'utilisation de systèmes prédictifs à leurs clients pour évaluer les mérites de leur affaire⁴⁵.

97. Dans le même temps, les décideurs publics voient l'opportunité de mieux réguler les flux de nouvelles procédures dans les tribunaux et de se doter d'un levier de réduction des coûts de fonctionnement de la justice. Il s'agirait ainsi d'inciter les justiciables à recourir par ce biais à des modes alternatifs de résolution de litiges (conciliation, médiation ou encore arbitrage)⁴⁶.

Les expérimentations menées en France

A l'initiative du Ministère de la justice, les deux cours d'appel de Rennes et de Douai ont accepté de tester sur divers contentieux d'appel au printemps 2017 une version d'un logiciel présenté comme de « prédictibilité de la justice » et en réalité d'analyse des décisions en matière civile, sociale et commerciale de l'ensemble des cours d'appel françaises.

Alors que ces données internes et exhaustives de jurisprudence leur étaient déjà accessibles gratuitement depuis de nombreuses années (base JURICA), et qu'elles avaient été mises spécialement à la disposition de la société éditrice par le Ministère, les deux cours s'étaient vues proposer d'expertiser l'intérêt d'une analyse chiffrée, présentée comme novatrice, des sommes allouées par les juridictions, ainsi que d'un classement géographique des écarts constatés pour des requêtes et des procès similaires.

L'ambition affichée du logiciel était donc la création d'un outil d'aide à la décision, aux fins de réduire si nécessaire une trop grande variabilité des décisions des tribunaux, et ce au nom du principe d'égalité des citoyens devant la loi. Le retour d'expérimentation, contradictoirement débattu entre les deux cours d'appel, le Ministère de la justice et la *legaltech* à l'origine du produit a malheureusement objectivé l'absence de plus-value de la version testée du logiciel pour le travail de réflexion et de prise de décision des magistrats utilisateurs.

Plus grave, des biais de raisonnement du logiciel ont été révélés qui aboutissaient à des résultats aberrants ou inappropriés, en raison d'une confusion entre de simples occurrences lexicales des motivations judiciaires et les causalités qui avaient été déterminantes dans le raisonnement étudié des juges.

⁴⁵ Le logiciel Predictice est proposé aux assurés d'Allianz par exemple.

⁴⁶ Voir à ce titre en France le rapport d'information n° 495 (2016-2017) au nom de la commission des lois du Sénat, déposé le 4 avril 2017 du sénateur Philippe BAS.

98. L'approche déjà existante dans de nombreux systèmes judiciaires, d'une harmonisation des décisions par des barèmes dans de nombreuses matières (divorce, licenciement, réparation du préjudice corporels) semble pouvoir être renouvelée par une approche probabiliste ou actuarielle⁴⁷. Des services alternatifs de résolution de litiges en ligne ont même vu le jour, entre autres pour aider à l'évaluation du montant de réparation pour des petits litiges. Ces approches intéressantes ne sont toutefois pas dénuées de biais et elles ne doivent pas priver le citoyen d'accéder à un juge ni remettre en cause le principe du contradictoire.

6.1 Une nouvelle puissance de calcul pour la constitution de barèmes

99. Les **modalités de calcul de barèmes** dans diverses matières civiles (indemnisation du préjudice corporel, prestation compensatoire, indemnité de licenciement par exemple) paraissent pouvoir être considérablement enrichies, en combinaison avec d'autres techniques de traitement et sous réserve de nombreuses précautions de conception et d'utilisation (effet dit « performatif »)⁴⁸.
100. En effet, rappelons ce que Jean-Paul Jean, président du groupe de travail évaluation de la CEPEJ qualifiait de « défi qualitatif » lors d'un colloque relatif à l'*open data* tenu en 2016 en France : le traitement opéré par le *machine learning* ou tout autre modalité de traitement devrait pouvoir se réaliser sur la base d'originaux certifiés, dont l'intégrité aura été vérifiée, et qui auront été enrichis pour distinguer l'important de l'insignifiant⁴⁹.
101. L'autre risque auquel ont déjà été confrontés les concepteurs de ces outils est celui de se livrer à du data-snooping, c'est-à-dire à ne sélectionner en amont que des données de manière à ce qu'elles puissent être significatives au regard de grilles d'analyse prédéterminées, par exemple en éludant de l'échantillon des décisions se prêtant mal à des corrélations de séquences linguistiques par le machine learning ou tout autre méthode (des décisions sans exposé du litige ou peu motivées par exemple).
102. Mais si l'on calcule un barème, ne fait-on pas déjà un peu de « prédiction » ? La frontière peut sembler relativement ténue sauf si l'on distingue clairement la finalité de la démarche : l'objectif ici n'est pas de délivrer une information prescriptive mais d'informer d'un état de fait.
103. Au prix de ces précautions méthodologies et d'utilisation, les barèmes apparaissent pour le traitement de certains contentieux un puissant outil d'harmonisation de la jurisprudence. Auparavant calculés sur des échantillons plus ou moins restreint de décisions, les outils issus de l'IA rendent possible l'exploration d'une quantité plus importante de décisions et sont susceptibles de permettre la constitution d'échelles plus précises, selon un rythme de mise à jour plus important.

6.2 Les résolutions de litiges en ligne

104. Les tribunaux européens font tous face, selon des degrés différents, à des contentieux civils répétitifs de faible valeur. L'idée de faciliter leur traitement par les technologies de l'information et/ou de les externaliser des tribunaux est assez largement partagée. La Grande-Bretagne, les Pays-Bas, la Lettonie représentent autant d'exemples de pays qui ont déjà mis en œuvre ou sont sur le point de mettre en œuvre ce type de solutions plus ou moins automatisées⁵⁰. Pour les

⁴⁷ S'agissant de l'approche actuarielle, l'offre de CaseLaw Analytics annonce réaliser plus une évaluation de risque que la prédiction d'une solution à un litige.

⁴⁸ L'effet performatif ou d'autoréalisation est le risque pour un système de produire progressivement en sortie les mêmes résultats en influant sur les producteurs d'information en entrée ; cet effet est souvent évoqué en ce qui concerne les barèmes en matière judiciaire qui, renseignés de décisions basées sur ces échelles, tendraient à n'être représentatifs que d'eux-mêmes.

⁴⁹ J-P. Jean, Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence, Colloque du 14 octobre 2016 tenu à la Cour de cassation, <https://www.courdecassation.fr/IMG///Open%20data,%20par%20Jean-Paul%20Jean.pdf>, page consultée le 17 mars 2018.

⁵⁰ Voir à ce titre le système de résolution de litiges en ligne en Grande-Bretagne - <https://www.judiciary.gov.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version1.pdf>. Voir également le système "AP" aux Pays-Bas, qui émet des décisions automatisées basées sur des permis précédemment accordés, et qui a donné lieu à des litiges au niveau national et devant la CJCE : deux affaires (c-293/17 et c-294/17) sont portées devant le Conseil d'Etat des Pays-Bas (agriculteurs / conservation de la nature contre les Pays-Bas) pour déterminer si un système (Programme Régulant l'azote) est autorisé ou non à décider

créances transfrontières, l'Union européenne a mis en place, par le règlement n° 524/2013, un cadre commun disponible sur internet (« European small claims »).

105. Le périmètre de ces services de résolution de litiges en ligne (*online dispute resolution – ODR*) semble toutefois avoir progressivement évolué : de services en ligne restreints à des mesures alternatives de règlement de litiges (*alternative dispute resolution*) pour un traitement avant la saisine de la juridiction, ces offres s'introduisent de plus en plus dans le processus juridictionnel lui-même au point d'offrir des services de « tribunal électronique »⁵¹ Elles ne concerne pas uniquement des litiges de faible valeur, mais s'appliquent également au contentieux fiscal ou à celui relatif aux services de sécurité sociale, ou encore aux procédures de divorce.
106. Pour les promoteurs de ces solutions, qui intéressent un certain nombre de professions judiciaires et le secteur privé, l'accès à la justice pourrait être significativement amélioré par une solution large combinant ODR et IA (ou à tout le moins des systèmes experts, voir *supra* chapitre 3 pour la distinction). L'idée est d'accompagner les plaignants dans un diagnostic automatisé du litige au travers d'un certain nombre de questions traitées par la machine, aboutissant à des propositions de solution. Les travaux du laboratoire de Cyberjustice de Montréal qui intègre les différentes phases précontentieuses et contentieuses dans un processus informatique pour des conflits de faible intensité (cour des petites créances du Québec par exemple) illustrent bien cette hybridation⁵². Les résultats en termes d'efficacité et de qualité paraissent selon les concepteurs exceptionnels.
107. Mais sur quelle base serait calculée l'éventuelle indemnisation proposée par un tel système ? Quelle méthode ? L'algorithme traite-t-il l'information de manière loyale ? Les propositions ont-elles vocation à pouvoir être discutées contradictoirement avec l'aide d'un tiers formé et certifié ? Un accès au juge est-il toujours possible ? Certains auteurs voient même dans la généralisation de ces modes de règlement de litiges une nouvelle manifestation du « solutionnisme » numérique, c'est-à-dire du recours systématique aux technologies pour tenter de résoudre des problèmes n'en relevant pas nécessairement⁵³. Rappelons aussi qu'en Europe, un cadre réglementaire contraignant pour les États membres et plus protecteur a été récemment mis en place : le RGPD, dans son article 22, prévoit explicitement la possibilité pour la personne concernée de refuser de faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, sauf exceptions⁵⁴.

si les agriculteurs et autres personnes violent la directive habitat. Récemment, la Cour de justice de l'Union européenne à Luxembourg a répondu aux demandes de décision préjudicielle concernant ces affaires (ECLI:EU:C:2018:882).

⁵¹ Darin Thompson, *Creating new pathways to justice using simple artificial intelligence and online dispute resolution*, Osgoode Hall Law School of York University

⁵² <http://www.cyberjustice.ca/projets/odr-plateforme-daide-au-reglement-en-ligne-de-litiges/>

⁵³ Evgeny Morozov, *Pour tout résoudre cliquez ici*, éditions FYP, cité par David Larrousserie, *Contre le « solutionnisme » numérique*, Le Monde, 6 octobre 2014, https://www.lemonde.fr/sciences/article/2014/10/06/contre-le-solutionnisme-numerique_4501225_1650684.html

⁵⁴ Article 22, 1 du Règlement UE 2016/679 : « *La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé* » ; des exceptions sont prévues (tels que le consentement de la personne concernée) mais des « *mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée* » doivent être mises en œuvre par le responsable du traitement dont « *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* ».

Voir, dans le même sens, la Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, telle qu'elle sera amendée par le Protocole adopté en mai 2018 quand ce dernier entrera en vigueur. L'article 9§1, sous a) prévoit le principe que « Toute personne a le droit de ne pas être soumise à une décision l'affectant de manière significative, qui serait prise uniquement sur le fondement d'un traitement automatisé de données, sans que son point de vue soit pris en compte. Par dérogation à ce principe d'interdiction, l'article 9§2 précise que « le paragraphe 1.a ne s'applique pas si la décision est autorisée par une loi à laquelle est soumis le responsable du traitement et qui prévoit également des mesures appropriées pour la sauvegarde des droits, des libertés et des intérêts légitimes de la personne concernée ».

