



HAL
open science

GENRE, ALGORITHMES ET DROIT

Sophie Sereno, Emmanuelle Bonifay

► **To cite this version:**

| Sophie Sereno, Emmanuelle Bonifay. GENRE, ALGORITHMES ET DROIT. 2022. hal-03736474

HAL Id: hal-03736474

<https://amu.hal.science/hal-03736474>

Preprint submitted on 22 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GENRE, ALGORITHMES ET DROIT

Projet de recherche porté
par Emmanuelle BONIFAY et Sophie SERENO¹

Cette publication constitue la présentation d'une recherche financée par le GIS Institut du genre dans le cadre de l'appel à projets 2021.

PRÉSENTATION DU PROJET

Résumé : Les usages des algorithmes foisonnent dans tous les domaines de l'activité humaine. La santé, l'éducation, l'emploi, l'action publique ou encore la justice n'y font pas exception. Certaines décisions vitales sont d'ores et déjà prises par des algorithmes². Pourtant, les enjeux de cette automatisation décisionnelle et leurs potentiels discriminants envers les personnes demeurent largement sous-estimés par les décideurs privés et publics, nonobstant l'ampleur de la révolution numérique à l'œuvre. Des études récentes ont mis en évidence l'existence de biais de genre affectant des algorithmes décisionnels. Cependant, ce champ d'étude n'est pas investi par les juristes. Les défis induits par le déploiement des technologies automatisées deviennent un enjeu central eu égard à l'importance des stéréotypes sexistes susceptibles d'être systématisés par le recours aux algorithmes. ***Il apparaît alors indispensable, dans un premier temps, de développer une recherche exploratoire et pluridisciplinaire afin d'identifier les biais de genre engendrés par les algorithmes, leurs formes et leur ampleur dans les processus décisionnels (étape 1). L'étude tendra dans un second temps à identifier les solutions techniques et juridiques susceptibles d'en réduire les effets néfastes éventuels (étape 2).***

Contexte de l'étude :

Les algorithmes : de multiples fonctions.- A l'heure actuelle, l'usage des algorithmes ne cesse de s'étendre. Le recours aux algorithmes répond à plusieurs objectifs. D'abord, les algorithmes sont parfois utilisés comme un soutien dans la réalisation de tâches fastidieuses de recherches, de tri ou de classement de données massives. Dans le domaine juridique, cette « algorithmisation » a été notamment renforcée par les phénomènes d'*open-data* et l'émergence de *legal tech*, qui participent par ailleurs à l'accès au droit pour tous. Ensuite, les algorithmes permettent de simplifier de nombreuses démarches et services, remplaçant l'intervention humaine pour la prise de certaines décisions administratives individuelles, comme en matière de prestations sociales, de logement ou d'accès aux formations. Dans cette hypothèse, l'intervention humaine est alors réduite, voire inexistante.

Cette perte de maîtrise fragilise les droits des personnes. En effet, **les algorithmes utilisent et exploitent bien souvent des données à caractère personnel qui résultent de l'état civil des personnes, dont leur sexe.** Certes, le « genre » et le « sexe » d'une personne ne peuvent plus continuer d'être confondus. Alors que le sexe est une donnée biologique, le

¹ Co-porteurs du projet de recherche.

² A titre d'illustration, le système *score cœur* fixe l'ordre des candidats proposés à la transplantation d'un greffon de cœur repose sur un algorithme (V. not. pour une analyse des biais du *score cœur* : C. Henin, « Confier une décision vitale à une machine », *Réseaux* 2021, n°225, p. 189).

genre s'apparente à une construction sociale variable et évolutive, qui peut s'affranchir de la binarité femme-homme. Toutefois, force est de constater que du point de vue du droit, et plus spécialement d'un point de vue administratif, les règles juridiques d'identification des personnes physiques perpétuent une binarité féminin-masculin, sans accorder de véritable reconnaissance juridique aux personnes dites intersexe.

