



HAL
open science

L'usage des algorithmes et de l'IA dans le recrutement : une occasion de (ne) plus discriminer ?

Marie Peyronnet

► **To cite this version:**

Marie Peyronnet. L'usage des algorithmes et de l'IA dans le recrutement : une occasion de (ne) plus discriminer ?. Patrice Adam; Martine Le Friant; Yasmine Tarasewicz. Intelligence artificielle, gestion algorithmique du personnel et droit du travail, Dalloz, pp.139-162, 2020, 978-2-2471-9608-1. hal-03736566

HAL Id: hal-03736566

<https://hal.science/hal-03736566v1>

Submitted on 27 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.

L'usage des algorithmes et de l'IA ■ dans le recrutement : ■ une occasion de (ne) plus discriminer ?

MARIE PEYRONNET

Maîtresse de conférences à l'Université de Tours

Récemment, la société *Tinder* a été pointée du doigt après qu'une journaliste a révélé¹ que l'algorithme utilisé pour mettre en relation des personnes en quête d'un ou une partenaire (pour une durée déterminée ou indéterminée) était imprégné de stéréotypes sexistes, racistes et sociaux. Pour s'assurer que chacun et chacune trouve « chaussure à son pied », avait été mis en place dans l'algorithme un score de « désirabilité ». Alors que le désir est propre à la personne qui observe, la désirabilité est attachée à la personne qui fait l'objet de l'observation. Or, si les goûts d'une personne peuvent être relativement stables, la désirabilité semble difficile à mesurer tant elle dépend de l'œil qui la juge. Pour faire de la désirabilité un objet mesurable, il faudrait donc s'accorder sur ce qui est désirable à tout moment, dans n'importe quel contexte, qu'importe l'observateur. Ceux qui ont conçu l'algorithme permettant de définir ce critère de désirabilité (le *Elo Score*²) l'ont conçu sur la base d'une compétition entre tous les profils, chaque profil soumis à l'utilisateur étant en réalité mis en concurrence avec un autre profil pour en définitive obtenir un classement général des utilisateurs les plus appréciés. Ce classement est ensuite pondéré par l'algorithme au regard du profil de l'utilisateur, parfois pour simuler le hasard et faire alors croire à l'utilisateur qu'il était destiné à rencontrer quelqu'un avec la même date d'anniversaire, les mêmes hobbies ou ayant fréquenté la même école

1. J. Duportail, *L'amour sous algorithme*, Goutte d'or, 2019.

2. Du nom du classement mondial des joueurs d'échecs.

primaire alors qu'il s'agit en réalité de la principale raison pour laquelle l'algorithme a mis en contact les deux personnes. Parfois, l'algorithme pondère l'*Elo Score* au regard de considérations plus critiquables – globalement basées sur le principe « qui se ressemble, s'assemble³ » : les beaux avec les belles/beaux⁴, les juristes avec les juristes, chacun jouant dans sa *league*. Mais les concepteurs de l'algorithme admettent quelques entorses à cette logique implacable de déterminisme. Il arrive en effet que des hommes diplômés, fortunés et plus âgés puissent accéder aux profils des jeunes femmes belles et moins diplômées alors que dans le même temps une cloison numérique empêchera tout contact entre les jeunes hommes beaux et les femmes aux hauts revenus ou très diplômées⁵...

En quoi *Tinder* est-il un exemple pertinent en droit du travail? L'objectif de l'algorithme est de permettre la mise en relation de personnes ayant des intérêts/besoins complémentaires et qui semblent compatibles, comme pourrait le faire un algorithme conçu pour optimiser le recrutement. L'employeur a des besoins en termes de compétences et l'algorithme aura pour mission de trouver les candidats à l'emploi qui les possèdent et sélectionner celui ou celle qui « *match* » le mieux avec l'employeur.

Cependant, contrairement à la relation amoureuse (totalement discrétionnaire), l'employeur ne peut opérer une sélection qui serait contraire au droit applicable en matière de recrutement. S'il dispose bien évidemment d'une grande liberté dans le choix de ses collaborateurs⁶, le Préambule de la Constitution de 1946 reconnaît également aux individus le droit d'obtenir un emploi⁷. S'il n'est pas possible de reconnaître à tous le droit effectif d'obtenir un emploi, *a minima*, ils ne peuvent être discriminés dans l'accès à

3. V. l'exemple d'Harry et Sally utilisé dans l'un des brevets déposé par *Tinder* et portant sur le système de mise en relation : « Matching process system and method », US9733811B2 United States, consultable sur *Google Patent* : [<https://patents.google.com/patent/US9733811B2/en>].

4. *Tinder* tente en effet de mieux atteindre toutes les communautés en permettant de choisir entre 9 sexualités différentes : hétéro, gay, lesbienne, bisexuel.le, asexuel.le, demisexuel.le, pansexuel.le, queer et « en questionnement ».

5. Pour une étude complète, v. J. Gonzalez-Pidoux, *Tinder : la rencontre à portée de main. Une sociologie de l'expérience du dating sur smartphone*, mémoire de maîtrise en sciences sociales (sociologie de la communication et de la culture) sous la direction d'Olivier Voirol, soutenu à l'Université de Lausanne, août 2015, disponible en ligne sur *Academia* :

[https://www.academia.edu/21957126/Tinder_la_rencontre_à_portée_de_main._Une_sociologie_de_l_expe_rience_du_dating_sur_Smartphone?auto=download].

L'un des brevets de *Tinder* explicite ce biais de manière très claire : « Le serveur de correspondance peut être configuré de cette manière, car des données empiriques ont montré que ces différences démographiques n'ont pas d'effet équivalent sur les choix que font les hommes et les femmes en ce qui concerne leurs partenaires. »

6. Décision n° 88-244 DC du 20 juill. 1988, Loi portant amnistie, cons. 22, *JORF* 21 juill. 1988, p. 9448, *Recueil*, p. 119, *Dr. soc.* 1988, note X. Prétot.

7. Sans pour autant que cela ne fasse de ce droit une liberté fondamentale selon la Cour de cassation : Soc. 21 sept. 2017, n°s 16-20.270 et 16-20.277, *RDT* 2017. 717, obs. Galy, *RJS* 2017/12, n° 786, *JCP* 2017. 1090, obs. Dedessus-le-Moustier.

l'emploi⁸ et après cela dans la relation de travail. Le principe de non-discrimination⁹ constitue donc une limite à la liberté de choix de l'employeur. On pourrait penser que l'algorithme – dont le développement ou l'achat coûteux aura été financé par l'employeur – n'a pas pour finalité d'améliorer l'accès à l'emploi des candidats injustement exclus¹⁰, et viserait au contraire à améliorer l'efficacité de l'entreprise et réduire les coûts que représentent les agents humains du recrutement¹¹. Cependant, la rhétorique souvent employée pour faire accepter dans les consciences le recours à la froideur des algorithmes et de l'IA est bien souvent leur absence de préjugés et de biais susceptibles d'engendrer des discriminations. L'argument est évidemment très critiquable. Mais ces méthodes présentent des atouts indéniables dans l'objectivation¹² des décisions patronales et donc dans la lutte contre les discriminations. Étudier les algorithmes et l'IA au prisme de la discrimination permet d'appréhender, sur un même sujet, à la fois les intérêts et les risques de ces outils. En effet, leur conception oblige l'employeur à verbaliser ses besoins, sa façon de gérer son personnel et de le recruter¹³. Dès lors qu'il extériorise, pour la conception un « modèle » de décision ou de gestion¹⁴, ses motivations deviennent apparentes et peuvent être systématisées par l'outil informatique. Si d'aventure les motivations ainsi verbalisées se trouvaient être contraires au droit de la non-discrimination, soit le concepteur décide de ne pas les introduire à l'algorithme, soit de les intégrer de manière plus ou moins dissimulée. Qu'importe la décision prise, soit certains biais auront été écartés à la conception, soit ils auront été gravés dans le « marbre » et pourront être utilisés comme mode de preuve à condition de pouvoir obtenir l'algorithme qui a présidé à la décision.

Étudier les algorithmes de recrutement est un peu différent de l'étude d'une IA dédiée au recrutement. La création de l'IA repose sur un algorithme. Cependant, là où l'algorithme est figé dans sa configuration initiale

8. A. Jeammaud et M. Le Friant, « L'incertain droit à l'emploi », *Travail, genre et sociétés* 1999, 2, p. 29.

9. C. trav., art. L. 1132-1.

10. Ce serait pourtant là une finalité intéressante à poursuivre.

11. D. Rotman, « How Technology is destroying jobs », *MIT Technology Review*, 12 juin 2013.

12. V. Ilieva, *L'exigence d'objectivité en droit du travail*, thèse soutenue sous la direction d'Antoine Lyon-Caen à l'Université de Nanterre – Paris X, 2018. Français. ffNNT : 2018PA100120ff. fftel-02165945f.

13. Les algorithmes peuvent aussi bien viser le recrutement comme l'identification des travailleurs susceptibles de faire l'objet d'une mesure de licenciement ou d'une formation (soit du fait de leur manque de productivité, soit de l'apparition d'une inadéquation des compétences avec l'évolution des postes de travail dans l'entreprise).