Le rapport explicatif précise (§75) : « Il est essentiel que toute personne susceptible d'être soumise à une décision purement automatisée ait le droit de contester cette décision en faisant valoir de manière effective son point de vue et ses arguments. En particulier, la personne concernée doit avoir la possibilité de prouver l'inexactitude éventuelle des données à caractère personnel avant leur utilisation, l'inadéquation du profil qu'il est prévu d'appliquer à sa situation particulière ou d'autres facteurs qui auront un impact sur le résultat de la décision automatisée. Tel est notamment le cas lorsque l'application d'un raisonnement algorithmique, en conduisant à la limitation d'un droit, au refus d'un avantage social ou à l'évaluation de leur capacité d'emprunt sur le seul

108. Il faut donc arriver à bien mesurer, au cas par cas, les potentiels bénéfiques d'un système d'ODR, son degré d'intégration dans un processus juridictionnel complet (du précontentieux au véritable contentieux) ainsi que le rôle plus ou moins décisif de l'IA dans l'exécution du processus.
109. En effet, pour les offres d'ODR existant déjà en amont des processus juridictionnels, leur place paraît relativement claire : il s'agit, hors du tribunal, de contribuer à la mise en œuvre de conciliation, médiation, d'arbitrage. Ces services peuvent également intervenir dans le cadre d'une procédure contentieuse sous le contrôle du juge avant qu'il ne tranche le litige sur le fond (une telle phase est parfois rendue obligatoire pour certains litiges).
110. D'autre part, l'apport effectif de l'IA est à mesurer. S'agit-il seulement d'utiliser de l'apprentissage machine pour établir des barèmes indicatifs ou prescrire une solution ? S'agit-il réellement d'IA ou de système expert ou du simple enchaînement de règles logiques ? En toute hypothèse, il conviendrait de pouvoir assortir ces systèmes d'impératifs de transparence, de neutralité et de loyauté⁵⁵.
111. Enfin, il faut également examiner la manière dont les plaignants sont incités à utiliser le système : existe-t-il une confusion ou non dans la dénomination même de l'offre ? Si l'on parle de tribunal, il doit s'agir de la forme d'organisation définie par la Convention européenne des droits de l'homme et non une simple justice privée aux simples apparences d'une justice étatique⁵⁶. Le recours au juge est-il clairement rendu possible ? Aux Pays-Bas des contrats privés d'assurance maladie semblaient prévoir par défaut un recours à un ODR avant toute saisine de la justice.
112. Le groupe de travail médiation de la CEPEJ (CEPEJ-GT-MED) a initié en 2018 des premières réflexions sur les apports des technologies de l'information pour la résolution de litiges avec des méthodes alternatives. Le CDCJ conduit actuellement des travaux approfondis sur les ODR afin d'identifier les potentialités de ces outils mais également les points problématiques au regard d'éventuelles violations des articles 6, 8 et 13 de la Convention européenne des droits de l'homme.

6.3 Les principales garanties à réaffirmer dans le cadre de la procédure civile, commerciale et administrative

Droit d'accès à un tribunal

113. La mise à disposition d'outils de résolution de litiges en ligne ne devrait pas remettre en cause le droit d'accès à un tribunal au sens de l'article 6, même si ce droit n'est pas absolu et se prête à des limitations implicites⁵⁷. En matière civile par exemple, chaque justiciable possède le droit de soumettre à un tribunal toute contestation relative à ses « droits et obligations de caractère civil »

fondement du logiciel, a pour effet de stigmatiser des individus. La personne concernée ne pourra néanmoins pas exercer ce droit si la décision automatisée est prévue par la loi à laquelle le responsable du traitement est soumis, qui prévoit des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée ».

Voir également l'article 9§1, sous c), de la Convention 108 modernisée, qui prévoit le droit de la personne concernée « d'obtenir, à sa demande, connaissance du raisonnement qui sous-tend le traitement de données, lorsque les résultats de ce traitement lui sont appliqués ». Le rapport explicatif de la Convention modernisée (§77) précise : « Les personnes concernées ont le droit d'obtenir connaissance du raisonnement qui sous-tend le traitement des données, y compris les conséquences de ce raisonnement et les conclusions qui peuvent en avoir été tirées, en particulier lors de l'utilisation d'algorithmes pour une prise de décision automatisée, notamment dans le cadre du profilage. Par exemple, dans le cas d'un système d'évaluation de leur solvabilité par notation, les emprunteurs ont le droit d'obtenir connaissance de la logique sur laquelle repose le traitement de leurs données et qui aboutit à la décision d'octroi ou de refus du crédit, au lieu d'être simplement informés

de la décision elle-même. La compréhension de ces éléments contribue à l'exercice effectif d'autres garanties essentielles comme le droit d'opposition et le droit de recours auprès de l'autorité compétente ». On retrouve également dans la RGPD ce « doit d'obtenir des informations utiles concernant la logique sous-jacente » (article 13§1, sous f ; article 14§2, sous g ; article 15§1, sous h).

⁵⁵ Charlotte Pavillon, Préoccupations au sujet d'un juge numérique, nrc.nl, <https://www.nrc.nl/nieuws/2018/01/19/zorgen-om-populaire-digitale-rechter-a1588963>

⁵⁶ Scarlett-May Ferrié, Les algorithmes à l'épreuve du procès équitable, document consulté sur <http://lexis360.fr>, téléchargé le 09/07/2018, § 27-38

⁵⁷ Art.6 §1 « 1. Toute personne a droit à ce que sa cause soit entendue (...) par un tribunal (...) qui décidera (...) du bien-fondé de toute accusation en matière pénale dirigée contre elle. (...) » ; pour les limitations voir Deweer c. Belgique, § 49 ; Kart c. Turquie [GC], § 67.

⁵⁸. L'Assemblée parlementaire du Conseil de l'Europe a adopté en 2015 une résolution relative à « L'accès à la justice et internet: potentiel et défis » dans laquelle elle appelle « à veiller à ce que les parties qui engagent une procédure de règlement de litige en ligne conservent un droit d'accès à une procédure d'appel devant les tribunaux qui satisfasse à l'exigence d'un procès équitable énoncée à l'article 6 de la Convention »⁵⁹.

Principe du contradictoire

114. S'agissant des barèmes, il semble impératif de rendre accessible aux citoyens et, en premier lieu, aux parties à un procès un certain nombre d'informations quantitatives (nombre de décisions traitées pour obtenir le barème par exemple) et qualitatives (origine des décisions, représentativité des échantillons sélectionnés, distribution des décisions entre différents critères comme le contexte économique et social) afin de saisir le sens des échelles construites, d'en mesurer les éventuelles limites et de pouvoir en débattre devant un juge.

Egalité des armes

115. L'utilisation de moyens technologiques ne devrait pas conduire à aggraver le déséquilibre entre les parties. L'utilisation de moyens numériques pourrait en effet faciliter les recours pour certains opérateurs (institutionnels, entreprises disposant de moyens, personnes alphabétisées numériquement) et, au contraire, poser des difficultés à certains types de population plus précarisées ou moins éduquées sur le numérique. Il est important que tout individu ne soit pas laissé seul derrière son écran, mais qu'il puisse être renseigné de la possibilité d'avoir accès à un conseil juridique et d'être accompagné dans sa démarche le cas échéant.

Impartialité et indépendance des juges

116. On peut se demander si « la norme issue du nombre » évoquée précédemment à la section 2.2. ne produirait pas d'effets indirects sur l'indépendance et l'impartialité des magistrats, et notamment dans des systèmes dans lesquels l'indépendance du système judiciaire n'est pas totalement acquise. Dans ces systèmes, le risque que cette « norme issue du nombre » puisse avoir un effet indirect de pression sur les magistrats lors de la prise de décision et favoriser une uniformisation des magistrats, voire un contrôle par l'exécutif de ceux qui dérogent à la norme ne peut pas être écarté.

Droit à l'assistance d'un avocat

Nous avons évoqué au début de ce chapitre les avantages dérivant de l'application des outils de justice dite prédictive pour les avocats, et notamment la possibilité d'apporter à leurs clients un conseil mieux éclairé en évaluant de manière empirique et systématique les chances de succès d'une procédure. Imaginons toutefois un cas où les chances de succès pour le justiciable soient extrêmement réduites : est-ce que cela pourrait avoir un effet sur la décision de l'avocat d'assister son client ? La pratique professionnelle devrait viser à minimiser le risque que des personnes nécessitant un conseil juridique puissent au final, en être privées.

⁵⁸ CEDH, *Golder c. Royaume-Uni*, §§ 28-36

⁵⁹ Résolution 2054 (2015) de l'Assemblée parlementaire du Conseil de l'Europe (APCE), 10 novembre 2015, <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-fr.asp?fileid=22245&lang=fr>

7. Les enjeux spécifiques à la justice pénale : prévention des infractions, du risque de récidive et évaluation de la dangerosité

Même sans être conçus pour être spécifiquement discriminatoires, le recours aux statistiques et aux IA en matière pénale a démontré un risque de résurgence de doctrines déterministes au détriment des doctrines d'individualisation de la peine, largement acquises depuis 1945 dans la plupart des systèmes judiciaires européens.

117. L'utilisation des sciences et technologies d'IA en matière pénale pose des difficultés spécifiques car leur application pourrait traduire certains débats publics actuels sur le caractère prétendument prédictible des comportements infractionnels. Ce débat avait pourtant paru tranché durablement durant une trentaine d'années dans un certain nombre de pays européens. En Italie par exemple, l'article 220, paragraphe 2, du Code de procédure pénale exclut expressément le recours à une expertise aux fins d'établir « *l'habitude ou le professionnalisme dans le crime, la tendance à commettre un crime, le caractère et la personnalité de l'accusé et, en général, les qualités psychologiques de l'accusé, indépendamment des causes pathologiques* ». En France, la doctrine dite de « défense sociale nouvelle » développée par Marc Ancel a été fondatrice du droit pénal : à une approche simplement punitive et déterministe s'est substitué un système de réadaptation sociale postulant prévenir la commission d'infraction en évitant que soient réunies les conditions de la délinquance. Cette approche est partagée par un nombre d'instruments européens sur les politiques criminelles qui mettent l'accent sur les objectifs de rééducation et de réinsertion des sanctions pénales⁶⁰.

118. Les outils en matière de justice pénale seraient donc à concevoir dans le respect de ces principes fondamentaux de réhabilitation⁶¹ et en rappelant le rôle du juge en matière d'individualisation de la peine, sur la base d'éléments objectifs de personnalité (formation, emploi, prise en charge médico-sociale régulière) sans autre forme d'analyse que celle opérée par des professionnels spécifiquement formés, comme les agents de probation par exemple. Les techniques de *Big Data Analytics* pourraient être utilisées par ces professionnels pour centraliser et collecter des informations sur la personne accusée d'un crime ou délit qui pourraient se trouver auprès d'institutions et organismes différents et qui auraient ensuite besoin d'être examinées par un juge, parfois dans des délais très courts (par exemple dans le cadre des procédures accélérées de jugement).

