

L'intelligence artificielle est-elle vraiment impartiale ?

L'évolution des nouvelles technologies et spécifiquement des intelligences artificielles révolutionnent aussi le domaine des ressources humaines et du recrutement en particulier. Ainsi, nous assistons à un fort développement de l'utilisation des algorithmes dans le processus de recrutement et d'ailleurs dans la poursuite de carrière. Cet usage expose à des risques majeurs. Le risque qui attire notre attention sont les « **biais** » de discriminations dont les candidats à l'emploi et salariés sont victimes.

La CNIL définit l'algorithme comme « la description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée. [À titre d'exemple,] les algorithmes mathématiques permettent de combiner les informations les plus diverses pour produire une grande variété de résultats ». En marge de l'algorithme, certaines notions méritent également une tentative de définition. Il s'agit entre autres du biais algorithmique et des biais d'automatisation.

Les biais d'automatisations surviennent lorsque les personnes sont portées à suivre « **aveuglement** » les recommandations d'un système automatisé, même lorsque celui-ci commet des erreurs évidentes. Cela peut s'expliquer par une confiance excessive dans la technologie et la présomption qu'elle a été programmée pour ne pas faire d'erreurs, contrairement aux êtres humains. Ce biais est particulièrement problématique dans des situations critiques où une erreur de jugement peut avoir des conséquences graves, comme dans le cas notamment des pilotes d'avion qui omettent de corriger les erreurs des systèmes de pilotage automatique.

« Un biais algorithmique est un défaut dans le fonctionnement de l'algorithme qui tend à traiter différemment et de manière [parfois] incohérente, voire injuste, les situations ou les individus. On parle aussi de « discrimination technologique » quand il s'agit d'injustice de traitement des êtres humains. Ces biais algorithmiques proviennent d'un biais dans les données d'apprentissage ou de calibration qui ne sont pas statistiquement représentatives, des critères ou des hypothèses faites sur la logique de l'algorithme. Dans tous les cas, ces biais proviennent souvent de nos biais cognitifs à nous êtres humains, qu'on soit concepteurs de ces algorithmes ou simples utilisateurs dont nos données souvent biaisées orientent maladroitemment l'algorithme. »

Quels sont les risques discriminatoires inhérents à l'utilisation des algorithmes ?

La neutralité de l'algorithme est en réalité très relative dans la mesure où elle dépend de la façon dont celui-ci est conçu, des données qui l'alimentent et de son évolution. Or, « dans la mesure où sa conception et son alimentation dépendent largement de l'être humain, l'algorithme en revêt nécessairement les faiblesses ». Aussi l'analyse de son évolution, qui échappe dans certains cas à la maîtrise de l'être humain, révèle-t-elle des biais discriminatoires imprévus.

Il est utile de préciser que la discrimination est proscrite par le législateur dans le code pénal et passible d'une peine d'emprisonnement de 3 à 5 ans et d'une amende d'au moins 45.000 euros.

Les risques discriminatoires inhérents à l'utilisation des algorithmes sont de deux ordres : d'une part, les biais inhérents aux données personnelles et d'autre part, les biais inhérents à l'algorithme lui-même. Il est impératif de se former aux questions des discriminations à l'œuvre dans le recrutement et la gestion de carrière. A Compétence Egale propose plusieurs modules dédiés à ces thématiques [pour en savoir plus](#).

Quels peuvent être les biais inhérents aux données personnelles ?

Il s'agit ici des données qui manquent d'objectivité dans la représentation des différentes catégories de personnes.

A titre d'exemple, dans le monde de l'emploi et spécifiquement dans certaines filières, la représentativité des femmes est moindre. Par conséquent, les profils types des candidats performants seront présumés masculins. Un algorithme nourri avec de telles données, mêmes si elles se fondent sur la réalité, sera plus enclin à privilégier des candidatures masculines et donc discriminer celles féminines.

C'était également le cas de l'algorithme utilisé par Amazon. En effet, étant donné que dans l'entreprise la représentation des métiers de la technologie était dominée par des hommes, l'algorithme a appris à favoriser les candidatures de sexe masculin au détriment des femmes.

Quels peuvent être les biais inhérents à l'algorithme ?

Trois (03) situations, non exhaustives, permettent de mettre en lumière les biais inhérents à l'intelligence artificielle dans le processus de recrutement :

- L'hypothèse d'un algorithme qui intègre dans son code des préjugés entraînant des discriminations.

L'exemple le plus parlant est celui d'une intelligence artificielle qui est utilisée pour rédiger une offre d'emploi et qui la rédige exclusivement au féminin. L'offre est discriminatoire parce qu'elle mobilise une caractéristique protégée par le droit, en l'occurrence le sexe, et exclut une catégorie de personne.

- L'hypothèse d'un algorithme qui intègre un critère de discrimination qui en apparence est neutre mais constitutif d'une discrimination au sens de la législation française.

C'est le cas par exemple d'un algorithme qui privilégie les candidats ponctuels.

C'est aussi le cas du litige qui oppose l'entreprise Deliveroo à l'un de ses employés. En effet, la décision du tribunal de Bologne en Italie a condamné Deliveroo pour discrimination liée à l'utilisation de son algorithme appelé "Frank" spécialisé dans la gestion des livreurs. En l'espèce, Frank attribuait un "score de réputation" aux livreurs en se basant sur deux critères : l'indice de fiabilité (lié aux absences) et l'indice de participation aux plages horaires très fréquentées. Ce score déterminait l'ordre d'accès des livreurs à la réservation des créneaux de travail chaque lundi (11h, 15h ou 17h). Les meilleurs scores accédaient en premier aux meilleurs créneaux. Cependant, Frank ne prenait pas en compte les absences ou retards des livreurs (maladie, grève, ...) et les traitait de manière discriminatoire en abaissant leur score.

Par conséquent, le tribunal a jugé que le système de Frank avait un "caractère essentiellement discriminatoire" en ne tenant pas compte des raisons légitimes d'absence des livreurs, les privant ainsi de leurs droits fondamentaux en tant que travailleurs. Deliveroo a été condamné pour violation du RGPD, notamment le manque de transparence et d'information préalable sur le fonctionnement de l'algorithme et pour discrimination. En somme, le tribunal a jugé que l'algorithme Frank de Deliveroo discriminait les livreurs de manière injustifiée et opaque, violant ainsi le RGPD et les droits des travailleurs.

[L'académie A Compétence Egale](#) a dédié plusieurs capsules sur ces sujets avec notamment des vidéos d'expert.e.s très éclairantes.

Rédigé par Rock I. IROKO

Relecture : Benoît BINACHON