14. C. O'Neil, *Algorithmes. La bombe à retardement*, Les arènes, 2018, p. 36. L'algorithme est comparé par l'auteure à la préparation d'un repas. Elle explique ce qu'est un modèle au regard de son expérience de mère de trois enfants. Si elle n'était pas là pour préparer le repas et devait expliquer à une tierce personne comment elle s'y prend, elle devrait, pour garantir un niveau de satisfaction équivalent, donner à cette personne tout un tas d'informations (les préférences de chacun de ses enfants, le jour et les activités de chacun qui ont une incidence sur leur appétit, etc.). À cet objectif premier peut-être adjoint d'autres objectifs (qualité nutritive, diversification des repas, etc.).

et ne fait qu'être appliqué rendant ainsi son contrôle plus facile, l'IA quant à elle évolue, va finir par déterminer ses propres critères de sélection en fonction des données dont elle dispose et des liens de corrélation statistique qu'elle constate. Ce sont ses variations permanentes qui peuvent empêcher d'obtenir des informations passées sur les raisons qui ont présidé à un recrutement.

Cette étude se fera donc au prisme particulier du respect du principe de non-discrimination dans le recrutement (incluant les promotions y compris interne)¹⁵, ce qui permet à la fois de décrire comment le droit peut venir conditionner la conception des algorithmes ou d'une IA (I) et comment de tels outils nécessitent à leur tour une importante adaptation du droit (II).

I. LA NÉCESSAIRE ADAPTATION DE L'ALGORITHME ET DE L'IA AU DROIT

Le potentiel des algorithmes et de l'IA n'a pour limites que celles de l'imagination et du droit. En matière de recrutement, les usages existants (A) demeurent limités et dépassent rarement le stade expérimental¹⁶. Si ces tests sont peu concluants, c'est bien souvent en raison des effets discriminatoires constatés et résultant de biais dans la conception et dans l'utilisation de ces outils (B). La phase de conception semble donc ne pas prendre en considération certaines variables pourtant indispensables pour rendre « juridiquement viables » de tels outils.

A. LES USAGES ACTUELS D'ALGORITHMES ET D'IA DANS LE RECRUTEMENT

Les algorithmes ou IA peuvent être utilisées de différentes manières dans le recrutement. Elles peuvent en effet être employées à des fins d'imitation du pouvoir patronal¹⁷ (1) ou, au contraire, à des fins de contrôle de ce dernier (2).

1. L'utilisation des algorithmes à des fins d'imitation du pouvoir patronal

L'immixtion de l'algorithme ou de l'IA dans la décision patronale peut être plus ou moins grande : une simple mise en relation, une aide à la décision, une prise de décision automatisée.

15. V. pour une étude sur la gestion du personnel l'article de L. Malfettes, « Gestion du personnel par algorithmes et droits du salarié », *Dr. soc.* 2019. 591.

16. V. l'autolicencement en vidéo de l'IA de L'Oréal : [<https://www.exclusiverh.com/articles/test-recrutement/l-oreal-limoge-son-robot-drh.htm>].

17. V. sur ce point dans cet ouvrage la contribution de Sébastien Fischman et Barbara Gomez.

C'est assurément en matière de recrutement que la comparaison avec les applications de rencontre est la plus pertinente. Un exemple, en 2007, une société française a créé *Meteojob*, un procédé de recrutement directement basé sur les logiques de « *matching* » des algorithmes de rencontre. L'algorithme compare le profil recherché par l'employeur avec celui du candidat à l'emploi, lorsqu'un « *match* » apparaît, la mise en relation est immédiate. Il ne s'agit ici que d'un simple algorithme de mise en relation, mais cela pose d'emblée des questions sur la rationalité que représente un « *match* » ponctuel par rapport à une véritable procédure de recrutement dans laquelle un panel de candidats aura été sélectionné. En effet, dans cette dynamique de mise en relation, l'employeur n'a techniquement pas émis d'offre d'emploi et la personne approchée n'a pas fait acte de candidature à cet emploi. L'une et l'autre des parties peuvent donc être déjà engagées et n'être qu'« ouvertes » aux propositions. Ce type d'application de rencontre entre professionnels peut donc permettre à un salarié qui souhaiterait quitter son emploi d'être approché par des recruteurs et à un employeur d'être à l'affût des « talents » se trouvant sur le « marché » de l'emploi tout en échappant aux contraintes légales que représente la constitution d'un vivier de candidatures¹⁸. On se trouve finalement ici dans une forme de cooptation digitale. L'employeur reçoit un profil, libre à lui de « *swiper* » vers le suivant ou de faire une proposition d'embauche ou de lancer une procédure de recrutement avec plusieurs profils qu'il juge intéressants et qui auraient accepté d'être présélectionnés de cette manière pour l'obtention d'un emploi.

L'algorithme peut s'immiscer encore un peu plus dans le recrutement en prenant la forme d'une véritable aide à la décision. L'exemple sur lequel on dispose du plus d'informations est celui du « recrutement » des futurs étudiants du supérieur via la plateforme *Parcoursup*. Les étudiants entrent dans la plateforme générale leurs données personnelles, note du bac, vœux, expériences, comme le ferait un candidat à l'emploi sur la plateforme de *Pôle emploi*. Les établissements (recruteurs donc), définissent chacun leurs critères pour classer les candidats, ce sont les fameux « Attendus ». Ces critères de sélection sont pondérés en fonction du degré d'importance que chaque filière leur attribue (maîtrise du français – coef. 5, des mathématiques – coef. 4, etc.)¹⁹. À l'issue du processus, les algorithmes locaux fourniront une liste ordonnée des candidatures, de la meilleure à celle qui sera jugée la moins bonne. Cette liste vient aider les agents humains à prendre leurs

18. Comme pour toute base de données, cela suppose une déclaration à la CNIL, une autorisation de la part des candidats et le respect des délais de conservations des données. Sans oublier l'interdiction de collecter certaines données sensibles.

19. C'est sur le choix de ces critères et leur pondération qu'il existe un problème de transparence. Les bacheliers n'étant pas avertis du contenu de ces algorithmes « locaux ». V. la décision 2019-021 du 18 janv. 2019 du Défenseur des droits relative au fonctionnement de la plateforme nationale de préinscription en première année de l'enseignement supérieur (Parcoursup).

décisions. Celles-ci peuvent être guidées par différentes valeurs : le mérite (les meilleurs sont prioritaires), l'équité (alors on introduira dans les paramètres un quota de boursiers par exemple), la diversité (notamment des parcours, en étudiant de manière spécifique le cas des candidats en reprise d'étude qui ne peuvent pas donner à l'algorithme les données dont il a besoin pour les classer), etc. L'algorithme pourrait parfaitement prendre en considération tous ces éléments de manière autonome et aboutir à une décision dès lors qu'on lui a appris les valeurs à prendre en considération et qu'on lui a fourni les données lui permettant de les satisfaire (ce qui sera délicat lorsqu'il s'agit d'apprécier une situation qu'aucune donnée ne vient matérialiser).

Enfin, le travail de recrutement peut être intégralement confié à une IA. C'est le choix qui a été fait par certaines GAFAM. Elles peuvent fonctionner sur des modèles différents. Le *deep learning* ou « apprentissage profond²⁰ » (utilisé dans la gestion du personnel par Google²¹) nécessite que l'IA soit continuellement approvisionnée en données pour lui permettre de modifier ses modèles (ex. : *tels salariés recrutés ont fait de bonnes performances sur tel emploi. L'IA veut reproduire ces bons recrutements. Quels sont les paramètres qu'ils ont en commun? La nature de leur diplôme. L'IA va donc chercher les personnes ayant les mêmes formations. Ça ne fonctionne pas. Elle constate un autre point commun [mais statistiquement moins courant que la nature du diplôme], beaucoup habitent près de leur lieu de travail. L'IA va donc privilégier les candidatures de ceux qui sont proches et qui ont suivi telles études et ainsi de suite*). L'employeur peut analyser les choix de l'IA, pour repérer des facteurs de productivité qu'il peut tenter d'améliorer en prenant des mesures spécifiques. À titre d'illustration le temps de trajet pour rejoindre l'entreprise est de plus de 2 heures pour la plupart des salariés, ce qui nuit à leur performance en comparaison de ceux qui vivent à proximité. L'employeur va donc instaurer un bus d'entreprise et un véritable « ramassage d'employés », équipé du wifi et de prises électriques pour que les salariés puissent commencer leur journée de travail dès leur montée dans le bus²² ou il peut aussi choisir de sacrifier « un peu d'efficacité au profit de l'équité » en neutralisant la donnée du lieu d'habitation²³. Ainsi, un algorithme optimisé pour le recrutement ne pourra

20. V. pour le mode de fonctionnement de ce type d'IA : O. Grisel (INRIA), « Peut-on faire confiance aux algorithmes d'apprentissage profond? », colloque amphithéâtre Marguerite de Navarre – Marcelin Berthelot, 23 avr. 2019, disponible sur le site du Collège de France, dans l'onglet « colloque » de la chaire « Algorithmes, machines et langages » (2007-2019) de Gérard Berry.