7.1 Les outils utilisés par les services d'enquête avant le procès pénal

119. Les instruments qualifiés de « police prédictive » (avant judiciarisation ou avant la saisine d'un tribunal) connaissent déjà un essor rapide et commencent à être connus par le grand public (pensons par exemple aux « no fly lists » qui sont en réalité une application de « big data analytics » rassemblant et analysant des données sur de potentiels terroristes dans le but de prévenir la commission d'actes ou encore aux algorithmes employés pour détecter des faits de fraude ou de blanchiment).

120. D'une manière générale, de nombreux outils informatiques sont couramment utilisés pour prévenir la commission d'actes criminels (en identifiant les lieux de prédilection des passages à l'acte ou leurs auteurs potentiels) ou pour les poursuivre de manière plus efficace⁶². On retrouve dans la première catégorie des instruments qualifiés de « police prédictive » qui servent à empêcher certains types d'infractions ayant des éléments de régularité dans leur occurrence comme le cambriolage, la violence dans la rue, le vol des/dans les véhicules. Cette qualification est liée à la capacité de ces technologies de déterminer précisément où et quand ces infractions pourraient être commises, et de reproduire ces informations sur une carte géographique, sous forme de « points chauds » qui sont visualisés en temps réel par les patrouilles de police. On parle d'ailleurs de « cartographie criminelle prédictive ». La plupart des logiciels utilisés dans ce

⁶⁰ See European Court of Human Rights, Grand Chamber, *Winter vs United Kingdom*, paras 114 – 118.

⁶¹ L'utilisation de l'IA à des fins de traitement et de réhabilitation (pour collecter par exemple des données sur les traitements administrés ou les méthodes de réintégration en prison) paraît en revanche devoir être encouragée.

⁶² Voir à ce propos Ales Završnik, "Big Data, crime and social control", page 194 et suivantes, qui énumère de manière détaillée une série d'instruments utilisés par les services de police en Europe et aux États Unis.

domaine se base sur des éléments de localisation historique des crimes, comme des rapports de police, mais des nouvelles technologies encore plus puissantes combinant des données variées et provenant de différentes sources sont également testées⁶³. Ces instruments, qui affichent des taux d'efficacité très persuasifs, auraient par ricochet des effets dissuasifs en ce qui concerne la commission d'infractions dans les zones entourant les « points chauds » et positifs pour les politiques publiques⁶⁴.

121. Il faut toutefois relativiser les capacités de « prédiction » de ces outils, qui montrent leurs limites à l'égard de crimes de nature moins régulière ou visant des lieux différents, comme le terrorisme. Par ailleurs, une de leurs faiblesses est l'effet de « cercle vicieux » et de « prophéties auto-réalisatrices » : les quartiers considérés à risque attirent davantage l'attention de la police et la police y détecte davantage de crimes, ce qui mène à une surveillance excessive par la police de ces communautés qui les habitent⁶⁵. Enfin, des interrogations quant à une possible « tyrannie de l'algorithme » qui pourrait minimiser, voire progressivement remplacer le jugement humain ne sont pas totalement absentes au sein des services de police eux-mêmes, même si pour l'instant, la technologie est présentée comme restant au service de l'humain pour mieux l'outiller dans la prise de décision⁶⁶.
122. En outre, les « Big Data Analytics » sont de plus en plus appliqués dans le cadre de la poursuite du crime. Des outils comme Connect, qui est utilisé par la police britannique et qui analyse des milliards de données produites dans le cadre des transactions financières pour retrouver des corrélations ou des schémas d'opération, ou encore la base de données internationale sur l'exploitation sexuelle des enfants (ICSE DB), gérée par Interpol, qui aide à identifier des victimes et / ou des auteurs par l'analyse, par exemple, des meubles et autres objets présents dans des images abusives, ou l'analyse des bruits de fond dans des vidéos, se sont montrés particulièrement efficaces dans le combat contre le crime. S'agissant de Connect par exemple, des recherches nécessitant autrefois des mois d'enquête peuvent désormais être effectuées en quelques minutes, avec un niveau de complexité et sur un volume de données très importants.
123. La doctrine s'interroge néanmoins sur la logique managériale de la réponse au crime apportée par ces outils de prédiction, dans laquelle « une analyse approfondie des raisons du crime devient moins important que le « faire quelque chose ici et maintenant » et dans laquelle, dans un contexte de réduction des budgets disponibles et dans une spirale vicieuse, la police doit assurer le même niveau de protection du public, mais avec un personnel, un équipement et des ressources limités⁶⁷.

7.2 Les outils durant le procès pénal

124. L'utilisation d'outils « prédictifs »⁶⁸ par le juge dans le cadre d'un procès pénal s'avère très peu répandue en Europe.

⁶³ Par exemple, dans le cadre du projet « E-Security – ICT for knowledge-based and predictive urban security » (<http://www.esecurity.trento.it/>), qui a été mené dans la ville de Trento en Italie de novembre 2012 à mai 2015, une base de donnée réunissant des informations relatives aux crimes dénoncés à la police, aux résultats d'enquêtes menés par la mairie portant sur la victimisation et la sécurité réelle et perçue par les citoyens, des éléments sur le désordre urbain physique et social provenant de la police, ainsi que d'autres variables relatives à la « Smart City (par exemple des informations relatives aux conditions sociodémographiques, le décor urbain, l'illumination nocturne, la présence de caméras de surveillance, de transports publiques etc) » a été créée pour mieux outiller le travail de prévention du crime et d'amélioration de la sécurité urbaine. Les responsables du projet ont témoigné de la fiabilité des techniques utilisés, qui permettraient de prédire la commission d'actes criminels avec un taux de succès d'environ 60-65 % et qui aiderait à mieux combattre le crime dans un contexte général de pénurie de ressources. Par ailleurs, de tests menés au Royaume Uni dans le cadre d'un projet pilote visant à prévoir les lieux de possibles cambriolages, vols, et assaut montrent que les projections du logiciel utilise, dénommé PREDPOL, étaient précises dans 78 % des cas, contre le 51 % des techniques traditionnelles.

⁶⁴ L'indication d'une concentration géographique du crime aiderait les forces de police à mieux s'interroger sur les facteurs environnementaux de prédisposition de celui-ci dans la zone identifiée (par exemple l'éclairage, la présence de commerces etc) et à prévoir des réponses adéquates en concertation avec d'autres acteurs.

⁶⁵ Predicting crime, LAPD style, The Guardian, 25 June 2014.

⁶⁶ How technology is allowing police to predict where and when crime will happen, The Independent, 7 October 2017.

⁶⁷ Aleš Završnik, « Big Data, crime and social control », page 196.

⁶⁸ Dans la doctrine, ces outils sont souvent qualifiés de « justice algorithmique » ou « justice automatisée », ou encore « justice simulée ».

125. HART – Harm Assessment Risk Tool, a été mis au point avec l'Université de Cambridge et est à présent testé au Royaume Uni. Cette technologie fondée sur l'apprentissage automatique a été entraînée à partir de cinq années d'archives de la police de Durham, comprises entre 2008 et 2012. En apprenant des décisions prises par les policiers pendant cette période, et de la récidive ou non de certains suspects, la machine est censée être capable d'évaluer le risque – faible, moyen ou élevé – des suspects, sur la base notamment d'une trentaine de facteurs, dont certains non liés au crime commis (par exemple le code postal, le genre).
126. Dans les tests initialement menés en 2013, dans lesquels le comportement des suspects a été observé sur une période de deux ans après la commission du crime, les prévisions de HART se sont révélées efficaces à hauteur de 98 % en cas de risque faible et de 88 % en cas de risque élevé de récidive. Dans cette phase d'expérimentation, HART aura une valeur purement consultative pour le juge; par ailleurs, des audits du fonctionnement de HART et de la fiabilité de ses conclusions seront régulièrement menés par la police.
127. Même s'il s'agit du seul outil « prédictif » recensé en Europe à ce jour, il est opportun de s'interroger sur les défis auxquels les décideurs publics pourraient se trouver confrontés dans un futur proche si ce type d'application était testé à une plus large échelle, au regard notamment de l'expérience américaine.
128. Aux Etats-Unis⁶⁹, l'ONG ProPublica a révélé les effets discriminatoires de l'algorithme utilisé dans le logiciel COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), dont l'ambition est d'évaluer le risque de récidive lorsque le juge doit déterminer la peine d'un individu.
129. Cet algorithme, qui a été développé par une entreprise privée et qui doit être utilisé obligatoirement par le juge dans certains états fédérés américains, regroupe 137 questions auxquelles répond la personne concernée, complétées par son cahier judiciaire. Les questions sont assez variées et incluent la présence d'un téléphone à la maison, la difficulté de paiement de factures, les antécédents familiaux, l'histoire criminelle du prévenu⁷⁰. L'algorithme note la personne sur une échelle de 1 (faible risque) à 10 (haut risque). Il représente une aide à la prise de décision judiciaire, ses conclusions n'étant qu'une des variables à considérer par le juge lors de la définition de la peine.
130. Or, les populations afro-américaines se sont vues attribuer par l'algorithme – sans que cet effet ne soit naturellement recherché par les concepteurs⁷¹ - un taux élevé de risque de récidive deux fois supérieur à celui des autres populations dans les deux ans qui suivaient l'application de la peine; inversement l'algorithme considérait que d'autres populations semblaient beaucoup moins susceptibles de réitérer une infraction⁷². Il doit être évidemment rappelé que ce type d'interprétation, fallacieuse, ne révèle en réalité que la fragilité sociale et économique de certains

⁶⁹ Une recherche de 2015 avait recensé une soixantaine d'outils prédictifs aux Etats Unis.

⁷⁰ Il y a d'autres algorithmes qui ont été élaborés sur la base d'observations critiques exprimés par la doctrine (voir le chapitre suivant) qui se basent sur des variables plus restreintes, plus directement liées au crime commis et moins à la race, le genre ou la condition socioéconomique. Citons à titre d'exemple le Public Safety Assessment Tool utilisé en 30 juridictions américaines.

⁷¹ Cet effet purement discriminatoire est en réalité explicable par le « calibrage » relativement permissif du modèle de l'algorithme qui crée beaucoup de « faux positifs ».

