«Il est préférable que des humains contrôlent le travail de l'IA»

L'avocat lausannois David Raedler, membre de la Commission fédérale pour les questions féminines, est spécialisé en droit du travail. Il a récemment accompagné une cliente victime de discrimination algorithmique. Il nous éclaire sur le litige en question et fait un point sur le rôle de l'intelligence artificielle dans la perpétuation de biais discriminants.

Entretien : Kiri Santer

Pourriez-vous nous expliquer comment ce cas vous est parvenu et comment votre cliente s'est rendu compte d'être victime de discrimination algorithmique ?
J'ai passablement de cas liés à la Loi fédérale sur l'égalité entre femmes et hommes (LEg), car je suis spécialisé en droit du travail et la LEg est malheureusement encore souvent enfreinte. Dans ce cas précis, ma cliente avait une grande ancienneté dans le domaine du marketing. Elle avait occupé des positions assez élevées et avait décidé de changer de domaine, toujours dans le marketing. Elle avait postulé à un poste qui l'intéressait particulièrement. Dans le cadre du processus d'embauche, l'employeur lui avait indiqué qu'il y aurait des tests de capacité générale à passer, dont un test psychologique, qui serait mené par voie algorithmique, c'est-à-dire avec l'aide d'une intelligence artificielle. Elle a effectué le test et a eu l'impression que cela s'était bien passé. Peu de temps après toutefois, il lui a été annoncé qu'elle ne serait pas retenue. C'était la première phase de l'entretien et il y avait encore trois phases de plus auxquelles elle n'a donc pas pris part. Compte tenu du fait qu'elle avait un très bon profil, cela l'a étonnée et elle a cherché à connaître les raisons de son élimination. La réponse de la part de l'employeur était qu'elle ne correspondait pas au profil cherché.

Quelque temps plus tard sur LinkedIn, elle a appris que l'entreprise en question avait engagé un homme qu'elle considérait avoir moins d'expérience qu'elle. Elle s'est alors souvenue du processus de recrutement et est venue me voir. Il faut ici préciser une chose, car il s'agit d'une information importante : la nouvelle Loi fédérale sur la protection des données (LPD) n'était pas encore entrée en vigueur.[1] Cela est pertinent, car la nouvelle LPD contient des règles spéciales sur ce que l'on appelle des décisions automatisées. Ces nouvelles règles permettent d'une part de demander que ces décisions automatisées soient revues par une personne physique, et, d'autre part, d'obtenir plus de transparence quant à la «logique» qui est derrière l'algorithme.[2] Étant donné que cette loi allait bientôt entrer en vigueur, j'ai proposé à ma cliente d'écrire à l'employeur pour lui demander sur quels critères de sélection s'était basé l'algorithme. Néanmoins, l'employeur n'a pas fourni ces informations, indiquant qu'il s'agissait d'un secret d'affaires. Nous nous sommes alors tournés vers le tribunal.

Comment cela s'est-il passé devant le tribunal ? Quels étaient vos arguments et ceux de l'employeur ?
J'ai basé mon argumentation sur la LEg, car c'est la seule loi qui permet de contester une non-embauche sur la base de discrimination. Ce qu'il faut savoir, c'est que la LEg contient un principe de présomption de violation. Pour les cas de licenciement tombant sous la LEg, s'il y a un élément qui indique que le licenciement est discriminatoire, la responsabilité revient à l'employeur de prouver que ce n'est pas le cas, et donc de fournir le motif réel de licenciement. C'est une forme de renversement du fardeau de la preuve. Cette présomption de violation ne s'applique cependant pas aux cas d'engagement. Cela signifie que pour une procédure d'embauche, il incombe à la personne qui

David Raedler

«Pour une procédure d'embauche, il incombe à la personne qui postule de prouver qu'il y a eu discrimination.»

