

Aurélien Bougeard

## **Journée de la jeune recherche en propriété intellectuelle - 2017**

Le « Big Data » à l'aube d'une réglementation : le cadre juridique des systèmes d'analyse prédictifs.

Thèse réalisé à l'Université de Strasbourg depuis 2016, sous la direction de Monsieur Franck Macrez.

### **Postulat des travaux de recherche**

L'intitulé du sujet de thèse est le suivant : Le « Big Data » à l'aube d'une réglementation : le cadre juridique des systèmes d'analyse prédictifs.

L'objectif est d'exposer ici, le cheminement d'une réflexion permettant une articulation logique des différents éléments. C'est pourquoi l'intitulé semble être scindé en deux.

Le droit norme, défini et délimite juridiquement des sujets complexes par le biais du droit ou d'un travail théorique juridique et pour le droit. Dans le cas précis de sujets portant sur les technologies de l'information et de la communication ou encore de l'informatique, manipuler de tels sujets par le prisme unique de rattachement à des connaissances juridiques est une impasse.

Il est nécessaire de se pencher sur la transdisciplinarité du droit et son application. L'avenir de la propriété intellectuelle repose sur cette transdisciplinarité, notamment en ce qui concerne le travail d'appréhension et de définition de ces sujets.

L'un des objectifs entrepris avec ces travaux est d'apporter des solutions juridiques en manipulant des outils para-juridiques ou techniques. Il est également important en l'absence d'une norme juridique de se pencher sur la norme technique pour comprendre comment les acteurs de ces domaines, qui ont pour activité l'exploitation de données massives ou la création d'algorithmes, pensent la norme technique pour se protéger, définir, exploiter et appréhender leur domaine d'activité au regard d'une norme juridique logiquement permissive.

Les praticiens, en dehors du champ juridique, l'ont très bien compris, « Code is Law » (Lawrence Lessig – 1999). Cette maxime est une réponse *a priori* à vocation architecturale, ayant pour objectif de structurer la norme juridique et la loi et non pas d'être une substitution permanente à ces dernières. C'est ce qu'on pourrait appeler du *soft law*.

L'ambition qui anime ces travaux de recherche est d'utiliser ce corpus de textes, de normes technique, de réflexions philosophique et de problématiques sociologique pour structurer une pensée juridique.

## Le sujet

Les systèmes d'analyse prédictifs, font partie de la famille des algorithmes et ont besoin d'une matière première pour fonctionner, pour avoir une finalité. Cette matière première se sont les données, des biens informationnels direct, indirect ou anonymes (définis en partie et pour la première fois dans la loi informatique et libertés de 1978).

Dans un objectif d'efficacité de ces systèmes d'analyse prédictifs, il faut s'intéresser nécessairement à cette matière première, dont la particularité repose sur la quantité. On parle alors de Big Data ou de données massives qui représentent *a minima* une quantité d'environ 76 téraoctet (Dominique Cardon). L'efficacité des algorithmes dépend donc de l'existence de ces données massives.

Ce qui est intéressant d'observer est qu'aujourd'hui les données massives comme les systèmes d'analyse prédictifs ne bénéficient d'aucun cadre juridique adapté, pour la simple et bonne raison que le droit ne sait pas appréhender momentanément ces nouveaux sujets. L'objectif de cette thèse est donc d'essayer de manipuler ces innovations dans un ensemble articulé permettant d'en faciliter la compréhension juridique.

Il est donc nécessaire, ici, de partir de ce que l'on connaît juridiquement, c'est à dire la notion de donnée qu'elle soit personnelle ou non, pour ensuite s'intéresser à cet agglomérat de données que sont les Big Data, prérequis substantiels et nécessaires à l'utilisation des systèmes d'analyse prédictifs.

La donnée à caractère personnel occupe une part importante dans l'espace normatif alors que celle-ci, au regard de la définition de 1978, ne représente qu'une infime quantité de celles qui transitent et qui sont extraites quotidiennement.

Cette logique qui consiste à placer cette catégorie de donnée au centre de la norme juridique, peut s'interpréter comme la mise à disposition, par le législateur, pour les personnes, d'un moyen pour se défendre quant à l'utilisation de ces données qui font leur identité.

L'exemple récent le plus probant quant à l'importance de la donnée à caractère personnel, au centre des préoccupations juridiques, est l'arrêt Google du 13 mai 2014 consacrant le droit à l'oubli. Celui-ci a laissé croire que les internautes avaient de nouveau un « droit de propriété » sur l'existence de certaines données à caractère personnel référencées sur les moteurs de recherche ou sites web.