⁷² Les populations noires étaient plus fréquemment classées à haut risque, alors qu'elles ne récidivaient pas dans les deux ans suivant la condamnation ; les populations blanches à l'inverse, qui étaient plus fréquemment classées à faible risque, commettaient des infractions dans les deux ans suivants. En bref, l'algorithme surestimait le risque de récidive pour les noirs et le sous-estimait pour les blanches (les « faux positifs » étaient majoritairement des noirs, alors que les « faux négatifs » étaient en majorité des blancs). En réponse aux allégations de ProPublica, NorthPointe (devenue Equivant depuis cette polémique) a répondu que les populations blanches et noires étaient également représentées lorsqu'on considérait les « vrais positifs », c'est-à-dire les personnes ayant effectivement récidivés. **La question de comment concilier à la fois la précision de l'algorithme dans la détection de la récidive et la nécessité d'éviter des effets discriminatoires vis-à-vis des populations noires a été la source d'un intense débat dans la doctrine** ; voir en particulier Chouldechova A (2016), «Une juste prédiction avec un impact disparate: une étude sur le biais dans les instruments de prédiction de la récidive », disponible sur : <http://arxiv.org/abs/1610.07524>; et aussi « Bias in criminal risks scores is mathematically inevitable, Researchers say », disponible sur <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>. **Ce débat traduit par ailleurs la question de la légitimité d'une société privée, en dehors de tout contrôle institutionnel, d'arbitrer entre deux exigences opposées : celle de défense de la société, d'une part, et celle du respect des droits des individus, d'autre part.**

groupes de populations qui ne sont évidemment pas criminogènes par nature. Des chercheurs du Dartmouth College ont par ailleurs démontré que ce type d'algorithmes ne produisait pas de valeur ajoutée puisque des humains sans aucune expérience judiciaire arrivaient à reproduire exactement la même évaluation en interprétant simplement le questionnaire.

131. Par ailleurs, l'opacité des processus de fonctionnement des algorithmes par les entreprises privées (qui revendiquent leur propriété intellectuelle) a été une autre source d'inquiétude. Si l'on tient compte du fait qu'elles tiennent leurs données sources des autorités étatiques elles-mêmes, leur absence d'esprit de responsabilité et de quant aux règles relatives à la protection des données (accountability) vis-à-vis des citoyens pose un problème démocratique majeur. Les chroniques ont montré que le public est informé des opérations relatives aux big data de manière accidentelle, sporadiquement, lorsqu'il y a des fuites ou des erreurs : exemplaire à cet effet est le cas de ProPublica qui a révélé les failles dans l'algorithme de COMPAS suite au refus de l'entreprise propriétaire de partager celui-ci, et qui a dû saisir à la fois les autorités publiques pour avoir accès aux données et engager son propre scientifique pour examiner l'algorithme.

7.3 Les défis du « prédictif » en matière pénale

132. Nous avons vu dans les sections précédentes que l'état de développement des outils prédictifs en Europe dans le domaine pénal est très varié. Si les instruments qualifiés de « police prédictive » connaissent un essor rapide et commencent à être connus même par le grand public, le constat n'est pas le même s'agissant de l'application de ce type d'outils par le juge dans le cadre du procès pénal. S'agissant des instruments à disposition des services de poursuite, des considérations sur leurs avantages et inconvénients ont déjà été exprimées précédemment. Etudions maintenant les outils spécifiques au procès pénal.
133. Tout d'abord, il est important de bannir une perspective unique d'efficacité ou, au contraire, d'inefficacité de ces outils. Les exemples illustrés précédemment montrent qu'il peut y avoir des opportunités formidables mais aussi de véritables risques dans l'application des nouvelles technologies, faute des précautions nécessaires. Les décideurs publics et les acteurs judiciaires doivent être particulièrement vigilants et jouer un rôle actif dans le développement de ces technologies : un regard continu est nécessaire pour déterminer la réelle efficacité, efficacité et éviter des conséquences imprévisibles. Cette exigence est encore plus forte en matière pénale en raison de l'incidence directe sur la liberté personnelle de l'individu⁷³.
134. Cela dit, les possibilités offertes mais aussi les inconvénients de l'application de tels outils dans le domaine judiciaire devraient être attentivement mesurés.
135. Leurs promoteurs font souvent valoir leur caractère neutre et le fait qu'ils se baseraient sur des moyens factuels et objectifs qui aideraient à rendre la justice plus précise et transparente. Un autre grand atout serait leur efficacité, parfois dépassant les capacités humaines, qui ne peut qu'être extrêmement précieuse dans un contexte général de réduction de fonds publics, voire de pénurie de ressources.
136. Leurs détracteurs expriment en revanche des préoccupations sur l'impact et le danger potentiel que ces outils auraient sur une justice individualisée et égale pour tous : l'inclusion de variables algorithmiques comme l'histoire criminelle et les antécédents familiaux a comme conséquence que le comportement passé d'un certain groupe peut décider du sort d'un individu, qui est, bien entendu, un être humain unique, avec un contexte social, une éducation, des compétences spécifiques, un degré de culpabilité et des motivations spécifiques qui le mènent à commettre un crime⁷⁴. Par ailleurs, ils font valoir que la décision humaine peut se fonder sur des valeurs et des considérations (par exemple sociétales) qui ne seraient pas retenues par la machine. Par exemple, un juge pourrait décider d'ordonner la liberté sous caution d'une femme délinquante tout

⁷³ Un extrait de la décision de la Cour Suprême de Winsconsin dans l'affaire *Winsconsin vs Loomis* peut offrir inspiration aussi au niveau européen : « Il est important de considérer que des outils tels que COMPAS continuent de changer et d'évoluer. Les préoccupations que nous abordons aujourd'hui pourraient très bien être atténuées à l'avenir. Il incombe au système de justice pénale de reconnaître que dans les mois et les années à venir, des données de recherche supplémentaires seront disponibles. Des outils différents et meilleurs peuvent être développés. À mesure que les données changent, notre utilisation d'outils basés sur des preuves devra également changer. Le système de justice doit suivre la recherche et évaluer continuellement l'utilisation de ces outils. »

⁷⁴ Aleš Završnik, "Big Data, crime and social control", page 196.

en ayant bien présent le risque de récidive, sur la base d'une hiérarchisation des valeurs, par exemple en donnant une plus grande importance à son rôle de mère de famille et de garde des enfants, alors que l'algorithme arriverait à déterminer de manière plus précise le risque de réitération de l'infraction mais n'arriverait pas à opérer une telle hiérarchisation d'impératifs.

137. En matière pénale, les risques de discrimination sont aussi possibles lorsqu'on considère que ces outils, qui sont construits et interprétés par l'humain, peuvent reproduire des inégalités injustifiées et déjà existantes dans le système de justice pénale concerné : au lieu de corriger certaines politiques problématiques, la technologie finirait par les légitimer. Comme évoqué précédemment, l'ONG ProPublica⁷⁵ a révélé les effets discriminatoires de l'algorithme utilisé dans COMPAS⁷⁶, qui prévoyait un risque de récidive des populations noires deux fois supérieur à celui des populations blanches dans les deux ans qui suivaient l'application de la peine, et inversement considérait ces dernières beaucoup moins susceptibles de réitérer l'infraction. Certains auteurs mettent toutefois en relief les potentialités d'utilisation des algorithmes pour corriger les inégalités du système pénal concerné : puisque ils reflètent l'humain, ils pourraient comme un miroir révéler les erreurs inhérentes à la prise de décision et faciliter leur correction par les acteurs⁷⁷.
138. A la lumière des considérations exprimées ci-dessus, lorsqu'on utilise l'algorithme dans le cadre d'un procès pénal il paraît essentiel de garantir pleinement le respect du principe d'égalité des armes et de présomption d'innocence établis par l'Article 6 de la CEDH: la partie concernée devrait avoir accès et pouvoir remettre en cause la validité scientifique, le poids donné aux différents éléments et les éventuelles conclusions erronées de celui-ci, et cela chaque fois que l'algorithme est proposé à un juge ou utilisé par lui avant qu'il ne prenne sa décision.
139. A cet égard, on remarque une différence entre l'Europe et les Etats Unis en ce qui concerne le droit d'accès à l'algorithme : si aux Etats Unis les autorités judiciaires sont encore réticentes à reconnaître pleinement ce droit et opèrent une mise en balance entre les intérêts privés (et notamment la protection de la propriété intellectuelle) et les droits de la défense, en Europe, le cadre est plus protecteur du fait du RGPD, qui établit un droit d'information sur la logique sous-jacente des décisions prises sur la base d'algorithmes⁷⁸.
140. Les considérations exprimées précédemment quant aux effets potentiellement négatifs de ces outils sur l'impartialité du magistrat sont également valables en matière pénale : un juge qui décide contre la prédiction d'un algorithme prend vraisemblablement des risques d'assumer une responsabilité accrue. Il ne semble pas irréaliste d'imaginer que les juges seraient récalcitrants à prendre cette charge supplémentaire, notamment dans les systèmes dans lesquels leurs mandats ne sont pas permanents mais assujettis au vote populaire⁷⁹, ou dans lesquels leur responsabilité personnelle (disciplinaire, civile voire pénale) est susceptible d'être recherchée, surtout si leurs garanties statutaires en matière disciplinaire sont insuffisantes.

⁷⁵ www.propublica.org/article/technical-response-to-northpointe

⁷⁶ D'autres algorithmes, se focalisent sur d'autres éléments plus directement liés à l'infraction commise.

⁷⁷ Mojca M. Plesnicar and Katja Sugman Stubbs, « Subjectivity, algorithms and the courtroom ».

⁷⁸ Article 15, 1. h) du Règlement UE 2016/679 : « La personne concernée a le droit d'obtenir du responsable du traitement « ... » les informations suivantes : « ... » l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée ».

⁷⁹ Mojca M. Plesnicar and Katja Sugman Stubbs, « Subjectivity, algorithms and the courtroom ».

8. Questions spécifiques relatives à la protection des données personnelles

L'utilisation des algorithmes pose la question de la protection des données personnelles traitées. Un simple principe de précaution devrait conduire à mener des politiques d'évaluation des risques.

141. Afin de permettre de saisir tout le potentiel offert par l'utilisation des algorithmes tout en veillant au respect des principes de protection des données, il conviendrait d'appliquer le principe de précaution et de mettre en place des politiques préventives concernant les risques potentiels liés à l'utilisation des données traitées par ces algorithmes et à l'impact de leur utilisation sur les personnes, et à plus large échelle sur la société.
142. Le principe de légitimité du traitement des données personnelles et l'obligation de prévenir ou de minimiser l'impact du traitement des données sur les droits et les libertés fondamentales des personnes concernées devraient imposer la réalisation préalable d'une évaluation des risques de l'impact potentiel du traitement des données par intelligence artificielle sur les droits et libertés fondamentales des personnes. Une telle évaluation devrait permettre de prévoir la mise en œuvre de mesures appropriées, notamment dès la conception (by-design) et par défaut (by default) destinées à atténuer les risques identifiés.
143. Rappelant que les données à caractère personnel doivent être traitées pour des finalités déterminées et légitimes, elles ne doivent pas être utilisées de manière incompatible avec ces finalités et elles ne doivent pas non-plus faire l'objet d'un traitement ultérieur que la personne concernée puisse considérer comme étant inattendu, inapproprié ou contestable (principe de loyauté). La problématique de la réutilisation des données personnelles qui auraient été rendues largement accessibles impose donc la plus grande prudence.
144. La conception du traitement de données effectué par les algorithmes devrait permettre de minimiser la présence de données redondantes ou marginales et d'éviter tout biais caché potentiel et tout risque de discrimination ou d'impact négatif sur les droits et libertés fondamentales des personnes concernées.
145. Dans cet environnement d'intelligence artificielle, les droits des personnes concernées revêtent une importance singulière et le contrôle que chacun de nous doit pouvoir garder sur ses informations personnelles implique que les droits suivants puissent être exercés : le droit de ne pas être soumis à une décision automatisée affectant de manière significative la personne concernée sans que son point de vue soit pris en compte, le droit d'obtenir connaissance du raisonnement qui sous-tend le traitement de données réalisé par les algorithmes, le droit de s'opposer à un tel traitement et le droit de disposer d'un recours.

