Open data des décisions de justice : la Cour de cassation mise sur l'intelligence artificielle

La loi de 2016 pour une République numérique prévoyait de publier gratuitement en ligne toutes les décisions de justice et renvoyait à un décret ultérieur le soin de préciser dans quelles conditions cette ouverture devait se faire. La loi de réforme de la justice de mars 2019 est venue rappeler l'ambition d'ouverture tout en resserrant son encadrement, s'agissant notamment de l'identité des personnes physiques.

Depuis janvier 2018 et sa reprise en main du processus d'anonymisation, alors qu'elle sous-traitait auparavant cette tâche au site Légifrance, la Cour s'appuyait déjà sur un premier logiciel, composé d'une interface d'annotation des décisions et d'un outil permettant d'occulter les éléments susceptibles d'identifier une personne.

Devant l'efficacité toute relative de ce logiciel toutes les décisions publiées en open data par la Cour font obligatoirement l'objet d'une relecture humaine.

Jusqu'à présent, c'est donc à une équipe de 3 à 4 personnes que l'on devait la publication de 12 000 à 15 000 décisions par an. Alors même que les bases Jurinet et Jurica de la Cour s'enrichissent annuellement de 15 000 et 150 000 décisions chacune. Au total, on estime à 4 millions le nombre de décisions rendues chaque année par les tribunaux de France.

Grace à une micro-équipe de spécialistes du numérique intégrée, la Cour de cassation est parvenue à concocter un nouveau logiciel d'anonymisation des arrêts de la Cour suprême et des décisions des cours d'appel beaucoup plus performant, et qui devrait entrer en action d'ici la fin de l'année.

À terme, la publication de ces décisions en open data pourrait enfin passer à la vitesse supérieure, en se passant en grande partie de toute intervention humaine.

Cependant le nouveau logiciel qui atteint 99,96 % de réussite sur l'anonymisation des noms et prénoms, est limité par le fait qu'il n'existe pas, à ce jour, d'harmonisation au niveau de l'ensemble des tribunaux de France de la manière de rédiger et structurer une décision. Or, les machines supportent difficilement tout ce qui n'est pas standardisé.

L'humain est et restera indispensable.