Audition devant la Commission nationale consultative des droits de l'homme (CNCDH)

Comment assurer l'efficacité du contrôle humain dans les systèmes de décision algorithmiques?

12 janvier 2022

Winston Maxwell

Telecom Paris, Institut Polytechnique de Paris
winston.maxwell@telecom-paris.fr

telecom-paris.fr/ai-ethics

Le contrôle humain dans les systèmes de décision algorithmiques

Que signifie "contrôle humain"?

A quoi sert-il?

Comment maximiser son efficacité?

Une pléthore de termes

Des travaux importants sur la définition du contrôle humain

Intervention humaine (RGPD) … significative (CEDP)	human-in-command (HIC) (idem)
pas sur le "fondement" ou "seul fondement" d'un traitement automatisé (LIL art. 47)	Meaningful Human Control (MHC); contrôle humain, jugement humain (droit international humanitaire)
contrôle humains significatifs, supervision humaine à tout moment (Parl. europ. résol. 20 oct 2020)	moyens non-automatisés (CJUE; Directive PNR; loi respect des principes de la République; projet règlement DSA)
contrôle humain, contrôle effectif (projet de règlement AI Act art. 14)	examen humain significatif (meaningful human review) (Loi de l'Etat de Washington)
human-in-the-loop (HITL) (lignes directrices HLEG)	garantie humaine (recommandations CCNE bioéthique)
human-on-the-loop (HOTL) (idem)	

Un contrôle humain pour quoi faire?

1. Détecter et corriger des erreurs
2. Garantir un processus de décision respectueux de valeurs humaines
3. Fixer les responsabilités et démontrer la conformité

Quels leviers pour agir sur la qualité d'une intervention humaine?

L'étalon d'or en termes de qualité	Les leviers pouvant affecter la qualité
La décision collégiale d'un tribunal qui prendrait en considération la recommandation algorithmique en tant qu'élément de preuve parmi d'autres, avec l'ensemble des garanties d'un procès équitable	Le temps disponible au décideur humain Les autres informations disponibles au décideur La formation du décideur L'indépendance et l'impartialité Une pluralité de décideurs et/ou la confrontation des points de vue (explications de points de vues alternatives) La participation des personnes affectées (procès équitable)

Quelles propositions pour l'IA Act? (1/2)

Le principe	Sa mise en oeuvre
1. Exiger une intervention humaine pour toute décision affectant les droits et libertés individuels	Imposer le principe directement sur l'utilisateur (actuellement les mesures de contrôle humain seront définies par le fournisseur et décrites dans la notice d'utilisation) L'idée d'une intervention humaine **obligatoire mais à géométrie variable** Lever la confusion liée au champ d'application de l'article 22 RGPD
2. Préciser les objectifs de l'intervention humaine	Les modalités de cette intervention seront fixées par le fournisseur et l'utilisateur afin de • minimiser les risques associés aux décisions erronées • préserver les valeurs humaines liées à la procédure • établir une chaîne de responsabilisation pour l'exploitation de l'algorithme

Une méthode inspirée de la régulation des communications électroniques

Préciser les objectifs à atteindre et fournir une boîte à outils + une feuille de route pour les atteindre

Quelles propositions pour l'IA Act? (2/2)

Le principe	Sa mise en oeuvre
3. En fonction du cas d'usage construire les modalités d'une intervention humaine efficace Une intervention humaine à géométrie variable	- une division claire entre les tâches confiées à l'algorithme et celles confiées à l'humain - définir les informations sur lesquelles l'humain s'appuie - définir le timing de l'intervention - préciser la formation des personnes - une explicabilité conçue pour favoriser l'intervention humaine et contrebalancer les biais de l'automatisation
4. Intégrer l'intervention humaine dans l'analyse d'impact et dans la démarche de co-régulation	- Intégrer le choix des modalités d'intervention humaine dans les analyses d'impact effectuées par le fournisseur et l'utilisateur de l'algorithme - Lignes directrices du comité européen de l'IA