21. Surtout connu pour les progrès spectaculaires de ses IA en matière de jeu : en 2016 AlphaGo a réussi à vaincre les meilleurs joueurs professionnels de go (on ne peut qu'encourager à visualiser les parties contre Lee Sedol qui apprennent beaucoup sur la manière de fonctionner de cette IA) et en 2019 AlphaStar a fait de même avec les joueurs professionnels de Starcraft II.

22. J. Marin, « Les "Google Bus" attaqués en justice », *Le Monde* 5 mai 2014.

23. C. O'Neil, *Algorithmes*, op. cit., p. 182. V. l'exemple de l'entreprise Xerox qui, ayant remarqué que les salariés les plus productifs étaient ceux qui habitaient le plus proche

L'usage des algorithmes et de l'IA dans le recrutement : une occasion de (ne) plus discriminer?

être considéré comme performant que s'il l'est sur le plan juridique et ne conduit pas l'employeur à être poursuivi pour non-respect de la législation en vigueur.

Une autre méthode consiste à entraîner une IA afin qu'elle reproduise un schéma donné. C'est le modèle sur lequel a fonctionné pendant un temps l'IA d'Amazon avant que l'expérience ne soit abandonnée²⁴. Ses concepteurs lui avaient fourni les données RH de l'entreprise des années précédentes avec pour objectif de dégager un modèle dans le recrutement et de le systématiser. Ce qui a parfaitement fonctionné. Cependant, les CV pris en exemple étaient principalement constitués de CV masculins. L'IA en a donc déduit que les hommes sont à privilégier et une fois ce modèle discriminatoire repéré, loin de le combattre, l'IA l'a logiquement systématisé et amplifié. Elle s'est donc mise à éliminer les femmes avec une redoutable efficacité, jusqu'à ce que la société mette un terme à cette expérimentation qui venait de lui échapper. La technique de l'apprentissage sur une base de données dédiée n'est pas mauvaise, à condition que l'échantillon d'entraînement soit exempt de tout biais et fasse l'objet d'un contrôle préalable très strict²⁵.

Les problématiques liées à la conception d'algorithmes et d'IA semblent infinies et conditionnées à la qualité du travail des êtres humains qui conçoivent ces outils. C'est pourquoi, avant de penser à remplacer le pouvoir patronal par une IA qui sera défaillante, à l'image de son ou ses auteurs, peut-être faudrait-il la perfectionner dans une autre activité dans laquelle l'identification de ces problématiques sera au cœur du procédé²⁶ : le contrôle du pouvoir patronal.

2. L'utilisation des algorithmes à des fins de contrôle du pouvoir patronal

D'une certaine manière, le contrôle du pouvoir patronal semble beaucoup plus facile à coder que ne l'est la volonté ou l'imitation de l'employeur. Le droit, qui est la norme de ce contrôle, semble fait pour être traduit en code

de l'entreprise, a choisi de modifier son algorithme pour qu'il ne tienne pas compte de cette donnée afin de ne pas pénaliser les candidats habitant les quartiers pauvres qui se trouvaient être plus éloignés.

24. J. Dastin, « Amazon scraps secret AI recruiting tool that showed bias against women », *Reuters* 10 oct. 2018. Cette expérimentation, de par l'ampleur de son échec, fait figure de cas d'école.

25. V. *infra*.

26. À noter que la nouvelle équipe chargée de concevoir l'IA d'Amazon a fait de la diversité un élément central du processus de conception : J. Dastin, « Amazon scraps secret AI recruiting tool that showed bias against women », préc. On ne doute pas que l'entreprise parvienne à résoudre ses difficultés de recrutement étant donné que ses équipes excellent sur la prédiction des besoins des consommateurs de la plateforme, au point que la manière dont ces algorithmes et leurs effets sont en permanence étudiés et révisés sont cités en exemple par Cathy O'Neil comme étant de « bons » algorithmes. Pour l'auteur, si les algorithmes de la justice prédictive faisaient l'objet d'autant d'attentions cela profiterait non seulement aux prisonniers, mais à toute la société. V. Cathy O'Neil, *op. cit.*, p. 153-155.

informatique²⁷. Le contrôle du pouvoir patronal par le biais des algorithmes ou de l'IA peut être à la fois opéré de manière interne à l'entreprise (a), par exemple dans le cadre d'une démarche d'auto-évaluation ou de contrôle de ceux à qui l'employeur délègue son pouvoir (les services de ressources humaines), et de manière externe à l'entreprise (b) par l'État ou les organismes sociaux.

a. Le contrôle interne de l'entreprise

Que ce soit dans une démarche vertueuse ou dans l'optique de réduire les risques de contentieux, l'employeur peut avoir à cœur de vérifier que la gestion des ressources humaines, généralement réalisée par des services dédiés, est respectueuse du droit et notamment du principe de non-discrimination. À cette fin, l'algorithme ou l'IA peut être employée pour dresser un bilan des politiques de recrutement, de promotion ou encore de départs de l'entreprise.

C'est le cas de certaines entreprises où il existe des processus de contrôle de la gestion des équipes. Si l'on constate par le biais d'un contrôle quantitatif (statistique) que certaines catégories de personnes, les femmes²⁸, ont un turn-over plus important que les autres catégories de salariés, qu'elles restent moins longtemps en moyenne dans l'entreprise, ou qu'elles sont moins souvent recrutées, alors un contrôle qualitatif est lancé. Ce dernier pourra révéler qu'un recruteur est discriminant, qu'un chef d'équipe harcèle certains profils, etc. Le contrôle qualitatif, mené sur la base des statistiques les plus troublantes et à l'aide d'entretiens, permet d'identifier l'origine du problème et de le corriger. Ce type de contrôle est utile car « l'algorithme est meilleur en termes de prédiction que de causalité », il est plus facile pour lui de prédire la productivité des individus (comment les femmes se comportent-elles sur tel poste dans telle entreprise?) que d'établir des liens de cause à effet (comment les femmes se comporteraient-elles si l'on modifiait l'environnement de travail?)²⁹.

Ce type de contrôles « préventif » ne peut, *a priori*, être mené qu'au sein de l'entreprise en raison de la confidentialité des données RH qu'il requiert. Pour autant, cela n'empêche pas l'État et les organismes de sécurité sociale de tenter de contrôler la gestion de son personnel par l'employeur.

b. Le contrôle par l'État ou les organismes de sécurité sociale

Deux exemples récents de contrôle³⁰ viennent à l'esprit. Tout d'abord une expérimentation prometteuse de l'assurance maladie et ensuite le nouveau,

27. Isaac Asimov en a d'ailleurs fait la base de la robotique dans le *Cycle des robots* (dont les différents tomes sont parus entre 1950 et 1985).

28. L'information peut être parfois encore plus précise avec une approche intersectionnelle : les femmes noires, les hommes latinos, etc. V. le rapport Google Diversity 2019 et les données concernant l'*attrition index*, p. 10.

29. J. Kleinberg, J. Ludwig, S. Mullainathan *et al.*, « Discrimination in the Age of Algorithms », *Journal of Legal Analysis* 2019, 10.

30. On fera immédiatement remarquer que ces contrôles ne relèvent pas à proprement parlé d'algorithmes ou d'IA, mais en revanche le traitement des données effectuées pourrait

et déjà célèbre, index sur l'égalité professionnelle entre femmes et hommes³¹. Alors que la première méthode consiste en une étude comparative de données statistiques pour détecter des comportements indésirables au sein des entreprises, la seconde prend la forme d'un indice, c'est-à-dire de l'agrégation de plusieurs indicateurs, pondérés les uns par rapport aux autres pour attribuer une notation, en général sur une échelle de 1 à 100 (ou de 0 à 1).

Dans le premier cas³², l'expérimentation est menée par l'Assurance maladie. Elle concerne les entreprises d'au moins 200 salariés³³ et vise à communiquer aux employeurs de ces dernières les taux d'arrêts maladie anormalement élevés portant sur certaines pathologies (trouble musculo-squelettique, lombalgies, risques psychosociaux). L'objectif est préventif et vise à aider l'employeur à réduire l'absentéisme en l'avertissant sur les problématiques spécifiques à son établissement afin qu'il adopte des mesures de prévention ou des correctifs améliorant la santé et sécurité des travailleurs. Cette logique serait en tout point transposable aux problématiques de recrutement. Les organismes collecteurs de la sécurité sociale reçoivent, par le biais de la DSN, des informations sur la composition des effectifs des entreprises et sur les rémunérations distribuées. Ils disposent donc matériellement des données pour signaler aux employeurs que leurs recrutements sont majoritairement masculins, notamment sur les postes les mieux rémunérés, que les plus âgés sont rarement embauchés et plus souvent licenciés, etc. Pourtant, ce contrôle n'est pas effectué par les organismes publics. Le législateur a privilégié la mise en place d'un index calculé par l'employeur lui-même et accompagné d'une sanction financière.

