

Sous quelles conditions les algorithmes de justice « prédictive » pourraient-ils réellement améliorer la prévisibilité de la justice ? par Sylvie THORON

Parmi les manifestations controversées de ce que l'on appelle la transformation numérique du système judiciaire, figurent, en bonne place, les algorithmes dits de justice prédictive, censés délivrer des prévisions sur les décisions de justice. Comment fonctionnent-ils et comment sont-ils utilisés? Jean-Marc Sauvé expliquait en 2018, alors qu'il était encore vice-président du Conseil d'État que « [...] les algorithmes prédictifs, fondés sur l'ouverture progressive, mais massive et gratuite des bases de jurisprudence à tous – l'open data –, visent à accélérer le règlement des litiges et à accroître la sécurité juridique, en améliorant la prévisibilité des décisions de justice. » En France, la loi pour une République numérique du 7 octobre 2016 prévoit en effet que « [...] les décisions judiciaires sont mises à la disposition du public gratuitement dans le respect de la vie privée des personnes concernées. » Des entreprises privées exploitent ces données codées pour développer des algorithmes d'apprentissage automatique et proposer des moteurs de recherche intelligents à des avocats ou à des compagnies d'assurance. L'algorithme nourri de décisions de justice codées apprend quelles décisions sont « similaires » les unes aux autres et fournit des statistiques sur des groupes de décisions similaires. Dans le cas d'un litige civil par exemple, une partie ou son défenseur peut entrer les caractéristiques de l'affaire qui le concerne dans l'algorithme et celui-ci produit des statistiques sur une sélection de décisions judiciaires antérieures. Ces statistiques sont alors interprétées comme la décision future attendue dans l'affaire.

Étant donné que ces prédictions sont donc essentiellement des extrapolations basées sur des décisions passées, il faut comprendre que leur fonctionnement dépendra du contexte. Lorsque la justice est stable et homogène, les extrapolations auront de bonnes chances d'être de bons indicateurs de décisions effectives. Des cas « similaires » aboutiront à des décisions « similaires ». Ainsi, en synthétisant les décisions, les algorithmes pourront aider à rendre le système juridique transparent. Lorsque le système juridique n'est ni stable ni homogène, les extrapolations risquent par contre d'être « très peu précises », c'est-à-dire très différentes d'une décision effective. Les « similitudes » seront difficiles à définir. Dans ce contexte, les algorithmes fonctionneront différemment, non pas en rendant la justice plus transparente mais en agissant sur le comportement de ses usagers. Pour des parties qui envisagent une médiation, les résultats de l'algorithme constitueront une référence qui aura de grandes chances d'orienter les résultats de la médiation. Même si une « moyenne » des décisions passées n'est pas forcément la meilleure décision dans ces environnements hétérogènes, les acteurs y verront une possibilité de diminuer l'incertitude, selon un effet d'ancre bien connu des psychologues.

Le risque serait alors que l'algorithme soit utilisé comme instrument pour gouverner des individus qui s'en remettent à une rationalité supposée mais non vérifiable. Il nous semble important, au contraire, de stimuler la rationalité des usagers du système judiciaire en offrant les moyens d'analyser les décisions passées sans prétendre être prédictif. Or tout ce qui contribue à en faire une boîte noire encourage les interprétations simplistes. Dans le cas précédant d'une grande hétérogénéité des décisions, celle-ci doit au minimum transparaître dans la façon dont les résultats sont présentés. Il faut en particulier réfléchir au minimum d'information nécessaire concernant la liste *exhaustive* des décisions qui ont été *choisies* par l'algorithme. Le principal avantage de l'algorithme n'est pas de rendre la justice humaine plus juste mais d'en révéler les faiblesses.

Il est cependant fort probable que les sociétés qui développent ces algorithmes ne soient pas prêtes à divulguer des informations qui sont conçues comme un secret de fabrication. Les algorithmes au sens strict de code informatique doivent rester publics, et les mécanismes d'apprentissage le sont, mais les sociétés ont bien d'autres choses à cacher. Elles n'ont pas à révéler comment elles ont nourri l'algorithme, en particulier les caractéristiques des données qu'elles ont choisies, ni comment l'algorithme a appris, en particulier les pondérations qu'il a attribuées aux différentes caractéristiques.

De manière paradoxale, dans le système actuel mis en place par le moyen de l'open data, les algorithmes de justice prédictive sont conçus tout à la fois comme rationnels et divinatoires. Censés rendre la justice prévisible et transparente, ils peuvent se révéler opacifiants. Pour nourrir la réflexivité des usagers il faut au contraire imaginer des algorithmes ouverts et interprétables. Cela pourra se faire si l'open source accompagne l'open data et si des collectifs de juges et d'avocats y travaillent, en collaboration avec des développeurs.