



AUTONOMISATION
DES ACTEURS JUDICIAIRES
PAR LA CYBERJUSTICE



RAPPORT D'ÉTAPE

Chantier 12 – Évaluation et indicateurs de qualité et de performance

CADRE D'ÉVALUATION

POUR L'IMPLÉMENTATION ET L'EFFICACITÉ DE L'IA
DANS LA PRATIQUE DES ACTEURS JUDICIAIRES

Cas d'étude: la jurimétrie

*Anne-Pauline DE CLER¹ & Fabien LECHEVALIER²
sous la direction du Pr. Dominique BOULLIER³*

¹ Doctorante en sociologie, Conservatoire National des Arts & Métiers (CNAM)

² Doctorant en droit, Université Paris-Saclay / Université Laval

³ Professeur des Universités en sociologie, SciencesPo Paris

« La justice reste l'effort constant et obstiné de rendre à chacun son dû, et non pas l'ajustement mécanique d'Hommes auxquels on voudrait épargner la peine d'avoir à se rencontrer »⁴ .

⁴ GARAPON A., LASSEGUE J., *La justice digitale*, PUF, 2018, p. 350.

RÉSUMÉ

La “mesure” est rapidement apparue comme un facteur clef du succès de l’autonomisation des acteurs judiciaires par leur transformation numérique. Dès lors, l’objectif de cette étude est de forger un cadre d’évaluation intégrant des indicateurs de qualité et de performance fiables ayant vocation à assortir les multiples outils numériques développés d’un cadre de gouvernance et d’une grille d’analyse des facteurs de succès de leur développement et implémentation.

Il ne s’agit pas de fournir un inventaire des cadres d’évaluation existants mais une analyse critique d’indicateurs recueillis sur le terrain et pertinents pour l’évaluation des technologies. Notre étude s’intéresse plus spécifiquement aux systèmes d’intelligence artificielle (SIA) dans la justice, incluant en particulier une évaluation de la contribution de l’IA aux fins de jurimétrie.

De nombreuses questions et difficultés peuvent surgir au cours des processus d’adoption et/ou de développement, d’utilisation et de mesure d’impact d’outils numériques. L’évaluation est importante pour mieux éclairer les stratégies et décisions. Elle permet d’établir des priorités, produire un cadre et surtout mesurer le succès, les résultats et la qualité du processus. Dans le domaine juridique, il existe un déficit de données et un manque de cadres et méthodes d’évaluation. Nous formulons, ici, une proposition de cadre d’évaluation applicable en particulier aux outils d’IA.

L’objectif de cette étude est, à terme, de donner les moyens aux acteurs d’identifier et de promouvoir a) les meilleures stratégies pour une mise en œuvre réussie de l’IA dans leurs pratiques, b) les facteurs qui peuvent influencer le résultat d’un projet d’IA et c) les indicateurs de qualité pour l’évaluation des services d’IA.

ABSTRACT

“Measure” rapidly emerged as a key factor in the successful autonomization of judicial actors through their digital transformation. Accordingly, the objective of this study is to forge an evaluation framework that integrates reliable quality and performance indicators, with the aim to provide the multiple digital tools developed with a governance framework and a grid for analyzing the success factors of their development and implementation.

The goal is not to provide an inventory of existing evaluation frameworks but a critical analysis of indicators collected in the field and relevant for the evaluation of technologies. Our study focuses more specifically on Artificial Intelligence Systems (AIS) in the field of justice, including in particular an evaluation of the contribution of artificial intelligence (AI) to jurimetry.

Numerous questions and difficulties can arise in the process of adopting and/or developing, using and measuring the impact of digital tools. Evaluation plays a crucial role in shedding light onto related strategies and decisions. It allows establishing priorities, producing a framework and above all measuring the success, results and quality of the process. In the legal field, there is a data deficit and a lack of evaluation methods and frameworks. What we put forward here is such an evaluation framework, applicable in particular to AI tools.

The objective of this study is, ultimately, to provide means for actors to identify and promote a) the best possible strategies for a successful implementation of AI in their practices, b) the factors that may influence the results of an AI project and c) the indicators of quality to evaluate AI services.

SOMMAIRE

RÉSUMÉ	3
SOMMAIRE	4
INTRODUCTION	5
JUSTIFICATION DE L'ÉTUDE	6
OBJET D'ÉTUDE	8
MÉTHODOLOGIE & TERRAIN DE L'ÉTUDE	11
RÉSULTATS : CADRE D'ÉVALUATION ET INDICATEURS	18
UTILISATION DU CADRE D'ÉVALUATION	44
LIMITES DU CADRE D'ÉVALUATION	45
PERSPECTIVES D'AMÉLIORATION	45
BIBLIOGRAPHIE	47
TABLE DES MATIÈRES	49

INTRODUCTION

La justice a par le passé été confrontée à de **multiples défis** : celui de son indépendance, celui de son efficacité et de sa qualité, celui de ses ressources, celui des technologies de l'information... Certains ont été surmontés, d'autres demeurent, parfois sous d'autres formes. De nouveaux défis, inédits et passionnants, se présentent aujourd'hui et annoncent peut-être le bouleversement de l'accès au juge et de son office, comme des méthodes de travail des magistrats, greffiers et auxiliaires de justice. Ainsi, à l'instar du monde économique, le monde juridique est indiscutablement en train de basculer dans l'ère du numérique.

Transformation numérique du droit & Cyberjustice

La transformation numérique du secteur juridique est un phénomène complexe qui fascine et inquiète. Porteuse de nombreuses promesses économiques, comme d'éventuelles destructions de métiers, sa mise en œuvre a pour conséquence de faire apparaître de nouveaux services numériques modifiant l'intermédiation classique de certains acteurs. Le monde du droit n'est pas indifférent à ces bouleversements et connaît de nombreuses évolutions. Malgré les multiples avancées technologiques, le **recours au numérique n'est pas encore complètement intégré à l'appareil juridique et judiciaire**. Le rattachement persistant au papier et à la présence physique de toutes les parties prenantes lors des procès en sont des exemples. D'après le Laboratoire de cyberjustice, ce retard s'explique par des facteurs humains qui sont liés à la complexité du système de justice, à la diversité des acteurs impliqués et surtout aux méthodes utilisées jusqu'à présent. C'est dans ce contexte que le projet Autonomisation des acteurs Judiciaires par la Cyberjustice et l'intelligence artificielle (AJC) a été imaginé.

Le projet “Autonomisation des acteurs Judiciaires par la Cyberjustice et l'intelligence artificielle (AJC)”

Partenariat de recherche unique dans le domaine de la recherche en intelligence artificielle (IA) appliquée à la justice, AJC vise à faciliter l'**autonomisation des acteurs judiciaires** en mettant l'IA à leur service pour améliorer la prévention et la résolution des conflits. AJC réunit une équipe composée de 52 chercheurs et de 44 partenaires représentant : les meilleurs Centres de recherche mondiaux dédiés à l'introduction et l'utilisation des technologies dans le champ de la justice (la cyberjustice) ; les justiciables et professionnels du droit (les acteurs judiciaires) ; ainsi que les principaux utilisateurs et développeurs de technologies pour la justice au Canada. Le projet est dirigé par le professeur Karim Benyekhlef, directeur du Laboratoire de cyberjustice et financé par le concours 2017 du Conseil de Recherche en Sciences Humaines du Canada (CRSH). Sur le plan organisationnel, la gouvernance et l'animation scientifique d'AJC sont structurées sur un système mettant continuellement en lien les différentes composantes du projet: chercheurs, partenaires, étudiants – à travers les 3 groupes de travail et 16 chantiers de recherche scientifique qui visent à répondre aux besoins des acteurs judiciaires: outils algorithmiques capables de prédire l'issue probable d'un procès pour les auxiliaires de justice, outils technologiques d'aide à la décision pour les autorités judiciaires et administratives, technologies contractuelles impliquant la substitution du tiers par l'outil informatique, plateforme d'aide au règlement des litiges en ligne, etc.

Mise en contexte du “Chantier 12 - Évaluation et indicateurs de qualité et de performance”

La mesure des progrès de la cyberjustice est rapidement apparue comme un facteur clef de réussite des prétentions du projet. C’est la raison pour laquelle le Chantier 12 a été constitué avec pour objectif de forger un cadre d’évaluation intégrant des indicateurs de qualité et de performance fiables qui recueillent l’assentiment de la communauté scientifique. Ce cadre d’évaluation a essentiellement vocation à assortir les multiples outils développés au sein des différents Chantiers d’un cadre de gouvernance et d’une grille d’analyse des facteurs de succès et d’efficacité de leur implémentation. Compte tenu de l’importance de convaincre les acteurs judiciaires des bienfaits de la cyberjustice, les méthodes d’évaluation des outils sont importantes, y compris dans une perspective d’implantation internationale dans différents systèmes juridiques comme en Europe. Ainsi, le chantier vise à fournir, au-delà d’un simple inventaire des cadres d’évaluation, une analyse critique des indicateurs recueillis et à suggérer des indicateurs pertinents pour un nouveau cadre spécifique à l’évaluation des technologies, et plus spécifiquement de l’IA, dans la justice, incluant une évaluation de la contribution de l’IA dans l’autonomie des différents acteurs. L’objectif est, à termes, de donner les moyens aux acteurs d’identifier et de promouvoir a) les meilleures stratégies pour une mise en œuvre réussie de l’IA dans la justice, b) les facteurs qui peuvent influencer le résultat d’un projet d’IA et c) les indicateurs de qualité pour l’évaluation des services d’IA.

JUSTIFICATION DE L’ÉTUDE

Répondre à un besoin des acteurs

L’évaluation des outils de la cyberjustice a pour objet de produire des connaissances sur leur qualité et leur performance, servant aux professionnels du droit pour s’en saisir de manière raisonnée et autonome. L’évaluation permettrait également de combler le déficit de données qui existe à ce jour dans le secteur légal⁵. Celle-ci reste peu utilisée et intégrée aux pratiques des professionnels du secteur, qui manquent encore à la fois d’un cadre d’évaluation unifié⁶ et de méthodes adaptées à leurs outils spécifiques. Le chantier 12 vise à participer aux efforts qui ont récemment été menés dans d’autres juridictions⁷ ainsi qu’au niveau européen⁸ pour pallier ces lacunes et contribuer aux débats portant sur la jurimétrie ou la mesure de la qualité de la justice. Ce projet porte spécifiquement sur la cyberjustice ou l’utilisation de l’IA dans la justice, soit sur des outils qui sont de plus en plus utilisés mais sans connaissances précises de leur efficacité. En vue des perturbations mais aussi du potentiel que ceux-ci représentent pour l’autonomisation des acteurs de la justice, il y a un besoin d’obtenir une meilleure compréhension de leur impact sur les usagers et leurs pratiques. Une évaluation a pour but de répondre à ce besoin, ainsi qu’à celui de promouvoir, de manière plus générale, une meilleure compréhension de l’impact de l’IA sur l’autonomie des acteurs de la justice. En plus de participer à la construction de ces connaissances, une évaluation doit permettre la comparaison de l’efficacité de différents outils, utilisant l’IA ou non, ce qui doit permettre à terme aux acteurs de naviguer de manière plus éclairée parmi les différentes ressources existantes. C’est un besoin qui a par ailleurs été exprimé par un des participants de l’étude, précisant qu’en vue de l’arrivée de nouveaux outils dans le marché il était utile de savoir “ce qui marche et ce qui ne marche pas”. Le but est ainsi de fournir aux acteurs des connaissances qui leur permettront d’implémenter ces nouveaux outils là où ils sont

⁵ GÉLINAS F. *et al.*, *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at. 14, forthcoming, p.40 ; PASCA, A., *Annotated Bibliography : Direct to Public Legal Services - Impact Evaluation* (ACT Project & CLEO), 2020 ; PASCA A., *Promising Evaluation Practices Guide: A Few Basic Tips*, Projet AJC : Chantier 8, 15p., document de travail n°27, 2021, p.2.

⁶ PASCA A., *Promising Evaluation Practices Guide: A Few Basic Tips*, Projet AJC : Chantier 8, 15p., document de travail n°27, 2021, p.2.

⁷ PASCA A., *Promising Evaluation Practices Guide: A Few Basic Tips*, Projet AJC : Chantier 8, 15p., document de travail n°27, 2021, p.2. ; Tim Roberts & Associates Consulting, “Learning from Justice metrics models in other jurisdictions”, A Report for the University of Victoria, Faculty of Law, Access to Justice Centre for Excellence, 2019.

⁸ Commission européenne pour l’efficacité de la justice (CEPEJ), *Mesurer la qualité de la justice - Guide*, 2017.

le plus efficaces⁹, tout en maîtrisant leurs divers effets. L'évaluation représente ainsi de nombreux avantages que ce chantier vise à promouvoir et mettre à la portée des acteurs concernés.

Tout comme l'évaluation de politiques publiques sert non seulement aux usagers ou aux citoyens concernés mais aussi à diverses parties prenantes, l'évaluation des outils de la cyberjustice permet aux professionnels de la justice mais aussi aux décideurs publics et aux citoyens de prendre des décisions informées quant à ces outils, et d'optimiser leur impact pour l'ensemble de la population. L'évaluation peut ainsi contribuer à la construction d'outils légaux plus efficaces et de meilleure qualité, tout en offrant aussi la possibilité de proposer des réformes et des politiques plus informées¹⁰. Le chantier 12 a ainsi pour but de promouvoir l'implication de l'ensemble des parties prenantes concernées dans l'évaluation des outils de la cyberjustice et de favoriser de fait une co-construction et un partage de connaissances dans le domaine, afin d'assurer une meilleure utilisation et maîtrise de ces outils.

Contribuer à un domaine de recherche en émergence

L'arrivée de l'IA dans les outils légaux est récente, aussi l'évaluation de ces outils est-elle un domaine de recherche en émergence. De manière générale, et pour diverses raisons¹¹, l'évaluation d'outils légaux est limitée et leur impact est rarement mesuré, malgré les besoins en la matière. C'est ce que l'on retrouve dans notre cas d'étude, où outils, méthodes et procédures d'évaluation étaient absents. Si les professionnels participant à l'étude expriment un intérêt certain pour l'évaluation, le manque de ressources dédiées à l'évaluation et les difficultés d'intégrer des processus d'évaluation à une profession qui ne se sert encore que marginalement des outils à évaluer marquent un départ prometteur mais contraint dans ce nouveau champ. En effet, peu d'études portent à ce jour sur les résultats en matière de qualité de ces nouveaux outils, et les évaluations restent principalement orientées sur l'efficacité¹². Notre modèle cherche ainsi à combler ces lacunes en proposant une évaluation non seulement de l'impact ou de la performance des outils de la cyberjustice mais aussi de leur qualité en termes de modification des pratiques et de résultats. De plus, les modèles existants sont pour la plupart relativement abstraits et difficiles à répliquer, en partie en raison de la difficulté d'appliquer des modèles généraux à des outils spécifiques. Nous participons ainsi à un effort collectif émergent pour construire un cadre, des indicateurs, des définitions et des protocoles communs dans l'évaluation d'outils légaux, comme l'ont fait dans d'autres domaines spécifiques les chercheurs Fabien Gélinas et Alexandra Pasca, dans les projets ACT et CLEO¹³, mais aussi dans le cadre d'AJC.

Précisément, ce chantier vise à contribuer au champ de l'évaluation des outils de la cyberjustice sur le plan théorique et pratique. Il participe au dialogue entre différentes disciplines dans ce domaine en y apportant, en particulier et de manière originale, des connaissances et des pratiques issues du domaine de l'évaluation de politiques publiques. Si cet apport disciplinaire est original, nous nous appuyons sur des travaux qui s'appuient eux-mêmes sur des cadres d'évaluations importés d'autres disciplines ou domaines voire d'autres juridictions, dans le but de construire un cadre d'évaluation

⁹ GÉLINAS F. *et al.*, *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at 14, forthcoming, p.3.

¹⁰ GÉLINAS F. *et al.*, *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at 14, forthcoming, p.5.

¹¹ CBA, "Reaching equal justice : an invitation to envision and act", Août 2013 ; CAMPBELL M., Access to Justice Annotated Bibliography, National Self-Represented Litigants Project, 2019 CanLIIDocs 1835, <<https://canlii.ca/t/sjq3>>, consulté le 28/12/2022.

¹² FARROW C. W. T. *et al.*, "Everyday Legal-Problems and the Cost of Justice in Canada - Cost of Justice Survey Data", 2018.

¹³ GÉLINAS F. *et al.*, "CLEO Case Study: Evaluation Steps for the Family Law Guided Pathways (FLGP)" CLEO & Montreal Cyberjustice Laboratory, forthcoming ; GÉLINAS F. *et al.*, *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at 14, forthcoming.

commun intégrant des approches diverses¹⁴. Notre approche se distingue légèrement en ce qu'elle part d'abord des données récoltées lors des entretiens pour construire à partir de celles-ci des variables et des indicateurs de mesure de la performance et de la qualité des outils évalués. Ces indicateurs sont ensuite formalisés et agencés dans un modèle plus général (par étape de procédures par exemple) qui lui s'inspire des modèles existants. L'originalité de nos propositions réside aussi en ce qu'elles portent sur la procédure de développement et de tests by design des outils utilisant des systèmes à base d'intelligence artificielle. En effet, le cadre d'évaluation proposé porte une attention particulière au design organisationnel de ces outils, approche encore peu commune en la matière.

OBJET D'ÉTUDE

Outils d'IA: de quoi parlons-nous?