9. Les potentialités et les limites des outils de justice dite prédictive

Le terme « justice prédictive » est à réfuter car il est ambigu et fallacieux. Ces outils relèvent de méthodes d'analyse de la jurisprudence, en s'appuyant sur des méthodes statistiques qui ne restituent en rien un raisonnement juridique mais peuvent tenter de le décrire. Les biais de l'analyse, s'ils ne peuvent être totalement supprimés, sont à identifier. Le processus de conception et d'utilisation doit s'intégrer dans un cadre éthique parfaitement clair.

146. Il a déjà été souligné, au chapitre 3, le caractère ambigu et fallacieux du concept de « justice prédictive » et comment celle-ci opère un lent glissement dans l'esprit collectif, en nous portant à croire que des machines, dénuées d'affect, seront un jour mieux à même de fiabiliser l'acte de juger. Les promesses de celle-ci ont besoin plus que jamais d'être examinées d'une manière objective et scientifique, en s'appuyant sur des solides bases de recherche fondamentale, pour en détecter d'éventuels limites. Rappelons à ce titre que les risques d'interprétations faussées du sens des décisions de justice s'avèrent extrêmement élevés sur la base des seuls calculs statistiques opérés ; ce constat est aggravé par l'absence de compréhension précise des liens opérés entre les données et la présence évidente de corrélations fallacieuses, non discernables dans des masses importantes de données.
147. Par ailleurs, la soi-disant neutralité des algorithmes est un mythe, leurs créateurs transférant, de manière consciente ou non, leurs propres systèmes de valeurs. Ainsi Eric Sadin, philosophe, rappelait que « les systèmes algorithmiques, sous leur allure efficace et impersonnelle, matérialisent de façon imperceptible les intentions de leurs concepteurs ou commanditaires, induisant un pouvoir agissant et asymétrique de personnes sur la vie d'autres personnes ». En ce sens, Aurélien Grosdidier, chercheur, estime que « un algorithme, en lui-même, n'est capable de rien d'autre que de nous permettre — au mieux — de saisir une partie de l'intention de son concepteur » et élargit le questionnement à l'ensemble de la chaîne de traitement de l'information (intention du concepteur, réalisation du code informatique, exécution du code informatique et contexte d'exécution puis maintenance). Ce constat est par ailleurs partagé par Aleš Završnik, criminologue, qui souligne comme les étapes de construction et d'interprétation des algorithmes sont effectuées par l'homme et pour l'homme et ne peuvent pas échapper aux erreurs, aux préjugés, aux valeurs et aux intérêts humains, et à une représentation humaine du monde, quelle que soit la manière dont ils sont conçus.
148. En dépit de ces fortes limites, doit-on pour autant négliger l'apport d'une technologie à la puissance inégalée ? Les mathématiciens Cr. S. Calude et G. Longo soulignent eux-mêmes dans leur étude relative au big data que « la portée limitative ou négative de [leurs] résultats, comme cela arrive souvent, ne détruit pas la science des données mais ouvre la voie pour de plus grandes réflexions », dont « le défi d'une nouvelle méthode scientifique, plus étendue, capable d'intégrer à la fois les nouveaux instruments algorithmiques et les outils classiques » en accompagnant les traitements d'une « évaluation rigoureuse des preuves ». Comme souligné en introduction, l'utilisation des IA est de nature à apporter une aide extrêmement significative aux professionnels, dont les magistrats et les avocats, mais également au grand public notamment si elles permettent un jour de forger des outils de recherche et d'analyse documentaire sans équivalent en matière législative, réglementaire, jurisprudentielle et doctrinale et de créer des liens dynamiques entre toutes ces sources . Mais ce type d'application sort du champ d'étude de cet article puisqu'il ne s'agit alors en aucun cas de « prédire » l'issue d'un contentieux mais d'analyser une jurisprudence dans un temps et un espace donné.
149. Nous avons également envisagé au chapitre 6, sous réserve d'une représentativité d'échantillons sélectionnés et traités, la construction de barèmes beaucoup plus précis de la moyenne ou la médiane des sommes allouées, mutatis mutandis, dans divers domaines (pensions alimentaires, prestations compensatoires, réparation du préjudice corporel, indemnités de licenciement, etc.). Ces barèmes qui relèvent plus d'une démarche de type « conférence de consensus » que d'une analyse moyenne de l'existant apportent d'ores-et-déjà une aide non négligeable à la décision et au conseil sans pouvoir se substituer à la loi elle-même. Comme évoqué précédemment, le risque est que, faute de représenter statistiquement la réalité ni de pouvoir prédire quoi que ce soit, les résultats des logiciels de « justice prédictive » ne soient érigés en normes en dehors de toute validation par le système juridique et en contradiction avec celui-ci.

150. Evoquons enfin une proposition qui consisterait à inclure dans les systèmes « prédictifs » la possibilité de pouvoir « dézoomer » à l'envi. Moins que d'enfermer l'utilisateur dans une probabilité (ou des groupes de probabilité), l'idée serait de lui permettre de naviguer dans les corrélations ayant conduit le système à proposer cette évaluation et de pouvoir s'en extraire par la sélection d'autres concepts ou groupes de mots plus pertinents ou d'exclure des corrélations fantaisistes. Pour reprendre l'exemple de l'UCL, cela consisterait à proposer une représentation graphique des différents termes retenus par le système (avec leurs poids respectifs) pour conclure à une violation (ou non-violation) et d'autoriser l'emprunt d'autres chemins en proposant la sélection d'autres termes ou groupes lexicaux.
151. Pour audacieuse et séduisante que soit cette proposition, elle suppose que les professionnels eux-mêmes (juges, avocats, universités) s'en emparent collectivement pour en éprouver la faisabilité et qu'ils ne laissent pas des opérateurs privés concevoir seuls, avec quelques scientifiques non contrôlés, des logiciels et des modes de raisonnement ou de calculs abscons ou verrouillés.
152. Les promesses ambitieuses (et non réalisées) de certains legaltechs ne doivent donc pas occulter les immenses potentialités des technologies et le besoin d'applications adaptées et construites en lien direct avec la recherche scientifique et universitaire ainsi qu'avec l'ensemble des professionnels du droit, tels les magistrats, greffiers, avocats, notaires, huissiers et experts du terrain.
153. Dans ce contexte dynamique, il paraît essentiel, premièrement, de ne pas prendre des décisions hâtives et de débattre préalablement des enjeux et des applications concrètes qui pourraient être faites de ces instruments dans les systèmes judiciaires, aussi bien que de les tester dans une première phase. Par ailleurs, une institution judiciaire adaptée à son temps deviendrait une institution en capacité de fonder, d'administrer et de garantir à la fois une véritable **cyberéthique** tant au secteur public que privé de même que d'imposer une totale transparence et loyauté de fonctionnement des algorithmes susceptibles de participer un jour à la prise de décisions judiciaires.

10. La nécessité d'un débat public approfondi autour de ces outils antérieurement à la mise en place des politiques publiques de développement de ceux-ci. L'impérieuse nécessité d'une cyberéthique pour encadrer le développement des algorithmes d'intelligence artificielle dans le respect des droits fondamentaux

Les enjeux de l'intégration de ces outils dans un processus décisionnel judiciaire justifient de vulgariser auprès de l'ensemble des publics concernés les concepts en cause. Il convient de mettre en place un cadre éthique favorable à un rapide développement de l'IA qui intègre, dans ses processus même de conception, des mécanismes prévenant les éventuels biais et discriminations.

10.1 L'importance de débattre, tester et continuellement revoir l'application de ces outils antérieurement à la mise en place de politiques publiques

154. Les enjeux liés à la mise en place d'outils de justice prédictive sont tellement nombreux et à multiples facettes qu'ils exigent une approche pondérée par les décideurs publics.
155. Tout d'abord, il paraît essentiel d'organiser un débat public autour de ces questions, qui réunisse à la fois les concepteurs des outils et les professionnels de la justice. Les conseils de la magistrature, les associations professionnelles des magistrats, les barreaux pourront sans doute y contribuer et faciliter l'identification des potentialités et des aspects plus controversés. Par ailleurs, les écoles des magistrats et des avocats peuvent jouer un rôle clé de sensibilisation des professionnels de la justice autour de ces questions, afin qu'ils puissent mieux comprendre et contribuer utilement aux évolutions en cours.
156. Entreprendre une recherche fondamentale sur les applications proposées et les tester – pour comprendre leurs potentialités et leurs faiblesses, mais aussi pour pouvoir ultérieurement les développer et les adapter aux besoins – semble aussi essentiel. Un droit de regard sur les éléments et les caractéristiques des instruments proposés par le privé (ou ceux développés par des instituts publics indépendants et spécialisés, une solution qui devrait être encouragée) semble tout aussi indispensable afin que le service public de la justice puisse assurer efficacement sa mission. Une évaluation rigoureuse des résultats du test devrait être menée avant un déploiement à plus large échelle et une intégration dans les politiques publiques. Une évaluation régulière de l'impact de ces outils sur le travail des professionnels de la justice semble aussi hautement recommandable.

10.2 La mise en place d'un cadre éthique

157. A titre liminaire, n'envisager qu'un encadrement législatif ou réglementaire des finalités de traitement des IA n'est pas suffisant dans un contexte numérique, par essence totalement transfrontalier. Une attention scrupuleuse à la nature des données en *open data* et à leur qualité est en revanche de nature à minimiser les risques de croisements inadéquats et à renforcer la pertinence des résultats des traitements automatisés. Spécifiquement en ce qui concerne le nom des professionnels, un principe de précaution simple consisterait à prohiber leur diffusion publique dans des bases brutes structurées au regard des risques de détournements de finalités. Il ne s'agirait pas de limiter l'accès à l'information déjà traitée (restitution unitaire de la composition de la formation de jugement par exemple) mais de filtrer les données brutes librement mises à disposition. Distinguer en somme l'accès à l'information de l'accès à la base de données, manipulable à l'envie.
158. Les chercheurs Buttarelli et Marr ont souligné comment les « big data » ont besoin d'un grand contrôle et d'une grande protection. D'autres chercheurs (Pasquale, Morozov) ont mis l'accent sur la nécessité de mettre en place de procédures transparentes de déploiement des big data, et de manière plus générale des IA dans le domaine judiciaire, car les solutions proposées ne peuvent jamais représenter la vie dans sa complexité.