« L'influence des développeurs – qui sont majoritairement des hommes – est très conséquente. »

postule de prouver qu'il y a eu discrimination. Vous pouvez imaginer que c'est donc très difficile. Néanmoins, cela valait la peine d'essayer et nous avons déposé une requête de conciliation.[3] Lors de l'audience de conciliation, à laquelle l'employeur était présent, j'ai indiqué que nous suspections que le logiciel utilisé contenait un biais de genre qui aurait influencé la décision de ne pas retenir ma cliente comme candidate. Puisque le fardeau de la preuve incombait à ma cliente et que nous n'avions pas d'informations précises quant au fonctionnement du logiciel, je prévoyais de demander l'expertise du logiciel dans le cadre de la procédure. Poussés par les juges, nous avons alors suspendu la procédure de conciliation et l'employeur a été d'accord que l'on teste le logiciel.

Comment l'avez-vous testé ?

J'ai créé cinq CV sur la base de celui de ma cliente. Ceux des femmes présentaient des parcours plus étoffés, incluant une expérience internationale ou des diplômes d'universités prestigieuses. En revanche, les CV des hommes affichaient des profils plus ordinaires et moins d'expérience. Pourtant, le système a de nouveau choisi un homme. L'employeur a alors demandé au fournisseur du logiciel, qui l'avait fabriqué sur-mesure, d'identifier les causes de ce résultat. Il en est ressorti que l'algorithme contenait bel et bien un biais de genre : plus de points étaient attribués aux candidatures qui contenaient la qualité « être un homme », car l'algorithme lui-même avait identifié qu'il fallait un homme pour la fonction recherchée, qui comprenait le marketing d'un produit qui était principalement destiné aux hommes. Naturellement, il ne s'agissait pas d'un critère décisif, en revanche, cela faisait gagner des points aux candidatures masculines.

Quelle solution a été trouvée ?

L'employeur a concédé qu'il avait commis une erreur. Ma cliente, elle, entre temps, avait retrouvé un travail. Elle a donc renoncé à l'indemnisation en disant qu'elle serait satisfaite si l'employeur prenait en charge ses frais d'avocats et s'engageait à changer le logiciel. Nous avons donc établi une convention dans laquelle l'entreprise en question a payé mes honoraires et s'est engagée à changer l'algorithme. Et l'affaire s'est arrêtée là.

Qu'avez-vous appris par le biais de ce cas ?

Avant, pour parler du biais algorithmique, je mentionnais souvent le cas du genre : je disais que si l'IA se base sur les données existantes qui indiquent que 90 pour cent des directeurs financiers sont des hommes, sans correcteur de biais, naturellement le logiciel arrive à la conclusion que d'être un homme est une qualité formidable. Lorsque je donnais des conférences sur le sujet, je donnais toujours cet exemple-là de discrimination de genre en disant que c'était assez simple d'identifier ce type de biais. Dans le cas auquel j'étais confronté, le logiciel n'avait pas identifié que 90 pour cent des directeurs de marketing étaient des hommes, mais avait pris en compte le domaine d'activité de l'entreprise. C'est comme si la machine s'était fait sa propre réflexion ou alors que lorsqu'elle a été entrainée avec des données, elle a développé le critère que « expérience dans le domaine » pouvait signifier aussi « utilisation du produit ».

Ce cas illustre bien les difficultés auxquelles l'on est confronté dans le cadre d'une discrimination à l'embauche : puisqu'il n'y a pas la présomption de violation de la LEg, il revient aux plaignants et plaignantes de prouver cette discrimination. Dans notre cas, si l'employeur n'avait pas accepté la suspension de la procédure et n'avait ensuite pas accepté que je crée moi-même de faux CV, je ne sais pas si nous aurions pu aller au tribunal. Il n'y a pas de frais judiciaires lorsque l'on agit sous la LEg et l'expertise n'aurait pas été à notre charge. Cependant, le risque aurait tout de même été de devoir financer les frais d'avocats de la partie adverse en cas de défaite. Si l'employeur avait été de mauvaise volonté, il aurait été très difficile de démontrer la présence de biais dans l'algorithme.

Avez-vous souvent travaillé sur des cas de discrimination à l'embauche ?