NB : cette étude a été conduite avant que les IA génératives soient observables sur le terrain. L'intelligence artificielle (IA) est une constellation de technologies différentes, qui fonctionnent de concert pour permettre aux machines de percevoir, de comprendre, d'agir et d'apprendre à des niveaux d'intelligence plus ou moins comparables à ceux des humains. Cela explique certainement pourquoi il y a une multitude de définitions de l'intelligence artificielle : l'IA ne se limite pas à une seule technologie. L'exercice de terminologie et de définition de l'IA constitue donc en soi un défi. En l'absence de définition universellement admise de l'IA, les réalités rattachées à ce vocable varient selon les interlocuteurs. La créativité terminologique et le démultiplication des concepts s'y rattachant créent, *de facto*, une forte incertitude sur les contours exacts de l'IA et du champ lexical associé. En outre, les imaginaires entourant la notion d'IA, la valeur commerciale attachée à la notion et les velléités d'instrumentalisation à des fins commerciales (parfois désignées comme des pratiques d'“IA-washing”) ou encore l'importance de disposer d'un langage commun maîtrisé par l'ensemble des acteurs (scientifiques, ingénieurs, juristes...) impliqués dans les politiques et projets qui s'y rapportent, appellent un effort de clarification terminologique. Aucune définition de l'IA ne fait consensus parmi l'ensemble des acteurs en dépit de nombreux travaux académiques menés pour recenser, comparer et synthétiser les dizaines de définitions proposées au fil du temps¹⁵. L'IA peut ainsi être décrite par référence à sa **finalité** (IA de traduction, de santé, de *trading*...) ou d'autres caractéristiques, comme le type de **modèle utilisé**. Des technologies comme le **Machine Learning (ML)** et le **traitement automatique du langage naturel (NLP)** font ainsi partie de l'IA. Aussi, par analogie avec le terme usuel de « systèmes d'information », la présente étude recourra à la notion, également utilisée au niveau européen, de système à base d'intelligence artificielle (SIA). Dans son acception la plus large, le SIA est généralement présenté comme un “traitement algorithmique” ou réduit plus simplement encore à l'**algorithme**, qui est en fait au principe de toute l'informatique¹⁶. On retiendra une autre définition des SIA, qui les ramène à leur version de type machine learning avec apprentissage automatique, réelle nouveauté contemporaine de l'informatique.

¹⁴ Les modèles utilisés par Alexandra Pasca et Fabien Gélinas sont le Drake Framework de Quinn et Raby; le Developmental Stage Approach de Panzer et Westhues, adapté par Rossi *et al.*, et le Assessment Framework for E-Justice Systems de Lupo.

¹⁵ Cf. par e.g. ce travail comparatif mené sur 55 documents émanant d'organismes publics ou parapublics : SAMOILI S., LÒPEZ COBO M., GÒMEZ E., DE PRATO G., MARTÌNEZ-PLUMED F. et DELIPETREV B., *AI Watch : Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, Publications Office of the European Union, 2020.

¹⁶ L'algorithme est classiquement défini comme une suite d'instructions et d'opérations permettant de résoudre une catégorie de problèmes.

En réalité, lorsqu'il est question de SIA, la principale difficulté de l'exercice de définition ne se loge pas tant dans le terme "artificielle"¹⁷ que dans l'autre terme de la notion, celui d'intelligence¹⁸. L'intelligence de l'IA renvoie à la modélisation de certaines fonctions cognitives humaines mises en œuvre par une machine. Pour autant, un SIA n'est **doté ni de bon sens, ni de conscience, ni de pleine autonomie**. Il se borne à faire ce que l'humain l'a programmé à faire. Néanmoins, certains SIA sont capables de modéliser plus de fonctions cognitives que d'autres et font davantage preuve d'autonomie. Ce constat a conduit les chercheurs et praticiens à distinguer les systèmes d' "IA Générale" de ceux d' "IA étroite"¹⁹. Les SIA que nous utilisons dans la vie quotidienne sont dans la majorité des cas des **systèmes d'IA étroite**, qui effectuent une seule tâche ou un ensemble de tâches étroitement liées, ex: les applications météo, les logiciels qui analysent les données afin d'optimiser certaines fonctions métier. Malgré la puissance de ces systèmes, leur champ d'actions est limité : ils sont généralement centrés sur l'amélioration de l'efficacité. Toutefois, lorsqu'elle est utilisée de façon intelligente, l'IA étroite dispose d'un immense pouvoir de transformation. Elle ne cesse d'ailleurs d'influer sur nos méthodes de travail et nos modes de vie. Les **systèmes d'IA générale** sont, quant à eux, plus proches des films de science-fiction, dans lesquels des machines sensibles imitent l'intelligence humaine. Ils sont capables de réflexion stratégique, abstraite et créative, tout en traitant un grand nombre de tâches complexes. Bien que les machines puissent se charger de certaines tâches plus efficacement que les humains (traitement des données, par exemple), cette vision d'une IA générale totalement accomplie n'est encore qu'une fiction. C'est pourquoi la collaboration Homme + Machine est essentielle : dans le monde que nous connaissons aujourd'hui, l'intelligence artificielle demeure une extension des capacités humaines. Elle ne les remplace pas.

L'IA dans la justice

Les SIA révolutionnent les possibilités d'automatisation des tâches, caractéristique majeure de la transformation numérique abordée plus haut. Ils permettent d'**accéder rapidement à de l'information** par une **veille automatisée**, de **modéliser facilement un contrat** ou bien encore d'**analyser des documents juridiques** pour prendre une décision en conséquence. Les SIA peuvent ainsi accompagner le professionnel du droit dans chacune de ses tâches et **l'aider dans sa prise de décision** sans fondamentalement remplacer la place de l'être humain dans la production de plus-values intellectuelles. Pour être efficaces, ces systèmes émergents nécessitent d'être alimentés massivement par des jeux de données afin d'apprendre et de s'adapter à la situation à laquelle ils sont confrontés, ainsi que de produire de nouvelles données. De la qualité et la quantité de données auxquelles les SIA auront accès découleront la précision et la valeur stratégique de son produit. Aussi, recourir à des SIA pour traiter automatiquement un volume massif de données s'est révélé particulièrement utile et puissant dans la tâche de quantification du risque et de l'aléa judiciaire. Sans être frappé de clairvoyance, il devient possible de prédire la réalisation d'événements futurs. Un SIA prédictif va être capable - en analysant et en comparant ce qui a déjà été produit - d'annoncer la **probabilité** d'un fait à venir, désigné sous le néologisme de "prédictivité". Plusieurs hypothèses d'application sont possibles: les SIA d'assurance prédictifs, les SIA de police prédictive ou bien encore les SIA communément nommés de "justice prédictive". Nous jugeons l'expression galvaudée et l'association des deux termes mal posée. Comme l'avait justement écrit Albert CAMUS, « mal

¹⁷ Selon l'inventeur, sinon du concept, du moins du terme, en 1956, le professeur en science informatique à Stanford John McCarthy, l'intelligence artificielle est « *la science et l'ingénierie de la fabrication de machines intelligentes* » (MACCARTHY J., *What is artificial intelligence*, Stanford University, Stanford, Novembre 2007). Cette définition évasive a le mérite de rattacher d'emblée l'IA à la machine, ou plus spécifiquement à un **ensemble d'entités automatisées** que l'homme fabrique ou programme. Le caractère « artificiel » de l'IA tient à ce qu'elle est due « *à la technique de l'homme, par opposition à ce qui a été créé et se développe naturellement* » (Dictionnaire de l'Académie française).

¹⁸ Étymologiquement, l'intelligence, du latin *intellego*, signifie « discerner, démêler, remarquer, se rendre compte, reconnaître », mais également « comprendre, entendre, saisir ».

¹⁹ « Un système d'IA générale est censé être un système capable d'effectuer la plupart des activités que les humains peuvent faire. Les systèmes d'IA étroite sont au contraire des systèmes qui peuvent effectuer une ou quelques tâches spécifiques. Les systèmes d'IA actuellement déployés sont des exemples d'IA étroite. » Groupe d'expert de haut niveau sur l'IA, Commission européenne, *Une définition de l'IA : principales capacités et disciplines scientifiques*, 2018.

nommer un objet, c'est ajouter au malheur de ce monde ». En réalité, cette expression a été choisie parce qu'elle porte en elle une dimension programmatique. Ce n'est pas tant le terme de "prédictive" qui pose problème, et ce malgré la confusion avec la prévisibilité. Ce qui trouble les auteurs est plus logiquement celui de "justice". Comment une machine statistique pourrait-elle **être** justice ? Nous y préférons le terme de jurimétrie, terme qui sera utilisé dans cette étude.

Cartographie des outils

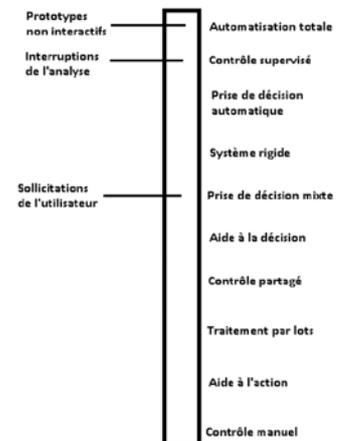
Il est certes matériellement impossible de recenser de manière exhaustive l'ensemble des cas d'usage actuels de l'IA dans la justice. Un tel travail encyclopédique serait nécessairement incomplet et obsolète dès sa publication et appellerait un lourd travail d'actualisation. En retenant une définition large des SIA comme celle qu'a proposée la Commission européenne, il ne présenterait du reste qu'un intérêt limité puisqu'il conduirait peu ou prou à répertorier la totalité des systèmes d'information juridiques. En revanche, la réalisation d'une **cartographie des cas d'usage innovants** répondrait à un besoin exprimé par nombre d'acteurs judiciaires désireux de mieux comprendre l'intérêt et l'utilité pratique de ces systèmes, favoriserait l'émergence de projets nouveaux, la mutualisation ou la duplication et fournirait aux justiciables une vue d'ensemble des utilisations de l'IA et des informations générales sur les systèmes qui les affectent. À défaut d'être en capacité de livrer une cartographie complète de cette nature, qui est l'objet des travaux d'autres chantiers du projet AJC, nous proposons un canevas - prenant la forme d'une boussole cosmopolitique - permettant de fixer un panorama illustratif et simplifié pour les besoins de cette étude et aidant à explorer les possibles.



- Abscisse

L'abscisse de la présente boussole tente de traduire l'échelle de contrôle que tend à conserver l'utilisateur sur l'outil. Cet axe pose la question de la maîtrise tant individuelle que collective des outils. À cet effet, le contrôle manuel, bien que rudimentaire, apparaît comme l'alternative implicite à des outils en parfaite autonomie bien que l'échelle que nous présentons pour compléter nos développements laisse apparaître une plus grande granularité dans l'Interaction Homme-Machine. C'est, en effet, le principal intérêt de cet axe: réfléchir aux interactions entre le juriste et l'algorithme. Quelle place pour l'Homme dans ce nouvel écosystème forgé par les algorithmes? Doit-il conserver l'ascendant et un contrôle purement manuel

Cette boussole²⁰ permet de cartographier les SIA dans la justice: du SIA aux fonctionnalités limitées jusqu'au Robot-Juge. L'étude attentive de la notion permet de constater son ambivalence, à la fois outil statistique et promesse terminologique d'un projet plus vaste, celui de rendre la justice sans intervention humaine. À sa double définition correspond une double friction: celle de sa finalité et celle de son contrôle.



²⁰ Cette boussole qui fait office de grille d'analyse est, en partie, inspirée de l'outil de cartographie nommé "boussole cosmopolitique" proposé par le Pr. Dominique BOULLIER. Cf. BOULLIER D., "Débousolés de tous les pays ...! : Une boussole écodémocrate pour rénover la gauche et l'écologie politique", *Cosmopolitiques*, éditions spéciales, 2003.

quitte à abandonner une partie de l'efficacité de l'outil ou doit-il accorder sa confiance en la technologie en laissant l'IA pleinement autonome dans ses choix?

- Ordonnée

L'ordonnée propose une analyse graduée des finalités des outils jurimétriques : de l'aide à la décision à la prise de décision. Aux extrémités des deux pôles, une promesse soit inachevée, soit dystopique. La première correspond à une réalité presque tangible, des premières expérimentations ont déjà été menées en ce domaine. La seconde est plus prospective : d'abord parce que les *legaltechs* à l'origine du développement de ces SIA n'ont pas encore eu accès aux nouvelles bases de données jurisprudentielles, ensuite parce que ces outils ne sont pas encore suffisamment aboutis, enfin parce que notre ordre juridique ne peut que très imparfaitement les recevoir.

ZOOM: les outils d'aide à la décision

Parmi les outils que nous pouvons répertorier dans cette grille, se trouvent les outils d'aide à la décision. La fonction d'aide à la décision des SIA est source d'enjeux importants. Leur mise en place, certes minimale aujourd'hui, est en forte croissance et devrait rapidement s'étendre après diffusion de l'essentiel des données judiciaires (notamment avec l'*open data*²¹ des décisions de justice), le développement de l'interopérabilité des systèmes et la formation de juristes aux pratiques "augmentées". Ce développement est à double tranchant. D'un point de vue positif, les capacités d'analyse surhumaines des SIA permettront une amélioration sensible de la compréhension du droit. La **fonction d'analyse des décisions et des documents juridiques** viendra au secours des professionnels et des justiciables, pour les guider vers une meilleure application de la norme, et une meilleure compréhension du droit. La **fonction de quantification** de ces outils devrait de son côté permettre des économies financières, par une rationalisation du contentieux et des tâches répétitives des professionnels. D'un point de vue négatif, ces outils présentent de nombreux dangers: d'abord parce que les techniques sont encore largement faillibles et méritent une amélioration technique, ensuite parce que leur utilisation pourrait catalyser les transformations du droit déjà en cours, que sont la factualisation et le renforcement de la place du juge dans l'édiction du droit. Le risque de performativité accentue encore ce phénomène. Les outils deviendraient alors indirectement des outils de prise de décision. C'est justement ces tensions qui animent le débat sur le développement et le recours aux outils de jurimétrie. Le choix de l'objet de notre étude s'est porté sur ces outils pour le potentiel de transformation qu'il représente dans l'autonomisation des acteurs judiciaires face à des outils de prise de décision qui ne sont pour l'heure que fantasmés.

MÉTHODOLOGIE & TERRAIN DE L'ÉTUDE

Cette étude prétend concilier un fort ancrage empirique avec une forte proposition de méthode d'évaluation, sans se contenter d'un inventaire de cas et sans prétendre fonder abstraitement les principes d'une évaluation hors-sol²². Sa méthodologie s'inspire donc de l'évaluation tout en ayant pour but de proposer un cadre d'évaluation, avec indications de méthode et propositions d'indicateurs, qui puisse s'adapter à d'autres objets. Des méthodes d'évaluation comme les expérimentations aléatoires ont été préalablement envisagées pour produire une première évaluation puis un cadre d'évaluation plus général. Elles ont finalement été mises de côté en vue de leur inadéquation à l'objet

²¹ Articles 20 et 21 de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique qui prévoient que toutes les décisions de justice doivent être mises à disposition du public à titre gratuit.

²² Nos observations, nos analyses et nos propositions ne reposent pas sur un benchmarking entre solutions techniques. Notre visée se focalise sur les enjeux d'appropriation de ces solutions techniques par les acteurs juridiques. La façon dont ces solutions sont "encastrées" (Polanyi) dans un "design organisationnel" (Boullier) à chaque fois spécifique est le cœur de l'observation du point de vue socio-technique qui est le nôtre.

de l'étude, celui-ci requérant une étude à composante qualitative, ainsi que des difficultés à les mettre en œuvre, notamment en raison du manque d'accès à la taille d'échantillon requise pour obtenir des résultats significatifs.

Le protocole d'étude choisi devait néanmoins partir d'un cas précis pour ensuite procéder à une comparaison contrôlée avec d'autres cas pour élaborer un premier cadre d'évaluation qui sera ensuite enrichi et surtout révisé par la confrontation à d'autres cas. Nous devons pour cela choisir une technologie (arbitrairement mais déjà opérationnelle), un domaine du droit (aisément accessible et utilisateur d'IA), dans un système juridique et dans une organisation spécifique d'acteurs judiciaires.

Phase 1 : Choix du cas d'usage de départ (terrain de recherche)

Le choix du terrain était dicté avant tout par la faisabilité, en termes d'accès, de possibilités d'observations, de connaissance a priori du domaine de façon à obtenir un matériau d'observation riche rapidement pour propulser notre construction d'un protocole avant de le confronter à d'autres situations et contraintes.

- Un système juridique: continental (français)

C'est la raison pour laquelle le cas d'étude initial est un SIA utilisé par des praticiens de droit continental, et plus précisément français (pour une extension future).

- Une organisation spécifique d'acteurs judiciaires (cabinet d'avocats)

En outre, l'étude nous laissait la possibilité d'envisager un vaste éventail de terrains possibles. Un premier choix devait s'opérer entre l'administration (services judiciaires) ou des structures privées. Nous avons arbitrairement décidé de nous intéresser aux acteurs privés dans un premier temps car il nous a semblé, après un rapide état de l'art, que le terrain semblait plus riche et les technologies employées plus matures. L'appropriation de ces outils par les administrations (notamment les juges) reste marginale. Parmi ces acteurs, nous avons dans un premier temps décidé de nous intéresser aux auxiliaires de justice²³ et plus précisément aux avocats car ces technologies, bien qu'elles ne devraient pas affecter le métier dans son essence, pourraient directement bouleverser le modèle d'affaires des cabinets. En effet, l'avocat libéré des tâches basiques et chronophages pourrait se concentrer sur l'activité au cœur de son métier: le conseil. Le conseil passe par l'écoute du justiciable et la mise en place d'une stratégie contentieuse adaptée pour résoudre le litige. Ces tâches ne sont pour l'heure pas automatisables. Ces technologies peuvent véritablement donner un nouveau souffle au métier d'avocat et c'est la raison pour laquelle il nous a semblé pertinent de commencer notre étude par ce terrain.