159. Le développement de règles de cyber-éthique pour guider l'activité des opérateurs dans le secteur et pour faire valoir les principes susmentionnés de transparence, loyauté et objectivité de l'outil devient essentiel. Une expertise indépendante et régulière devrait pouvoir assurer que les « moteurs » des intelligences artificielles employés pour l'aide à la décision des juges ne sont pas biaisés. Il n'est pas incongru en effet d'anticiper la mise en œuvre, discrète ou non, de dispositifs de référencement payant (sur le modèle du *Search Engine Advertising* de Google) permettant à certains opérateurs d'accorder moins de poids à des décisions leur étant défavorables. Ces règles semblent impératives pour renforcer la confiance des citoyens dans leurs systèmes judiciaires.
160. A ce propos, la qualité reconnue des meilleures réalisations pourrait être matérialisée par la délivrance d'un label ou d'une certification. Doivent notamment être garanties une totale transparence et une parfaite loyauté des modalités de traitement de l'information, tant en direction des professionnels que des citoyens, pour éviter que des exemples comme celui précédemment mentionné de l'algorithme COMPAS puissent se reproduire. En effet, une implication très étroite des professionnels de la justice est essentielle pour pouvoir bien mesurer les enjeux et l'impact de ces applications sur les systèmes judiciaires.
161. L'ensemble des experts participant au développement des IA, notamment les chercheurs, ingénieurs et développeurs informatiques, se trouvent aujourd'hui en charges de responsabilités tout à fait exceptionnelles et inédites. Leur formation pourrait être assortie d'un renforcement encore plus important des « humanités ». L'exemple de certaines écoles novatrices de développeurs informatiques montre que derrière la volonté de « *hacker le système* » se cache, selon certains observateurs, en réalité un pragmatisme dénué de toute contextualisation de la responsabilité qui pèse désormais sur des techniciens aux pouvoirs quasi-démiurgiques ». Le serment d'Hippocrate a certainement ses limites en médecine mais ritualise la prise de responsabilités et offre un cadre éthique.
162. Enfin, la cyberéthique doit s'accompagner d'une formation à large échelle des acteurs, depuis les concepteurs des algorithmes et les legaltechs jusqu'à leurs utilisateurs. De nouvelles humanités, transdisciplinaires, devraient être partagées entre tous pour l'IA devienne le vecteur de développements positifs pour l'humanité et renforce les garanties de la démocratie et de l'État de droit.

Annexe II : Quelles utilisations pour l'IA dans les systèmes judiciaires européens ?

Cette Annexe de la Charte passe en revue différentes utilisations de l'IA dans les systèmes européens et encourage à un degré différent leur application à la lumière des principes et des valeurs énoncés dans la Charte éthique.

L'utilisation de l'apprentissage automatisé (*Machine Learning*) pour constituer des moteurs de recherche valorisant le patrimoine jurisprudentiel est une opportunité à développer pour l'ensemble des professionnels du droit. D'autres applications (constitution de barèmes, appui à des mesures alternatives de règlements de litige) sont à considérer en prenant des précautions (qualité de la source des données notamment et non traitement de masse de l'ensemble du contentieux concerné). Les autres applications (« justice dite prédictive ») seraient à renvoyer au champ de la recherche et d'un développement ultérieur (en consultation avec les professionnels de la justice afin d'assurer une pleine correspondance aux besoins), avant d'en envisager une exploitation publique sérieuse.

En matière pénale, la question apparaît comme extrêmement sensible mais ne doit pas être ignorée. A la lumière des nombreux questionnements existants quant à leur compatibilité avec un certain nombre de droits fondamentaux, l'utilisation d'algorithmes de calcul de potentiels risques de récidive d'un individu traduit en justice devrait être envisagée avec les plus extrêmes réserves. Le traitement de données quantitatives globales pour la prévention de la délinquance est en revanche une piste à approfondir avec ces techniques nouvelles, en prenant en compte les biais connus (effet performatif, qualité des données). De même, l'utilisation d'algorithmes pour mieux faire le lien entre l'offre de travail d'intérêt général et la personnalité d'un individu constituerait peut-être un facteur d'efficacité de ce type de mesure.

➤ Utilisations à encourager

- **Valorisation du patrimoine jurisprudentiel** : les techniques de *Machine Learning* ont été de plus en plus utilisées dans le domaine du traitement du langage naturel au cours des dernières années (ce qui comprend les efforts initiaux de compréhension du langage naturel) et constituent un atout considérable pour bâtir des solutions de recherche complémentaires aux actuels mots-clés ou recherche « plein texte ». Ces outils pourraient lier plusieurs sources (Constitution et conventions, lois, jurisprudence, doctrine). Des techniques de visualisation de données pourraient illustrer le résultat des recherches.
- **Accès au droit** : sans remplacer une intervention humaine, des chatbots pourraient être mis en place pour faciliter l'accès, en langage naturel, aux différentes sources d'information existantes. Des modèles d'actes (saisine d'un tribunal, contrat de bail par exemple) pourraient aussi être générés en ligne.
- **Construction de nouveaux outils de pilotage** : l'utilisation des techniques de datascience et d'intelligence artificielle sur les données d'activité des juridictions peut contribuer à améliorer l'efficacité de la justice en permettant notamment de procéder à des évaluations quantitatives et qualitatives et à construire des projections (anticipation des moyens humains et budgétaires). Des indicateurs clés de performance pourraient être constitués sur cette base. L'implication des professionnels de justice, et notamment des juges, dans la mise en place de ces outils est recommandée, tant sur le plan de l'appropriation de ceux-ci que sur celui de l'analyse des résultats et leur pondération avec des éléments relatifs aux spécificités de la juridiction ou de qualité de la justice (par exemple, la nécessité de préserver l'accès à la justice).

➤ **Utilisations à envisager avec de fortes précautions méthodologiques**

- **Aide à la construction de barèmes/échelles dans certains contentieux civils** : l'analyse de la masse totale des décisions judiciaires n'a statistiquement pas grand sens si l'on n'identifie pas tous les facteurs causatifs (explicites et implicites dans les décisions). Savoir que la moyenne de la prestation compensatoire dans une certaine zone géographique est plus haute que dans une autre est explicable non du fait du comportement des juges, mais au vu de la caractéristique du territoire. Le *Machine Learning* peut donc être utile pour identifier des décisions (cf valorisation du patrimoine jurisprudentiel supra) mais le traitement automatisé ne saurait seul produire du sens. La construction d'un échantillon pertinent de décisions à traiter (cf technique de sondages par exemple) est un préalable indispensable.
- **Appui à des mesures alternatives de règlement de litiges en matière civile** : dans certains pays européens, des outils de justice dite « prédictive » sont utilisés par des compagnies d'assurance pour évaluer les chances de succès d'un contentieux et détourner le justiciable vers un autre mode de règlement de son litige lorsque les chances de succès sont considérées faibles. Par ailleurs, certains systèmes à l'étranger proposent des montants d'indemnisation sans réelle transparence des règles de calcul. Toutefois, l'évaluation de ces systèmes ne peut être considérée comme impartiale et fiable (cf point sur les techniques de Machine Learning) : on fait prendre une décision au citoyen sur des bases tronquées. Dans d'autres cas encore, une mesure alternative de règlement de litige peut être conseillée au justiciable par un agent relationnel (chatbot) après un examen préliminaire des critères renseignés par le justiciable lui-même, lorsque celui-ci se rend sur le site d'une juridiction ou recherche sur le web une information juridique. Cet agent relationnel peut, le cas échéant, également recommander au justiciable de demander le conseil d'un avocat ou d'un service de médiation. Dans toutes ces hypothèses, la présence d'un tiers formé (médiateur s'appuyant non seulement sur des techniques mais peut-être des barèmes comme calculé plus haut, ou un avocat) semble être la solution la plus appropriée à ce stade.
- **Règlement des litiges en ligne** : lorsqu'il se rend sur une plateforme de résolution de litige en ligne, le justiciable devrait être informé d'une manière claire et compréhensible si le traitement de son litige se fait de manière entièrement automatisée ou avec l'implication d'un médiateur ou d'un arbitre. Par ailleurs, l'information au justiciable devrait aussi être loyale, et éviter de lui donner l'impression de saisir un tribunal (à cet égard, on note que la dénomination « tribunal en ligne » est souvent utilisée pour ce type de plateformes, alors que techniquement elles visent à fournir des services de règlement alternatif des litiges). Il s'agit de deux éléments essentiels pour permettre au justiciable de faire un choix éclairé et décider éventuellement de s'y opposer et des recourir à un vrai tribunal, au sens de l'article 6 de la CEDH. Par ailleurs, eu égard aux exigences des articles 6 et 13 de la CEDH, il faudrait toujours envisager des formes de contrôle de la procédure de règlement des litiges en ligne et de son issue par les juridictions de l'Etat, surtout lorsque le justiciable a consenti au règlement des litiges en ligne en forme entièrement automatisée.
- **Utilisation des algorithmes en matière d'enquête pénale afin d'identifier des lieux de commission d'infractions** : ce genre d'application pourrait non-seulement concerner la police mais également les parquets dans le cadre des instances de prévention de la délinquance auxquelles elles participent. Là encore des systèmes ont été utilisés aux Etats-Unis afin de diriger en temps réel des patrouilles de police vers des lieux possibles de commission d'infraction. Ce type d'approche quantitative est toutefois susceptible de générer un fort effet « performatif » (en un lieu, on a plus de chance de découvrir une infraction et l'on autoalimente ainsi le système). L'analyse criminelle par des approches couplées entre des systèmes d'information géographique (SIG) et des masses de données des procédures pourraient être mieux partagées avec les parquets et bénéficier certainement d'un fort apport du *Machine Learning*. Des unités de lutte contre le blanchiment utilisent déjà des systèmes « prédictifs » pour identifier des flux financiers suspects, mais s'agissant d'information quantitative (monétaire), les machines sont plus à même de produire des analyses fiables. La recherche devrait aussi avoir un meilleur accès à ces données pour produire des études pertinentes à destination des décideurs publics.

➤ **Utilisations à envisager au terme de travaux scientifiques complémentaires**

- **Profilage des juges** : la quantification de l'activité d'un juge rendra moins compte de ses possibles biais de raisonnement que de facteurs externes ayant pesé sur sa décision. L'activité juridictionnelle dans un territoire paupérisé ne produit pas les mêmes résultats que dans un autre territoire, et cela indépendamment de la personnalité du juge. Lorsque la décision est rendue de manière collégiale et sans possibilité pour un juge d'exprimer une opinion divergente, il est vain de profiler chacun des juges de la chambre. Offrir en revanche aux juges une évaluation quantitative et qualitative plus fine de leurs activités, grâce à de nouveaux outils, mais dans un but purement informatif d'aide à la décision et à leur usage exclusif pourrait en revanche être encouragé.
- **Anticipation des décisions de justice** : le seul traitement statistique de groupes lexicaux révèle la fréquence de l'utilisation de certains groupes de mots mais n'identifie par les causes réelles d'une décision car il ne procède pas d'un raisonnement juridique (cf. étude conduite sur les décisions de la CEDH par l'*University College of London* qui obtient des résultats meilleurs sur la partie fait que sur la partie analyse du droit). Des systèmes hybrides, fondés sur la construction de modèles mathématiques censés représenter la diversité de raisonnement des juges, ne sont pas plus efficaces car ils restent limités aux biais de l'échantillon de données qu'ils ont traités et nécessitent d'être totalement réécrits en cas de modification de la loi ou de revirement de jurisprudence.

