

Doctrine.fr : l'intelligence artificielle au service du droit

Par Antoine DUSSÉAUX
Cofondateur de Doctrine.fr
et Hugo RUGGIERI
Responsable juridique de Doctrine.fr

Où est le Google du droit ?

Il y a vingt ans, deux chercheurs de Stanford découvraient que l'information disponible sur Internet pouvait être rendue plus accessible. Le web commençait alors à peine à se développer, mais il était déjà possible d'entrevoir les bouleversements profonds qu'il permettrait.

La recherche des informations disponibles sur le web était alors dominée par les moteurs AltaVista et Yahoo!, dont le fonctionnement reposait principalement sur l'analyse de mots-clés : la page contenant le plus d'occurrences de la requête était la première. Si ce principe de base a initialement suffi, les limites d'un tel fonctionnement sont rapidement apparues : il suffisait de cacher un mot-clé dans une page pour être automatiquement mieux classé dans les résultats de recherche.

La grande réussite de Larry Page et Sergey Brin, les deux cofondateurs de Google, fut de mettre de l'ordre dans la masse de données disponibles sur Internet⁽¹⁾. L'algorithme de Google consiste à déterminer la réputation d'un site en prenant en compte une multitude de critères, parmi lesquels figurent principalement les liens vers ce site provenant de sites tiers, associés à la réputation de ces sites tiers eux-mêmes. Google définit ainsi un maillage réputationnel qui permet d'afficher non pas le résultat comprenant le plus d'occurrences du mot-clé recherché, mais celui qui est le plus reconnu par le public comme étant pertinent pour ce mot-clé.

L'arrivée de Google a bouleversé la façon dont Internet fonctionnait, puisque la quantité d'informations disponibles sur le web, gigantesque et en perpétuelle croissance, était désormais triée et organisée. La promesse de l'ouverture de l'information au public est d'offrir à tous le même accès à la bibliothèque mondiale de l'information et de permettre le développement de nouveaux services aux moyens des connaissances glanées.

Vingt ans ont passé, et un grand nombre de services sont apparus en exauçant cette promesse dans de nouvelles industries. Amazon, eBay, Wikipédia, Spotify, Netflix : dans chaque cas, une masse conséquente d'informations a été analysée par le service en question, permettant un accès renouvelé à la connaissance, à la culture, ou à la consommation.

Il reste pourtant un pan de la société qui demeure enveloppé d'une aura de mystère tant l'accès à la connaissance y est privilégié, et ce alors qu'il s'agit d'une des institutions les plus vitales au bon fonctionnement de la vie en société.

Il est en effet patent que le monde de la justice a jusqu'à présent peu bénéficié des avancées des nouvelles technologies. La doctrine juridique est rédigée par une poignée de grands éditeurs

(1) Voir par exemple l'article du journal britannique *The Economist*, "How PageRank works", publié le 16 septembre 2004.

centenaires qui publient également les codes annotés, essentiels à l'office de l'avocat et du juge. Peu de décisions de justice sont mises en ligne⁽²⁾ et sont rendues accessibles au justiciable. Les moteurs de recherche juridiques traditionnels sont ceux de ces mêmes grands éditeurs, et fonctionnent plus comme un AltaVista qu'un Google : il s'agit essentiellement de recherche par mot-clé.

La conséquence : le monde de la justice n'est accessible que par ses initiés, ce qui dessert son bon fonctionnement et va à l'encontre même de ses valeurs de publicité et d'ouverture⁽³⁾. À l'inverse, le développement d'un moteur de recherche adapté aux besoins de la recherche d'aujourd'hui permettrait au justiciable de mieux comprendre l'institution garante de la vie en société, et au praticien de se concentrer sur les aspects les plus gratifiants de sa profession. La question est alors de déterminer pourquoi ce secteur est resté jusqu'à présent si cloisonné, ou, formulé autrement : « *Où est le Google du droit ?* »

Intelligence artificielle : émergence d'une nouvelle technologie pour analyser la langue juridique

L'un des facteurs les plus importants pour l'émergence d'un tel moteur de recherche est qu'il est nécessaire de *comprendre* les textes de lois et les décisions de justice pour pouvoir les indexer de manière pertinente.