Toutefois, même si la donnée à caractère personnel est très encadrée voir harmonisée (Règlement GDPR), et que son extraction était, hypothétiquement, interdite demain, les Big Data n'en seraient que faiblement impactés pour la simple raison que ce qui a déjà été extrait ne peut pas être supprimé et que l'outil d'analyse que sont les algorithmes n'a pas nécessairement besoin de données à caractère personnel pour cibler, analyser et prévoir un comportement.

Ce qui est nécessaire de mettre en avant, est de souligner l'importance d'une protection des données personnelles, sans pour autant « géocentrer » notre attention sur ces données. C'est une mise en

garde qu'il est important de faire, au regard de la récente jurisprudence du Conseil d'État en date du 8 février 2017, l'arrêt JcDecaux. Cet arrêt et sa portée démontrent la méconnaissance que l'on a de ce domaine. En se raccrochant systématiquement à la notion de données à caractère personnel, on est en train de méconnaître et brimer juridiquement un secteur économique en pleine expansion, celui de l'exploitation de ces données massives par le biais des algorithmes.

Ces derniers commencent à s'insérer dans la vie quotidienne des individus, notamment avec l'usage d'Internet, mais également avec tous les appareils électronique comme les téléphones, les aspirateurs autonomes, les voitures, les applications. Ces algorithmes ont différentes finalités, comme par exemple les systèmes d'analyse prédictifs, dont l'objectif est d'anticiper certains résultats au regard d'une masse de données à l'entrée. Il y a également ce qu'on appelle le machine learning, c'est à dire un algorithme qui va sous supervision ou non, déduire des résultats des données fournies à l'entrée et qui peut mettre en lumière des corrélations entre différentes masses de données, que le data-mining ou la statistique n'auraient pu exposer. Nous n'aborderons pas ici, l'étape la plus avancée qui est celui du deep learning qui consiste en une architecture neuronale d'algorithmes dans une finalité, entre autre, de cognition.

L'immixtion de cette technologie soulève un certain nombre de questions et de problématiques juridiques, comme le régime de protection de ces algorithmes qui visiblement aujourd'hui ne les conçoit que comme de « simple succession d'opérations ne traduisant qu'un énoncé logique de fonctionnalités » et de ce fait insusceptible de protection (cour d'appel de Paris - arrêt du 08/09/2017) ; ou encore l'impact de ces derniers sur le régime de la responsabilité et sur son adéquation et son respect des droits fondamentaux.

Le point sur lequel il faut insister est que les implications juridiques que soulève la démocratisation de ces algorithmes vont trouver une solution nouvelle ou de rattachement à la norme juridique, malgré l'absence, d'une définition, d'un cadre juridique de l'objet même. Et c'est à ce titre, qu'il faut se pencher sur la norme technique.

Si la norme juridique se refuse à encadrer, à appréhender ces innovations, c'est l'appropriation privée de la norme qui fera le droit.

## **Conclusion**

Le travail d'appréhension de ce sujet de thèse a permis de mettre en avant plusieurs courants. Le premier est celui d'un droit dépassé par l'innovation, sa complexité et ses implications. Ce droit juridico-juridique n'essaye pas de comprendre cet objet à travers d'autres prismes que la jurisprudence, la norme et la glose.

Le second courant est celui de l'appréhension juridique de ces innovations par le biais d'autres domaines comme la philosophie (Antoinette Rouvroy), la sociologie (Dominique Cardon) l'histoire (Alain Supiot).

Le dernier courant est celui d'une réflexion juridique réalisée *a priori* par les acteurs eux-mêmes qui permet aux autres acteurs de comprendre où se trouvent les intérêts économiques et juridiques de celui-ci et d'en connaître les limites (Rand Hindi, Yann Le Cun, Paul Duhan, Mounir Mahjoubi)

L'évolution du droit des NTIC réside dans sa capacité à pouvoir diversifier son savoir pour créer de la norme juridique. Aujourd'hui, cette diversification passe par la transdisciplinarité du droit mais également par l'appropriation privée de la norme et ses conséquences juridiques mais aussi sociologiques et économiques.

### **Échange constructif et débat informel**

- La privatisation d'une norme *a priori* comme moyen d'appréhender la norme juridique ;
- La fonction des métiers du droit à l'heure des systèmes d'analyse prédictifs et de la justice prédictive ;
- L'impasse d'une définition juridique du Big Data, l'intérêt d'une pluralité d'approches juridique, d'encadrement juridique en fonction de leur finalité, de leur utilisation ;
- L'évolution de notre rapport aux travaux de recherche extérieurs au droit, dans un but d'insertion et d'appréhension de leur objet dans l'espace juridique ;