L'index portant sur l'égalité professionnelle entre les femmes et les hommes³⁴ use d'objectifs chiffrés (nombre de femmes dans les plus hautes rémunérations, nombre d'augmentation pour les femmes en retour de congé

à l'avenir faire l'objet d'une automatisation par le biais d'un algorithme (l'index sur l'égalité professionnelle fait d'ailleurs déjà l'objet d'une feuille de calcul mise à disposition par le ministère). Or, les problèmes que soulèvent les algorithmes sont nombreux à provenir en réalité des données chiffrées qu'ils mettent en jeu. V. sur le glissement progressif de la statistique vers les indicateurs : A. Desrosières et E. Didier, *Prouver et gouverner*, La Découverte, 2014.

31. C. trav., art. L. 1142-7 s.

32. Plus amplement décrit dans un précédent article : M. Peyronnet, « Harcèlement au travail et diversité », actes de l'intervention lors de la table ronde « Réseaux sociaux et harcèlement discriminatoire (sexuel et raciste) », Journée d'étude sur les Nouveaux modes de détection et de prévention de la discrimination et accès au droit, organisée par Trans Europe Expert et le CNAM, CNAM, Paris, le 10 déc. 2018, à paraître.

33. Seuil permettant, selon l'assurance maladie, de préserver l'anonymat des salariés et donc le secret médical, mais jugé trop bas par une sénatrice lors d'une session de question au gouvernement : « Expérimentation de la CPAM sur les arrêts maladie », 15^e législature, Question d'actualité au gouvernement n° 0193G de M^{me} Sonia de la Provôté (Calvados – UC) et la Réponse du ministère des solidarités et de la santé, publiées dans le *JO Sénat* du 2 févr. 2018, p. 844.

34. Institué par la loi n° 2018-771 du 5 sept. 2018 pour la liberté de choisir son avenir professionnel et dont le décret n° 2019-15 du 8 janv. 2019 a fixé la méthodologie de calcul

maternité, etc.) pour définir le contenu de l'obligation de l'employeur³⁵. Le choix des indicateurs correspond à des choix politiques. En l'occurrence en limitant le nombre d'indicateurs à 4 ou 5 selon la taille de l'entreprise, portant principalement sur les rémunérations pour un travail égal ou de valeur égale et surtout à durée de travail équivalente, le législateur a choisi de concentrer l'index sur l'écart « inexpliqué » de 10,5 points³⁶ existant entre les rémunérations des femmes et des hommes dont les conditions de travail sont en tout point similaires par ailleurs. En limitant le champ de la comparaison, en passant sous silence la part « expliquée » des inégalités, l'index vient « légitimer » cette part en permettant aux entreprises de valider l'index avec un score avoisinant les 100/100 malgré la persistance d'un grand nombre d'inégalités notamment liées à la non-mixité des métiers et à la disproportion de femmes occupant un emploi à temps partiel. La partie des inégalités qui n'était alors qu'expliquée est devenue, sous l'autorité de l'argument mathématique, « justifiée ».

B. LES BIAIS

Qu'il s'agisse des algorithmes mis en œuvre pour imiter le pouvoir patronal ou pour le contrôler, il existe des biais à éviter. Les biais les plus récurrents concernent d'abord la conception des algorithmes (1) et se repèrent également au stade de leur usage (2).

1. Les biais de conception

Les biais, en particulier discriminatoires, peuvent apparaître à différents stades de la conception de l'algorithme ou de l'IA³⁷. Ils peuvent résulter d'une définition incomplète des objectifs (a), de la détermination des critères de prédiction (b) ou encore des données qui alimentent l'algorithme ou l'IA (c).

a. La définition des objectifs

La définition de l'objectif est évidemment déterminante dans la façon dont va être conçu l'algorithme. Si l'on reprend l'exemple de Parcoursup, l'objectif affiché par le ministère de l'enseignement supérieur et de la recherche est

des indicateurs à utiliser et les conditions de fixation de la pénalité financière encourue en cas de score inférieur à 75/100 points pendant trois années consécutives.

35. V. la controverse consacrée à cet outil : C. Coron, E. Boussard-Verrecchia, K. Berthou et D. Meurs, « Quelle efficacité attendre des indicateurs pour mesurer les écarts de rémunération entre les femmes et les hommes? », *RDT* 2019. 147.

36. Pour rappel, les rémunérations des hommes sont en moyenne 25,7 % supérieures à celles des femmes. Sur ces 25,7 %, environ 15,2 points sont expliqués par la nature différente des métiers exercés par les femmes et les hommes, par le travail à temps partiel très majoritairement féminin, etc. et 10,5 points restent « inexpliqués » et seraient de la discrimination directe en raison du sexe. V. les chiffres de l'Insee de 2015.

37. V. pour une présentation générale sur les origines des biais l'intervention d'Olivier Grisel (INRIA), « Peut-on faire confiance aux algorithmes d'apprentissage profond? », *op. cit.*

celui de garantir une « meilleure » orientation pour les lycéens. Pour que l'orientation soit la « meilleure » possible, « on peut écrire un algorithme qui privilégie le premier vœu de l'élève, ou s'assurer que tous les élèves aient au moins leur 3^e choix, ou encore que les notes jouent un rôle prépondérant dans la décision³⁸ ». Le choix qui sera opéré entre ces trois formules est politique et l'algorithme sera le reflet de l'intention de son créateur.

Dans le monde du travail, il est également possible d'opérer ce type de choix. Après le fiasco d'Amazon, l'équipe devant travailler sur une nouvelle IA a mis l'accent sur la recherche de diversité dans les recrutements et plus seulement sur la reproduction de l'existant³⁹. Une telle recherche ne manquera pas cependant de poser quelques difficultés. L'employeur désireux de promouvoir un profil particulier au sein de son entreprise, un profil généralement discriminé⁴⁰, se trouvera la plupart du temps en infraction au regard de l'interdiction d'opérer toutes formes de discrimination à l'embauche. On peut se demander s'il est en revanche plus acceptable, d'un point de vue légal, que l'employeur se contente de chercher à diversifier son recrutement en tenant compte des profils des personnes qu'il a déjà recrutées par le passé, la présence au sein de l'entreprise et dans les recrutements passés pouvant peut-être justifier le choix volontaire d'un autre profil⁴¹. Un monitoring de la composition actuelle et passée de l'entreprise pourrait ainsi permettre à l'employeur d'avoir conscience de ses propres biais lors des recrutements passés et de tenter de les supprimer, voire de les compenser dans ses recrutements futurs.

b. La détermination des critères de prédiction

Si l'on part du principe que l'on veut un algorithme de recrutement capable d'identifier les salariés qui seront les plus « productifs » au sein d'un vivier de candidatures, plusieurs jeux de données seront utilisés pour prédire quelle sera la productivité de tel ou tel candidat s'il était recruté.

Tout d'abord, celles permettant de définir ce qu'est un salarié « productif⁴² ». Par hypothèse, la définition du salarié « productif » se fait sur la base des données générées par les salariés actuellement dans l'entreprise (ex. : taux de présence, rendement, ventes effectuées, etc.) et non celles du vivier de

38. A. Bernheim et F. Vincent, *L'intelligence artificielle, pas sans elles!*, Belin, 2019, p. 41.

39. J. Dastin, « Amazon scraps secret AI recruiting tool that showed bias against women », *Reuters* 10 oct. 2018.

40. C'est ainsi que la recherche d'une personne appartenant à « une famille issue de l'immigration » dans le cadre d'une mesure se voulant positive a été jugée discriminatoire : Nîmes, 22 nov. 2002, *D.* 2003. Somm. 2920. Cité par M. Miné, *Droit des discriminations dans l'emploi et le travail*, Bruxelles, Larcier, 2016, p. 272.

41. C'est là une justification admise par la Cour suprême américaine dans une décision *Regents of the University of California v. Bakke*, 438 U.S. 265 (1978) et plus récemment *Fisher v. University of Texas at Austin (Fisher II)*, 579 U.S. (2016).

42. V. sur ce point les méthodes d'évaluation des salariés : M. Roussel, *L'évaluation professionnelle des salariés*, LGDJ, 2019.

candidats. Au même titre que la définition des « objectifs » de l'algorithme, on va avoir ici des choix politiques et subjectifs à faire. La définition de ce qu'est un bon employé doit être minutieusement étudiée, car si elle est trop restreinte (focalisation par exemple sur la rapidité), la recherche systématique et automatisée d'une compétence déterminée chez les prochains travailleurs sera excessivement uniforme et contre-productive (la qualité sera sacrifiée sur l'autel de la rapidité)⁴³.

À ces données d'évaluation s'ajoute un autre jeu de données servant quant à lui à effectuer la prédiction et qui portera sur des indicateurs propres à chaque salarié (sexe, âge, diplômes, expérience, etc.). L'algorithme ou l'IA va établir des liens de corrélation statistiques entre ce type de critère personnel et la productivité. Ainsi, si la définition de ce qu'est un « bon » employé est basique, le recrutement va être « formaté » et favorisera un certain type de profil ce qui conduira potentiellement à automatiser une discrimination indirecte.