- Une technologie basée sur l'IA (jurimétrie)

La « jurimétrie » est sur toutes les lèvres de ceux qui font la justice. Nombre de colloques en témoignent²⁴. L'actualité du sujet en fait bien entendu son intérêt, mais cela est d'autant plus vrai que l'essentiel des grands bouleversements qu'elle augure sont encore, dit-on, prospectifs. Cette évolution

²³ Les métiers du droit que nous ciblons sont les auxiliaires de justice à l'exclusion des métiers de l'authentification (huissiers, notaires), car ces cas d'usage nous paraissent difficilement comparables.

²⁴ Cf. not. à la Cour de cassation, "La jurisprudence dans le mouvement de l'open data", le 14 octobre 2016, à l'Université catholique de Lille "La justice prédictive" le 19 mai 2017, à la Cour de cassation "La justice prédictive", le 12 février 2018, à l'Université de Laval, "La Justice prédictive : risques et avenir d'une justice virtuelle" le 6 avril 2018, à l'Université de Montpellier "Voulez-vous être jugé par un algorithme ? Réflexions autour de la justice prédictive", le 4 mai 2018, à la Cour de cassation "Le juge et le numérique : un défi pour la justice du XXIème siècle", le 8 juin 2018 pour n'en citer que quelques-uns.

“révolutionnaire”²⁵ s’inscrit dans un large mouvement de transformation sociale et organisationnelle lié au développement de la technique. Mais de quoi parle-t-on au juste ? Le rapport dit “CADIET” la définit comme l’« ensemble d’instruments développés grâce à l’analyse de grandes masses de données de justice qui proposent, notamment à partir d’un calcul de probabilités, de prévoir autant qu’il est possible l’issue d’un litige »²⁶. La jurimétrie est fondée sur l’ouverture progressive, mais massive des données judiciaires – l’*open data* –, visent à accélérer le règlement des litiges et à accroître la sécurité juridique, en améliorant la prévisibilité des décisions de justice. Par son utilisation les juges peuvent, en effet, mieux connaître les pratiques juridictionnelles de leurs collègues et les parties pourront, quant à elles, déterminer avec plus de finesse les chances de succès d’une procédure juridictionnelle tout en relevant les moyens les plus pertinents à soulever. En retour, cette justice augmentée devrait favoriser l’accès au droit et l’égalité devant la justice ainsi que la stabilisation et l’harmonisation de la jurisprudence. Le recours à des algorithmes pour le traitement des dossiers les plus répétitifs et les moins complexes, ceux par exemple qui ne nécessitent que l’évaluation d’un dommage, l’application d’un barème ou d’une trame prédéterminée, encouragerait aussi le règlement de nombreux litiges en amont même du recours au juge, par le développements des MARD²⁷, comme la médiation ou la conciliation. En réduisant le temps passé aux recherches fondées sur des éléments de fait et de droit comparables, l’utilisation des algorithmes permettrait ainsi aux juges et autres professionnels du droit de se décharger des tâches les plus chronophages au profit de l’examen des questions nouvelles ou plus complexes²⁸ *id est* dans les dossiers où leur expertise apporte une plus grande valeur ajoutée. Il découlerait de ces évolutions une plus grande confiance dans la justice, les jugements pouvant être purgés de leur part d’aléa et les juges étant délestés de tâches répétitives ou moins complexes, le tout au profit d’une justice plus rapide, sûre et efficace²⁹. Déjà intégrée dans le quotidien de nombre de professionnels du droit, la jurimétrie est envisagée comme un outil d’aide à la décision à l’avenir prometteur. C’est la raison pour laquelle, nous avons souhaité modéliser notre cadre d’évaluation en partant de cet **objet spécifique de recherche**³⁰.

²⁵ Tel que le prouve Antoine GARAPON dans son article: GARAPON A., “Les enjeux de la justice prédictive”, *JCP G* 2017, n° 01-02, p. 47.

²⁶ CADIET L. (dir.), *L’open data des décisions de justice*, Rapport au ministre de la Justice, nov. 2017, p. 14. Il s’inscrit en cela dans la droite ligne des écrits du Pr. Bruno DONDERO, dans lesquels l’auteur considère que l’expression « désigne non la justice elle-même, mais des instruments d’analyse de la jurisprudence et des écritures des parties, instruments qui permettraient de prédire les décisions à venir dans des litiges similaires à ceux analysés. » Cf. DONDERO B., “Justice prédictive : la fin de l’aléa judiciaire ?”, *D.* 2017, n° 10, p. 532.

²⁷ Abv. Modes alternatifs de règlement des différends.

²⁸ CASSUTO T., “La justice à l’épreuve de sa prédictibilité”, *AJ Pénal*, 2017, p. 334. Voir aussi les propos de Cédric Villani rapporté par RONFAUT L. dans l’article “Cédric Villani avance sur l’intelligence artificielle”, *Le Figaro*, 29 novembre 2017.

²⁹ GARAPON A., “Les enjeux de la justice prédictive”, *JCP G.*, 9 janvier 2017, doct. 31.

³⁰ Il est, toutefois, important de noter que les progrès et l’enthousiasme qui entoure cette technologie ne doivent pas masquer les risques qu’elle fait peser sur l’office du juge et l’accès à la justice. Au premier rang de ces risques se trouve celui où le juge, sous l’effet de la surveillance résultant d’un traitement massif des décisions de justice, pourrait perdre sa liberté d’appréciation et son indépendance en préférant se ranger à l’opinion dominante ou majoritaire de ses pairs. Nous l’avons vu avec le logiciel COMPAS. Or, le propre de la justice est que chaque affaire soit examinée pour ce qu’elle est, avec sa part d’originalité et d’irréductible complexité qui ne saurait être systématisée par un logiciel, aussi puissant soit-il. Par ailleurs, si la prévisibilité du droit est nécessaire, elle ne doit pas pour autant figer la jurisprudence. Parce que les avocats sauront plus sûrement demain quels sont les moyens fondés ou pas, et parce que les juges pourraient être dissuadés de s’écarter de la tendance majoritaire des décisions de justice, les résultats produits par les algorithmes risquent d’être répétés et amplifiés et toute décision “atypique”, même justifiée, risquera de paraître inacceptable, si elle n’est pas spécialement et très fortement motivée (v. BUAT-MÉNARD E., GAMBIASI P., “La mémoire numérique des décisions judiciaires. L’*open data* des décisions de justice de l’ordre judiciaire”, *Recueil Dalloz*, 2017, p. 1483.) Les algorithmes risquent ainsi de cristalliser la jurisprudence, alors que celle-ci doit au contraire être non pas rétrospective, mais apporter une solution concrète à un litige présent et, plus largement, accompagner les évolutions législatives, économiques et sociales. Ils risquent en outre de conférer une force excessive à des solutions majoritaires, mais pas forcément pertinentes : le nombre est une chose, la justice en est une autre. Ainsi, si l’évolution se confirmait et venait à acquérir la stabilité qui lui fait encore partiellement défaut aujourd’hui, la jurimétrie pourrait transformer profondément le monde du droit dans

- Un domaine du droit (contentieux commercial)

Enfin, afin de restreindre le champ de recherche de notre cas de départ, il nous semblait utile de nous limiter à un seul domaine du droit dans lequel l’outil de mesure jurimétrique semblait intéressant pour soutenir le praticien dans l’élaboration de sa stratégie. L’outil est principalement conçu/utilisé pour le contentieux et nous remarquons après un premier inventaire des outils développés mis sur le marché que ces derniers concernent principalement les contentieux de droit privé. Dès lors, nous avons arbitrairement décidé de nous intéresser aux contentieux privés et plus spécifiquement encore au contentieux commercial. D’un regard extérieur, il nous a semblé, peut-être à tort, que les acteurs judiciaires seraient plus enclins à collaborer à la réalisation d’une étude sur les indicateurs de qualité et de performance de l’IA judiciaire dans un domaine du contentieux privé où les clients ont de forts intérêts pour l’émergence et la diffusion de ce type de prestations.

- Conclusion: le terrain de recherche

In fine, la technologie de jurimétrie du cas d’usage de départ choisie fut une technologie **issue de la recherche publique** et disponible sur le marché depuis quelques années, ce qui lui permet de présenter plusieurs cas d’usage assez divers avec des clients de taille et de statut différents. Le domaine juridique associé est celui du **contentieux commercial, et plus spécifiquement de la propriété intellectuelle**. Notre terrain d’entrée est constitué par **un cabinet d’avocat, de taille moyenne**, spécialisé dans le contentieux et notamment dans les questions de propriété intellectuelle, mais pas uniquement. Le choix du cabinet s’est fait en fonction des contacts favorables de l’équipe.

1. Outil de jurimétrie appliqué au contentieux en droit d’auteur, des dessins et des modèles
2. Outil de jurimétrie appliqué au contentieux en droit des marques

Phase 1 bis : Observations du cas d’usage de départ

La première phase de l’étude consistait en la contextualisation et l’observation du cas d’usage de départ. Elle a commencé par une documentation générale sur les méthodes d’évaluation existantes, sur les offres techniques de solutions IA et toute autre littérature académique pertinente. Des contacts furent pris avec des membres d’autres chantiers des projets AJC, notamment des chantiers 2 et 8, afin de discuter de leurs travaux en cours sur l’évaluation et bâtir des ponts avec ces autres projets. L’observation de terrain du cabinet d’avocat mentionné se fit à l’occasion d’une **rencontre sur place** et de **trois entretiens semi-directifs** menés en visioconférence. Pour des raisons expliquées ci-dessous, aucune observation participante ou d’activité et d’utilisation ni de collecte de documentation technique n’ont pu être effectuées. Sans contraintes temporelles et matérielles et avec un meilleur accès au terrain, d’autres entretiens auraient aussi dû être menés. En effet, il était prévu que cette observation du cas d’usage soit étendue à d’autres clients de la même solution technique, en faisant varier les types de clients et leur taille, ce qui aurait fait au moins quatre observations supplémentaires. Aucun contact avec de tels clients n’a finalement pu être fait. Des efforts pour reprendre ce qui n’a pu être accompli dans le cadre de notre recherche restent néanmoins souhaitables. Il était également prévu que l’observation soit étendue à des cas d’acteurs juridiques voisins et comparables qui n’ont pas adopté cette solution technique. Chaque observation doit être accompagnée d’une documentation des protocoles d’activité existant avant l’introduction des solutions d’IA. Ceci a été fait lors des

son ensemble : ses acteurs, le raisonnement et la logique propres à la discipline, et à terme le droit lui-même. Tout dépend de ses fonctions, c’est-à-dire de ce qu’elle pourrait accomplir si on le lui permettait. Plus grande sera la confiance accordée dans ces outils, plus ce potentiel de transformation sera fort et les risques associés probables. Cette étude n’entend pas revenir sur les débats qui agitent la doctrine et la pratique à propos de ces outils. Néanmoins, pour la clarté du propos et pour une pleine compréhension de l’objet qui nous occupe, il nous apparaissait essentiel de revenir sur les controverses associées.

entretiens dès que possible, et ce dans le but d'obtenir une description des changements réels, au-delà d'une évaluation succès/échec.

Notre approche a pour but d'offrir un cadre d'évaluation flexible, adapté à chaque étape de développement d'un outil. Ceci est particulièrement important dans la mesure où les outils d'IA ou de cyberjustice sont encore nouveaux et peu utilisés. Le cadre d'évaluation proposé a ainsi pour but de se développer avec ces nouveaux outils, à mesure qu'ils se démocratisent et s'ancrent plus profondément dans les usages des professionnels du droit concernés. La grille d'entretien que nous avons suivie est donc elle aussi décomposée en plusieurs étapes, allant de la genèse de l'outil, son développement, son implémentation/utilisation, aux résultats et perspectives sur les utilisations et développements futurs de l'outil.

Phase 2 : Comparaison, modélisation et protocole d'évaluation

La deuxième phase de l'étude consiste en la comparaison, la modélisation et la construction du protocole d'évaluation. Une exploitation systématique de la comparaison devait être faite en construisant des grilles de comparaison selon un modèle d'analyse fonctionnelle. La comparaison prend en compte de façon extensive tous les traits qui pourraient être pertinents, afin d'élaborer une première version d'un modèle des stratégies d'appropriation. Cette comparaison, faite à partir des données qualitatives récoltées lors des entretiens, fut transposée dans une première version d'un protocole d'évaluation, avec une proposition d'indicateurs.

Phase 3 : Extension à d'autres solutions exploitant l'IA

Comme indiqué plus haut, la troisième phase devait consister en l'extension de notre observation à d'autres solutions exploitant l'IA. Précisément, une étude de cas d'usages dans le même domaine du contentieux mobilisant d'autres solutions techniques (déjà installées en organisation, hors des annonces, des présentations publicitaires ou des expérimentations en laboratoire) était prévue. Avec d'autres observations et selon les résultats obtenus, le protocole d'évaluation devrait être revu et ses indicateurs, facteurs et pondérations modifiés. Le modèle proposé a ainsi pour ambition d'être utile non seulement aux professionnels du domaine du contentieux commercial mais à d'autres domaines où sont utilisées des solutions exploitant l'IA. L'idée serait donc de mener à terme plus d'une étude de cas, afin de tester davantage nos indicateurs et généraliser au maximum notre modèle, tout en préservant une adaptabilité à des cas précis.

Phase 4 : Extension à d'autres systèmes juridiques et à d'autres pays

Afin de renforcer le protocole d'évaluation proposé dans cette étude, nous proposons qu'elle soit étendue en collectant de la documentation sur des cas d'usages (solutions installées en organisation) dans des pays de droit continental autres que la France et dans des systèmes de *common law* (ou mixte, comme au Canada). Le protocole d'évaluation devrait ensuite être révisé selon les résultats (indicateurs, facteurs, pondération).

Phase 5 : Extension à d'autres domaines du droit

Dans une phase encore plus avancée, les observations pourraient être étendues à d'autres domaines du droit, en sélectionnant quelques domaines où des solutions mobilisant l'IA concernent des acteurs judiciaires du même type. À la manière dont des évaluations menées dans le cadre d'autres domaines du droit ont informé cette étude *a posteriori*, sans qu'elles portent précisément sur des outils utilisant l'IA, d'autres évaluations pourraient être menées selon le protocole proposé ici pour fortifier un cadre spécifiquement adapté aux outils d'IA.

Phase 6 : Fourniture du modèle final de protocole d'évaluation

Enfin, une fois le protocole d'évaluation validé après itérations entre chaque extension d'observations (ce qui n'a pu être fait seulement jusqu'à la phase 3 dans le cadre de notre étude), une proposition finale d'indicateurs peut être faite. Les indicateurs que nous proposons ici doivent être

pris comme points de départ à adapter (avec d'autres facteurs et la pondération) en fonction de l'extension des observations que nous laissons à d'autres chercheurs ou professionnels du droit. Les indicateurs ne sont pas seulement des indicateurs de performance mais des points de veille pris en compte ou non par les usagers eux-mêmes, comportant plusieurs aspects pour traiter le problème éventuel. Il est ici question de promouvoir un pluralisme non-normatif des solutions ainsi qu'une dimension participative de l'évaluation, avec une implication active des usagers au long du protocole d'évaluation et de sa construction.

Limites du protocole d'étude et adaptation au terrain

En vue de la nature de l'objet et du terrain de recherche et des moyens à disposition pour conduire cette étude, la méthodologie proposée n'a pu être mise en œuvre que de manière relativement limitée. De nombreux obstacles ont été rencontrés au moment des premiers entretiens, qui malgré nos efforts nous ont empêchés d'accomplir toutes les phases d'étude prévues. Nous tirons néanmoins de ces obstacles des enseignements qui nous permettent de proposer des pistes de recherche plus réalistes pour continuer à étudier cet objet. Les limites rencontrées sur le terrain offrent aussi des éléments de compréhension sur la cyberjustice et les outils de jurimétrie, leur évolution et leur appropriation par les professionnels concernés.

Les principaux obstacles rencontrés concernent le manque de réactivité des premiers acteurs interrogés pour la poursuite d'observations participantes et d'entretiens supplémentaires ainsi que l'ouverture vers d'autres utilisateurs et d'autres solutions exploitant l'IA, soit *in fine* l'ouverture vers d'autres systèmes juridiques dans d'autres pays et domaines de droit. Notre protocole d'étude et notamment son volet comparatif à partir d'un cas précis n'a ainsi pu être porté plus loin que l'étude d'un cas de départ, soit la phase 3 du protocole. La comparaison est ainsi très limitée ; s'arrêtant à celle de deux outils reposant tous deux sur une unique solution exploitant l'IA. Au niveau des entretiens, seuls quatre entretiens ont pu être conduits. Alors qu'elles avaient été faites lors des entretiens puis renouvelées par écrit dans la limite du raisonnable, aucune suite n'a été donnée à nos demandes concernant la conduite d'observation participante (sur le terrain, des acteurs maniant les outils dans leur environnement professionnel) et la prise de contact avec d'autres utilisateurs (de la même solution d'IA dans d'autres domaines). Alors que des accords avaient été donnés à l'oral et qu'un certain enthousiasme avait été montré à l'égard de notre projet de recherche lors des entretiens, aucune réponse permettant sa poursuite ne fut donnée.