Le droit est une matière particulièrement propice à la mise en œuvre d'un tel système de recherche : comme sur Internet, les textes de lois et les décisions de justice reposent sur un maillage étroit de liens et de références. Cependant, au contraire d'Internet, ces entrelacements ne sont pas des liens cliquables visibles aisément dans le code d'un site web⁽⁴⁾ : il s'agit de citations, de références, de renvois incorporés dans le corps des textes de loi ou dans les décisions de justice. Même lorsque ces liens sont explicites, comme un renvoi à un autre article du même code ou le visa d'une décision de justice, ils n'en restent pas moins rédigés en langage dit « naturel » par parallélisme au langage informatique, le code, qu'il faut donc faire déchiffrer à la machine. Et, très souvent, les références les plus essentielles sont implicites ou du moins peu évidentes pour un œil non initié⁽⁵⁾.

Par ailleurs, il ne suffit pas seulement d'identifier la référence à un autre texte ou une autre décision : il faut également comprendre cette référence et le contexte dans lequel elle s'inscrit. De nombreuses décisions vont ainsi citer un texte pour en exclure l'application, ou faire référence à une jurisprudence pour mettre en exergue des différences. Dans de tels cas, il serait malvenu de lier les textes en question par un lien de connexité, or c'est précisément ce qui arriverait lors d'une analyse par un moteur de recherche classique par mot-clé.

Le développement d'un moteur véritablement pertinent pour la recherche juridique dépend donc non seulement de l'identification des liens entre les textes et décisions de justice, mais également de la compréhension du contexte.

C'est avec le développement récent de l'intelligence artificielle, ce domaine de l'informatique qui consiste à traiter des problèmes mathématiques non pas par une série de conditions comme en

(2) Selon le rapport « L'open data des décisions de justice » de Loïc Cadiet, publié en novembre 2017 en réponse à la Mission d'étude et de réfiguration sur l'ouverture au public des décisions de justice confiée par le ministère de la Justice, 2 677 253 décisions ont été rendues en matière civile en 2016, pour la publication de 13 360 décisions sur Légifrance, soit environ 0,5 % des décisions rendues cette année.

(3) Rappelons ainsi que la publicité de la justice est un principe cardinal à son bon fonctionnement, notamment consacré par l'article 6.1 de la Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales.

(4) Un lien est identifié en html par la balise « <a> ».

(5) Ainsi, une référence à la « jurisprudence Chronopost » est une notion qui a énormément de sens pour un praticien de droit privé, mais sera difficilement identifiée par un algorithme comme une référence à l'arrêt n°93-18.632 du 22 octobre 1996 de la chambre commerciale de la Cour de cassation.

informatique traditionnelle mais avec des méthodes tentant d'imiter un raisonnement humain, que le traitement du langage naturel a véritablement pris son envol. En particulier, les méthodes dites d'apprentissage machine (machine learning) et d'apprentissage profond (deep learning) ont permis des avancées sans précédent, en permettant à des algorithmes « apprenants » de s'améliorer au fur et à mesure des essais.

Les progrès algorithmiques réalisés ces dernières années dans l'intelligence artificielle permettent désormais d'appréhender le langage juridique et le maillage de liens et de références issu de la lecture attentive des textes et des décisions de justice. Il est désormais possible de créer un véritable Google du droit, un moteur de recherche juridique intelligent, et d'utiliser l'intelligence artificielle pour améliorer la relation entre la justice, le justiciable et les praticiens du droit.

Une multitude de possibilités au service du droit

Le champ des possibles qu'ouvrent de telles avancées est immense. Il s'agit tout d'abord d'une contribution à la vie démocratique : la justice est au cœur du fonctionnement du contrat social, au nom duquel on ne renonce à se faire justice soi-même qu'à la condition qu'une institution neutre, indépendante et transparente rende cette justice de manière impartiale et ouverte. L'ouverture des décisions de justice et des textes de loi, et leur mise à disposition du public grâce à un moteur de recherche, ne peuvent que contribuer au renforcement de la confiance envers cette institution.