Par conséquent, les actes administratifs étant construits sur cette binarité, les algorithmes s'appuieront *ipso facto* sur deux données possibles d'identification de la personne : soit il s'agit d'une femme, soit il s'agit d'un homme. Dès que des algorithmes utilisent dans leur codage la donnée liée au sexe de l'individu, on peut déjà craindre un premier biais de genre à l'égard des personnes de sexe ou de genre indéterminé. L'utilisation des algorithmes, qui ouvre le champ à de nombreux biais possibles de genre, est susceptible de produire *in fine* des discriminations liées au sexe, à l'identité de genre ou à l'orientation sexuelle en reproduisant les stéréotypes sexistes préexistants dans notre société³. Il s'agit d'une réalité déjà mise en lumière notamment dans le domaine de l'assurance, dans lequel il a pu être constaté que les algorithmes utilisés pour calculer le risque d'accident de la circulation reproduisaient des stéréotypes sexistes.

Pourtant, en parallèle, les protagonistes de l'intelligence artificielle ne cessent de louer les avantages du recours aux algorithmes en raison de leurs promesses d'efficacité et de rapidité. En outre, l'idée s'est répandue que les algorithmes offrent une garantie d'objectivité des procédures et décisions⁴. Leur usage permettrait justement d'évincer les biais du jugement humain et ainsi écarter tout risque de discrimination dans les processus décisionnels. Autrement formulé, les systèmes d'algorithme ne pourraient pas procéder à des sélections genrées et donc à des discriminations au sens de la loi⁵. Mais la neutralité algorithmique est une illusion. Comment affirmer que les algorithmes sont totalement neutres, quand l'on sait que ces algorithmes ne sont rien d'autre que des outils conçus par des êtres humains ? Il est indéniable qu'au stade de la conception de l'algorithme, le concepteur peut être amené à reproduire des biais cognitifs, et ce bien souvent de façon inconsciente. Un double écueil existe. D'une part, les professionnels du secteur informatique sont majoritairement des hommes⁶. D'autre part, ils ne sont pas sensibilisés aux stéréotypes. Il en résulte un risque important d'intégration de biais de genre lors de la conception du code source et par conséquent un risque d'exploitation ultérieure de données biaisées. De manière indicible donc,

³ A titre d'illustration, certaines décisions de justice ont évalué le préjudice sexuel subi par les femmes âgées de 50 ans au plus comme étant moindre que celui subi par des hommes de même âge d'autant plus si elles avaient déjà eu des enfants. Cette appréciation, reposant sur le postulat que la sexualité a une moindre importance pour les femmes que pour les hommes, a été jugée discriminatoire par la Cour européenne des droits de l'homme (CEDH, 25 juill. 2017, req. n°17484/15, *Carvalho Pinto de Sousa Morais c. Portugal*). Or, un décret du 27 mars 2020 (n°2020-356) a autorisé l'utilisation d'un algorithme, *Datajust*, pour les juges lors de la fixation des préjudices corporels. Celui-ci propose un référentiel basé sur l'analyse de décisions de justice rendues antérieurement. Il convient donc d'évaluer le risque de reproduction des stéréotypes de genre éventuellement présents dans certaines évaluations passées.

⁴ V. not. : A. Rouvroy et T. Berns, « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux* 2013, n°177, p. 163.

⁵ L'opération de *discriminer* consiste à « distinguer, établir une différence entre deux ou plusieurs personnes, deux ou plusieurs choses, d'après des caractères distinctifs » (« Discriminer », *Dictionnaire de l'Académie française*, 9^{ème} éd., 1992). Une telle mesure constitue une discrimination au sens de la loi lorsqu'elle repose sur un critère prohibé (V. not. : C. pén., art. 225-1).

⁶ V. not. : H. Stevens, « Mais où sont les informaticiennes ? », *Travail, genre et sociétés*, 2016/2, n°36, p. 167. Pour une étude de la composition sexuée de cette catégorie professionnelle : I. Collet, *L'informatique a-t-elle un sexe ? Hackers, mythes et réalités*, L'Harmattan, 2006.

des discriminations liées au genre peuvent se trouver systématisées, voire amplifiées par les algorithmes.

De manière générale le présent projet vise à étudier les biais générés induits par l'usage des algorithmes dans l'accès au(x) droit(s) ainsi qu'à saisir les risques de discrimination liée au sexe, à l'identité de genre et à l'orientation sexuelle susceptibles d'en résulter. En fonction des premiers constats obtenus, la question sera ensuite posée des modes de régulation offerts par le droit pour faire face à ces nouvelles formes de discrimination.