Les raisons qui expliquent la limite de notre étude et de son application concrète sont multiples. Les premières résistances du terrain furent remarquées par le délai de réponse de nos interlocuteurs et le besoin de parfois re-programmer des rendez-vous à la dernière minute. Les justifications données de la part des professionnels contactés portent sur leur charge de travail et les imprévus de leurs emplois du temps. Cela révèle, malgré leur intérêt annoncé pour notre projet et donc pour la production d'un cadre d'évaluation de leur outil, à laquelle ils participeraient par ailleurs, un certain manque d'adhésion, matérialisé par l'absence de dispositifs intégrant l'évaluation à leur pratique. Le manque d'intérêt effectif ou réalisé pour notre étude et à la fois l'évaluation ne s'explique pas entièrement par l'absence de procès d'évaluation pré-existant ou apte à accueillir et réaliser notre proposition de recherche et d'évaluation. Les résistances du terrain se trouvent aussi dans le développement encore très primaire des outils d'IA étudiés et leur faible utilisation auprès des acteurs interrogés. En effet, nos entretiens nous ont révélé, de manière plus ou moins explicite, que les outils que nous aurions aimé étudier voire observer, dans leur utilisation effective, n'étaient que très peu utilisés par des utilisateurs autres que ceux ayant participé à leur développement, ainsi qu'en dehors de ce cadre de développement. Autrement dit, les outils n'ont été "utilisés" que par leurs développeurs, et ce durant la période de développement de ces outils. Ce qui explique que nous n'ayons pas eu accès à une observation de ces outils en utilisation, tient non pas à ce que le temps manquait à nous faire rentrer dans ce terrain, mais qu'il n'existe pas à ce jour de moments où ces outils sont effectivement utilisés. Certaines des raisons pour cela, par exemple le temps de formation requis pour former des jeunes professionnels à l'utilisation de ces outils, ou encore leur manque de

performance et l'absence de données de qualité pour les nourrir, sont explorées dans nos résultats ci-dessous et forment une partie intégrante du cadre d'évaluation proposé. À ce stade où les outils ne sont pas parfaitement intégrés aux organisations, ils servent plutôt d'instruments de différenciation concurrentielle, qui attendent néanmoins idéalement d'être développés en vue de servir à l'efficacité et à l'autonomie des acteurs et de leurs pratiques. Si l'arrivée d'outils de jurimétrie sur le marché est réelle et reconnue par les acteurs interrogés, celui-ci n'en reste qu'à ses premiers balbutiements. Leur intégration profonde dans les pratiques rencontre ainsi des barrières relativement fortes liées à la culture de ces organisations.

Notre étude connaît quelques limites liées d'une part à l'état de développement des outils d'IA dans notre cas d'étude, ceux-ci étant encore en phase de développement, et d'autre part liées aux ressources temporelles et matérielles de l'équipe mobilisée sur ce projet. En effet, bien que la mobilisation de deux chercheurs occupés à temps partiel sur ce projet peut être considérée comme suffisante, elle aurait pu être plus efficace sur un temps plus long. En vue de l'ampleur et de la temporalité de l'objet d'étude, il aurait fallu la conduire sur plusieurs années, afin qu'elle puisse s'étendre au fil du développement des pratiques étudiées et s'investir dans des pistes alternatives d'entrée sur le terrain. Nous la comparons à ce titre au travail mené par Alexandra Pasca et Fabien Gélinas³¹ (projet CLEO), qui pour le développement d'un cadre d'évaluation similaire dans le contexte des outils légaux en ligne pour les plaideurs auto-représentés, ont un contact avec leur terrain de recherche plus régulier, depuis plusieurs années et avec d'autres supports matériels (e.g. des données plus accessibles, des outils de collecte de données, un objet de recherche plus propice à la construction de données, etc.). Cette comparaison a également fait ressortir qu'il était parfois question de créer des données ou des pratiques d'évaluation là où il n'y en a pas, ce qui nécessite de convaincre des professionnels dont l'intérêt pour la recherche n'est qu'indirect d'évaluer leurs outils, ce qui requiert en soi le travail de plusieurs personnes mobilisées sur le temps long. Si l'énergie dépensée à la recherche de contacts supplémentaires et d'observation de terrain demeure à ce jour sans réponse, nous ne pensons pas qu'un renouvellement du contact dans le moyen ou long-terme, selon les avancées de la cyberjustice, soit vain.

Ainsi, les pistes que nous proposons pour continuer et persévérer dans cette étude sont multiples : nous suggérons en premier lieu une attention soutenue aux développements de son objet, à savoir les outils d'IA dans la justice, afin de saisir toute occasion d'entrée sur un terrain où leur usage est significatif. De fait, le développement du cadre d'évaluation en question ne se fera que conjointement au développement de son objet, des pratiques et des institutions liées à leur usage. Nous avons ainsi à cœur de continuer à promouvoir le développement de cadres et de pratiques d'évaluation dans le domaine de la cyberjustice, dans le but général d'assurer la maîtrise et l'autonomie des acteurs dans l'utilisation de ses outils. Enfin, nous proposons que le cadre d'évaluation ici proposé soit utilisé dans d'autres cas d'études afin de tester et le cas échéant adapter ses indicateurs. Nous invitons donc tout acteur (chercheur ou professionnel du droit) à s'en saisir et, à partir de la méthodologie proposée, à continuer à l'adapter. L'apport d'autres cadres d'évaluation, provenant d'autres disciplines, sont les bienvenus afin de le consolider. Si ce travail de production d'un cadre d'évaluation est en cours de progression et qu'il a seulement été entamé dans le cadre de ce projet, il appelle à la mobilisation collective et à l'implication de diverses parties prenantes, ainsi qu'à leur engagement dans le temps long. L'objectif serait à terme d'avoir produit un cadre d'évaluation aisément modifiable et suffisamment général pour pouvoir vivre en autonomie et être adapté à l'usage de chaque profession et outil concerné. Il s'agit donc de résultats et d'un cadre "ouverts", dont le but est le développement futur dans un terrain qui s'ouvrira progressivement sans aucun doute.

³¹ Ces sujets ont été abordés avec eux lors d'une discussion au cours du projet de recherche (le 30 août 2022), avec celui de la difficulté de la création d'un cadre d'évaluation à la fois relativement universel et adapté à des outils spécifiques.

RÉSULTATS : CADRE D'ÉVALUATION ET INDICATEURS

Cette section présente le cadre d'évaluation et les indicateurs construits à partir des données qualitatives recueillies lors des entretiens semi-directifs menés auprès des acteurs judiciaires appartenant au terrain décrit ci-dessus. Différents types d'évaluation existent ; le cadre proposé ici ne correspond pas à une évaluation de seul l'impact qu'auraient les outils d'IA sur la performance des acteurs judiciaires. Celui-ci correspond également à une évaluation de la qualité de ces outils en tant qu'ils s'intègrent à et modifient la nature de leurs pratiques. Il est donc à la fois un cadre d'évaluation de processus ou d'implémentation et d'évaluation de résultat ("*outcome evaluation*"), en ce qu'il porte respectivement sur la qualité de l'intégration d'outils à des pratiques professionnelles et à celle de leur effet. Ce cadre d'évaluation propose donc des indicateurs quantitatifs (liés à la performance des outils) et des indicateurs qualitatifs (liés à la qualité des outils et de leurs effets). Par ailleurs, sans qu'elle fasse l'objet d'un indicateur explicite parmi les différents indicateurs de performance et de qualité proposés, l'autonomisation des acteurs est une valeur qui sous-tend l'ensemble du cadre d'évaluation.

Principes directeurs de l'évaluation

Nous posons ici quelques principes directeurs du cadre d'évaluation proposé, soit des principes qui le fondent et qui ont pour but de diriger ses évolutions futures. Parmi ceux-ci il y a les questions d'évaluation, qui permettent d'orienter l'évaluation vers certains objectifs et de construire des indicateurs pertinents pour ceux à qui ils vont servir. Afin de veiller à ce que l'évaluation, ses indicateurs ainsi que les résultats qu'elle produit soient de qualité, il est suggéré que certains critères d'évaluation, présentés ici, soient observés. Un autre principe directeur de l'évaluation est l'intégration de ses parties prenantes à chaque étape de sa réalisation, y compris de sa conception à l'utilisation de ses résultats.

Questions d'évaluation

La qualité d'une évaluation dépend de la clarté des questions qu'elle pose et auxquelles elle est censée répondre. Comme noté dans de nombreux guides d'évaluation de politiques publiques³², la bonne définition des questions d'évaluation est importante dans la mesure où elle détermine la nature des résultats que l'évaluation va générer. Ces questions doivent être posées au stade où le cadre d'évaluation est construit, et doivent continuer à l'informer à mesure qu'il se modifie. Dans le cadre des politiques publiques au Royaume-Uni par exemple, un guide du gouvernement sur la conception d'une évaluation³³ suggère que les personnes en charge de l'évaluation ou de sa réalisation se posent les questions ci-dessous, dans lesquelles nous substituons aux politiques publiques les outils d'IA :

- **Peut-il fonctionner ?** Comment l'outil a-t-il été délivré ? Comment a-t-il été mis en place ? A-t-il été mis en place correctement ? Quels ont été les obstacles à son implémentation et comment ont-ils été traversés ? Ces questions correspondent à une évaluation de processus ou d'implémentation, ce qui dans notre cas se rapporte à l'évaluation (des conditions) de la mise en place voire l'intégration des outils d'IA au sein des pratiques professionnelles dans le domaine du droit.
- **Fonctionne-t-il ?** Cette question se rapporte à l'évaluation d'impact soit à celle des impacts d'un outil dans le cadre d'une pratique juridique.

³² Pour une vue d'ensemble des guides d'évaluation existant dans le domaine des politiques publiques, voir le chapitre 2 du rapport de l'OCDE, *Improving Governance with Policy Evaluation: Lessons from Country Experiences*, OECD Public Governance Reviews, OECD Publishing, Paris, 2020 <https://doi.org/10.1787/89b1577d-en>.

³³ Voir le *Magenta Book* (publié en 2011 et dernièrement mis à jour en Mars 2020) du Trésor du Royaume Uni ("*HM Treasury*") et de son groupe de travail sur l'évaluation ("*Evaluation Task Force*"). Accessible [en ligne](#) (consulté le 10 janvier 2023), et le chapitre 2 du rapport de l'OCDE (2020) cité ci-dessus (Boîte 3.3).

D'autres questions proposées qui nous semblent moins pertinentes dans le cadre du présent projet mais qui peuvent toujours représenter un intérêt dans le cadre d'une autre évaluation ou dans son prolongement concernent la théorie du changement³⁴ qui sous-tend ou explique la manière dont un outil produirait un impact (comment l'outil fonctionne-t-il ?) ainsi que l'efficacité en termes de coûts d'un outil soit son rapport coût-bénéfice, ce qui se rapporte à son évaluation économique³⁵ (les coûts sont-ils justifiés par les avantages que l'on en retire ?).

Dans le cadre d'évaluation proposé par Alexandra Pasca et Fabien Gélinas³⁶, construit à partir de plusieurs modèles d'évaluation sur lesquels nous revenons ci-dessous, des questions sont posées à différentes étapes du "programme" ou, dans notre cas, de la mise en place à l'utilisation d'un outil d'IA dans une certaine pratique juridique. Leur grille d'évaluation associe donc les questions suivantes à chaque étape et au type d'évaluation correspondant³⁷ :

1. **Avant l'implémentation de l'outil** : dans quelle mesure le besoin pour lequel l'outil a été mobilisé est-il satisfait ? Qu'est-il possible de faire pour faire face à ce besoin ? Le type d'évaluation correspondant à cette étape est l'analyse ou l'évaluation des besoins ("*needs assessment*"), le besoin pouvant être défini comme l'écart entre un état actuel des choses et un état désiré³⁸. Cette étape implique de déterminer les besoins de la population pour laquelle l'outil sera conçu, ainsi que sa taille, sa socio-démographie, ses connaissances, ses compétences et ses barrières à l'accès ou à l'utilisation de l'outil (e.g. ressources, alphabétisation, culture)³⁹.
2. **Aux débuts de son utilisation**, soit quand l'outil est possiblement encore en voie de développement et d'amélioration : l'outil opère-t-il comme prévu ? Cette question se rapporte à l'évaluation de processus ou d'implémentation, comme celles présentées plus haut.
3. **Dès que l'outil est bien établi**, à savoir quand il est disponible à un large public et est utilisé⁴⁰ : l'outil accomplit-il ses objectifs ? Ce qui correspond à l'évaluation des résultats.
4. Enfin, **dès que l'outil atteint un stade de maturité**⁴¹ : quels sont les impacts prévus et imprévus de l'outil sur une population plus large d'acteurs ?

³⁴ La théorie du changement est une notion propre au domaine des politiques publiques et de leur évaluation. C'est pour sa valeur heuristique plutôt que dans le but d'une application systématique que nous la transposons ici au domaine des outils d'IA dans le droit.

³⁵ Si l'évaluation économique des outils d'IA dans le droit n'est pas traitée de manière directe dans le cadre de ce projet (qui porte principalement sur l'évaluation de leur performance et de leur qualité), les sections suivantes montreront que les bénéfices économiques, sous forme d'avantage commercial, de ces outils ne sont pas négligés par les acteurs qui les utilisent. Cette dimension économique est donc prise en compte dans le cadre proposé, mais pas sous la forme d'une évaluation économique en soi.

³⁶ Voir GÉLINAS F. *et al.*, *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at. 14, forthcoming. p.23.

³⁷ Nous adaptons les questions proposées dans leur grille à notre objet d'évaluation, dans la mesure où celles-ci sont pertinentes dans la formulation des principes directeurs de notre cadre d'évaluation. Aussi adaptons-nous les étapes auxquelles correspondent ces questions.

³⁸ BURKE JOHNSON, R. & DICK W., "Evaluation in Instructional Design: A Comparison of Evaluation Models (chapter 10)", *Trends and Issues in Instructional Design and Technology*, 2012, pp.96-98.

³⁹ GÉLINAS F. & PASCA A., *A Step-by-Step Evaluation Model for Legal Tools*, forthcoming, pp.23-24.

⁴⁰ Dans le modèle d'Alexandra Pasca et Fabien Gélinas (forthcoming : 23), cette étape est aussi décrite comme celle durant laquelle l'utilisation et la disponibilité de l'outil font qu'elles permettent la collecte de données sur ses usages et ses utilisateurs. Dans notre cas, si les entretiens menés constituent une collecte de données sur les usagers et leurs usages de l'outil d'IA étudié, elles se limitent à des cas rares et peu diffus dans l'espace et dans le temps. C'est ce qui limite notre propre évaluation mais n'empêche pas la proposition de recommandations pour la poursuivre.

⁴¹ Cette étape est définie dans le modèle d'Alexandra Pasca et Fabien Gélinas comme étant celle à laquelle l'outil a été utilisé suffisamment longtemps et de manière adéquate que pour fournir des données sur son impact sur la communauté dans un sens large (cf. GÉLINAS F. & PASCA A., *A Step-by-Step Evaluation Model for Legal Tools*, Figure 2: "Evaluation Chart", forthcoming, p.23).

Le modèle d'évaluation proposé ci-dessous s'est construit notamment en référence aux deux premières étapes, compte tenu de l'état dans lequel se trouve l'objet de notre étude. Si nos indicateurs sont construits à partir des données récoltées sur le terrain soit sur les besoins et les considérations exprimées par les acteurs eux-mêmes, ils peuvent ensuite être catégorisés comme se rapportant à ces étapes respectives. Si l'objet lui-même n'est pas à son stade final de développement, des réflexions concernant son développement futur soit son établissement et son éventuelle démocratisation (étapes 3 et 4) sont faites également. En cohérence avec la portée du chantier 12, les étapes de développement de l'outil sur lequel porte notre cadre d'évaluation correspondent à des évaluations de processus/implémentation et résultat (qualité) et d'impact (performance).

Critères de qualité

En deuxième lieu, un des principes directeurs que suit le cadre d'évaluation proposé consiste en l'assurance de la qualité de l'évaluation et de ses résultats. Si la qualité de l'évaluation est habituellement reconnue comme enjeu clé et parfois systématiquement traité dans le domaine des politiques publiques, il ressort moins explicitement dans celui de la justice. De fait, nous nous appuyons ici sur les critères de qualité d'évaluation mis en avant dans le domaine des politiques publiques pour aiguiller l'assurance et le contrôle de qualité de notre cadre d'évaluation. Si des institutions, des dispositifs et des professions existent dans ce domaine pour mettre en œuvre des mécanismes d'assurance et de qualité de l'évaluation, nous ne mettrons ici en avant que certains critères de qualité, tels que ceux présentés dans les guides ou documents officiels concernant l'évaluation de politiques publiques.

