

24 octobre 2023

Recommandations

L'accès à la justice par l'IA

Par et pour les communautés marginalisées et/ou sous-représentées

Rapport produit dans le cadre des travaux du sous-projet 11/Axe 3



**AUTONOMISATION
DES ACTEURS JUDICIAIRES
PAR LA CYBERJUSTICE**

CRSH  SSHRC
Conseil de recherches en sciences humaines
Social Sciences and Humanities Research Council

Cette recherche a été financée par des fonds du CRSH

Éditeur Laboratoire de cyberjustice



Participation de groupes marginalisés et/ou sous-représentés :

Guide pour soutenir l'inclusion de leurs besoins et valeurs

Karine Gentelet, Professeure agrégée, UQO

Marie Zumstein (Candidate au doctorat, UdeM) et Lily-Cannelle Mathieu (Candidate au doctorat, McGill)

TABLE DES MATIÈRES

1. Introduction	p.3
2. Projet de recherche, méthodologie et résultats	p.3
3. Principes directeurs	p.4
4. Recommandations	p.6
Annexe 1 : Incidences générales de l'IA sur l'accès à la justice	p.9

1. Introduction :

Des travaux de recherche menés dans le cadre du Chantier sur l'autonomisation des groupes et communautés marginalisés, ont permis de s'intéresser aux raisons pour lesquelles des tensions peuvent survenir lors du déploiement d'un outil d'IA dédiée à l'accès à la justice auprès de populations marginalisées. Une synthèse d'initiatives déjà mises en place sur des thèmes entourant des enjeux de justice sociale a ainsi permis de dégager plusieurs éléments à considérer quant aux conditions d'implication des populations dans le design, le déploiement ou l'utilisation d'outils d'IA :

- Justice par le design
- Témoignage par la donnée et activisme par la preuve
- Éducation populaire et accès à l'information
- Mobilisation citoyenne
- Contre-surveillance
- Contre-cartographie
- Représentativité, inclusivité et souveraineté des données

En effet, ces outils développés par et pour les groupes et communautés marginalisés ou/et peu représentés nous ont permis, à partir de leurs représentations, de déterminer les conditions importantes de l'inclusion effective et efficace de ces groupes et communautés au sein des projets d'accès à la justice par l'IA.

2. Projet de recherche, méthodologie et résultats

L'objet de cette recherche était concentré sur l'étude des initiatives pour la justice sociale déjà mises en place dans les différentes régions du monde, à destination des communautés marginalisées et/ou sous-représentées. Il s'agit en effet d'atténuer les incidences potentiellement négatives de l'IA sur les biais et discriminations (voir annexe 1 – incidence 1), une préoccupation récurrente en matière d'éthique sociale de l'IA. Les populations marginalisées et sous-représentées vivant déjà en marge de la société, il est important de ne pas aggraver la situation par la création et le déploiement de nouveaux outils d'IA pour la justice qui pourraient aller à l'encontre de leurs droits à la dignité et la non-discrimination.

Pour ce faire, la question qui se posait alors pour nous était la suivante : « Comment les groupes et communautés marginalisés abordent-ils les problèmes de justice par l'intermédiaire des technologies numériques? » Partir de leurs propres conceptions nous a alors permis d'apprendre sur leurs priorités, leurs besoins et leurs méthodes pour y remédier.

Dans un premier temps, des recherches qualitatives ont été menées sur internet entre 2019 et 2021, au moyen de mots-clés en anglais et en français. Il s'agissait de recenser les initiatives de justice sociale qui utilisaient et développaient des technologies numériques et/ou d'IA *par* et *pour* les communautés et groupes sous-représentés, marginalisés ainsi que par les organisations de la société civile. S'en est ensuite suivi pour la période 2021-2022 la tenue d'une douzaine d'entretiens semi-structurés. Les intervenants, de genres différents, provenaient également de contextes socioculturels et géographiques eux aussi différents afin d'assurer une représentativité la plus large possible.