* *
*

Méthodologie de la recherche :

Dans le cadre de la présente étude, il apparaît indispensable de débiter par une clarification terminologique⁷. Le terme *algorithme* désigne les processus de traitement informatique automatisé de données à partir de modèles préétablis par le concepteur. L'intelligence artificielle (IA), également désignée sous l'expression *machine learning*, vise spécifiquement les algorithmes d'apprentissage. Le fonctionnement de ces systèmes algorithmiques consiste à reproduire les capacités de réflexion. L'IA vise ainsi la « *capacité des machines à prendre des décisions et à adopter des comportements attribués généralement à des humains ou à des animaux, mais avec un moindre niveau de performance* »⁸. Ces premiers aspects terminologiques et techniques seront ainsi particulièrement intéressants à mettre en perspective avec les neurosciences (grâce à l'expertise de Laura Pignatelli). En effet, les interfaces technologiques reposent sur des modèles automatisés d'apprentissage capables d'opérer des sélections et des prédictions à partir des informations fournies par l'utilisateur. L'IA va alors adapter les règles prédéfinies par le concepteur en fonction de ces données d'apprentissage.

A partir de ces définitions, le projet de recherche tendra à mettre en lumière la diversité des usages des algorithmes d'automatisation et d'apprentissage exploitant des données sensibles sur les individus. Cette mise en évidence est particulièrement complexe à effectuer car elle implique de disposer des compétences techniques pour comprendre le fonctionnement des algorithmes. A ce titre, la *retro-ingénierie* sera mobilisée. Il s'agit d'une méthode scientifique qui consiste à étudier un objet pour en déterminer les modalités de fonctionnement interne, sans avoir accès au code source. Une approche interdisciplinaire est indispensable pour mettre en exergue les éventuels biais algorithmiques dans chaque secteur et proposer des outils de régulation adaptés.

➤ **Première étape : Recherche visant à l'identification de biais générés dans l'usage des algorithmes en matière juridique**

Certains algorithmes fournissent une aide pour le traitement de données. D'autres interviennent directement dans la prise de décision, de manière plus ou moins autonome. À titre d'illustration, l'accès à nombre de droits sociaux est laissé au soin d'algorithmes, avec ou sans intervention humaine. A bien y réfléchir, l'être humain n'est plus seulement assisté mais

⁷ Il existe un flou sémantique qui nuit à l'appréhension de ce sujet d'étude. Pour une analyse approfondie des différentes définitions et des évolutions terminologiques, v. not. : A. Bensoussan et J. Bensoussan, *IA, Robots et Droit*, coll. Lexing – Technologies avancées et droit, Larcier, 2019, § 357.

⁸ Y. Lecun, « Apprendre aux machines à penser : un défi majeur pour l'économie ? Un progrès pour l'humanité ? », *Défis* 2017, n°8, p. 9.

bien remplacé par la machine. Les stéréotypes de genre sont alors susceptibles d'être systématisés par les algorithmes. Mais le risque existe également qu'ils soient amplifiés par l'IA qui ajuste son fonctionnement par corrélation et induction à partir de données biaisées. A titre d'illustration, quelques études ont déjà mis en exergue cet écueil dans le secteur des ressources humaines⁹, au stade du recrutement¹⁰, sous couvert d'une standardisation des profils recherchés. Or, les algorithmes se sont implantés sans que le risque de reproduction des stéréotypes de genre en résultant n'ait été anticipé. Les biais genrés ne sont pas annihilés. Pire, sous couvert de l'apparente objectivité de la machine, ils sont susceptibles d'être systématisés, voire amplifiés.

Dans ce premier temps, l'analyse des opérations algorithmiques et des résultats produits permettra de déterminer si les algorithmes neutralisent les biais de genre éventuellement présents dans les données exploitées ou, à l'inverse, s'ils en (re)produisent de manière directe ou indirecte.

Ce travail exploratoire sera réalisé en deux étapes :

1) *Identifier les traitements automatisés déployés dans l'emploi et la protection sociale (S. Sereno), la santé (G. Nicolas), la matière pénale (L. Pignatel), la matière civile concernant les droits des personnes et de la famille (E. Bonifay) ainsi que la matière administrative (C. Blanchon).*

2) *Le cas échéant, déterminer si l'exploitation de données sensibles - dont des éléments sexués - par les algorithmes a pour effet de neutraliser les biais de genre ou, à l'inverse, de les reproduire voire de les amplifier.*