Les biais peuvent donc apparaître à deux moments : lors de la définition des critères d'évaluation de la productivité et lorsque la machine va lier productivité et caractéristiques individuelles. L'établissement d'un « indice de productivité » doit donc être défini dans une optique à la fois d'efficacité : une fois au sein de l'algorithme le résultat obtenu est-il satisfaisant en termes de productivité ? Mais il faudra également veiller à le « contenir » pour qu'il n'y ait pas de recrutement discriminant : ce qui exclut *a priori* l'usage de caractéristiques prohibées (sexe, origine...) comme prédicteurs. Ces caractéristiques, si elles doivent être employées dans le cadre d'algorithmes, ne pourraient l'être qu'au sein d'algorithmes de contrôle (étatique)⁴⁴ du respect du principe d'égalité et de non-discrimination et éventuellement au sein d'algorithmes de correction œuvrant à la réduction de discriminations indirectes ou systémiques.

Un moyen de lutter contre les effets discriminatoires de l'algorithme et de l'IA serait donc d'implémenter le droit au sein même du code informatique. Il devrait être possible d'intégrer au sein d'un algorithme des règles particulières indiquant que dans la recherche du candidat le plus performant, l'usage d'un critère de prédiction (neutre en apparence) ne peut pas avoir pour effet d'éliminer de manière disproportionnée⁴⁵ les femmes, les jeunes, etc. Ou,

43. V. la définition de ce qu'est un bon officier de police par J. Kleinberg, J. Ludwig, S. Mullainathan *et al.*, « Discrimination in the Age of Algorithms », préc. La productivité d'un policier ne peut se résumer au nombre d'interpellations effectuées. Il faut prendre en compte également les liens qu'il a avec la communauté, le respect des procédures, la prévention, etc.

44. *A priori* seule la statistique publique serait à même d'effectuer ce type de mesure en toute légalité. Et encore, l'étude devrait user de données de remplacement telles que le lieu de naissance, la nationalité des parents, pour établir des données statistiques sur l'« origine » des individus...

45. Ce qui supposerait que le concepteur de l'algorithme détermine un « seuil » numérique à la disproportion.

si l'usage de ce critère est légitime, que la sélection intègre d'autres critères qui cette fois seront plus favorables à la catégorie qui a été éliminée de manière disproportionnée. En d'autres termes, il est possible d'apprendre à l'algorithme à ne pas discriminer. Cela n'est cependant possible que si le concepteur de l'algorithme a connaissance des biais que peuvent comporter certains critères « neutres » et proposera d'en contourner les biais en choisissant d'autres critères ou en les diversifiant ou en fixant des contraintes quant à l'automatisme de leur utilisation.

c. Les données utilisées

La qualité des données utilisées dans le cadre d'un algorithme ou d'une IA est d'une importance capitale.

Tout d'abord, si l'on utilise des évaluations de salariés pour déterminer ce qu'est un salarié productif, il faudra impérativement veiller à ce que les agents humains ayant opéré cette évaluation n'aient pas été discriminants dans leur manière de mener les évaluations. S'ils ont des biais, s'ils sont ouvertement sexistes ou racistes, lorsque les mauvaises évaluations des femmes, des personnes racisées, des jeunes, des vieux, seront intégrées à l'algorithme celui-ci les traitera comme n'importe quelle donnée objective⁴⁶. Donc si les évaluations des femmes sont mauvaises parce que le manager est sexiste, l'algorithme ne retiendra comme information que le fait que les femmes sont moins productives et doivent donc être écartées à l'avenir du recrutement.

Ensuite, Cathy O'Neil, mathématicienne et autrice d'un ouvrage décryptant les défauts de conception de nombreux algorithmes, alerte sur le risque que comporte l'utilisation de « données de rechange », aussi appelées données de « remplacement » ou « *proxy data*⁴⁷ ». Une donnée de rechange consiste, face à l'impossibilité d'obtenir « la » donnée utile pour le fonctionnement de l'algorithme, à lui substituer une autre donnée qui lui soit proche, indirecte ou supplétive. Prenons un exemple récent apparu dans le cadre de la gestion des urgences hospitalières⁴⁸. Un algorithme devant prendre en considération les antécédents des patients pour déterminer un ordre de priorité calculait sa prédiction, non pas sur la nature et la gravité des antécédents de santé des patients, mais au regard du coût de leurs précédentes hospitalisations pour l'assurance santé qui a fourni les données. Or, les personnes pauvres sont moins souvent titulaires d'un contrat d'assurance santé et ont donc des dépenses de santé moindre parce qu'elles se soignent moins et non parce qu'elles seraient moins malades. Ce type de difficultés aurait été évité en France du fait de l'existence de la Sécurité sociale et plus particulièrement

46. Qu'il pourra écartier si elles conduisent à éliminer certains profils de manière disproportionnée à condition de lui avoir appris à le faire...

47. C. O'Neil, *Algorithmes*, op. cit., p. 35.

48. Z. Obermeyer, B. Powers, C. Vogeli et S. Mullainathan, « Dissecting racial bias in an algorithm used to manage the health of populations », *Science*, 25 oct. 2019, vol. 366, p. 447-453, DOI : 10.1126/science.aax2342.

de la PUMA⁴⁹ qui permet d'avoir une base de données de santé unique au monde⁵⁰ et un accès, pour les patients qui l'autorisent, à l'ensemble du dossier médical, rendant ainsi inutile de passer par un intermédiaire (assurance privée) dont le filtre d'analyse (sa propre base de données) dégrade la qualité de la donnée (information sur le coût des soins et non sur l'état de santé). Compte tenu de la sensibilité de l'information, il est probable que l'assureur privé ne pouvait légalement pas communiquer librement les données médicales du patient. Le besoin de données fiables plaide bien souvent pour que la collecte et le traitement des données soit effectué de préférence par des organismes étatiques offrant *a priori* de meilleures garanties qu'un opérateur privé. C'est d'ailleurs la conclusion à laquelle arrive Dominique Meurs sur la question des données de l'égalité entre femmes et hommes⁵¹.

L'IA de recrutement d'Amazon évoquée précédemment a été abandonnée précisément en raison du jeu de données d'entraînement défectueux. Entraînée pour déterminer quels sont les travailleurs les plus productifs, notamment au regard de leur taux de présence à leur poste, elle a établi un lien de corrélation entre le sexe et le taux d'absentéisme, les femmes étant statistiquement plus susceptibles d'être absentes du fait de charges familiales ou d'une grossesse. Seulement deux jeux de données alimentaient son apprentissage : les personnes embauchées par les recruteurs (plus d'hommes que de femmes) et les données liées à l'absentéisme. L'IA en a déduit, comme les êtres humains, que les hommes étant moins absents, ils sont donc plus efficaces et les femmes doivent donc être écartées lors du recrutement. Personne ne lui avait enseigné qu'il ne fallait pas discriminer.

2. Les biais d'utilisation

Dès lors que l'on automatise un processus de recrutement ou de contrôle, on expose le système, s'il n'est pas régulièrement modifié, à des stratégies qui seront élaborées par les candidats pour les contourner. C'est ce que les ingénieurs désignent comme des boucles de « rétroaction ». La personne évaluée modifie son comportement pour réussir l'évaluation. C'est un comportement repéré depuis des centaines d'années en matière de concours et

49. Protection universelle maladie.

50. Interview de Delphine Jaafar, « La France dispose d'une base de données de santé unique en Europe », Actualité pro, *Gaz Pal.* 31 juill. 2019.

51. « À partir du moment où l'État impose une obligation assortie d'une sanction, il me semble que c'est à la statistique publique de calculer les inégalités de rémunération, en étant bien sûr transparente sur les modes de calcul. C'est ce que nous prônions dans une tribune du *Monde* du 25 avr. 2018, signée par vingt économistes spécialistes du domaine. Les entreprises sont déjà tenues de transmettre à l'État des données sur les salaires et les heures des salariés (DSN). Ces bases de données pourraient très bien servir à calculer des indicateurs d'inégalités par les services de la statistique publique qui transmettraient ensuite à chacune son score », in C. Coron, E. Boussard-Verrecchia, K. Berthou et D. Meurs, « Quelle efficacité attendre des indicateurs pour mesurer les écarts de rémunération entre les femmes et les hommes? », préc., p. 147.

d'examens⁵², l'algorithme de recrutement n'y fera pas exception. Dès lors que les candidats, en premier lieu ceux qui sont les mieux informés et qui bénéficient d'informations internes (ce qui réintroduit un biais social), auront compris le mode de fonctionnement de l'algorithme, ils pourront mettre en place des stratégies pour le satisfaire : privilégier certains mots-clés, éviter la féminisation des mots si l'algorithme est sexiste, utiliser telle police d'écriture plus lisible pour un programme informatique, etc. Le premier biais viendra donc de la connaissance qu'ont certains⁵³ candidats de la machine à laquelle ils vont être confrontés lors du recrutement.