Un moteur de recherche juridique dopé à l'intelligence artificielle peut également avoir de nombreuses possibilités d'exploitation concrètes par le justiciable et le praticien. Ainsi que le souligne la Commission nationale de l'Informatique et des Libertés dans son « Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle » : « les logiciels algorithmiques peuvent être d'une utilité stratégique en optimisant l'identification des solutions statistiquement les plus probables pour un contentieux donné ». Cela offrirait aux praticiens un véritable gain de temps, qui pourrait notamment permettre aux avocats de « se consacrer à des tâches plus gratifiantes d'analyse juridique et de contact humain ». De l'autre côté, le justiciable sera en mesure de mieux juger de l'opportunité de s'engager dans une procédure en comparant son cas aux cas similaires.

La mise à disposition facilitée des sources du droit apparaît ainsi comme un facteur d'amélioration de la justice à tous égards : facilitation du travail du juge, qui peut plus aisément vérifier les arguments qui lui sont soumis ; facilitation du travail de l'avocat, qui dispose de meilleurs outils pour affiner ses plaidoiries ; facilitation du travail de l'universitaire et du justiciable qui peuvent accéder à des résultats de recherches pertinents leur permettant de mieux comprendre les enjeux en question.

L'intelligence artificielle peut également servir à personnaliser les outils juridiques à la pratique de chaque personne, comme par exemple par le biais d'alertes personnalisées lors de nouveaux développements sur des points de droit précis, identifiés en fonction des recherches d'une personne.

De tels outils existent en complémentarité de l'humain, afin de l'assister dans son métier ou dans la compréhension du monde juridique, et non en opposition avec lui ou dans une optique de remplacement par des algorithmes « prédictifs ». Un moteur moderne offre en effet une promesse de justice *assistée* : l'humain y est un élément essentiel, son travail étant facilité et rendu plus performant par des outils d'intelligence artificielle qui le complètent.

Grâce à son moteur de recherche basé sur le traitement du langage naturel et de l'intelligence artificielle, grâce à son système de recommandations auto-apprenant, Doctrine.fr vise justement à assurer cette complémentarité entre l'humain et l'intelligence artificielle, avec un accès au droit facilité pour tous.

L'accès à la donnée juridique brute, l'essence de l'intelligence artificielle et la garantie de la compétitivité nationale

De tels outils ne peuvent cependant exister sans accès aux données juridiques, et ce pour principalement deux raisons. Tout d'abord, il est nécessaire que ces outils disposent des données les plus exhaustives possibles, afin de ne pas induire les utilisateurs en erreur : il est donc essentiel que toutes les données soient disponibles et accessibles afin que le service ne desserve pas ses utilisateurs en leur fournissant une information incomplète.

Par ailleurs, l'amélioration des modèles d'intelligence artificielle nécessite d'avoir à disposition une base de données aussi large que possible, afin de confronter le modèle d'intelligence artificielle à un très grand nombre de cas et pouvoir ainsi le peaufiner en fonction de la diversité des situations existantes.

Les promesses de la justice assistée par l'intelligence artificielle ne peuvent donc être remplies qu'à condition que les décisions de justice soient publiées de la même manière que les textes de loi, à savoir de manière systématisée et sans modification du contenu. Cette ouverture, qui relève d'un mouvement nommé open data, a été consacrée textuellement⁽⁶⁾.

La mise en œuvre concrète de cet open data jurisprudentiel est une nécessité pour la compétitivité de la France sur la scène internationale et pour le développement des moteurs de recherche juridique à la française.

En effet, le mouvement d'open data est initialement apparu outre-Atlantique, où il continue de se développer et donc de faire apparaître de nouvelles entreprises dans le secteur du droit. Des moteurs de recherche intelligents sont ainsi déjà en train de naître aux États-Unis. Par exemple, la startup Judicata a récemment levé plusieurs millions de dollars pour soutenir le développement de son moteur de recherche juridique et promouvoir l'adoption d'un modèle de moteur de recherche juridique à l'américaine⁽⁷⁾.

Dans ce cadre, la mise en œuvre de la politique d'open data juridique est souhaitable non seulement pour favoriser l'accès à la justice, mais également dans une optique de compétitivité nationale, à la fois économique, juridique et politique.