3. Principes directeurs pour une représentation et inclusion efficace dans les projets d'accès à la justice

Avant de s'intéresser aux recommandations en tant que telles, il nous apparaît essentiel de présenter les résultats de la recherche. À partir des analyses du corpus des données collectées, nous avons identifié un certain nombre de principes mis en avant par les participants à la recherche dans leurs représentations d'outils qui favorisent la justice. Ces principes constituent en ce sens les fondements sur lesquels doit reposer toute action d'inclusion :

- Considérer les enjeux de justice sociale et ce bien avant de traiter d'enjeux d'accès aux tribunaux. Les groupes et communautés marginalisés représentent des citoyens sous-représentés dans les espaces publics, mais surreprésentés dans le système de justice. Ce déséquilibre sociétal ne peut pas être ignoré dans les projets d'accès à la justice par l'IA.
- Traiter des enjeux de représentations dans les données en ce que ceux-ci constituent un facteur d'exclusion sociale et d'iniquité structurelle. Les enjeux de représentations dans les données constituent des enjeux déterminants qui sont structurellement antérieurs aux développements de l'IA. Ces enjeux de représentations sont sociétaux bien avant d'être techniques.

- Instaurer un sentiment de confiance envers les initiatives, leurs développeurs et en leurs opérateurs en incluant des personnes de la communauté dès le démarrage des projets ou au sein des structures de gouvernance.
- Développer une compréhension appuyée des besoins, des valeurs, des objectifs et de la dynamique interne des groupes et des communautés desservis, en amont de leur engagement dans le projet/initiative/outils, est essentiel. Cela peut se faire au moyen de rencontres régulières avec le groupe en question ou par le développement d'outils de collaboration, de manière à ce que chaque partie puisse exprimer ses besoins et limites. Ces connaissances et compréhensions permettront de favoriser une approche collaborative, préventive ainsi qu'une vision à long terme des problèmes sociétaux potentiels touchant à la communauté ou au groupe de la population concerné (par opposition à l'approche réactive des secteurs publics et privés).
- Proposer une réponse locale aux besoins énoncés par les populations concernées (et ce, indépendamment des impératifs économiques, politiques ou technologiques : les besoins sociaux du groupe et les droits humains doivent primer). Il convient de garder à l'esprit que les besoins de chaque groupe ne vont pas forcément de pair avec les besoins de la société en générale ou même, des autres communautés et minorités.
- Soutenir objectivement plusieurs principes présentés comme fondamentaux pour tous les groupes que nous avons interviewés lors de la conception des outils: respect de l'autodétermination des personnes, de l'autonomie, et du consentement dans le processus de décision. Cela doit aussi inclure un droit de retrait.
- Mettre en place des structures horizontales de gouvernance afin de miser sur des interactions de proximité.
- Inclure de manière continue ces groupes et communautés dans la gouvernance de leurs données et des outils déployés à leur intention de façon qu'ils participent à toutes les prises de décision.
- Déterminer conjointement les choix de certains partenaires pour le projet et respecter celui d'en exclure certains du fait de contextes sociohistoriques préexistants de discrimination ou de surcriminalisation. Sans ceci, la participation active de certains groupes spécifiquement visés pourrait être compromise.

Dans ce qui suit, nous allons nous présenter une liste de recommandations à déployer en vue de mettre en place des outils numériques par le biais de l'éducation aux droits. En effet, il est important que les différentes parties et partenaires soient sensibilisés aux meilleurs moyens d'inclure et de faire participer les groupes et communautés marginalisés et/ou sous-représentés lorsque sont développés et déployés des outils d'IA à leur égard. Ceci, dans une optique collaborative et horizontale. C'est de cette manière que leurs besoins et droits seront respectés dans cette nouvelle ère technologique qui s'amorce.

4. Recommandations pour une participation effective des groupes et communautés marginalisés dans les projets en IA d'accès à la justice

Les résultats de la recherche nous ont permis de dégager plusieurs recommandations qui vont permettre d'opérationnaliser les principes présentés plus haut. Ces pratiques devraient permettre de développer des outils d'IA d'accès à la justice plus pertinents, inclusifs et respectueux des groupes et communautés marginalisés et/ou sous représentés:

1. *Identifier* clairement le groupe ou la communauté potentiellement concerné par l'outil d'IA à développer en prenant en compte le contexte historique et systémique de discrimination et/ou de violations des droits humains sur la population concernée.
2. *Développer* un modèle d'outil spécifiquement pour ce groupe ou communauté. Il n'est pas envisageable de penser qu'un même modèle puisse s'appliquer de manière homogène à toutes les communautés.
3. *Procéder* à un audit approfondi des données qui vont être utilisées pour le développement et le fonctionnement dans l'algorithme de l'outil afin que la population visée soit effectivement et adéquatement représentée, ceci répondant aux problématiques de biais et discriminations. En effet, la question du recours aux données n'est pas indépendante de celle de la représentation et de la visibilité des groupes et populations sous-représentées. C'est en outre un élément fondamentalement important pour des groupes qui sont encore systématiquement mal représentés dans les données.
4. *Consulter*, préalablement à la conception du projet, les représentants du groupe ou de la communauté identifiée et pour lequel l'outil d'IA est destiné afin d'inclure leurs besoins, valeurs et objectifs ainsi que de tenir compte de la dynamique interne du groupe. Cette

étape est d'autant plus cruciale puisqu'elle permet d'apporter une réponse adaptée et locale, et respectueuse des expériences vécues et structures internes de gouvernance.

5. *Inclure*, pour la suite de la mise en œuvre du projet des représentants de la communauté ou du groupe impacté par la création, par exemple, d'un comité de pilotage citoyen. Il convient de veiller à ce que ce comité soit représentatif de la composition, de la structure et du fonctionnement de la communauté en question. Cette initiative permet non seulement de s'assurer de la reconnaissance de la valeur des structures organiques, d'un processus inclusif, mais également de la mise en place d'actions et de décisions ancrées sur des connaissances expérientielles et concrètes, et répondant à des besoins réels et utiles.
6. *S'assurer* ensuite d'un processus de consultation continue et de la participation effective de ce comité de pilotage citoyen à tous les stades du développement et du déploiement de l'outil d'IA. La participation effective de ce comité favorise en effet des modèles de gouvernance horizontaux fondés sur l'inclusion de la communauté desservie dans le processus décisionnel. Ceci s'avérant crucial pour rendre l'initiative d'IA déployée ainsi que ses opérateurs dignes de confiance aux yeux de la population concernée.
7. *Prendre en compte* les avis et besoins énoncés par le comité de pilotage citoyen, et ce même lorsqu'ils ne représentent pas forcément les objectifs poursuivis initialement. En effet, les principes d'autodétermination, d'autonomie, et de pouvoir effectif dans le processus décisionnel sont décisifs pour favoriser l'utilisation du système d'IA par les membres de la population concernée. Cette démarche s'assure non seulement que le bien-fondé de l'IA déployé répond au contexte socioculturel en l'espèce, mais également que le modèle de structure de gouvernance horizontal est respecté.
8. *Suivre* les impacts du déploiement de l'outil d'IA par le comité de pilotage citoyen et procéder aux réajustements nécessaires, et ce à n'importe quel moment de la mise en œuvre. Ce processus ne devrait en aucun cas être unilatéral ou unidimensionnel. La mise en place d'un système d'IA au sein d'une communauté doit demeurer un processus flexible et continu où les rétroactions sont possibles à chaque stade.
9. *Permettre* aux représentants ou structures de la communauté le moyen de s'approprier la technologie d'IA déployée afin de s'assurer de la pérennité à long terme du projet par un ancrage local. Cela pourrait être par exemple au moyen d'une technologie ouverte ou d'une formation technique. Les enjeux d'autonomie et d'émancipation des communautés sous-

représentées et/ou marginalisées vont permettre de s'assurer de leur inclusion effective dans la mise en œuvre de l'outil d'IA qui leur est destinée.

- 10.** *Réévaluer* de manière continue l'outil et la procédure de participation s'il est envisagé de déployer ce même outil d'IA vers un autre groupe qui serait marginalisé ou sous-représenté. Il serait ainsi important de *reprendre* chacune des recommandations énoncées ci-dessus dans leur intégralité.

Annexe 1: Incidences générales de l'IA sur l'accès à la justice

La recherche nous a également permis de dégager 6 incidences plus générales et potentiellement négatives de l'IA sur l'accès à la justice. Nous les incluons dans le présent document en annexe, en ce qu'elles sont importantes à considérer lors de développements d'IA en matière d'accès à la justice et d'autant plus cruciales lorsque l'on cherche à déployer des projets, initiatives ou outils d'IA dans les populations et communautés marginalisées et/ou sous-représentées.