Cela vaut également lorsque ce qui est évalué n'est pas une personne, mais une institution. L'objet mesuré sera modifié sous l'effet de la mesure. L'exemple le plus parlant se trouve une nouvelle fois dans l'enseignement supérieur⁵⁴ et les célèbres classements universitaires. Ces classements, qui visaient initialement à mesurer l'excellence des universités, viennent modifier en profondeur et uniformiser les pratiques des universités devenues soucieuses de satisfaire les critères « d'excellence » du classement au détriment de la diversité des méthodes, des priorités de recherche, etc. Cet effet performateur du classement est une boucle de rétroaction aux effets négatifs (exacerbation de la concurrence entre universités au détriment de la collaboration, augmentation des frais d'inscription, communication à outrance sur l'« excellence » de l'Université au détriment des conditions de travail des chercheurs et de l'accueil des étudiants les plus modestes...) ⁵⁵.

Cet effet performateur est problématique lorsque l'algorithme entend mesurer et saisir le réel, car il vient modifier ce qu'il est censé mesurer. Il est en revanche bienvenu lorsque l'algorithme est utilisé à des fins de modification du réel. Lorsque l'on crée un index pour mesurer le respect de l'égalité entre les femmes et les hommes, le but recherché est bien la modification des

52. M. Walzer, *Sphères de justice*, Seuil, 2013, p. 201-206. L'auteur décrit le dévoiement au fil des siècles du système chinois de recrutement des fonctionnaires mis en place par l'empereur T'ai-Tsung (627-649) qui était un formidable ascenseur social et un exemple de promotion de la méritocratie. Le délitement du concours est intervenu du fait d'une routinisation des examens, d'une collaboration entre les examinateurs et les candidats qui apprenaient à passer les épreuves plutôt qu'à véritablement comprendre le sens des savoirs qu'on leur transmettait.

53. C. O'Neil, *Algorithmes*, op. cit., p. 176-177.

54. *Ibid.*, p. 84.

55. On retrouve cet effet en matière de prédiction de la récidive dans le système carcéral américain, ou dans la politique répressive de certaines villes usant du logiciel *Predpol* ou *ComptStat* ciblant la délinquance des quartiers pauvres (facilité par la politique du *stop and frisk* dont il a été démontré qu'elle était raciste par un Hackatlon auquel l'auteure a participé et qui a donné lieu à un procès contre le maire de New York) permettant de faire du chiffre facilement alors que dans le même temps les délits perpétrés par les riches ne sont pas l'objet du logiciel, conduisant à incarcérer toujours plus de personnes pauvres, majoritairement noires ou latinos, ce qui conduit le logiciel de prédiction de la récidive à considérer que la pauvreté est un facteur de récidive et donc à maintenir en détention ces catégories... *Ibid.*, p. 133-162.

pratiques des entreprises. Évidemment cela n'empêchera pas les stratégies pour tenter de contourner les exigences découlant de la mise en place de l'index. Il est donc nécessaire que ces index soient régulièrement modifiés ou perfectionnés pour éviter les stratégies de contournement. À noter que si l'index est mal conçu, le fait d'obtenir une « bonne note » peut donner aux entreprises un faux sentiment d'avoir en tout point respecté leurs obligations légales en matière d'égalité professionnelle et de non-discrimination.

Les algorithmes font naître des besoins sur le plan juridique, besoin en « données », pas toujours accessibles, besoin d'un cadrage numérique (qu'est-ce que la disproportion dans la définition de la discrimination indirecte? quel écart fait naître la disproportion? à partir de quelle donnée de référence?), besoin de sécurité juridique. Si les employeurs ne veulent pas voir automatiquement engager leur responsabilité en raison des recrutements ou promotions discriminantes décidés par un algorithme, ils doivent intégrer les règles de droit à l'algorithme, et le droit doit en partie être adapté pour permettre cette intégration, du moins si le législateur entend accompagner cette révolution de l'automatisation.

II. LA NÉCESSAIRE ADAPTATION DU DROIT À L'ALGORITHME

L'enjeu sur le plan juridique est de permettre au droit de devenir un paramètre implémentable au sein du code informatique. Par exemple (et de manière tout à fait approximative), si l'on cherche à respecter le principe d'égalité de traitement au sein d'une entreprise, on peut tenter d'automatiser le traitement dans le cadre d'un algorithme qui, en simplifiant, poserait que pour un avantage x , si A et B ont un même emploi ou un emploi similaire (que l'employeur devra lister comme étant similaire : *emploi 1*, *emploi 2*, *emploi 3*) alors le traitement de A doit être égal au traitement de B . La difficulté est alors bien connue des juristes et porte sur les différentes interprétations pouvant être faites de certains termes. Cependant, à mesure que le juge interprète la loi (notamment sur ce qu'est un emploi similaire ou équivalent) ou que le législateur vient la préciser, l'algorithme peut intégrer ces interprétations. À défaut de précisions, le concepteur devra effectuer des choix d'interprétation au même titre que n'importe quel employeur, particulier ou autre futur justiciable, qui pouvait penser être dans son droit avant d'être contredit par un juge. Cette part d'aléa, souvent qualifiée d'insécurité juridique, résulte bien moins de l'instabilité jurisprudentielle (faible et prévisible)⁵⁶ que de celle du législateur. Cependant, la stabilité des règles de non-discrimination permettrait

56. V. Pontif, *La sécurité juridique et le droit du travail*, thèse de doctorat soutenue sous la dir. de Marie-Armelle Rotschild-Souriac, à l'Université Paris X Nanterre-La Défense, le 2 avr. 2011.

L'usage des algorithmes et de l'IA dans le recrutement : une occasion de (ne) plus discriminer?

la mise en place systématique d'une analyse d'impact d'un algorithme ou d'une IA au regard de cette exigence, en amont de toute utilisation. S'il existe des règles encadrant l'usage des données, nous verrons qu'elles peuvent être parfois en contradiction avec les exigences de qualités que l'on peut attendre d'un algorithme ou d'une IA (A). Une modification de la législation n'est donc pas à exclure tout comme des procédures de contrôle plus efficaces des algorithmes (B).

A. L'ENCADREMENT JURIDIQUE DES ALGORITHMES ET DE L'IA DANS LE RECRUTEMENT

L'encadrement des algorithmes et des IA dédiées au recrutement est intégralement contenu dans la loi Informatique et libertés et le règlement général pour la protection des données (1) et les règles classiques de responsabilités patronales (2).

1. Le respect de la loi Informatique et libertés et du RGPD

On retiendra dans ce corpus de textes trois règles principales venant guider la conception et l'utilisation des algorithmes et des IA. Tout d'abord, l'interdiction d'user de données sensibles (a) qui constitue une entrave importante au développement d'algorithmes de contrôle des discriminations. Ensuite l'obligation de conserver une intervention humaine dans le processus de décision (b). Enfin, l'obligation de transparence (c) qui vise à obtenir la confiance des personnes qui confient leurs données personnelles.

a. L'interdiction d'user de données sensibles

Beaucoup des travers constatés aux États-Unis lors de recrutements⁵⁷ devraient être évités en France grâce aux protections offertes par le RGPD⁵⁸ et la loi Informatique et libertés⁵⁹. Chaque personne dont les informations sont collectées doit en principe avoir été informée de la finalité de la collecte et de la possibilité de consulter, rectifier voir demander l'effacement des informations qui le concernent⁶⁰. Théoriquement donc, les entreprises ne

57. Par exemple sur l'usage des scores de crédit, censés attester de la fiabilité des personnes, dans le cadre de la procédure de recrutement dans le secteur bancaire (qui accède à ces scores de crédit dans le cadre des demandes de prêts bancaires). V. sur les conséquences d'un tel usage C. O'Neil, *op. cit.*, p. 22.

58. Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avr. 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

59. V. en particulier les droits de la personne au regard du traitement de ses données, énumérés à l'article 4 de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés.

60. Art. 49 à 51 de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés.

peuvent effectuer des recrutements sur la base de données obtenues sans que le candidat en ait été informé. Certaines données font l'objet d'une protection particulière puisque, au titre de l'article 6 (I) de la loi Informatique et libertés, « il est interdit de traiter des données à caractère personnel qui révèlent la prétendue *origine raciale* ou *l'origine ethnique*, les *opinions politiques*, les *convictions religieuses* ou *philosophiques* ou *l'appartenance syndicale* d'une personne physique ou de traiter des *données génétiques*, des *données biométriques* aux fins d'identifier une personne physique de manière unique, des données concernant la *santé* ou des données concernant la *vie sexuelle* ou *l'orientation sexuelle* d'une personne physique ». Dans cette hypothèse, « la personne concernée [doit avoir] donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques⁶¹ ». On remarque immédiatement la difficulté qui va naître de cette interdiction de traiter massivement de ces données dans le cadre d'algorithme ou d'IA.

C'est là l'un des paradoxes que posent les algorithmes. Apprendre à l'algorithme à ne pas discriminer suppose de lui apprendre ce qu'est la discrimination et quels sont les motifs de discrimination prohibés. Or, l'interdiction de collecter certaines données pose problème dès lors que l'on entend lutter contre les effets discriminatoires des algorithmes⁶². On interdit à l'algorithme de traiter défavorablement certaines catégories, et on lui refuse dans le même temps les données lui permettant de mesurer s'il traite différemment certaine catégorie. Il faut donc impérativement distinguer l'usage par l'algorithme ou l'IA de certains critères discriminatoires comme prédicteur pour prendre une décision, la mesure de l'impact de ses décisions en apparence « neutre » sur ces catégories identifiées par un motif discriminatoire et l'éventuelle prise en compte de ces critères discriminatoires à des fins de rééquilibrage des discriminations indirectes.