➤ **Seconde étape : Le droit français dispose-t-il des outils pour lutter contre les éventuelles discriminations de genre liées à l'usage des algorithmes ? A défaut, quel(s) outil(s) de régulation proposer ?**

Les algorithmes n'évoluent pas dans une zone de non-droit. Tant les concepteurs que les utilisateurs des machines doivent respecter les normes en vigueur. Bien sûr, la législation actuelle offre des outils pour lutter contre les discriminations, qui peuvent être mobilisés à l'égard des algorithmes. Pour autant, cette lutte ne pourra être effective que si l'on parvient, après avoir relevé l'existence éventuelle de biais genrés, à comprendre comment ces derniers se sont intégrés dans les systèmes algorithmes. Aussi, cette réflexion doit nécessairement être transdisciplinaire afin d'aboutir à d'éventuelles propositions de régulation. À défaut, les normes proposées se réduiront à de simples incantations. C'est pourquoi le travail de recherche associera des personnalités scientifiques disposant d'une expertise dans d'autres branches disciplinaires (cf. la présentation de l'équipe de recherche).

⁹ C. Lévy, « L'IA dans les RH : une standardisation qui pose problème », in P. Adam, M. Le Friant et Y. Tarasewicz (dir.), *Intelligence artificielle, gestion algorithmique du personnel et droit du travail*, coll. Thèmes et commentaires, Dalloz, 2020, p. 16.

¹⁰ V. not. : J. Porta, « Algorithme et risques discriminatoires », in M. Mercat-Bruns (dir.), *Nouveaux modes de détection et de prévention de la discrimination et de l'accès au droit*, coll. Trans Europe Experts, Ed. Société de législation comparée, 2020, p. 61 ; J.-F. Amadiou et A. Roy, « Stéréotypes et discriminations dans le recrutement », *Hermès* 83 2019, p. 166.

Dans ce second temps, à partir de l'exploration des stéréotypes de genre intégrés ou reproduits par les systèmes algorithmiques, le projet de recherche tendra ensuite à s'interroger sur l'adéquation du droit positif pour saisir et sanctionner les discriminations liées au genre en résultant. Ces propositions pourraient être prolongées en menant une réflexion sur la possibilité d'inverser la tendance afin de recourir à l'IA comme un véritable outil de lutte contre les discriminations liées au genre.

Ce travail de réflexion est décomposé en deux étapes :

1) *Analyser l'adéquation des règles et outils juridiques mobilisables en matière de lutte contre les discriminations liées au genre pour saisir et sanctionner les discriminations générées par les biais algorithmiques de genre.*

2) *Déterminer les adaptations normatives éventuellement nécessaires pour lutter contre les discriminations liées au genre, tout particulièrement en présence de systèmes algorithmiques automatisés. Il s'agira d'identifier les modes de régulation juridiques et techniques adaptés pour faire face à ces nouvelles formes de discrimination.*

* *
*

COMPOSITION DE L'EQUIPE DE RECHERCHE (par ordre alphabétique) :

Jessica ATTALI-COLAS, Maître de conférences, Université Paul Valéry – Montpellier 3, Faculté des sciences sociales, des organisations et des institutions.

Thierry BENOIT, sociologue et secrétaire général de l'association La Boucle.

Rodolphe BIGOT, Maître de conférences, Le Mans Université

Emmanuelle BONIFAY, Maître de conférences, Aix-Marseille Université, Faculté de droit et de science politique.

Nicolas CATELAN, Maître de conférences, Aix-Marseille Université, Faculté de droit et de science politique.

Margot CHAMBON, Doctorante, Aix-Marseille Université, Faculté de droit et de science politique.

Arthur CHARPENTIER, Professeur, Université du Québec & Université de Rennes 1 - Faculté de sciences économiques.

Samir MERABET, Professeur, Université des Antilles.

Marie MERCAT-BRUNS, Maître de conférences HDR, Conservatoire National des Arts et Métiers.

Pierre MICHEL, Docteur qualifié aux fonctions de Maître de Conférences, Université de Montpellier, Faculté de droit et de science politique.

Guylène NICOLAS, Maître de conférences HDR, Aix-Marseille Université, Faculté de droit et de science politique.

Laura PIGNATEL, Maître de conférences, Aix-Marseille Université, Faculté de droit et de science politique.

Sophie SERENO, Maître de conférences, Aix-Marseille Université, Faculté de droit et de science politique.