Les critères de qualité d'évaluation peuvent s'appliquer tout au long du processus de l'évaluation, de sa conception à l'utilisation de ses résultats, en passant par la collecte et le traitement des données. Deux catégories de critères qui se retrouvent dans de nombreux documents officiels liés à l'évaluation de politiques publiques⁴² portent sur la qualité technique des évaluations et sur leur gouvernance. À ce stade, s'il semble y avoir plus de matière à juger de la qualité technique des évaluations dans le domaine de la présente étude, la gouvernance des évaluations ne saurait être laissée de côté. À titre d'exemple et sans prétendre à l'exhaustivité, le tableau suivant présente certains des critères de qualité permettant de juger et de s'assurer de la qualité technique et de la gouvernance du cadre d'évaluation proposé :

Figure 1. Critères de qualité d'évaluation

<p>Critères de qualité technique de l'évaluation</p>	<ul style="list-style-type: none"> ● Identification et conception de l'approche à l'évaluation choisie (e.g. en établissant un plan d'évaluation exposant les besoins liés à l'évaluation, les questions de l'évaluation, les informations nécessaires à la construction d'indicateurs, les méthodes de collecte de données, etc.) ● Définir l'objectif de l'évaluation et l'usage prévu de ses résultats ● Définition du processus de commission de l'évaluation (le cas échéant) ● Élaboration d'un calendrier d'évaluation (suffisamment à l'avance) ● Planifier les ressources (financières et humaines) nécessaires et mobilisées pour l'évaluation ● Définition et justification des méthodes de collecte de données utilisées (qui doivent être représentatives et suffisamment nombreuses)
-------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

⁴² Ceux-ci sont répertoriés dans OCDE, *Improving Governance with Policy Evaluation: Lessons from Country Experiences*, Chapitre 2, OECD Public Governance Reviews, OECD Publishing, Paris, 2020.

	<ul style="list-style-type: none"> ● Définition de la méthode d'évaluation utilisée ● Mise en place d'un suivi de l'évaluation
Critères de qualité de la gouvernance de l'évaluation	<ul style="list-style-type: none"> ● Indépendance de l'évaluation (non-soumission à des pressions politiques ou organisationnelles) ● Implication des parties prenantes dans les décisions prises à l'égard de l'évaluation ● Transparence du processus d'évaluation ● Éthique dans la conduite de l'évaluation (e.g. s'assurer du consentement dans les données collectées ou de leur anonymisation, et de la vie privée des agents concernés, éviter les conflits d'intérêts, etc.)

Sources : OCDE, Tableau 3.2. "Quality standards included in evaluation guidelines", 2020.

Outre les critères de qualité présentés sous formes de standards à remplir, il existe d'autres mécanismes tels que la revue par les pairs, la méta-évaluation, et l'acquisition de compétences ou la formation en matière d'évaluation. Si les conditions matérielles facilitant la mise en place de ces mécanismes ne sont pas nécessairement accessibles, nous recommandons toutefois une simple veille vis-à-vis des critères de qualité présentés ci-dessous, voire une assurance et une vérification de leur satisfaction par quiconque engagé dans une évaluation.

En matière de qualité d'évaluation, le cadre des *key performance indicators* (KPI), mobilisé dans le domaine de la gestion de la performance organisationnelle, offre aussi des indications utiles. Trois étapes servent à s'assurer de la qualité d'indicateurs : la première est l'identification des besoins auxquels ces indicateurs répondent. La réflexion en amont étant essentielle, elle doit être suffisamment approfondie pour déterminer des indicateurs pertinents. En deuxième étape, il faut s'assurer que les indicateurs (d'évaluation) sont bien alignés avec la stratégie de l'entreprise et ses objectifs. Enfin, et en troisième lieu, les indicateurs doivent fournir des plans d'action : seuls les indicateurs actionnables se révéleront utiles.

Engagement des parties prenantes

La qualité d'une évaluation dépend de l'intégration des parties prenantes à l'ensemble de son processus, car elle assure que l'évaluation produise des résultats utiles à leur égard et qu'ils les utilisent effectivement. Si une section ci-dessous porte spécifiquement sur l'utilisation du cadre d'évaluation, celle-ci se penche sur les différentes manières dont les parties prenantes (et lesquelles en particulier) peuvent s'engager dans les étapes successives d'une évaluation. Selon les différents acteurs concernés, il est utile de s'informer au préalable sur leurs besoins quant à l'évaluation et à son objet. Dans notre cas, les parties prenantes à considérer sont⁴³ :

- les développeurs du système d'IA
- les avocats et autres professionnels du cabinet qui les utilisent (dont stagiaires, assistants juridiques ou juristes)
- les pairs appartenant à d'autres cabinets (dans d'autres domaines du droit)
- les clients (directs et potentiels)
- les conseils juridiques d'entreprises
- les magistrats
- les pouvoirs publics à même de pouvoir modifier le cadre ou les normes juridiques

⁴³ Cette liste, qui répertorie les parties prenantes rencontrées dans notre cas, est non-exhaustive dans la mesure où elle n'est pas universelle et peut-être amenée à être modifiée selon les cas précis d'application et/ou leur évolution.

- la société civile de manière plus large, en tant qu'elle peut-être affectée par la cyberjustice (e.g. par la gestion et la protection des données qu'elle implique ou par l'effet sur la nature des décisions prises)

Tous sont mobilisés à différentes étapes et de manière plus ou moins directe dans le développement de l'outil en question et dans le processus d'évaluation. L'implication des parties prenantes peut se faire à l'étape de conception de l'évaluation, y compris celle de la définition de sa méthode et de son application, dans la participation à la collecte de données ou encore dans la vérification de leur qualité et de celle des résultats de l'évaluation. Les résultats de l'évaluation doivent enfin être communiqués clairement à ses parties prenantes au fur et à mesure de sa progression, de manière à favoriser leur appropriation des résultats et à en faire un usage utile à leur pratique et d'assurer que l'évaluation réponde continuellement à leurs besoins.

Conclusions tirées des entretiens

Pour éviter d'adopter une posture hors-sol, la conduite d'observations de terrain devait nous permettre d'identifier les véritables besoins, difficultés et questionnements des acteurs ayant recours à des systèmes d'IA dans le cadre de leur pratique. Notre série d'entretiens devait ainsi nous permettre de relever les premiers indicateurs et variables. Ces entretiens ont été semi-dirigés.

Il s'agit ici de rendre compte des grandes conclusions tirées des entretiens selon les différentes thématiques qui les ont structurés : la phase d'initiation ou de la mise en place de l'outil d'IA, la phase de développement de l'outil, son exploitation à des fins organisationnelles, et son exploitation à des fins commerciales. Nous synthétisons donc nos principaux apprentissages quant à ces diverses phases de développement et d'utilisation de l'outil d'IA par les praticiens interrogés, et nous appuyons sur ceux-ci dans la partie suivante pour développer les indicateurs et le cadre d'évaluation.

Phase d'initiation

Nous avons pu déduire des échanges la place fondamentale que tenait le client dans l'initiative du développement de tels outils. Les outils de quantification du risque et de mesure de la performance sont régulièrement employés dans le cadre d'activités commerciales. Ainsi, les cabinets d'avocats sont eux aussi confrontés à cette tendance et nombre de clients souhaitent obtenir de leur conseil une évaluation en pourcentage de succès ou de risques encourus de leur dossier. Évaluer en pourcentage est une tâche impossible à l'échelle humaine, de sorte que les avocats que nous avons interrogés se sont intéressés au développement d'outils qui permettent d'automatiser cette évaluation.

Plus qu'une simple demande de la clientèle, l'utilisation de ce type d'outils algorithmique a pu se justifier du point de vue de la stratégie commerciale du cabinet en ce qu'elle permet de se différencier de concurrents tout en renforçant l'argumentation commerciale par l'objectivisation de la stratégie et donc de la prise de décision.

Concomitamment le développement de ce types d'outils apparaissait aux yeux de certains comme un moyen de contrôler la qualité et la performance de leur activité de juriste *i.e.* la fiabilité de leur raisonnement: "se faire une idée de la situation", "se donner un sentiment sur ses chances de succès" ou "se convaincre de la force de son dossier".

Le recours aux systèmes d'IA a pu, enfin, être perçue comme un moyen permettant de dégager du temps que les avocats pourraient consacrer à des activités qui constituent "le coeur de leur métier", comme de "changer la jurisprudence" par exemple⁴⁴, plutôt que de l'analyser⁴⁵. Les praticiens

⁴⁴ « Le but ou le travail des avocats, c'est changer la jurisprudence » (Entretien)

⁴⁵ « Le but n'est pas que les avocats fassent l'analyse de la jurisprudence » (Entretien) « Si la machine ou un langage peut nous aider à faire l'analyse c'est super. » (Entretien)

peuvent ainsi se consacrer à des activités jugées comme plus nobles, et ne pas “perdre leur temps”⁴⁶ sur d’autres activités comme, par exemple, “la vérification de références”⁴⁷.

Phase du développement

Le développement des outils s’est fait en interne avec le soutien de prestataires extérieurs. Les relations tant avec la machine que les développeurs se sont faites sur la base d’une confiance mutuelle. La confiance est angulaire dans la phase de développement de l’outil sans pour autant qu’elle soit aveugle. La pertinence des premiers résultats était contrôlée. De même le rôle qu’a pu tenir le praticien dans le choix des critères du barème algorithmique fut central. La sélection par le praticien des critères (80 à plus de 100 selon l’outil) était essentielle pour la compréhension des résultats. Cette sélection s’est faite grâce à un fastidieux travail d’analyse des décisions de justice. C’est, en tout, une vingtaine d’avocats juniors ou seniors, de collaborateurs et stagiaires qui ont contribué au développement de l’outil. La sélection des critères n’est toutefois pas suffisante pour obtenir un résultat satisfaisant. Les praticiens interrogés nous rappellent la nécessité d’alimenter l’algorithme de données rigoureusement sélectionnées manuellement (“La machine n’apprend qu’avec ce qu’on lui donne!”). La lecture manuelle des décisions peut paraître laborieuse mais le praticien affirme qu’il s’agit de la seule fiable. Les avocats interrogés ont bien conscience que la lecture humaine comporte des failles mais estiment la lecture automatique utopique. Cet exercice de lecture peut s’avérer quelquefois difficile notamment sur certaines disciplines comme le droit d’auteur. Par exemple, les données liées à l’originalité de l’œuvre et à ce qui fait l’empreinte de la personnalité de l’auteur ne sont pas objectives ou facilement objectivables. Sur l’outil de jurimétrie consacré au design, les résultats sont donc à prendre avec le plus grand recul.

Exploitation à des fins organisationnelles

Toutefois, le principal intérêt que pouvait présenter le développement d’un tel outil pour le cabinet était d’ordre organisationnel. Il devait faciliter la pratique de l’avocat en l’assistant dans les tâches qu’il devait accomplir. Le premier constat a été paradoxalement que les praticiens seniors se sont davantage approprié l’outil comparativement aux jeunes avocats. L’un des praticiens que nous avons interrogé avance l’hypothèse que les jeunes praticiens ont une tendance naturelle à effectuer une recherche manuelle telle qu’enseignée à l’Université car ils sont encore en phase d’apprentissage. Le praticien remarque, d’ailleurs, que leurs jeunes collaborateurs maîtrisent parfaitement les outils de recherches et bases de données documentaires mais ont été insuffisamment sensibilisés aux outils d’IA ou de justice simulative comme ceux développés par le cabinet. Au fur et à mesure du temps l’outil a perdu en attractivité et est aujourd’hui peu utilisé dans le cadre des missions pour lesquelles il a été conçu. Ce délaissement peut être justifié de plusieurs manières. Tout d’abord, seuls ceux qui ont participé à l’élaboration de l’outil le connaissent et l’utilisent. La formation et la sensibilisation à l’outil ne sont pas insérées au processus d’intégration des nouveaux collaborateurs. Par ailleurs, les données de l’outil (qui doivent être manuellement intégrées) ne sont pas mises à jour régulièrement. L’un des avocats interrogés nous indique que pour optimiser le SIA, ils auraient dû rester en contact avec ses développeurs. Ce n’est actuellement plus le cas. Enfin, l’outil développé pour la discipline “design” (droit d’auteur, des dessins et des modèles) ne semble pas adapté à la subjectivité inhérente à la discipline. L’outil ne peut être pleinement efficace si des visuels ne sont pas intégrés aux décisions de justice. Cela supposerait une modification formelle des décisions mais également l’intégration de technologies de reconnaissance visuelle au sein de l’outil. À contrario, l’outil reste utilisé par certains praticiens mais davantage comme outils de recherche documentaire affinés. L’un des avocats nous explique que les outils généralistes qui fonctionnent sur la base de l’*open data* des décisions de justice ne permettent pas d’aboutir au même niveau de précision des résultats de recherche que ceux obtenus avec l’outils développé en interne car l’efficacité des outils dépend de la façon dont ils ont été conçus

⁴⁶ (Entretien)

⁴⁷ (Entretien)

et des données qui ont été intégrées. Or la sélection des critères qui s'est faite grâce à l'expertise des avocats spécialisés et l'alimentation manuelle en données de l'outil leur permet d'atteindre une plus grande qualité et précision des résultats. Dans ce cadre, le recours à l'outil développé en interne par un avocat spécialisé en capacité d'en décrypter les résultats peut être particulièrement efficace et opportun. Un dernier aspect auquel les acteurs interrogés semblent accorder une certaine importance est le potentiel de systématisation des pratiques professionnelles que représente l'utilisation d'un outil numérique. Pour un des participants à l'étude, ce qui compte c'est de "nous rendre le travail plus facile". La systématisation des pratiques relèverait d'un indicateur de performance, d'impact ou d'efficacité de l'outil numérique en question.

Exploitation à des fins commerciales

L'outil ainsi développé n'a pas fait l'objet d'une exploitation commerciale exclusive par le cabinet qui a souhaité le mettre à disposition du plus grand nombre. L'exploitation commerciale de l'outil s'est indirectement faite *via* son intégration au discours commercial et à la stratégie marque employeur. D'une part, l'outil s'est révélé être un formidable accélérateur du processus d'acquisition de clients en permettant de "mettre le pied dans la porte". Il était ainsi utilisé à la manière d'un produit d'appel pour permettre d'engager une première approche auprès d'un potentiel client qui se trouve davantage mis en confiance par un professionnel assisté d'un système d'IA capable de "vérifier" semi-objectivement le travail manuel qu'il effectue⁴⁸. La confiance accordée par le client aux professionnels du droit peut ainsi être liée à la légitimité scientifique que l'outil leur donne. Cette légitimité scientifique relève en partie de l'ordre de l'apparence⁴⁹, mais reste ancrée dans un réel appui que les acteurs prennent sur l'IA pour donner de meilleurs conseils. Nos entretiens révèlent malgré tout que le facteur humain reste essentiel et au centre du développement d'une relation commerciale vertueuse. D'autre part, l'outil est apparu intéressant à exploiter dans le cadre de la stratégie marque employeur en ce qu'il permet d'attester le dynamisme et l'esprit d'innovation du cabinet. À terme, le recours à ce type d'outils peut ainsi permettre d'attirer les talents, en particulier les nouvelles générations sensibilisées à ces nouvelles méthodes du "juriste augmenté".

Sélection des indicateurs

Nous proposons de fonder l'évaluation sur un contrôle de qualité aux différentes phases de conception et d'utilisation des systèmes d'IA parmi les professions de justice. Les concepteurs peuvent alors choisir d'accentuer plus ou moins les dimensions centrées utilisateur mais leurs choix sont rendus explicites par ce guide de description des phases, des tâches et des allocations de rôle. Ce guide permet de rendre visible (et donc discutable) le design organisationnel, parfois délibéré mais trop souvent tacite, qui est encapsulé dans un processus de conception d'IA. Selon les secteurs et les situations, la combinaison des différentes phases et tâches peut varier mais d'une façon ou d'une autre, on retrouve tous ces éléments incontournables pour produire un algorithme et au bout du compte un programme fiable.

Nous reprenons ainsi les quatre thématiques autour desquelles se sont structurés les entretiens, à savoir la phase d'initiation, la phase de développement, l'exploitation à des fins organisationnelles, ainsi que l'exploitation à des fins commerciales, et les étayons en étapes ou tâches correspondantes qui suivent ce même mouvement afin d'en faire ressortir les indicateurs ou critères de qualité qui permettent d'évaluer chacune de ces étapes. Ces phases et les étapes correspondantes sont résumées dans la figure 2 ci-dessous.

⁴⁸ « On le fait parce que c'est le client qui l'a demandé » (Entretien)

⁴⁹ « La machine [...] donne à notre pratique un vernis scientifique » (Entretien)

Figure 2. Phases et mise en oeuvre d'un outil d'IA

<i>Phase d'initiation</i>				
1. Sonder les attentes				2. Sonder les moyens
1.1. Demandes	1.2. Offres	1.3. Cible	1.4. Interface Design	2.1. Barrières
<i>Phase préparatoire</i>				
3. Constitution de la base de données d'apprentissage	4. Preprocessing et nettoyage		5. Feature extraction	
<i>Phase de développement</i>				
6. Méthode de calcul	7. Présentation/visualisation des données			
<i>Phase d'utilisation</i>				
8. Protocole de traitement permanent en routine	9. User experience		10. Design organisationnel	11. Modèle économique
12. Résultats				

Les grilles développées ci-dessous, contenant les indicateurs correspondant à chaque phases et étapes définies ci-dessus ont une portée indicative : elles présentent une liste d'indicateurs préconisés pour chacune de ces étapes ainsi que, le cas échéant, des sous-indicateurs ainsi que les valeurs possibles que ceux-ci peuvent prendre. Chaque grille est à prendre comme objet à construire et à alimenter, selon le contexte et l'outil des acteurs en question. Elles n'ont ainsi aucune prétention à l'exhaustivité et nous invitons, en tant que non-spécialistes de leurs domaines, les acteurs et actrices du droit à se les approprier et à les modifier à leur gré et au fil de leurs usages. Par ailleurs, comme l'indiquent les cases de gauche apposées à chaque grille (contenant les questions "Qui évalue ?", "Qui renseigne les valeurs possibles de chaque indicateur ?"), nous préconisons que leur usage soit accompagné d'une indication de ou des personne(s) qui se chargeront du renseignement des indicateurs d'évaluation et leurs valeurs. A la discrétion des usagers, ces choix ont pour visée une certaine formalisation et intégration des processus d'évaluation dans les process et pratiques des acteurs concernés.