Si les algorithmes ne peuvent intégrer certaines données relevant de critères de discriminations (dans la limite où leur collecte et utilisation ne constituent pas une violation de la vie privée⁶³), alors il est impossible de leur apprendre à ne pas discriminer de manière indirecte ce qui pose problème en matière de responsabilité de l'employeur⁶⁴.

61. Art. 44 de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés renvoyant aux exceptions énumérées à l'article 9 (II) du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avr. 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

62. V. *infra*.

63. V. M. Peyronnet, *La diversité : étude en droit du travail*, thèse réalisée sous la direction de C. Radé, soutenue le 4 déc. 2018 à l'Université de Bordeaux, en ligne sur Hal-SHS, tel-02279370.

64. V. *infra*.

b. L'obligation de conserver une intervention humaine dans le traitement automatisé

Dans sa fonction d'imitation du pouvoir patronal, l'algorithme ne sera bien souvent qu'un outil d'aide à la décision alors qu'à l'inverse l'IA visera à se substituer à la décision patronale en fixant les moyens pour remplir les objectifs qui lui ont été assignés (recruter le travailleur le plus compétent, le plus productif...) ou qu'un jour elle s'assignera elle-même. Dans une telle configuration, l'IA peut rapidement tomber dans l'illégalité puisque la loi informatique et libertés, modifiée suite à l'adoption du RGPD, impose qu'« aucune [...] décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise *sur le seul fondement d'un traitement automatisé* de données destinées à prévoir ou à évaluer certains aspects personnels relatifs à la personne concernée⁶⁵ ». Mais lorsqu'un algorithme trie des dossiers de candidature, il en écarte certains, en sélectionne d'autres, tout l'intérêt de l'outil est de « dégrossir » le travail du recruteur. Donc, si des CV ont été écartés par l'algorithme, il y a déjà eu des décisions individuelles sur lesquelles l'être humain n'est, à aucun moment, intervenu et qui se trouvent, de ce fait, avoir fait l'objet d'un traitement entièrement automatisé. À moins que l'on considère que l'intervention humaine, lors de la dernière phase du recrutement pour départager les candidats retenus ou valider le recrutement proposé par l'algorithme, soit une décision de refus implicite des dossiers des candidats non retenus. Une telle approche serait critiquable car l'opérateur humain n'aura jamais eu entre ses mains ou sous son regard le dossier qu'il a rejeté en validant la candidature d'un autre.

Avec l'IA, le problème sera encore plus complexe, car elle a vocation à prendre la décision de recrutement et l'humain devient paradoxalement⁶⁶ son exécutant⁶⁷. Jouant ainsi un rôle de « cooptation » artificielle, elle affirme au recruteur que ce sont là les meilleurs profils et le recruteur choisi de la croire, ou non. La responsabilité *a priori* reste la sienne même s'il ne dispose pas nécessairement de toutes les informations utiles pour motiver « sa » décision.

c. L'obligation de transparence

Il ressort de l'étude de Cathy O'Neil que les algorithmes qui présentent le moins d'effets néfastes sont ceux qui sont les plus ouverts avec les données

65. Considérant 71 du RGPD et art. 95 de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés telle que modifiée par l'art. 1 de l'ordonnance n° 2018-1125 du 12 déc. 2018 prise en application de l'article 52 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles et portant modification de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel.

66. Paradoxalement car l'employeur reste quoi qu'il en soit le seul responsable juridique des décisions algorithmiques. V. *infra*.

67. À noter que lorsqu'il s'agit de l'octroi d'une prime, l'IA peut potentiellement tout gérer elle-même, de la détermination des critères pour son octroi jusqu'à sa mise en paiement.

les plus directes et continuellement mises à jour⁶⁸. C'est donc assez logiquement que l'obligation de transparence est l'un des piliers du RGPD⁶⁹.

Pour qu'une personne soit responsable de l'usage qui est fait d'un algorithme, il est important au préalable que ceux pouvant engager sa responsabilité soient conscients d'avoir été soumis à l'examen d'un algorithme, et conscients que ce dernier a eu une influence sur les décisions individuelles ou collectives qui les ont affectés. La question de la connaissance et de l'accès au contenu de l'algorithme (notamment les données personnelles ayant été utilisées) est donc fondamentale pour quiconque entend contester les décisions qui lui sont opposées. Pour certains auteurs, l'être humain est une boîte noire, alors que l'algorithme a le potentiel d'être bien plus transparent⁷⁰. Mais le potentiel seulement. Par transparence on parle bien évidemment de l'accès au code source du logiciel, mais également aux explications qui l'accompagnent et à l'algorithme qui a dirigé la prise de décision⁷¹. La transparence dans le cadre de l'IA est beaucoup plus insaisissable comme on a pu le voir précédemment du fait de son caractère évolutif. L'obligation de transparence dans ce cas devrait donc être renforcée.

68. C. O'Neil, *op. cit.*, p. 49. Elle vise tout particulièrement les algorithmes utilisés en matière de recrutement sportif (baseball principalement : où les données sont issues de la pratique du sport : nombre de home-runs, de balles perdues, etc. Ces données sont à la vue de tous les acteurs et les joueurs peuvent contester leur non-recrutement, transfert ou autre en argumentant sur les statistiques prises en considération), mais donne également l'exemple de la solution avortée du gouvernement de Barack Obama qui avait pour ambition de remplacer le classement privé des universités. Dans l'incapacité de créer un algorithme approprié, une plateforme avec l'ensemble des données collectées a été mise en place et les futurs étudiants peuvent classer les universités en fonction de leurs propres critères d'excellence. Ce qui est finalement le meilleur des systèmes.

69. Art. 5 du RGPD et art. 4 de la loi n° 78-17 du 6 janv. 1978 relative à l'informatique, aux fichiers et aux libertés.

70. J. Kleinberg, J. Ludwig, S. Mullainathan *et al.*, « Discrimination in the Age of Algorithms », *op. cit.*

71. C'est sur la base de cette obligation de transparence qu'a été demandée la publication des algorithmes locaux des universités. À noter d'ailleurs que dans la décision 2019-021 du 18 janv. 2019 du Défenseur des droits relative au fonctionnement de la plateforme nationale de préinscription en première année de l'enseignement supérieur (Parcoursup) le débat porte notamment sur la communication « préalable » des attendus et de leur pondération. Ce qui pose une question de prévisibilité : pour que les étudiants puissent au mieux préparer leur dossier, il faut connaître les exigences très en amont de leur demande, dès la fin du collège pour certaines options qu'ils risqueraient, à tort, d'abandonner. Cela rentre en contradiction avec la nécessité de régulièrement actualiser et modifier l'algorithme en fonction des nouvelles données disponibles et des nouvelles exigences des universités (peut-être tel critère n'était-il pas pertinent et n'était pas un gage de réussite et devrait donc être modifié pour les prochaines années?).

2. La responsabilité de l'employeur

L'employeur sera toujours responsable de ses décisions de recrutements ou de promotions à l'égard de ses salariés et des candidats à l'emploi, quand bien même un algorithme ou une IA aurait fait la majeure partie du travail de sélection à sa place. *In fine*, c'est toujours lui qui recrute (ou refuse de recruter)⁷².

La question plus délicate concerne l'employeur qui, conscient des biais de recrutement humains et algorithmiques, voudrait les repérer et les éliminer *via* l'outil algorithmique. L'un des moyens serait d'implémenter le droit applicable à l'algorithme pour lui donner des limites⁷³. Par exemple, l'algorithme peut user de n'importe quelles données prédictives (nature et niveau de diplôme, expérience, compétences, etc.), mais si l'utilisation d'une donnée conduit à écarter de manière disproportionnée les femmes (par exemple) par rapport aux hommes, l'algorithme peut être conçu pour privilégier dans ce cas une autre donnée prédictive ou pour modérer son utilisation en prenant en considération les recrutements passés et futurs à effectuer. Apprendre à l'algorithme à ne pas discriminer les femmes suppose de lui expliquer ce qu'est une discrimination dans le recrutement, notamment indirecte (à partir de quel seuil considère-t-on qu'il y a disproportion ?), et de lui donner plusieurs informations : combien de femmes candidates, combien de femmes éliminées, et évidemment comment déterminer si le candidat est une femme ou un homme. Ce dernier point peut devenir extrêmement compliqué si au lieu de discriminer les femmes l'algorithme discrimine indirectement les personnes racisées. En effet, cette donnée n'existant pas, l'algorithme ne peut pas limiter son action discriminante sur un motif racial. Cette information que le recruteur perçoit en un battement de cils est invisible pour l'algorithme qui respecterait l'interdiction de collecter ces données sensibles. C'est là la limite juridique majeure pour l'algorithme : pour lui apprendre à ne pas discriminer, il faut lui apprendre ce qu'est la discrimination au sens juridique et surtout lui permettre de la mesurer. L'algorithme ne comprend que les données. Sans données, il ne peut donc pas s'empêcher