➤ **Utilisations à envisager avec les plus extrêmes réserves**

- **Utilisation des algorithmes en matière pénale afin de profiler les individus** : les expériences étrangères (COMPAS aux Etats-Unis et HART en Grande-Bretagne) ont été dénoncées par des ONG (cf travaux de *ProPublica* aux Etats-Unis ou de *Big Brother Watch* en Grande Bretagne). En raison des limites des méthodes employées, cette approche purement statistique a conduit à des résultats erronés : au constat que certains individus afro-américains seraient plus souvent impliqués dans des faits criminels, il en a été déduit un facteur de risque plus élevé pour toute la population afro-américaine. Donc, même pour des faits mineurs, ces systèmes ont pondéré négativement des accusés afro-américains et ont conduit à un injuste alourdissement du quantum de leurs peines. Une orientation différente, plus respectueuse des standards européens concernant la sanction pénale, et qui doit offrir des chances de réhabilitation et de réintégration à tout individu, doit remplacer cette approche aux effets discriminatoires et déterministes. Si des systèmes algorithmiques arrivent à contribuer à une meilleure collecte d'information au profit des services de probation par exemple, et à leur offrir de manière plus rapide les informations pertinentes pour un traitement respectueux des droits de l'Homme, de telles initiatives peuvent être utiles (notamment pour les procédures de jugement rapide). Toute autre utilisation est en revanche sujette à des biais qui heurteront de nombreux principes fondamentaux nationaux et supranationaux.
- **« Norme issue du nombre »** il ne s'agit pas seulement de produire des barèmes, ce qui pourrait être légitime, mais de fournir à chaque juge le contenu des décisions produites par l'ensemble des autres juges et de prétendre enfermer son futur choix dans la masse de ces « précédents ». Or le nombre ne saurait ajouter à la loi ou s'y substituer. Pour les raisons évoquées ci-dessus, la masse ne fait pas nécessairement sens. L'étude de la CEPEJ a aussi souligné les risques existants en matière de cristallisation de la jurisprudence et de ses effets potentiellement négatifs sur l'impartialité et l'indépendance des juges.

Annexe III : Glossaire

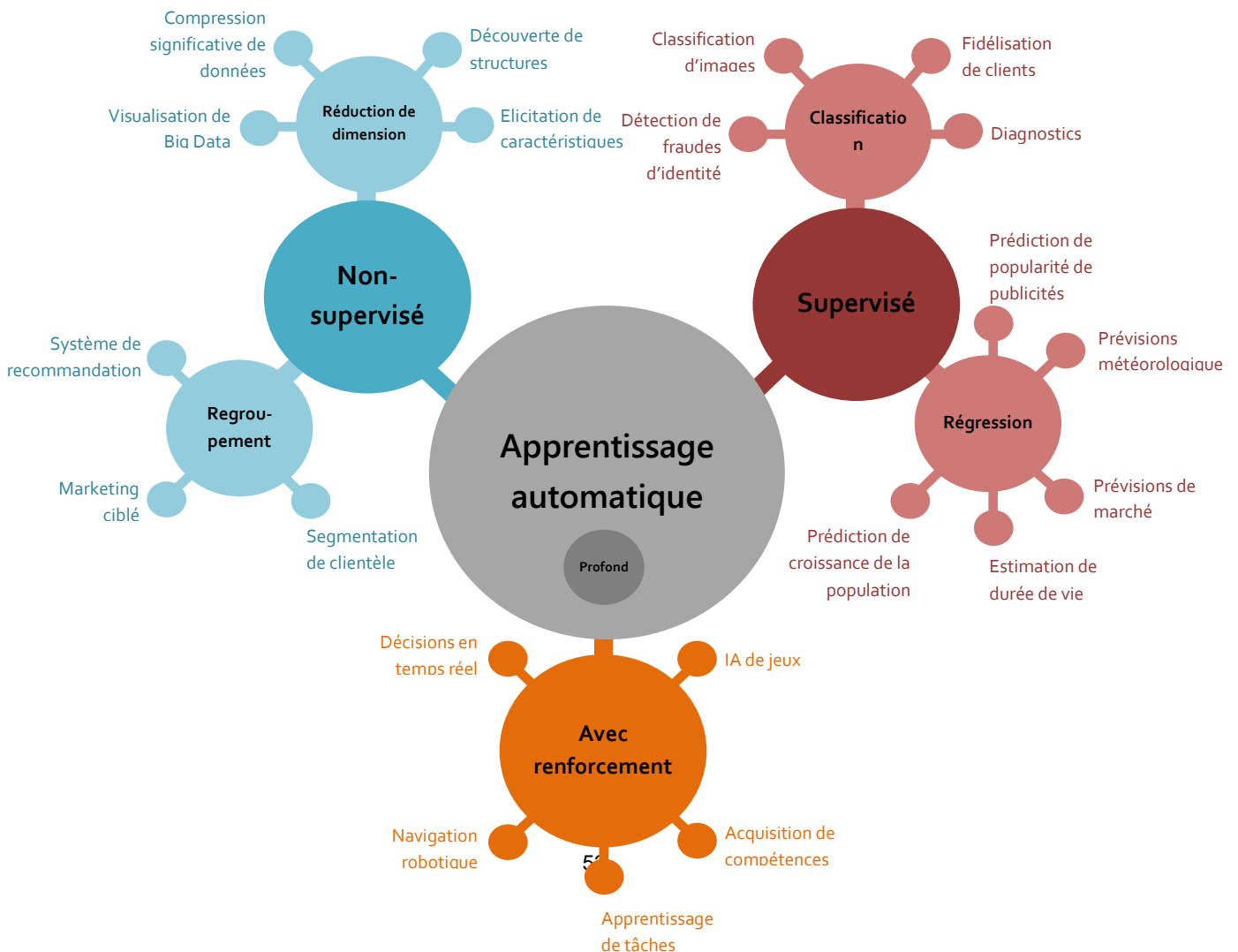
Le présent glossaire vise à définir les termes employés par la charte éthique et le document d'études. Il a été préféré une acception étroite à l'ensemble du vocabulaire employé. L'ensemble des documents doivent être lus et compris à la lumière de ces définitions

A

ALGORITHME Suite finie de règles formelles (opérations logiques, instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Cette suite peut être l'objet d'un processus automatisé d'exécution et s'appuyer sur des modèles conçus par le biais d'apprentissage machine.

APPRENTISSAGE MACHINE (*machine learning*) L'apprentissage machine permet de construire un modèle mathématique à partir de données, en incluant un grand nombre de variables qui ne sont pas connues à l'avance. Les paramètres sont configurés au fur et à mesure lors d'une phase d'apprentissage, qui utilise des jeux de données d'entraînement pour trouver des liens et les classifie. Les différentes méthodes d'apprentissage machine sont choisies par les concepteurs en fonction de la nature des tâches à accomplir. Ces méthodes sont habituellement classées en 3 catégories : apprentissage supervisé par un humain, apprentissage non supervisé et apprentissage par renforcement. Ces 3 catégories regroupent différentes méthodes dont les réseaux de neurones, l'apprentissage profond etc.

Le graphique ci-dessous illustre les différentes catégories d'apprentissage machine :



ANONYMISATION Processus consistant à traiter des données à caractère personnel afin d'empêcher totalement et de manière irréversible l'identification d'une personne physique ou morale. L'anonymisation suppose donc qu'il n'y ait plus aucun lien possible entre l'information concernée et la personne à laquelle elle se rattache. L'identification devient alors totalement impossible¹. Les principes relatifs à la protection des données relatives à une personne physique identifiée ou identifiable n'ont donc plus lieu à s'appliquer si ces données sont totalement anonymisées.

B

BASE DE DONNÉES Une base de données est un « conteneur » stockant des données telles que des chiffres, des dates ou des mots, pouvant être retraités par des moyens informatiques pour produire une information ; par exemple, des chiffres et des noms assemblés et triés pour former un annuaire.

BIG DATA (mégadonnées, grand jeux de données) Le terme « big data » désigne un grand ensemble de données, de source hétérogène (open data, données propriétaires, données achetées commercialement). Pour les données issues de l'activité judiciaire, le *big data* pourrait être la réunion des données statistiques, des traces de connexion dans les logiciels des métiers (logs des applications), des bases de données des décisions des tribunaux etc.

C

CHATBOT (agent conversationnel) Agent conversationnel qui dialogue avec son utilisateur (par exemple : les robots empathiques à disposition de malades, ou les services de conversation automatisés dans la relation au client)².

D

DATAMINING (Analyse et fouille de données) Le *datamining* permet d'analyser un grand volume de données et d'en faire ressortir des modèles, des corrélations, des tendances.

DEEP LEARNING (Apprentissage profond) Voir *Apprentissage machine* et *Neurones*

DONNÉES Représentation d'une information en vue d'un traitement automatique. Lorsqu'on dit que les algorithmes peuvent « s'appliquer » à des réalités les plus diverses dans le monde du droit ou ailleurs, on présuppose le caractère numérisable de toute réalité sous forme de « données ». Mais on voit bien dès le cas de la physique, que rien ne nous dit que les processus physiques peuvent être traduits adéquatement en termes de « données » (et s'intégrer au cycle entrée / sortie des algorithmes). Si c'est déjà le cas en physique, il n'y a pas de raison que ce ne soit pas non plus le cas dans les rapports sociaux. Il faut donc être prudent avec l'idée de « données » qui suppose toujours que la réalité que l'on cherche à décrire possède un format tel qu'il soit naturellement traitable algorithmiquement.

¹ Avis 05/2014 du G29 sur les techniques d'anonymisation. Voir aussi considérant n° 26 du règlement européen (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016.

² Rapport Commission Nationale de l'Informatique et des Libertés, Décembre 2017 : Comment permettre à l'homme de garder la main ? les enjeux éthiques des algorithmes et de l'intelligence artificielle.

DONNÉES À CARACTÈRE PERSONNEL Toute information concernant une personne physique identifiée ou identifiable (dite « personne concernée »), directement ou indirectement. Parmi celles-ci, les données sensibles concernent les données génétiques, les données biométriques identifiant un individu de façon unique, les données concernant des infractions, des procédures et des condamnations pénales et des mesures de sûreté connexes ainsi que toutes données pour les informations qu'elles révèlent sur l'origine raciale ou ethnique, les opinions politiques, l'appartenance syndicale, les convictions religieuses ou autres convictions, la santé ou la vie sexuelle.



INTELLIGENCE ARTIFICIELLE (IA) Ensemble de sciences, théories et techniques dont le but est de reproduire par une machine des capacités cognitives d'un être humain. Les développements actuels visent à pouvoir confier à une machine des tâches complexes auparavant déléguées à un humain.