1. Sonder les attentes

Développer ou adopter un outil de traitement quantitatif des décisions de justice à des fins de jurimétrie nécessite, dans la phase d'initiation du projet, de sonder les besoins. Quels sont mes

besoins? En ai-je besoin pour ma pratique? pour l'optimisation organisationnelle? ou commerciale ? L'adoption ou le développement de l'outil est-il orienté client ou praticien? Ce travail engage une réflexion sur les cibles qu'il faudra par la suite poursuivre. Il est nécessaire de sonder les usages dès la conception de l'outil pour qu'il puisse être adapté aux besoins de l'organisation ou dès l'adoption pour identifier l'outil le plus adapté aux besoins de l'organisation. Ce sondage des attentes du client et/ou du praticien est à faire par un professionnel du droit⁵⁰.

<i>Phase d'initiation</i>				
1.1. Demandes				
	Indicateur	Sous-indicateurs	Valeurs possibles	Faisabilité⁵¹
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Techniques (Problèmes juridiques)	[Dépend des cas]	Qualification d'un besoin technique, de l'objectif technique de l'outil	
	Commerciales	Stratégie marketing	Dépend de l'organisation et du ou des marché(s) dans le(s)quel(s) elle se situe	
		Stratégie marque-employeur		
		Stratégie de communication		
	Organisationnelles	Délégation de tâches (quantité/temps)	Dépend de l'outil en question	
		Vérification de tâches (qualité)		

En fonction des usages, il faudra définir les fonctionnalités de l'outil et le degré d'automatisation des tâches attendu. Il convient d'identifier précisément ces fonctionnalités et les valeurs attendues en résultat. Ai-je besoin d'un outil qui fournit un résultat unique et donc à fort potentiel d'automatisation des tâches (par exemple, le pourcentage de succès d'une stratégie contentieuse en fonction de faits)? Suis-je plutôt à la recherche d'un outil pour traiter un grand volume de données et m'en faire un résumé sans nécessairement produire un résultat unique? Quelles sont les valeurs attendues: des probabilités ? des statistiques? des données textuelles?

⁵⁰ HOMOKI P., "Guide on the use of Artificial Intelligence-based tools by lawyers and law firms in the EU", *Council of Bars and Law Societies of Europe*, Brussels, 2022, p.50.

⁵¹ Faisabilité de la mesure des indicateurs (légende) : en **vert**, ceux qui sont facilement mesurables ; en **orange**, ceux qui nécessitent la mise en place de nouvelles pratiques dans l'organisation concernée ; en **rouge**, ceux qui sont difficilement mesurables, qui nécessitent la mise en place de nouvelles pratiques et l'accès à des données ni envisageables ni accessibles en l'état. Ces différents niveaux de faisabilité sont indicatifs ; c'est aux acteurs de juger de la faisabilité selon leur contexte particulier.

<i>Phase d'initiation</i>			
1.2. Offres			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Informations juridiques	Dépend de l'outil et de l'organisation en question	
	Aide à la décision	- Routine (automatisation) - Seuils critiques (avis d'experts) - Exploration de possibles	
	Prise de décision ⁵²	- Nature des décisions, - Qui les prend - Sur quelle base de données	

Concomitamment il est nécessaire de mener une étude sur les cibles à “atteindre”, tant à l’interne qu’à l’externe. En cartographiant ces cibles et en sondant leurs attentes par le biais d’enquêtes, d’entretiens semi-directifs et toutes méthodes d’engagement des parties prenantes issues de la littérature sur la co-conception ou conception participative⁵³, il sera alors possible de développer ou sélectionner les outils adaptés aux usages particuliers de chaque cible. Si l’outil est développé en interne, il est essentiel que l’ensemble des parties prenantes puissent participer aux échanges et phases de test. Un outil de qualité et performant est nécessairement centré utilisateur.

<i>Phase d'initiation</i>				
1.3. Cibles & parties prenantes				
	Indicateur	Sous-indicateurs	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Cartographie des cibles à l’interne	Les praticiens seniors	Énumération et catégorisation des cibles internes : - nombre - statut - poste - ancienneté - rapport à l’outil (plus ou moins intégré)	
		Les praticiens juniors		
		Les stagiaires juridiques		
		Les para-juristes		

⁵² Cf. partie introductive sur le rôle de l’IA dans l’aide à la prise de décision.

⁵³ En parallèle, dans le domaine de l’évaluation de politiques publiques, il y a un consensus dans la littérature académique (PATTON M., *Utilization-focused evaluation*, Thousand Oaks: SAGE, 2008 ; KUSTERS C. et al., “Making evaluations matter: a practical guide for evaluators”, Centre for Development Innovation, Wageningen University & Research Centre, 2011) sur le fait que l’engagement des parties prenantes dès le début d’une évaluation permet d’améliorer son design, sa pertinence, sa transparence ainsi que l’utilisation de ses résultats.

		Le personnel administratif		
Cartographie des cibles à l'externe	Les clients (profils)	Énumération et catégorisation des cibles externes : <ul style="list-style-type: none"> - nombre - activité - rapport à l'organisation - besoins (voir ci-dessus) 		
	Les autres praticiens (captation de l'attention des talents)			
	Les étudiants			
	Autre			
Socio-démographie ou socio-organisation des cibles	À établir selon les cibles en faisant appel à un design organisationnel (voir plus bas)	Selon la cible (personne/population/organisation), reporter : <ul style="list-style-type: none"> - Sa localisation (géographie) (e.g. un cabinet parisien vs. un cabinet dans une petite commune) - Sa hiérarchie - Sa culture (e.g. culture de l'entreprise) - Ses compétences - Son genre (ou distribution des genres) - Sa taille (nombre d'employés) - Son âge (ou la distribution des âges) 		
Construction des personas	A établir pour chaque catégorie identifiée	Pour chaque cible, construire les personas selon les caractéristiques socio-démographiques identifiées ci-dessus		

Naturellement une conception orientée utilisateur nécessite de prendre en considération son expérience: expériences passées avec des outils numériques, comprendre sa pratique quotidienne du droit et ses attentes vis-à-vis d'un nouvel outil. Cela inclut la prise en compte du ressenti des utilisateurs ainsi que leurs comportements face à une interface digitale. Cette étape est extrêmement importante pour l'implémentation des outils dans la pratique des utilisateurs (avocats, juristes, assistants...). L'objectif est alors d'optimiser le parcours de ces derniers pour le rendre le plus agréable, accessible et intuitif possible. L'interface et plus globalement le design des outils adoptés ou développés doivent correspondre aux attentes des utilisateurs et être adaptés à leurs pratiques

professionnelles. Ces attentes envers le design de l'interface doivent être cartographiées avant tout projet d'adoption d'un outil ou idéalement dès son développement.

<i>Phase d'initiation</i>				
1.4. Interface Design				
	Indicateur	Sous-indicateurs	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Choix des architectures	Médium	À définir selon les choix opérés (le plus précisément et exhaustivement possible)	
		Options		
		Présentation		
	Complexité de la tâche	Instructions		
		Tâches		

--	--	--	--	--

2. Sonder les moyens

La dernière étape à effectuer avant d'initier un projet d'adoption ou de développement d'un outil est celle de la cartographie des obstacles qui pourraient empêcher la réalisation et la réussite du projet. Il convient dès lors de sonder les ressources à disposition et de s'assurer d'un contexte favorable, aussi bien à l'interne qu'à l'externe.

<i>Phase d'initiation</i>				
2.1. Barrières				
	Indicateur	Sous-indicateurs	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i>	Ressources	Ressources financières disponibles	À définir de la manière la plus précise et exhaustive possible	
		Ressources "temps" disponibles		
Ressources humaines disponibles (quantité/capacité)				
<i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Contexte interne	Culture de la structure (histoire, raison d'être, stratégie, capacité au changement)	À définir de la manière la plus précise et exhaustive possible	
		Modèle organisationnel		
		Soutien des décideurs		
		Soutien des utilisateurs primaires		
		Soutien des utilisateurs secondaires (para-juristes, documentalistes, secrétaires, équipes administratives...)		

	Contexte externe	Juridique (instabilité du droit i.e. projet réforme/codification à droit constant...)	À définir de la manière la plus précise et exhaustive possible (selon les données disponibles, actuelles et prospectives)	
		Organisationnel (accès aux données, forme/qualité des données, adaptabilité de l'outil à l'organisation juridictionnelle...)		
		Économique (étude de marché, instabilité économique générale...)		
		Social (acceptation de l'innovation..)		
		Technologique (limites techniques au besoin)		
		Environnemental (limites environnementales à la technique...)		

CONCLUSION PHASE D'INITIATION: L'introduction d'une application issue de l'IA ne peut se contenter de réaliser un benchmark sur les performances des systèmes proposés: elle doit reposer sur une description et une compréhension des formats organisationnels explicites (organigrammes, workflows, protocoles, etc.) et implicites (ce qui relève de la gageure quelquefois). Cette calculabilité nouvelle permet en effet de rendre plus explicite des traditions anciennes dans l'organisation et d'en débattre pour décider, en intégrant les parties prenantes. Ainsi, les flux de demandes d'avis révèlent souvent des préférences dans les réseaux internes d'un cabinet ou de tout autre collectif: si l'IA est supposée répondre à certaines de ces demandes, qu'en est-il de l'influence exercée par ceux qui répondaient auparavant? Ne peuvent-ils pas avoir un rôle particulier de contrôle de qualité des réponses proposées par l'IA par exemple? En réalité, c'est tout le "design organisationnel" qui mérite d'être repris, explicité et discuté. Non pas pour exiger que tout le monde s'adapte mais pour tenir compte de la réallocation des rôles, des ressources, du temps, etc. Une stratégie *d'appropriation*, qui plie l'offre technique aux propriétés et aux choix de l'organisation, est celle qui comporte les meilleures chances de succès et d'intégration réussie de certaines fonctions des IA dans les process, à la satisfaction des différentes parties prenantes.

3. Constitution de la base de données d'apprentissage

Durant la phase préparatoire nous nous intéresserons principalement aux données et à leur traitement. Une attention toute particulière doit être accordée à cette phase quand on souhaite développer un outil. Bien entendu, si on adopte un outil d'ores et déjà développé ces grilles pourront être utiles pour

comprendre la façon dont l’outil a été développé ainsi qu’à la façon dont l’outil est ou sera alimenté. Nous nous intéresserons dès lors tant à la nature des données, à leur quantité, à leur qualité qu’à leur traitement. Le développement d’un outil nécessite, d’abord, la constitution d’une base de données, laquelle permettra à l’algorithme d’apprendre et donc *in fine* de répondre aux besoins de chaque cible formulés durant la phase d’initiation.

<i>Phase préparatoire</i>			
3. Constitution de la base de données d’apprentissage			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ? Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Sources (et validation des sources)	Présence (ou non) d’un répertoire des sources	
	Hétérogénéité	Nombre de bases de données utilisées Type de base (images, etc.)	
	Interopérabilité	Selon quel langage ? Quels standards ? Portée de l’interopérabilité	
	Conversions	Des formats notamment, avant traitement	
	Extension (degré de couverture du domaine)	À définir selon l’outil et l’objectif de l’organisation	
	Adaptation aux potentiels d’exploration	À définir selon l’outil et l’objectif de l’organisation	
	Adaptation aux visées stratégiques	À définir selon l’outil et l’objectif de l’organisation	
	Volume suffisant	Degré d’exhaustivité	
	Actualisation suffisante	Fréquence à laquelle la base de donnée est alimentée/mise à jour À jour ou non	
	Granularité suffisante	Niveau de décomposition des classes/agrégats, accès aux sources primaires	
	Comparaison avec les bases juridiques classiques	Disponibilité des autres bases pertinentes	
	Identification des biais possibles et tests pour vérifier	Traçabilité de la constitution de la base de données Identification des degrés de risque de ces biais	
	Partition du dataset pour apprentissage et test	Définition des critères de partition	
	Propriétés sémiotiques requises	Quelle visualisation notamment	

4. Preprocessing et nettoyage

Une telle base de données ne peut naturellement pas être utilisée afin de produire des résultats de qualité, et donc développer un outil performant, si elle ne subit pas un “nettoyage”, précédant le calcul.

Phase préparatoire			
4. Preprocessing et nettoyage			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Doublons, incohérences, contradictions	Nombre Présence (oui/non) et correction possible (ou non) le cas échéant	
	Valeurs aberrantes	Nombre Présence (oui/non)	
	Valeurs manquantes	Nombre Quel traitement pour les valeurs manquantes?	
	Formats incompatibles	Oui/non	
	Instabilité des flux	À déterminer selon les sources et les process de nettoyage	
	Modèles de langue pour conversion, traduction et NLP	À faire de manière exhaustive	
	Validation des métadonnées (dates, lieux, sources, historiques, etc.)	Absentes/présentes Complètes (ou non)	
	Réitération du nettoyage	Par qui ? À quelle fréquence ? Selon quels critères ?	
	Intégration à un processus de contrôle de qualité	Ce processus doit être défini selon les critères suivant : - Qui le conduit - À quelle fréquence - Selon quels moyens - Avec quels objectifs/selon quels standards (de qualité) - Quels sont les résultats de ce contrôle ? - Feuille de route pour la correction des données	

5. Feature extraction (détermination des propriétés pertinentes des problèmes pour engendrer les calculs)

Ici il s'agit d'établir un barème algorithmique. Le barème, sous réserve qu'il soit largement utilisé, est un outil qui tend à garantir une meilleure répétabilité et donc une meilleure prévisibilité des décisions. Il permet de formaliser le raisonnement suivi : il sélectionne les critères jugés pertinents auxquels il associe un résultat. Ce faisant, il hiérarchise les critères retenus et délaisse d'autres critères, jugés secondaires. On s'intéresse également ici à la manière dont ces critères, classes ou clusters, sont créés. Il faut alors s'interroger sur les méthodes d'extraction et de validation.

<i>Phase préparatoire</i>			
5. Feature extraction			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ? Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Méthode d'extraction	À définir selon la méthode choisie, éventuellement à partir d'un benchmark sur les différentes méthodes d'extraction	
	Terminologie prédéfinie	Source	
	Ontologie complète ou pas (IA symbolique ou connexionniste)	Oui/non	
	Construction a priori de classes	Oui/non, selon les besoins	
	Nombre de classes (ou de clusters)	Selon les besoins	
	Labellisation des clusters	À faire de manière exhaustive	
	Réduction ou augmentation des dimensions (ou variables)	Selon les besoins	
	Conversion précise des features en indicateurs calculables	À faire de manière exhaustive	
	Degré de validation par les experts	À définir selon des critères (voir conclusion phase préparatoire ci-dessous)	
	Qualité des experts	Compétences, expérience	
	Nombre des experts	À définir selon les capacités mobilisables	
	Format de validation	À définir selon l'organisation et les personnes en charge de l'évaluation	

	Process formel de validation	Oui/non	
	Itérations	Oui/non, selon quels besoins, selon quels critères	

CONCLUSION PHASE PRÉPARATOIRE: La qualité des résultats de l’outil d’IA, dépendra de la quantité de données disponibles pour l’alimenter⁵⁴, de leur qualité et de leur traitement. Mais si s’interroger sur la qualité et la quantité des données nécessaires au développement de l’outil d’IA est certes important, encore faut-il avoir accès aux dites données. Comme l’indique un acteur interrogé, cet accès n’est pas toujours aisé. Par exemple, celui-ci peut dépendre de la base de données du client, qui est parfois inexistante ou inaccessible⁵⁵. Elle peut aussi dépendre, selon les pays, de la politique d’ouverture des données de justice. Sans accès aux données donc, l’automatisation est difficile. Mais les données seules ne sont pas suffisantes. Un traitement est nécessaire. D’abord un traitement de la base de données. Ce travail a été effectué manuellement par les juristes interrogés dans le cadre de notre étude. Ensuite, il est nécessaire de sélectionner des critères pour la constitution d’un barème algorithmique. On parle ici de classes ou clusters, lesquels influencent les résultats. Or, et en particulier lorsqu’il s’agit d’adopter un outil d’ores et déjà développé, il est important de garder à l’esprit que par la sélection et hiérarchisation des critères jugés pertinents, et donc par l’éviction de critères jugés secondaires, les barèmes algorithmiques tendent à ignorer certains des déterminants. Les concepteurs et juristes associés à la conception des algorithmes vont sélectionner un certain nombre de critères « objectivés » mais qui comportent toujours une part de subjectivité. Les critères retenus le sont en fonction de la pratique et de l’expérience de chaque praticien. Cette pratique et cette expérience est différente d’un praticien à un autre. La construction même du barème relève donc d’une décision pleinement politique. Et pourtant ce choix de construction du barème est nécessaire au moment de son élaboration. Le barème propose des interprétations des données mobilisées et remplit les silences desdites données. On voit ici l’importance de l’explicitation des choix de construction des outils pour apprécier l’opportunité d’adopter et d’utiliser l’outil dans tel ou tel cas d’espèce. À défaut on serait face à une boîte noire, ce qui peut porter atteinte à l’intégrité des résultats.

6. Méthode de calcul

Ces étapes préliminaires accomplies, il est alors possible de poursuivre sur la phase de développement de l’outil. Si l’outil n’est pas développé en interne mais adopté, il est également utile de s’intéresser à son processus de développement. Il est essentiel de comprendre le fonctionnement des algorithmes qui permettent de produire le résultat attendu. Il faut s’intéresser dès lors à la méthode de calcul. Le choix ne peut se résumer à l’adoption d’un système à base d’IA. Il faut pouvoir justifier et évaluer le type d’approche utilisée en IA, ce qui entraîne nécessairement plus d’opacité en IA connexionniste (réseaux de neurones et LLM) et plus de temps de paramétrage en IA symbolique (cas des arbres de décisions par exemple). La reprise des briques dans des bibliothèques existantes est une pratique fréquente qui doit cependant être explicitée et validée.