- Incidence 1 : Discrimination du fait de biais structurels

L'existence même des technologies d'IA se fonde sur l'assimilation des données aux algorithmes qui les composent. Or, ces données sont loin d'être neutres et prennent appui sur un tissu social déjà biaisé, ainsi que sur des rapports de pouvoir inéquitables entre les différents acteurs de la société. Les systèmes d'IA, dès lors, peuvent reproduire ou même amplifier ces biais et renforcer des discriminations existantes. Ceci pourrait s'avérer problématique, voire dangereux, si de tels systèmes venaient à être déployés dans des communautés déjà sous-représentées ou marginalisées.

- Incidence 2 : Uniformisation des valeurs socioculturelles

Le déploiement et l'utilisation de systèmes d'IA composés de caractéristiques similaires se font de manière homogène, et ce malgré les différences de contextes et de cultures. Or, l'application et l'utilisation de ces technologies dans certains contextes sociaux culturels peuvent se révéler problématiques dans la mesure où les valeurs, idées ou priorités ne sont pas toujours les mêmes que ceux dans lesquels l'IA a été pensée et créée. Ainsi, il est important de développer des systèmes d'IA adaptés aux situations des populations concernées afin de ne pas tomber dans ce que certains auteurs qualifient de « colonisation par l'IA ».

- Incidence 3 : Enjeux liés aux données

Il arrive dans certains cas que les citoyens ne considèrent pas leurs données personnelles comme privées, surtout lorsqu'il ne s'agit pas d'informations estimées comme étant sensibles. Or, ces données, même si elles ne sont pas forcément sensibles, ont une incidence non seulement sur le droit fondamental à la vie privée, mais plus largement sur la manière dont ces données seront utilisées pour ensuite représenter, voire même définir un individu, puis indirectement, des groupes

d'individus. Ces données n'étant, contrairement à ce que l'on peut estimer, pas toujours neutres, les incidences qu'elles peuvent avoir sur les communautés marginalisées ne sont pas négligeables.

- Incidence 4 : Prise de décision automatisée

La grande confiance généralement octroyée aux nouvelles technologies et particulièrement à l'IA, et la perception selon laquelle ces systèmes seraient objectifs et infaillibles semblent faire en sorte que les résultats ou données sortantes (*outputs*) de ces systèmes d'IA sont trop peu monitorés, vérifiés, évalués, ou ajustés, ce qui fait que certaines problématiques peuvent malheureusement passer inaperçues. Les répercussions négatives peuvent pourtant être importantes sur les personnes (particulièrement les populations sous-représentées ou les minorités) et contrevenir aux droits humains ou favoriser la discrimination et les biais. Toute prise de décision et les responsabilités qui l'accompagnent ne doivent jamais être déléguées à un seul système d'IA, une personne humaine devant être capable de se l'approprier et de maintenir une surveillance sur son fonctionnement.

- Incidence 5 : non-pertinence du recours aux technologies d'IA

Le portrait positif dressé de l'IA est parfois survalorisé. Il est pourtant essentiel de porter une réflexion critique concernant la pertinence du recours à une technologie d'IA, des apports potentiels ou de bien-fondé de l'utilisation de celle-ci dans une population ou un groupe de population, relativement à l'objectif à atteindre. Ceci de manière encore plus importante lorsque l'on connaît les incidences potentiellement négatives que l'IA peut engendrer non seulement sur les droits fondamentaux des individus, mais aussi sur l'environnement. Il convient de se rappeler que l'outil est un moyen pour arriver à ses fins, pas une fin en soi : n'y a-t-il pas une réponse sociale, économique ou politique à appliquer avant une solution technologique?

- Incidence 6 : Manque de clarté et de recours juridique direct

Il est régulièrement répété que le manque de clarté sur les données intégrées aux systèmes d'IA, l'opacité de ces derniers, leur complexité et leurs capacités d'apprentissage (particulièrement pour le cas des IA à apprentissage profond) font que les mécanismes habituels d'octroi de responsabilité peuvent être parfois difficiles à appliquer en cas de préjudice causé par une technologie d'IA. Une double incidence potentiellement négative peut alors s'exercer : comment préserver et faciliter les mécanismes de réparation dans le cas de dommages causés par un système d'IA et comment donner

aux populations, notamment les plus marginalisées et sous-représentées les moyens d'obtenir efficacement réparation. Cela se répercutant plus largement, sur le maintien d'une justice sociale effective.