Le terme d'intelligence artificielle est toutefois critiqué par les experts qui réalisent une distinction entre des IA qualifiées de « fortes » (en capacité de contextualiser des problèmes spécialisés très différents de manière totalement autonome) et des IA « faibles » ou « modérées » (pourtant extrêmement performantes dans leur domaine d'entraînement). Les IA « fortes » nécessiteraient, selon certains experts, des progrès significatifs en recherche fondamentale pour être en capacité de modéliser le monde en son entier et non de simples améliorations de performance des systèmes existants.

Les outils recensés dans le cadre de ce document sont développés en utilisant des méthodes d'apprentissage machine, donc des IA « faibles ».



JUSTICE dite « PRÉDICTIVE » On entend par « justice prédictive » l'analyse de grandes masses de décisions de justice par des technologies d'intelligence artificielle afin de construire, pour certains types de contentieux spécialisés, des prévisions sur l'issue des litiges (montant des indemnités de licenciement ou de la pension alimentaire par exemple).

Le terme « prédictif » utilisé par les *legaltechs* provient des sciences (essentiellement statistiques) qui permettent de « prédire » des résultats futurs par une analyse inductive. Un traitement est fait à partir de l'analyse des décisions judiciaires qui vise à identifier des corrélations entre des données en entrée (les critères de la loi, les faits de l'affaire, la motivation) et des données en sortie (dispositif de jugement tel que le montant d'une indemnité). Les corrélations jugées comme pertinentes permettent de bâtir des modèles qui, utilisés avec de nouvelles données d'entrée (nouveaux faits ou précisions nouvelles en paramètre, telles que par exemple la durée d'une relation contractuelle), produisent selon leurs concepteurs une prévision de décision (fourchette d'indemnisation par exemple).

Cette approche est contestée sur la forme et le fond par certains auteurs. Ceux-ci avancent en effet que, de manière générale, la modélisation mathématique de certains phénomènes sociaux n'est pas une tâche comparable à d'autres activités plus aisément quantifiables (isoler les facteurs réellement causatifs d'une décision judiciaire est infiniment plus complexe que de jouer au jeu de Go ou reconnaître une image par exemple) : les risques de corrélations fallacieuses seraient bien plus importants. De plus, en théorie du droit, deux décisions contradictoires peuvent s'avérer valide si le raisonnement juridique n'est pas contestable. Dresser des prévisions, dès lors, serait un exercice à la vocation purement informative sans aucune prétention prescriptive.

L

LEGALTECH Entreprises exploitant les technologies de l'information dans le domaine du droit afin de proposer des services juridiques innovants. Ces entreprises sont des startups spécialisées dans le domaine juridique. D'autres termes dérivés des secteurs d'activités sont aussi apparus comme les « *fintech* » pour les startups déployant des services financiers, les « *medtechs* » dans le domaine médical, etc.

LOGICIEL EN OPEN SOURCE Logiciel dont le code source est ouvert. Le logiciel peut ainsi être utilisé, modifié et redistribué librement.

M

MACHINE LEARNING (*apprentissage machine ou apprentissage automatisé*)
Voir *apprentissage machine*.

MÉTADONNÉES Données qui permettent de définir, de contextualiser ou de caractériser d'autres données. Dans la plupart de ses usages informatiques, le préfixe méta signifie « définition ou description de référence ».

Les métadonnées synthétisent des informations élémentaires sur les données, elles facilitent la recherche et la manipulation d'instances de données particulières. *L'auteur, la date de création, la date de modification et la taille du fichier* en sont des exemples. Les métadonnées et leur corollaire, le filtrage des données, aident à localiser un document spécifique.

N

NEURONES / RESEAU DE NEURONES Les réseaux neuronaux sont des systèmes informatiques vaguement inspirés des réseaux neuronaux biologiques qui constituent le cerveau des animaux. Ces systèmes « apprennent » à effectuer des tâches en considérant des exemples, généralement sans être programmés avec des règles spécifiques. Par exemple, dans la reconnaissance d'images, ils peuvent apprendre à identifier les images qui contiennent des chats en analysant des images d'exemple qui ont été étiquetées manuellement comme « chat » ou « sans chat » et en utilisant les résultats pour identifier les chats dans d'autres images. Ils le font sans aucune connaissance préalable des chats, par exemple, qu'ils ont de la fourrure, des queues, des moustaches et des visages de chats. Au lieu de cela, ils génèrent automatiquement des caractéristiques d'identification à partir du matériel d'apprentissage qu'ils traitent.

Un NN est basé sur un ensemble d'unités ou de nœuds connectés appelés neurones artificiels qui modélisent de façon lâche les neurones dans un cerveau biologique. Chaque connexion, comme les synapses d'un cerveau biologique, peut transmettre un signal d'un neurone artificiel à un autre. Un neurone artificiel qui reçoit un signal peut le traiter puis signaler d'autres neurones artificiels qui lui sont connectés.

Le but initial de l'approche NN était de résoudre les problèmes de la même manière qu'un cerveau humain le ferait. Cependant, avec le temps, l'attention s'est portée sur l'exécution de tâches spécifiques, ce qui a entraîné des écarts par rapport à la biologie. Les réseaux neuronaux artificiels ont été utilisés pour une variété de tâches, y compris la vision par ordinateur, la reconnaissance vocale, la traduction automatique, le filtrage des réseaux sociaux, les jeux de société et les jeux vidéo et le diagnostic médical.

O

OPEN DATA Le terme désigne la mise à disposition publique, par téléchargement, de bases de données structurées. Ces données sont ré-employables de manière non-onéreuse dans les conditions d'une licence spécifique, pouvant notamment préciser ou prohiber certaines finalités de réemploi.

L'*open data* n'est pas à confondre avec de l'information publique unitaire disponible sur des sites internet, dont l'intégralité de la base n'est pas téléchargeable (par exemple les bases de données des décisions des tribunaux). Il ne se substitue pas aux modes de publicité obligatoire de certaines mesures ou décisions administratives ou judiciaires déjà prévues par certaines lois ou règlements.

Enfin, une confusion est parfois réalisée entre les données (*open data* à proprement parler) et leurs moyens de traitement (apprentissage machine, *data science*) pour différentes finalités (moteurs de recherche, aide à la rédaction d'actes, analyse des tendances décisionnelles, anticipation des décisions de justice).

P

PROFILAGE Une technique de traitement automatisé des données qui consiste à appliquer un « profil » à une personne physique, notamment afin de prendre des décisions à son sujet ou d'analyser ou de prévoir ses préférences, comportements et attitudes personnels.

PSEUDONYMISATION Traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable³.

S

SCIENCES DES DONNEES Regroupement large des mathématiques, des statistiques, des probabilités, de l'informatique, de la visualisation des données afin d'extraire des connaissances d'un ensemble hétérogène de données (images, son, texte, données génomiques, liens dans les réseaux sociaux, mesures physiques, etc.).

Les méthodes et outils issus de l'intelligence artificielle entrent dans cette famille.

SYSTEME EXPERT Il s'agit de l'une des voies tentant d'aboutir à l'[intelligence artificielle](#). Un système expert est un outil capable de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier. Plus précisément, il s'agit d'un [logiciel](#) capable de répondre à des questions, en effectuant un raisonnement à partir de faits et de règles connues. Il se compose de 3 parties :

- une base de faits ;
- une base de règles ;
- un moteur d'inférence.

Le moteur d'inférence est capable d'utiliser faits et règles pour produire de nouveaux faits, jusqu'à parvenir à la réponse à la question experte posée.

La plupart des systèmes experts existants reposent sur des mécanismes de logique formelle (logique aristotélicienne) et utilisent le raisonnement déductif.

³ Article 4 du Règlement européen (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016.

T

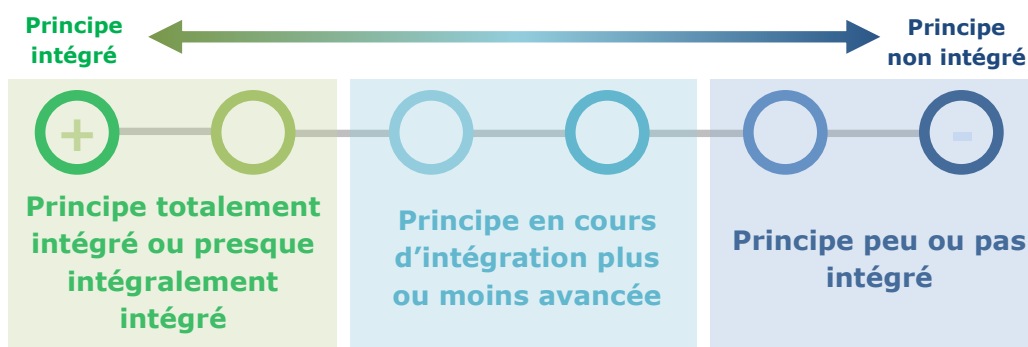
TRAITEMENT DE DONNÉES À CARACTÈRE PERSONNEL Selon l'Article 2 de la Convention 108 modernisée, toute opération ou ensemble d'opérations effectuées sur des données à caractère personnel, telles que la collecte, l'enregistrement, la conservation, la modification, l'extraction, la communication, la mise à disposition, l'effacement ou la destruction des données, ou l'application d'opérations logiques et/ou arithmétiques à ces données.

Annexe IV : Checklist d'intégration des principes de la Charte dans vos traitements

Afin d'évaluer la compatibilité de votre traitement avec la Charte, une échelle d'autoévaluation vous est proposée pour chacun des principes d'action énoncés.

Pour chaque principe, cochez l'une des 6 cases vous paraissant adaptée au contexte de votre traitement.

La case la plus à gauche indique une intégration totale, la case la plus à droite l'absence d'intégration.



Additionnez en bas de la feuille d'évaluation (ligne « total ») le nombre de cases cochées. Le nombre majoritaire vous indique le degré de compatibilité de votre traitement avec la charte.

Cette évaluation est bien entendu purement indicative et n'équivaut pas à délivrer un label ou une certification.

Checklist d'évaluation de votre traitement

1. **Principe de respect des droits fondamentaux :**
Assurer une conception et une mise en œuvre des outils et des services d'intelligence artificielle qui soient compatibles avec les droits fondamentaux, notamment le droit la protection des données personnelles



2. **Principe de prévention :**
Prévenir spécifiquement la création ou le renforcement de discriminations entre individus ou groupes d'individus



3. **Principe de qualité et sécurité :**
En ce qui concerne le traitement de décisions juridictionnelles et des données judiciaires, utiliser des sources certifiées et des données intangibles, avec des modèles conçus d'une manière disciplinaire, dans un environnement technologique sécurisé



4. **Principe de transparence, de neutralité et d'intégrité intellectuelle :**
Rendre accessible et compréhensible les méthodologies de traitement des données, autoriser les audits externes



5. **Principe de maîtrise par l'utilisateur :**
Bannir une approche prescriptive et rendre l'utilisateur un acteur éclairé et maître de ses choix



TOTAL



Compatible avec la charte	Mesures à prendre pour être compatible	Non compatible avec la charte
---------------------------	--	-------------------------------