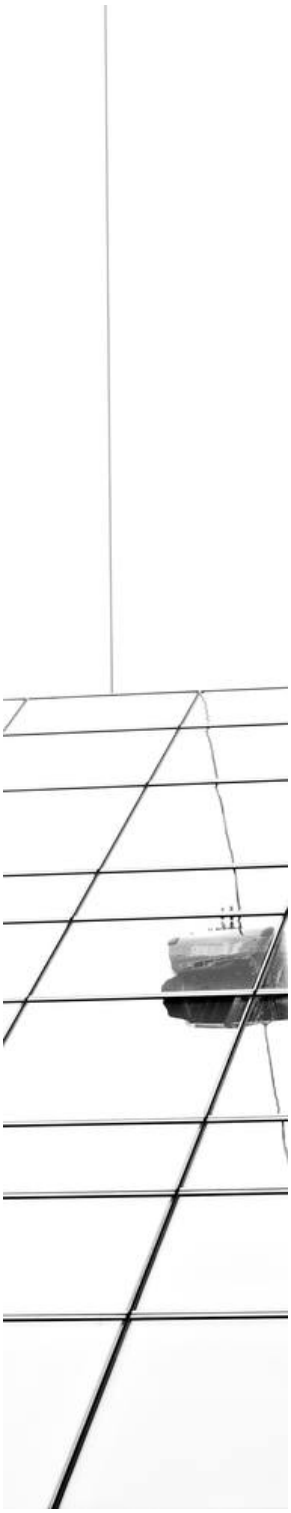

ARTICLE

IA : CINQ QUESTIONS À WINSTON MAXWELL

Droit de la propriété intellectuelle, média et art | 22/03/23 |

TECH & DIGITAL

Winston Maxwell, Directeur d'Etudes, Droit et Numérique à Télécom Paris, Institut polytechnique de Paris, ancien avocat et docteur en économie

Le projet de règlement européen sur l'IA protège-t-il suffisamment les droits fondamentaux ?

Tout le monde parle du projet de règlement sur l'IA, mais la décision de la Cour de justice de l'Union européenne du 21 juin 2022 est tout aussi importante. Dans l'affaire « La Ligue des Droits Humains », la Cour devait évaluer la compatibilité de la directive PNR, qui autorise les autorités de police et de renseignement à utiliser des algorithmes pour analyser les données des passagers, avec la Charte des droits fondamentaux de l'Union européenne. La Cour devait déterminer si ces mesures de surveillance étaient proportionnées et nécessaires dans une société démocratique. La Cour a estimé qu'elles l'étaient, mais a posé un certain nombre de conditions. Ces conditions s'appliqueront, avec plus ou moins d'intensité, à tout traitement algorithmique ayant un impact significatif sur les droits fondamentaux. C'est pourquoi je pense que l'arrêt de la Cour du 21 juin 2022 est au moins aussi important que le futur règlement sur l'IA lorsqu'il s'agit de cas d'usage de l'IA créant un risque élevé pour les droits fondamentaux.

Quelles sont les conditions imposées par la Cour ?

L'une des conditions les plus radicales est d'exclure totalement l'apprentissage automatique de la détection algorithmique des risques terroristes. Seules des règles humaines prédéterminées peuvent être utilisées.

Pourquoi la Cour a-t-elle exclu l'apprentissage automatique ?

Le manque d'explicabilité : un évaluateur humain examinant le niveau de risque généré par l'apprentissage automatique ne serait pas en mesure de comprendre le "motif" de cette évaluation, le rôle de l'évaluateur humain deviendrait inefficace. De même, un citoyen qui souhaiterait ultérieurement contester une décision fondée sur ce niveau de risque ne disposerait pas d'un recours efficace, car les raisons de l'évaluation resteraient cachées.

Quelles sont les autres conditions imposées par la Cour ?

La Cour a déclaré que les modèles utilisés pour identifier les risques terroristes devaient tenir compte à la fois des facteurs de "culpabilité" et d'"innocence". J'ai trouvé cela fascinant parce que la Cour essaie d'insérer certaines caractéristiques d'équité dans l'algorithme lui-même !

La Cour a déclaré que chaque alerte algorithmique devait être examinée par un être humain, afin de détecter d'éventuels « faux positifs » (fausses alertes) et de déceler toute discrimination, même si je ne vois pas très bien comment l'examen d'une alerte individuelle pourrait permettre de déceler une discrimination. La Cour a souligné que le système algorithmique ne devait pas conduire à une discrimination indirecte ("disparate impact" dans le langage américain) à l'encontre des groupes protégés. La Cour a exigé que l'efficacité du système dans la détection des risques de terrorisme soit prouvée, mais n'a pas précisé ce qu'il fallait entendre par "efficace". La mesure de l'efficacité est une question importante dans le domaine de la détection criminelle en raison du problème de déséquilibre des classes (erreur de taux de base), qui conduit inévitablement à une proportion élevée de faux positifs.

Le projet de règlement sur l'IA couvre-t-il l'un ou l'autre de ces points ?

Pas à l'heure actuelle. Le projet de règlement sur l'IA fait référence à la "surveillance humaine", à la "partialité" et aux risques pour les droits fondamentaux, mais ne fournit pas actuellement de cadre en matière de droits de l'homme pour analyser la proportionnalité des risques liés à l'IA. L'arrêt de la Cour du 21 juin 2022 le fait, bien que pour un cas d'utilisation très sensible - les algorithmes antiterroristes. D'autres cas d'utilisation seront moins sensibles que l'affaire PNR, mais l'analyse progressive de la proportionnalité effectuée par la Cour devrait fournir un modèle pour l'évaluation de toute application de l'IA qui crée des risques pour les droits fondamentaux, même dans le cas de mesures prises par le secteur privé.