72. Ce qui ne veut pas dire que la responsabilité (notamment pénale) d'autres intervenants dans le processus discriminatoire ne puisse être recherchée : Crim. 23 juin 2009, *Garnier, Adecco et a.*, n° 07-85.109, *Bull. crim.*, n° 126, *AJ pénal* 2009. 408 s., obs. J. Lasserre-Capdeville, *Dr. pénal* 2009, comm. n° 121, obs. Véron; Délib. Halde n° 2006-98 du 9 mai 2006 : concernant une offre d'emploi d'animatrices pour une campagne de vente des laboratoires Garnier donnant pour consigne à Adecco de recruter des personnes « BBR » (pour Bleu, Blanc, Rouge). *Idem* pour la responsabilité civile du développeur de l'algorithme, on peut imaginer engager sa responsabilité sur le fondement des produits défectueux (C. civ., art. 1245 s.) ou s'il s'agit d'un prestataire de service sur le fondement du contrat de prestation. Selon l'hypothèse il faudra alors déterminer si les biais sont liés à la conception (produit défectueux), aux données utilisées par l'employeur (cause d'exclusion du régime des produits défectueux, C. civ., art. 1245-10 2°), ou si le cahier des charges a été respecté ou non par le prestataire de service.

73. Commission européenne pour l'efficacité de la justice, Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement, adoptée lors de la 31^e réunion plénière de la CEPEJ, Strasbourg, 3-4 déc. 2018, p. 37.

de discriminer. On peut adopter le même raisonnement pour une IA. Si en analysant ses pratiques une IA n'a pas les moyens de voir qu'elle discrimine tel ou tel groupe, elle ne pourra pas se corriger. Et quand bien même un être humain verrait le problème, il ne serait légalement pas possible de fournir des données à l'IA pour lui permettre de se corriger. Cette limite vaut aussi bien pour les algorithmes imitant le pouvoir patronal que pour ceux visant à le contrôler. Le débat sur les statistiques ethniques ne manquera donc pas d'être rouvert compte tenu de l'ampleur que prend l'usage de ces outils.

Au-delà de la collecte des données, c'est peut-être le droit de la non-discrimination lui-même qui devrait être adapté à la réalité algorithmique notamment en reconnaissant en droit français la notion de discrimination systémique. Ainsi, pour Jérôme Porta, « l'identification de la discrimination systémique aurait tout d'abord pour mérite de clarifier la notion de discrimination indirecte qui semble *a priori* demeurer associée à une décision identifiée, quand la discrimination systémique prend en compte un ensemble de pratiques sans qu'il faille qu'à elle seule aucune de ces pratiques n'ait de caractère discriminatoire identifiable⁷⁴ ». Cela permettrait en effet d'appréhender les situations de discriminations indirectes découlant de la prise en compte d'une multiplicité de facteurs sans que l'on ne puisse déterminer lequel – ou même si l'un d'eux – a été prépondérant dans l'exclusion d'une catégorie d'individus. Cette notion permettrait également de compléter la notion de discrimination. Celle-ci n'est pas qu'individuelle, elle doit aussi pouvoir être appréhendée de manière collective. Lorsqu'un collectif de travail est majoritairement masculin et que l'employeur entend à l'avenir « privilégier », à compétences égales, les candidatures féminines, on devrait pouvoir écarter l'argument d'une discrimination individuelle à l'égard d'un homme (qui va faire « les frais » de cette mesure de rattrapage) en raison de l'absence de discrimination systémique envers les hommes et du fait qu'elle est justifiée par l'existence d'une discrimination systémique envers les femmes.

Le droit existant mériterait donc quelques amendements, mais ce n'est rien en comparaison du véritable arsenal de contrôle du contenu des algorithmes et de l'IA qu'il faudrait idéalement mettre en place.

B. RENFORCER LE CONTRÔLE DES ALGORITHMES ET DE L'IA

Tous les algorithmes devraient faire l'objet d'un contrôle, *a priori*, dès leur conception, et *a posteriori* de leur mise en œuvre afin de vérifier leurs effets sur les moyen et long termes, notamment pour contrôler l'absence de boucle de rétroaction. Cependant, pour rendre possibles ces contrôles, il est impératif de renforcer l'obligation de transparence.

74. J. Porta, « Algorithme et risques discriminatoires », actes de la journée d'étude sur les Nouveaux modes de détection et de prévention de la discrimination et accès au droit, organisée par Trans Europe Expert et le CNAM, CNAM, Paris, le 10 déc. 2018, à paraître.

L'usage des algorithmes et de l'IA dans le recrutement : une occasion de (ne) plus discriminer ?

Le RGPD impose aux concepteurs de systématiquement procéder à une étude d'impact des traitements de données personnelles sur le respect de la vie privée des individus (*privacy impact assessment - PIA*)⁷⁵. Le rapport Villani⁷⁶ préconise, conformément aux orientations retenues par le groupement européen des autorités de protection des données (G29), que le PIA s'accompagne d'un DIA (*discrimination impact assessment*) pour « amener les concepteurs d'IA à s'interroger sur les conséquences sociales des algorithmes qu'ils produisent ». Il s'agit ici d'une précaution essentielle visant à garantir l'« acceptabilité sociétale de l'IA ».

Il semble cependant préférable, plutôt que de laisser les entreprises faire elles-mêmes leur propre étude d'impact, que les algorithmes fassent l'objet d'un dépôt obligatoire⁷⁷. En matière de recrutement, plusieurs institutions pourraient garder un œil sur ces algorithmes : la CNIL compte tenu des données sensibles qui seront utilisées pour son fonctionnement et l'utilisation qui sera faite de ces données ; l'inspection du travail et/ou le défenseur des droits qui pourraient intervenir lorsqu'un algorithme enfreint les règles en matière de non-discrimination et enfin Pôle emploi qui pourrait ainsi utilement informer et conseiller les demandeurs d'emploi qui vont se soumettre à l'examen de ces machines.

À défaut de dépôt possible, notamment s'agissant d'une IA qui évoluerait en permanence, on pourrait *a minima* exiger que chaque entreprise dispose d'un référent dans l'une ou plusieurs des institutions mentionnées précédemment et qui serait informé des modifications apportées et pourrait à tout moment accéder à l'algorithme d'apprentissage, aux données utilisées et à l'historique des corrélations effectuées⁷⁸ dans le cadre du recrutement et la gestion du personnel. Autre acteur incontournable en la matière : le Comité social et économique. En effet, le nouvel article L. 2312-38 du Code du travail prévoit dans les attributions du CSE que celui-ci « est informé,

75. V. sur l'intégration de cette obligation dans le droit français : CNIL, « Ce qu'il faut savoir sur l'analyse d'impact relative à la protection des données (AIPD) », site internet de la CNIL, 22 oct. 2019.

76. C. Villani, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, rapport rendu dans le cadre de la mission confiée par le Premier ministre du 8 sept. 2017 au 8 mars 2018, p. 147-148. On a pu constater qu'il existe quelques discussions à l'Assemblée nationale sur la propagation des discriminations du fait des algorithmes à l'occasion de certains débats, v. Rapport fait au nom de la commission des lois constitutionnelles, de la législation et de l'administration générale de la République sur la proposition de loi visant à lutter contre la haine sur internet (n° 1785) présenté par M^{me} Laetitia Avia le 19 juin 2019.

77. À l'image de l'obligation de dépôt de l'accord ou de la convention collective négociée (C. trav., art. L. 2231-6).

78. L'IA établit des corrélations. Son algorithme pose comme objectif d'augmenter la vente de chocolat. Pour augmenter les ventes on lui fournit des données. Elle établit une corrélation entre le nombre de prix Nobel et la consommation de chocolat et estime que le meilleur moyen de vendre plus de chocolat est cibler les pays ayant le plus de prix Nobel. Il n'y a aucun lien de cause à effet entre le prix Nobel et le chocolat, seulement une corrélation.

préalablement à leur utilisation, sur les méthodes ou techniques d'aide au recrutement des candidats à un emploi ainsi que sur toute modification de celles-ci. Il est aussi informé, préalablement à leur introduction dans l'entreprise, sur les traitements automatisés de gestion du personnel et sur toute modification de ceux-ci ».

Il devrait également être possible, s'agissant d'une IA, de la contraindre, pour chacune de ses décisions de rejet ou d'admission, à envoyer automatiquement un rapport au candidat expliquant pourquoi sa candidature a ou n'a pas été retenue dans le cadre du recrutement ainsi que l'inventaire de toutes les données et corrélations utilisées dans le cadre de cette procédure. Le plus efficace serait d'y ajouter le profil professionnel de la personne recrutée pour que le salarié fasse lui-même la comparaison. Ainsi, si le candidat est automatiquement informé qu'il n'a pas été retenu pour le poste par un courriel lui indiquant les motifs objectifs de ce refus, on pourra dire que le recrutement n'aura jamais été aussi transparent!