Comme le précise le professeur Loïc Cadet dans le rapport qui lui a été confié sur l'ouverture des données de justice⁽⁸⁾, « les perspectives d'amélioration de la prévisibilité du droit ouvertes par l'open data constituent, sous certaines conditions, un facteur d'amélioration de la compétitivité de l'ensemble des entreprises et de l'attractivité de la France ». Il ne fait en effet aucun doute que les entreprises américaines de justice assistée ne tarderont pas à s'implanter en France : en témoigne ainsi la récente arrivée à Paris de l'entreprise californienne Rocket Lawyer, qui propose de faciliter l'accès aux avocats⁽⁹⁾.

(6) Articles 20 et 21 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique.

(7) Il est à noter que l'un des investisseurs ayant participé à cette levée de fonds n'est autre que Peter Thiel, cofondateur de PayPal, ayant récemment fait partie de l'équipe de transition de Donald Trump. "Judicata Raises \$2M From Peter Thiel, Keith Rabois And Others To Give Lawyers Better Research And Analytics Tools", TechCrunch, accessible à l'adresse suivante :

<https://techcrunch.com/2012/12/11/judicata-raises-2m-from-peter-thiel-keith-rabois-and-others-to-give-lawyers-better-research-and-analytics-tools/>

(8) Ce rapport était explicitement prévu par les articles 20 et 21 de la loi pour une République numérique et a été remis à la garde des Sceaux le 9 janvier 2018. Il est accessible à l'adresse suivante :

<http://www.justice.gouv.fr/publications-10047/rapports-thematiques-10049/remise-du-rapport-sur-lopen-data-des-decisions-de-justice-31165.html>

(9) « Lancement de Rocket Lawyer France, la plateforme qui rend le droit accessible à tous », article accessible sur le site Carrières-Juridiques :

<https://www.carrieres-juridiques.com/actualites-et-conseils-emploi-juridique/lancement-de-rocketlawyer-france-la-plateforme-qui-rend-le-droit-accessible-a-tous/1488>

Il n'appartient qu'à la France de se doter de ses propres réussites dans le domaine du droit, en favorisant le développement de ses propres pousses. L'alternative est simple : accepter la domination des États-Unis dans ce secteur, et ce alors que l'intelligence artificielle a été érigée comme domaine d'investissement privilégié par le gouvernement⁽¹⁰⁾.

Outre le rayonnement économique de la France, le développement de l'open data juridique permettra de « préserver l'indépendance de son système juridique et l'attractivité de son droit⁽¹¹⁾ ». La facilitation de l'accès au droit est en effet susceptible de renforcer la compétition entre systèmes juridiques : là où la France est de tradition civiliste, le Royaume-Uni et les États-Unis sont des pays de *common law*, et accordent donc une attention renforcée à la décision de justice. Il est probable que des moteurs de recherche juridiques vont y croître rapidement, facilitant l'accès à leur droit non seulement pour les justiciables, mais également pour tout tiers ou étranger intéressé.

Le développement de moteurs de recherche juridiques est donc également vecteur de diffusion d'une pensée et d'un modèle juridique, domaine dans lequel la France est historiquement déjà pionnière. Les conditions d'application de l'open data juridique impacteront donc directement l'attractivité de notre droit. Par exemple, sur le sujet de l'anonymisation des professionnels de la justice, le rapport Cadiet rappelait que « dans un contexte de compétitivité, tout spécialement entre les systèmes de droit civil et ceux de *common law* [...] le maintien du nom des professionnels du droit est un gage de sérieux et de fiabilité, et donc de crédibilité et d'attractivité à la fois du droit français et des juridictions étatiques françaises ». Il ne tient donc qu'à la France de maintenir la préséance de son modèle juridique sur la scène internationale en favorisant les initiatives permettant d'y accéder.

En somme, l'intelligence artificielle peut permettre le développement d'outils de justice, qui élargiront l'accès au droit et faciliteront le travail des avocats. Où est le Google du droit ? En France, il n'attend que l'ouverture des décisions de justice, avec le moteur de recherche Doctrine.fr et toutes les legaltechs qui analysent ces données.

(10) Le site AI for Humanity résume bien les intentions de la France en la matière : devenir un leader en intelligence artificielle

<https://www.aiforhumanity.fr/>

(11) Extrait du rapport Cadiet.