

Convergences du Droit et du Numérique – Édition 2020  
Anonymization des décisions de justice

Louis Béziaud

*Univ Rennes, CNRS, IRISA, France*  
*Université du Québec à Montréal, Canada*

`louis.beziaud@irisa.fr`

## 1 Communauté

Je suis doctorant en *informatique*. Mon sujet d'étude porte sur les problématiques de vie privée et d'éthique liées à l'utilisation d'intelligence artificielle en droit.

## 2 Contribution

On assiste depuis peu à la mise à disposition massive de la jurisprudence par le biais de services en ligne (e.g., Légifrance, CanLII ou CourtListener). Ce phénomène est poussé par les mouvements *Open Data*<sup>1,2</sup> ainsi que par la nécessité d'accessibilité et de transparence de la justice. Combinées aux avancées en intelligence artificielle et en traitement automatique du langage naturel, ces données permettent l'émergence de technologies juridiques (Legaltechs) offrant de nombreux services facilitant par exemple la recherche de jurisprudence ou l'analyse de contrats. Cependant, la distribution de données juridiques pose de nombreux problèmes de vie privée, dus à la sensibilité du contenu et au caractère identifiant des jugements. La nouvelle simplicité d'accès est un avantage en matière de transparence et d'utilité mais n'offre plus la protection par "obscurité pratique" du format papier à diffusion restreinte.

Les technologies de caviardage (censure) utilisées actuellement pour l'anonymisation des décisions se limitent à la détection et suppression d'entités telles que les noms, dates de naissance, lieux de naissance. Cette limitation offre une protection très limitée et non quantifiable si l'on considère les traitements massifs mis en œuvre par les Legaltechs. Le caractère non-structuré inhérent au langage naturel avec lequel sont rédigées les décisions juridiques empêche d'utiliser les approches formelles développées ces dernières années en matière de protection de la vie privée (e.g., differential privacy).

Nous proposons de contourner la difficulté d'anonymiser les décisions de justice, et donc la tension entre vie privée et transparence, en offrant deux modes de publication : une publication traditionnelle du texte, dont l'accès serait limité, combinée à une publication "massive" dans un format offrant des garanties en matière de vie privée. On peut, pour ce dernier point, profiter du fait que le traitement massif de documents implique toujours une première étape d'extraction d'informations structurées. Il est donc envisageable de diffuser, pour l'usage des Legaltechs, des données juridiques pré-traitées (structurées) et anonymisées.

---

<sup>1</sup>Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique (1), Article 20

<sup>2</sup>Declaration on Free Access to Law <http://www.worldlii.org/worldlii/declaration>