⁵⁴ “Il faut des données en masse” (Entretien)

⁵⁵ L’exemple donné portait en particulier sur la connaissance des entreprises de leur propre patrimoine contractuel.

<i>Phase de développement</i>			
6. Méthode de calcul			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ? Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Choix du type d'approche d'IA et par qui ?	À définir, idéalement à partir d'un benchmark des différentes approches possibles et selon les besoins (voir ci-dessus 1. Sonder les attentes)	
	Sélection des algorithmes	À choisir, idéalement à partir d'un benchmark des différents algorithmes existants et accessibles	
	Pré-entraînement des algorithmes	Où ? Dans quel contexte ? Estimation des risques de biais	
	Définition des seuils significatifs	Expliciter la méthode de calcul des seuils et les justifier (en fonction du volume, de la nature des données)	
	Définition des pondérations	Si possible, si nécessaire	
	Apprentissage supervisé	Oui/non, selon quels besoins	

7. Présentation/visualisation des données

Réfléchir à la présentation des données et à leurs modes de visualisation est au moins aussi important. Une bonne visualisation de données permet de voir et de comprendre des tendances ou des valeurs inhabituelles dans les données, de manière très accessible. La visualisation des données (ou *data visualization* en anglais) désigne le fait de représenter visuellement des données pour pouvoir déceler et comprendre des informations, les données brutes étant difficilement interprétables et exploitables. Ce processus se fait par des outils analytiques spécifiques et se matérialise par des tableaux, des graphiques, des cartes visuelles ou même des infographies regroupées dans des *dashboards*. En classant, segmentant et scénarisant les données visuellement, les praticiens du droit peuvent déceler des informations auparavant inaccessibles. S'intéresser à la visualisation des données est essentiel pour toute organisation qui souhaite piloter son activité plus efficacement en adoptant une stratégie *data-driven* et agile. C'est notamment en fonction des perceptions proposées dans la visualisation que l'utilisateur pourra se faire ou non une idée du modèle conceptuel du système⁵⁶, ce qui est une condition pour son appropriation future. Par ailleurs, selon les contraintes des situations de travail, des priorités différentes peuvent être accordées à certains signaux.

⁵⁶ Cf. NORMAN, D. *The Design Of Everyday Things* (Revised édition). Basic Books, 2013.

<i>Phase de développement</i>			
7. Présentation/visualisation des données			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i>	Niveau d'interprétabilité	À définir selon le code, les résultats	
<i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Finalisation de la "ground truth" ⁵⁷	Quelle validation par les experts ?	
	Validation sur dataset de tests	Oui/non, fréquence, satisfaction	

8. User experience

L'expérience utilisateur (UX) désigne la qualité de l'expérience vécue par l'utilisateur dans toute situation d'interaction. L'UX qualifie l'expérience globale ressentie par l'utilisateur lors de l'utilisation d'une interface, d'un appareil digital ou plus largement en interaction avec tout dispositif ou service. Il s'agit ici de poursuivre la réflexion amorcée dans la phase d'initiation sur les interfaces. Plus précisément, il convient alors d'obtenir des retours des utilisateurs à chaque étape du développement de l'outil et/ou de s'assurer de l'adéquation de l'interface de l'outil aux attentes des utilisateurs.

<i>Phase de développement</i>			
8. User Experience			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i>	Modèles utilisateurs	Observations en input ? Degré de précision ? Degré de variation ?	
<i>Qui renseigne les valeurs possibles</i>	Participation des publics concernés	Oui/non, lesquels	
	Tests utilisateurs formels	Oui/non, quels tests, quels utilisateurs, résultats ?	

⁵⁷ "Données et labels constituent ensemble ce que Florian Jatton nomme "ground-truth" (2017) : la "vérité" catégorielle que l'IA cherchera ensuite à reproduire. L'obtention, ou la constitution, de la ressource que représentent ces jeux de données annotés constitue, de ce fait, un enjeu principal pour la conception d'un outil d'IA. Ceux-ci sont qualifiés, à la Cour de cassation, comme dans d'autres fabriques d'IA, de jeux de données [...] "étalon-or" – l'ensemble informationnel de référence décrivant avec le plus de précision la réalité. De leur taille, composition, précision, sophistication, dépendent en grande partie les performances de l'outil et la qualité des résultats produits." GIRARD-CHANUDET, "Mais l'algo, là, il va mimer nos erreurs !" Contraintes et effets de l'annotation des données d'entraînement d'une IA", *Réseaux*, 2023/4, n°240, p.120.

de chaque indicateur ?	Plateforme d'accès à l'application	Oui/non, quel logiciel (propriétaire, open-source, etc.) ?	
	Terminaux acceptés	À lister	
	Contrôle d'accès	Oui/non, selon quelles modalités ?	
	Inspection RGPD	À vérifier selon les normes en vigueur	
	Normes accessibilité	À définir selon utilisateurs et données (sensibles ou non)	
	Référence du design (standard, spécialisé métier, etc.)	À expliciter	

CONCLUSION PHASE DE DÉVELOPPEMENT: Les décisions encodées dans des algorithmes ne sont jamais neutres. Des chercheuses comme Virginia Eubanks⁵⁸ ou Cathy O'Neil⁵⁹ ont de longue date montré qu'automatiser un choix peut faire courir le risque de reproduire plus ou moins consciemment certains biais, voire des injustices. C'est pourquoi depuis quelques années, tout un champ disciplinaire se structure autour de la responsabilité algorithmique et plus largement, de l'éthique du numérique. Seulement, c'est bien souvent a posteriori – une fois les systèmes conçus et déployés – que les questions clés sont posées. Nous nous attachons au contraire à montrer que c'est bien dès les toutes premières étapes de conception qu'un système algorithmique doit être pensé, avec ceux qui l'utilisent ou qui sont « calculés » à travers lui. C'est dans la phase de développement que la prise en compte de l'ensemble des parties prenantes prend tout son sens. Sans entrer dans les détails, notons que des méthodes de co-conception d'objets numériques de tous genres existent déjà : *value-sensitive design*, *human centered design*, design participatif (sans oublier le “Design justice” qui interroge plus en profondeur encore la dimension communautaire et politique des démarches de co-conception). L'intérêt ici est d'appliquer une méthode de co-design à un algorithme, qui n'est pas un objet numérique anodin.

9. *Protocole de traitement permanent en routine*

Une fois développé l'outil doit être intégré à la pratique et répondre aux attentes pour lesquelles il a été initialement adopté ou développé. Pour permettre la meilleure intégration de l'outil dans la pratique, il est d'abord nécessaire de s'assurer qu'il sera toujours opérationnel, c'est à dire mis à jour, et que la nature des résultats générés soit en conformité avec les besoins exprimés dans le temps. Ces besoins peuvent, en effet, évoluer avec les pratiques, la discipline et l'environnement (interne et externe). Il est nécessaire que l'outil soit également alimenté en données récentes pour que puisse être exprimée la réalité juridique à chaque utilisation.

⁵⁸ EUBANKS V., *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New-York, 2018.

⁵⁹ O'NEIL C., *Algorithmes, la bombe à retardement*, Les Arènes, Paris, 2018.

<i>Phase d'utilisation</i>			
9. Protocole de traitement permanent en routine			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i>	Fréquence	À définir	
<i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Responsabilité	De qui ? De quoi ?	
	Retours utilisateurs	Fréquence, selon quels utilisateurs ?	
	Révisions	Fréquence	
	Optimisation	Temps de calcul ou volume de données	

10. Design organisationnel

Encore aujourd'hui, il est considéré que lorsque l'on dispose de solutions techniques, le "solutionnisme" selon l'expression d'Evgeny Morozov⁶⁰, cela suffit. Mais comme dans tout autre domaine, le "Machine Learning" seul n'est ici pas suffisant. Pour que les outils développés ou adoptés puissent contribuer à l'amélioration de la qualité et de la performance des pratiques, il est essentiel d'avoir un "Human-Machine Learning", interprétable et associé à un design organisationnel pertinent. Dans la phase de développement des architectures techniques, nous avons déjà abordé l'importance de développer une boucle vers les utilisateurs. De la même manière, dans la phase d'utilisation il est nécessaire de veiller à l'encastrement des nouveaux outils dans le design organisationnel. Il ne suffit pas de développer un outil et d'ensuite laisser les praticiens se débrouiller et s'adapter. Il faut implémenter l'outil et développer des systèmes d'accompagnement pertinents pour permettre une utilisation optimale. L'encastrement des outils dans le design organisationnel doit certes s'amorcer dès la conception de l'outil, avec des moments de couplage, mais doit se poursuivre dans la phase d'utilisation. Cette tâche d'implémentation doit être l'objet de discussions à l'interne, i.e il est préférable d'éviter de la déléguer à des prestataires externes. De la même manière, l'outil doit également s'intégrer aux écosystèmes desquels il dépend (écosystème client, organisation juridictionnelle, politiques d'innovation publique, communautés d'acteurs...).

⁶⁰ MOROZOV E., *Pour tout résoudre, cliquez ici : l'aberration du solutionnisme technologique*. Fyp éditions, Limoges, 2014.

<i>Phase d'utilisation</i>				
10. Design organisationnel				
	Indicateurs	Sous-indicateurs	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Encastrement design organisationnel interne	Taux d'utilisation (stagiaire/junior/senior)	Pourcentage	
		Éducation/Formation	Qualitative - oui/non	
		Intégration à la stratégie organisationnelle (pratique)	Échelle	
		Intégration à la politique générale (stratégie)	Échelle	
		Intégration à la stratégie commerciale (marketing)	Échelle	
		Intégration à la stratégie de communication (image interne/externe)	Échelle	
	Encastrement design organisationnel externe	Intégration à l'organisation juridictionnelle/alignement avec les stratégies d'innovation du secteur public (open data, nature des données...)	Échelle	
		Couplage avec les écosystèmes clients	Échelle	
		Appropriation par les communautés (concurrents, associations professionnelles...)	Échelle	

11. Modèle économique

On ne peut faire l'impasse sur la question du modèle économique de l'outil. C'est une question à laquelle seront confrontés toutes les structures qui feront le choix de développer un outil adapté à leurs besoins. Si cela permet de bénéficier de l'avantage de la personnalisation, il faut garder à l'esprit que le coût en temps, argent et énergie est important. Aussi l'outil pourrait bénéficier à une communauté de praticiens plus large que celle de la structure dans laquelle il a été développé. Plusieurs modèles économiques sont possibles et devront faire l'objet de discussions avec le prestataire technique. Les choix dépendront de la stratégie propre à chaque structure. L'avantage de l'open source est de bénéficier de l'effet de réseau qui permet de rendre l'outil plus performant, précis et moins coûteux sur le long terme (la mise à jour peut être collaborative). Au contraire rendre l'outil moins accessible peut permettre de bénéficier d'un avantage concurrentiel. Il est important de lier la réflexion sur le modèle économique à la stratégie plus globale de la structure.

<i>Phase d'utilisation</i>			
11. Modèle économique			
	Indicateur	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs possibles de chaque indicateur ?</i>	Abonnement	Oui/non, quels tarifs (selon quelles offres) ? À choisir en cohérence avec le modèle économique de l'organisation	
	Licence	Propriétaire ? Libre ?	
	À l'acte	Oui/non, quels tarifs (selon quelles offres) ? À choisir en cohérence avec le modèle économique de l'organisation	
	Open source	Selon la licence, selon la sensibilité des données, de leurs usages ainsi que de ceux des résultats	
	Liens avec communauté, avec base clients	Lesquel(le)s ?	

12. Résultats

Enfin, et il s'agit sans doute de l'étape la plus évidente lorsqu'on imagine un protocole d'évaluation d'un outil, qu'en est-il des résultats? Quels sont les indicateurs de qualité et de performance d'un outil? S'intéresser aux résultats c'est en réalité s'intéresser à l'ensemble du processus d'adoption ou de création de cet outil: des besoins initiaux ayant fait naître l'idée à l'utilisation dans la pratique quotidienne, en passant par le développement de l'algorithme ou la constitution de la base de données qui permet de l'alimenter. Ce que nous souhaitons dire ici c'est que la grille d'évaluation des résultats dépendra très fortement du succès des autres phases.

<i>Phase d'utilisation</i>				
12. Résultats				
	Indicateur	Sous-indicateurs	Valeurs possibles	Faisabilité
<i>Qui évalue ?</i> <i>Qui renseigne les valeurs</i>	Satisfaction utilisateur	Expérience	Qualitative	
		Confiance		
		Facilitation de la pratique		
		Encapacitation		

<i>possibles de chaque indicateur ?</i>		Temps gagné observable		
	Satisfaction client	Conversion prospect/client	Pourcentage	
		Fidélisation client		
	Performance	Résultat suivi - succès/échec	Pourcentage	
		Résultat non suivi - succès/échec		
		Succès/échec global		
	Qualité	Prévisibilité		
		Clarté		
		Information exacte/inexacte		
		Information complète/incomplète		

CONCLUSION PHASE D'UTILISATION: Intégrer un nouvel outil dans une pratique c'est s'engager dans une démarche plus globale de transformation de l'organisation. Il s'agit d'encadrer le nouvel outil dans l'ensemble du design organisationnel. Il est alors essentiel de prendre en compte la conduite du changement. Un travail complémentaire serait le bienvenu car se situe au-delà de la portée de cette étude. Les résultats que produira l'outil seront le reflet du succès de l'encastrement de l'outil dans ce design organisationnel. Mieux l'outil sera intégré au design organisationnel et plus ses résultats seront satisfaisants pour l'ensemble des parties prenantes. La qualité des données est au moins aussi importante. Mais encore une fois l'accès aux données de qualité dépendra de l'encastrement dans un design organisationnel plus global. L'organisation participe à un ensemble, à un écosystème. La qualité des données, et donc la qualité des résultats en fin de chaîne, dépend énormément d'acteurs qui ne prennent pas part actuellement au développement des algorithmes de jurimétrie. Par exemple, les juridictions ne communiquent pas de données visuelles dans leurs décisions. Cela a inévitablement des conséquences négatives sur la qualité des résultats d'un outil de jurimétrie en droit des dessins et modèles ou en droit d'auteur. Or la confiance ne vaut que si la qualité des résultats de l'outil est appréciée. De plus, la systématisation des pratiques est souhaitable si les utilisateurs peuvent avoir confiance en leur outil et que leur autonomie est préservée.

Modélisation

La figure 3. ci-dessous reprend les indicateurs pour chaque phase et étape associée de la mise en oeuvre de l'outil d'IA. Elle sert de référence par rapport aux grilles ci-dessus qui développent les indicateurs, sous-indicateurs et valeurs possibles.

Figure 3. Indicateurs d'évaluation

<i>Phase d'initiation</i>				
1. Sonder les attentes				2. Sonder les moyens
1.1. Demandes	1.2. Offres	1.3. Cible	1.4. Interface Design	2.1. Barrières
Techniques (Problèmes juridiques) Commerciales Organisationnelles	Informations juridiques Aide à la décision Prise de décision	Cartographie des cibles à l'interne Cartographie des cibles à l'externe Socio-démographie ou socio-organisation des cibles Construction des personas	Choix des architecteurs Complexité de la tâche	Ressources Contexte interne Contexte externe
<i>Phase préparatoire</i>				
3. Constitution de la base de données d'apprentissage	4. Preprocessing et nettoyage		5. Feature extraction	
Sources (et validation des sources) Hétérogénéité Interopérabilité Conversions Extension (degré de couverture du domaine) Adaptation aux potentiels d'exploration Adaptation aux visées stratégiques Volume suffisant	Doublons, incohérences, contradictions Valeurs aberrantes Valeurs manquantes Formats incompatibles Instabilité des flux Modèles de langue pour conversion, traduction et NLP Validation des métadonnées (dates, lieux, sources, historiques, etc.) Réitération du nettoyage Intégration à un processus de contrôle de qualité		Méthodes d'extraction Terminologie prédéfinie Ontologie complète ou pas (IA symbolique ou connexionniste) Construction a priori de classes Nombre de classes (ou de clusters) Labellisation des clusters Réduction ou augmentation des dimensions (ou variables) Conversion précise des features en indicateurs calculables Degré de validation par les experts Qualité des experts Nombre des experts Format de validation	

<p>Actualisation suffisante</p> <p>Granularité suffisante</p> <p>Comparaison avec les bases juridiques classiques</p> <p>Identification des biais possibles et tests pour vérifier</p> <p>Partition du dataset pour apprentissage et test</p> <p>Propriétés sémiotiques requises</p>		<p>Process formel de validation</p> <p>Itérations</p>	
<i>Phase de développement</i>			
6. Méthode de calcul	7. Présentation/visualisation des données		
<p>Choix du type d'approche IA et par qui ?</p> <p>Sélection des algorithmes</p> <p>Pré-entraînement des algos</p> <p>Définition des seuils significatifs</p> <p>Définition des pondérations</p> <p>Apprentissage supervisé</p>	<p>Niveau d'interprétabilité</p> <p>Finalisation de la "ground truth"</p> <p>Validation sur dataset de tests</p>		
<i>Phase d'utilisation</i>			
8. User experience	9. Protocole de traitement permanent en routine	10. Design organisationnel	11. Modèle économique
<p>Modèles utilisateurs</p> <p>Participation des publics concernés</p> <p>Tests utilisateurs formels</p> <p>Plateforme d'accès à l'application</p> <p>Terminaux acceptés</p> <p>Contrôle d'accès</p>	<p>Fréquence</p> <p>Responsabilité</p> <p>Retours utilisateurs</p> <p>Révisions</p> <p>Optimisation</p>	<p>Encastrement design organisationnel interne</p> <p>Encastrement design organisationnel externe</p>	<p>Abonnement</p> <p>Licence</p> <p>À l'acte</p> <p>Open Source</p> <p>Liens avec communauté, avec base client</p>

Inspection RGPD Normes accessibilité Références du design (standard, spécialisé métier, etc.)			
12. Résultats			
Satisfaction utilisateur Satisfaction client Performance Qualité			

UTILISATION DU CADRE D'ÉVALUATION

Un enseignement tiré de l'évaluation de politiques publiques transposé ici est celui du rapport positif entre l'utilisation d'une évaluation et sa qualité. Plus une évaluation va être utilisée, meilleure va être sa qualité, et vice-versa. Le cadre d'évaluation proposé ici s'inscrit ainsi dans l'approche qui promeut des évaluations focalisées sur leur utilisation, telle que l'a développé Michael Quinn Patton⁶¹ dans le champ de l'évaluation des politiques publiques. Selon celle-ci, l'évaluation doit se planifier et se conduire de manière à augmenter à la fois l'utilisation de ses résultats et celle de son processus afin d'informer les décisions et améliorer les performances⁶². À cette fin, les premiers usagers visés de l'évaluation doivent être clairement identifiés et personnellement engagés, dès le début du processus d'évaluation⁶³. Dans le cadre de ce projet, les premiers usagers visés sont les professionnels du cabinet de contentieux commercial utilisant l'IA, qui ont été informés du projet de construire un cadre d'évaluation dès les premiers entretiens. Il reste bien entendu à les maintenir engagés tout au long du projet, donc de leur faire part des avancements dans le développement du cadre d'évaluation, et surtout dès lors que ce dernier sera utilisable. Nous invitons donc ceux qui seront impliqués dans l'usage d'outils d'IA d'utiliser et de développer le cadre d'évaluation proposé, de manière intégrée et concomitante à leurs activités.

Qui sont ou seront les utilisateurs de ce cadre d'évaluation ? Autrement dit, qui sont les *évaluateurs* de l'utilisation de l'outil ? Dans notre cadre, il s'agit de tout professionnel et chercheur impliqués dans les domaines où l'IA s'intègre aux outils de justice utilisés. Il s'agit en premier lieu des concepteurs qu'ils soient développeurs ou praticiens du droit, des personnes impliquées dans la stratégie d'adoption ou de développement de l'outil (selon les cas l'équipe Legal Ops ou la direction stratégique et commerciale) et l'ensemble des utilisateurs (du praticien expérimenté au stagiaire). Il s'agit en second lieu de tout acteur extérieur au processus de développement et d'adoption de l'outil (principalement des chercheurs mais également des évaluateurs extérieurs). Les évaluateurs peuvent donc être internes ou externes, selon les besoins et les moyens et matière de ressources humaines et

⁶¹ PATTON M. Q., *Utilization-Focused Evaluation: 4th edition*. Thousand Oaks, CA : Sage Publications, 2008.

⁶² Voir le billet du blog de la plateforme *BetterEvaluation* sur l'évaluation focalisée sur son utilisation ("*utilization-focused evaluation*") écrit par Alan Mountain en 2012 ([Lien](#) ; dernier accès le 9 janvier 2022).

⁶³ *Ibid.*

de financements disponibles et dédiés à l'évaluation⁶⁴. Néanmoins, cette étude a été rédigée avec l'idée que le cadre d'évaluation servirait d'abord les professionnels souhaitant adopter ou développer un outil de jurimétrie. Dans tous les cas, les évaluateurs se doivent d'assurer que les usages visés de l'évaluation par ses premiers utilisateurs orientent toutes autres décisions prises quant au procès d'évaluation⁶⁵. Il s'agirait, par exemple, pour ceux qui viendraient à modifier les indicateurs ou les variables du cadre d'évaluation proposé, de s'assurer que ceux-ci correspondent toujours aux besoins des utilisateurs de l'évaluation, à savoir les professionnels du droit qui utilisent les nouveaux outils d'IA. Un autre exemple serait de s'assurer, dans les choix liés au contenu du cadre d'évaluation ainsi qu'au calendrier de l'évaluation et à sa durée, que ceux-ci tiennent compte des changements qui ont lieu dans le domaine de droit concerné et ses professions. En somme, il est recommandé que l'évaluateur ou les personnes en charge de l'évaluation vise à faciliter la prise de décision par les utilisateurs de l'évaluation et de ses résultats⁶⁶.

LIMITES DU CADRE D'ÉVALUATION

Ce cadre d'évaluation a été élaboré sur la base d'entretiens menés auprès de professionnels du droit. Nous avons tenté de donner forme à des indicateurs en partant de leur expérience dans le développement de deux outils de jurimétrie. Cette expérience leur est propre et donc l'un des premiers biais est sans doute d'avoir élaboré ce cadre à partir d'une faible diversité d'expériences. Pour élaborer un cadre d'évaluation plus neutre nous avons donc dû tirer de la littérature d'autres indicateurs (de la littérature scientifique en droit, sciences de gestion, économie, sociologie, informatique...). Enfin, pour proposer un cadre plus complet nous avons également élaboré des indicateurs indirectement tirés des entretiens en nous basant sur notre "intuition" et sur nos expériences personnelles. De ce fait, ce cadre d'évaluation peut comporter certains biais ou tropismes disciplinaires. C'est la raison pour laquelle nous invitons les lecteurs à considérer ce cadre comme un canva ou prototype de cadre d'évaluation dont il convient de s'approprier et d'adapter en fonction des particularités propres à chaque contexte, organisation, discipline et outils. Ce cadre d'évaluation n'est donc pas à considérer comme un produit fini et c'est la raison pour laquelle nous invitons la communauté à contribuer à l'amélioration de cette "boussole".

PERSPECTIVES D'AMÉLIORATION

Il serait intéressant de faire évoluer sa "trame" par la contribution d'experts de disciplines différentes. Des contributions de designers, ingénieurs et spécialistes des sciences de gestion seraient particulièrement les bienvenues. De plus, ce cadre d'évaluation mériterait d'être enrichi par l'apport de compte-rendus d'autres expériences de terrain. Il serait utile de compléter les indicateurs en menant par exemple d'autres études de cas et/ou en confrontant ce cadre à l'expérience de professionnels ayant eu à développer ou adopter des outils similaires. Il serait notamment utile de bénéficier du regard de professionnels agissant dans d'autres disciplines et d'autres contextes (direction juridique en entreprise, autre système juridique...).

Par ailleurs, afin d'améliorer les indicateurs existants il serait utile que ce cadre d'évaluation puisse être utilisé et confronté au réel. Ces retours d'expérience et l'étude par des acteurs extérieurs (des chercheurs principalement) des cas d'utilisation du cadre d'évaluation permettraient de vérifier

⁶⁴ Si les réflexions sur les contraintes matérielles de l'organisation et la conduite d'une évaluation sont essentielles pour son bon déroulement et des résultats de qualité, elles dépassent le cadre de ce projet. Nous invitons donc les utilisateurs du cadre d'évaluation proposé à les considérer de leur côté.

⁶⁵ Voir le billet du blog de la plateforme *BetterEvaluation* sur l'évaluation focalisée sur son utilisation ("*utilization-focused evaluation*") écrit par Alan Mountain en 2012 ([Lien](#) ; dernier accès le 9 janvier 2022).

⁶⁶ *Ibid.*

certaines de nos hypothèses et la qualité des indicateurs. De plus, l'utilisation du cadre d'évaluation et les retours permettraient de déceler d'autres indicateurs et d'affiner certains d'entre-eux.

Toutefois, même si un grand nombre d'utilisations du cadre d'évaluation, de retours d'experts et d'étude cas de variés peuvent améliorer substantiellement sa "trame", il demeurera toujours que ce cadre nécessite d'être adapté aux particularités de chaque contexte. Encore une fois, ce cadre doit être considéré comme un canevas qu'il est nécessaire d'adapter à chaque situation.

Enfin, nous souhaitons contribuer, en conclusion de cette étude et pour répondre aux interrogations plus générales du projet AJC, à la réflexion sur l'existence éventuelle d'indices d'autonomie des acteurs de la (cyber)justice. Il existe des indices (canadiens et internationaux) d'accès à la justice, tels que ceux de Gramatikov *et al.* présentés dans le *Handbook for Measuring the Costs and Quality of Access to Justice* auxquels se réfèrent (entre autres) Alexandra Pasca et Fabien Gélinas dans leur travail de construction d'un modèle d'évaluation dans le domaine de l'accès à la justice⁶⁷. Si de tels indices peuvent sembler difficiles à construire, en particulier compte tenu de la difficulté de mesurer l'autonomie, ils pourraient néanmoins s'avérer très utiles compte tenu du développement des outils d'IA et de leur impact sur les pratiques. C'est dans cette direction que se situe ce travail ainsi que le reste du projet AJC.

⁶⁷ Des indices d'autonomie des acteurs de la (cyber)justice pourraient-être considérés comme des "méta-indices" dans leur cas d'étude, par rapport aux indices d'accès à la justice.

BIBLIOGRAPHIE

BOULLIER D., “Déboussolés de tous les pays ...! : Une boussole écodémocrate pour rénover la gauche et l'écologie politique”, *Cosmopolitiques*, éditions spéciales, 2003.

BOULLIER D., “Les études d’usages : entre normalisation et rhétorique ”, *Annales des Télécommunications*, 57, n°3-4, 2002, pp.190-209.

GAPENNE O. et BOULLIER, D. (dirs.), Systèmes d’aide opératoire. Enjeu pour les technologies cognitives, *Intellectica*, 2006/2, 44.

BOULLIER D., *Sociologie du numérique*, Paris: Armand Colin (collection U), 2016. (2nde édition 2019)

BOULLIER, D. et EL MHAMDI E. M., “Le *machine learning* et les sciences sociales à l’épreuve des échelles de complexité algorithmique”, *Revue d’anthropologie des connaissances* [En ligne], 14- , 2020, mis en ligne le 01 mars 2020.

BOULLIER, D., “Le problème n’est pas le numérique en lui-même mais le capitalisme financier numérique”, *Entretien dans la Revue des conditions de travail* (ANACT), n°6, 2017 (Mieux travailler à l’ère du numérique. Définir les enjeux et soutenir l’action), pp. 56-62.

BUAT-MÉNARD E., GAMBIASIP., “La mémoire numérique des décisions judiciaires. L’*open data* des décisions de justice de l’ordre judiciaire”, *Recueil Dalloz*, 2017.

BURKE JOHNSON, R. & DICK W., “Evaluation in Instructional Design: A Comparison of Evaluation Models (chapter 10)”, *Trends and Issues in Instructional Design and Technology*, 2012, pp.96-98.

CAMPBELL M., Access to Justice Annotated Bibliography, National Self-Represented Litigants Project, 2019.

CBA, “Reaching equal justice : an invitation to envision and act”, Août 2013.

CADIET L. (dir.), *L’open data des décisions de justice*, Rapport au ministre de la Justice, nov. 2017.

CASSUTO T., “La justice à l’épreuve de sa prédictibilité”, *AJ Pénal*, 2017.

Commission européenne pour l’efficacité de la justice (CEPEJ), *Mesurer la qualité de la justice - Guide*, 2017.

DONDERO B., “Justice prédictive : la fin de l’aléa judiciaire ?”, *D.* n° 10, 2017.

EUBANKS V., , *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin’s Press, New-York, 2018.

FARROW C. W. T. *et al.*, “Everyday Legal-Problems and the Cost of Justice in Canada - Cost of Justice Survey Data”, 2018.

GARAPON A., LASSEGUE J., *La justice digitale*, PUF, 2018.

- GARAPON A., “Les enjeux de la justice prédictive”, *JCP G*, n° 01-02, 2017.
- GÉLINAS F. & PASCA, A., *A Step-by-Step Evaluation Model for Legal Tools*, CLEO & Montreal Cyberjustice Laboratory at. 14, forthcoming.
- GÉLINAS F. *et al.*, “CLEO Case Study: Evaluation Steps for the Family Law Guided Pathways (FLGP)”, CLEO & Montreal Cyberjustice Laboratory, forthcoming.
- GÉLINAS F. *et al.*, *Foundations of Civil Justice: Toward a Value-Based Framework for Reform*, Cham: Springer, 2015.
- GIRARD-CHANUDET, “‘Mais l’algo, là, il va mimer nos erreurs !’ Contraintes et effets de l’annotation des données d’entraînement d’une IA”, *Réseaux*, 2023/4, n°240, p.120.
- Groupe d’expert de haut niveau sur l’IA, Commission européenne, *Une définition de l’IA : principales capacités et disciplines scientifiques*, 2018.
- HM Treasury, *The Magenta Book*, 27 avril 2011.
- HOMOKI P., “Guide on the use of Artificial Intelligence-based tools by lawyers and law firms in the EU”, Council of Bars and Law Societies of Europe, Brussels, 2022.
- KUSTERS C. *et al.*, “Making evaluations matter: a practical guide for evaluators”, Centre for Development Innovation, Wageningen University & Research Centre, 2011.
- RONFAUT L., “Cédric Villani avance sur l’intelligence artificielle”, *Le Figaro*, 29 novembre 2017
- MACCARTHY J., *What is artificial intelligence*, Stanford University, Stanford, Novembre 2007.
- MOROZOV E., *Pour tout résoudre, cliquez ici : l’aberration du solutionnisme technologique*. Fyp éditions, Limoges, 2014.
- MOUNTAIN A., “Utilization-Focused Evaluation”, BetterEvaluation blog, 2012.
- NORMAN, D. *The Design Of Everyday Things* (Revised édition). Basic Books, 2013.
- OCDE, *Improving Governance with Policy Evaluation: Lessons from Country Experiences*, OECD Public Governance Reviews, OECD Publishing, Paris, 2020.
- O’NEIL C., *Algorithmes, la bombe à retardement*, Les Arènes, Paris, 2018.
- PASCA A., *Annotated Bibliography : Direct to Public Legal Services - Impact Evaluation* (ACT Project & CLEO), 2020.
- PASCA A., *Promising Evaluation Practices Guide: A Few Basic Tips*, Projet AJC : Chantier 8, 15p., document de travail n°27, 2021.
- PATTON M. Q., *Utilization-Focused Evaluation: 4th edition*. Thousand Oaks, CA : Sage Publications, 2008.
- POLANYI, K., *La grande transformation : Aux origines politiques et économiques de notre temps*. Gallimard, 2007.

Tim Roberts & Associates Consulting, “Learning from Justice metrics models in other jurisdictions”, A Report for the University of Victoria, Faculty of Law, Access to Justice Centre for Excellence, 2019.

SAMOILI S., LÒPEZ COBO M., GÒMEZ E., DE PRATO G., MARTÌNEZ-PLUMED F. et DELIPETREV B., *AI Watch : Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, Publications Office of the European Union, 2020.

TABLE DES MATIÈRES

RÉSUMÉ	3
ABSTRACT	3
SOMMAIRE	4
INTRODUCTION	5
Transformation numérique du droit & Cyberjustice	5
Le projet “Autonomisation des acteurs Judiciaires par la Cyberjustice et l’intelligence artificielle (AJC)”	5
Mise en contexte du ”Chantier 12 - Évaluation et indicateurs de qualité et de performance”	5
JUSTIFICATION DE L’ÉTUDE	6
Répondre à un besoin des acteurs	6
Contribuer à un domaine de recherche en émergence	7
OBJET D’ÉTUDE	8
Outils d’IA: de quoi parlons-nous?	8
L’IA dans la justice	9
Cartographie des outils	10
ZOOM: les outils d’aide à la décision	11
MÉTHODOLOGIE & TERRAIN DE L’ÉTUDE	11
Phase 1 : Choix du cas d’usage de départ (terrain de recherche)	12
Phase 1 bis : Observations du cas d’usage de départ	15
Phase 2 : Comparaison, modélisation et protocole d’évaluation	15
Phase 3 : Extension à d’autres solutions exploitant l’IA	15
Phase 4 : Extension à d’autres systèmes juridiques et à d’autres pays	16
Phase 5 : Extension à d’autres domaines du droit	16
Phase 6 : Fourniture du modèle final de protocole d’évaluation	16
Limites du protocole d’étude et adaptation au terrain	16
RÉSULTATS : CADRE D’ÉVALUATION ET INDICATEURS	18
Principes directeurs de l’évaluation	18
Questions d’évaluation	19
Critères de qualité	20
Engagement des parties prenantes	22
Conclusions tirées des entretiens	22
Phase d’initiation	23
Phase du développement	23
Exploitation à des fins organisationnelles	24
Exploitation à des fins commerciales	24
Sélection des indicateurs	25
1. Sonder les attentes	26

2. Sonder les moyens	30
3. Constitution de la base de données d'apprentissage	31
4. Preprocessing et nettoyage	32
5. Feature extraction (détermination des propriétés pertinentes des problèmes pour engendrer les calculs)	33
6. Méthode de calcul	35
7. Présentation/visualisation des données	36
8. User experience	36
9. Protocole de traitement permanent en routine	38
10. Design organisationnel	38
11. Modèle économique	39
12. Résultats	40
Modélisation	41
UTILISATION DU CADRE D'ÉVALUATION	44
LIMITES DU CADRE D'ÉVALUATION	45
PERSPECTIVES D'AMÉLIORATION	45
BIBLIOGRAPHIE	46
TABLE DES MATIÈRES	48