Non pas vraiment. Les seuls cas que l'on voit parfois sont ceux où il y a un critère de genre évidemment discriminatoire indiqué dans l'annonce, par exemple, une recherche de vendeuse, car les femmes vendraient mieux. Là, il s'agit de mon premier cas de discrimination algorithmique sous la LEg. J'ai eu d'autres cas en lien avec les algorithmes, mais pas en lien avec l'égalité. Les cas en lien avec les algorithmes tombent plutôt sous la Loi sur la protection des données (LPD).

Comment voyez-vous la discrimination algorithmique évoluer notamment en lien avec l'utilisation de l'IA de façon plus générale ?

Nous sommes aux débuts de l'IA, et à mon avis, les problèmes de discrimination continueront à augmenter. L'influence des développeurs – qui sont majoritairement des hommes – est très conséquente. Il faudrait donc que les employeurs utilisant des systèmes d'IA réfléchissent eux-mêmes aux possibles biais. Je pense qu'il est préférable que des humains contrôlent le travail de l'IA, en tout cas pour les décisions finales, car le risque de biais est très élevé. Rien que dans le contexte du travail, l'IA peut être utilisée à tous les stades, aussi bien à l'embauche que dans l'analyse de la performance. Pour l'embauche, une autre mesure à prendre pour éviter le genre de cas comme celui de ma cliente serait l'éventuelle anonymisation des CV sous l'angle de critères non-pertinents (genre, origine ethnique, etc.). Malheureusement, cette mesure n'exclut pas qu'il y ait de la discrimination algorithmique à d'autres niveaux. Par exemple, en termes de performance, une femme qui allaiterait pourrait être pénalisée par un logiciel pour avoir pris trop de pauses s'il n'y a pas de prise en compte de la différence de genre. J'ai en tête un autre cas de discrimination à l'embauche en France, cette fois pas de genre, mais de classe : le logiciel attribuait une mauvaise note aux candidat·e·s·x venant d'un quartier défavorisé. Le logiciel faisait une équivalence de critère sans contexte. Le quartier en question était défavorisé dans le système, car statistiquement il n'y avait pas beaucoup d'universitaires qui y vivaient. Ces algorithmes restent des systèmes binaires qui ont du mal à prendre en compte des données sociologiques plus complexes.

Kiri Santer est chercheuse en sciences sociales. Actuellement post-doctorante à l'Université de Berne, ses recherches portent sur les politiques migratoires et climatiques de l'Union européenne.

Notes

1 La nouvelle LPD est entrée en vigueur le 1er septembre 2023.

2 Ces dispositions valent pour n'importe quelle situation, de la contraction d'une assurance à une procédure d'embauche.

3 La conciliation est la première étape de la procédure judiciaire. Les étapes sont : requête de conciliation, audience de conciliation et puis après, en cas de non-accord, dépôt de demande qui est suivi par un procès.

« Si l'employeur avait été de mauvaise volonté, il aurait été très difficile de démontrer la présence de biais dans l'algorithme. »

Abstracts

«Es ist besser, wenn Menschen die Arbeit der KI kontrollieren»

Der Lausanner Rechtsanwalt **David Raedler** berichtet von einem konkreten Fall von algorithmischer Diskriminierung: Eine Frau wurde nicht zur zweiten Runde in einem Bewerbungsverfahren eingeladen, weil ein Algorithmus «beschlossen» hatte, dass für die betreffende Stelle ein Mann besser geeignet sei. Dieser Fall zeigt, dass algorithmische Verfälschungen weit verbreitet sind, unter anderem deshalb, weil die Entwickler einen grossen Einfluss haben. Und diese sind – auch heute noch – mehrheitlich Männer.

«È meglio che gli umani controllino il lavoro dell'IA»

L'avvocato losannese **David Raedler** ripercorre un caso concreto di discriminazione algoritmica in cui a una donna è stato precluso l'accesso al secondo turno di una procedura di assunzione, perché un algoritmo aveva concluso che il genere maschile costituiva un titolo preferenziale per la posizione in questione. L'autore dell'articolo ci ricorda che i bias dell'intelligenza artificiale sono numerosi anche a causa del ruolo preponderante degli sviluppatori che ancora oggi sono per la maggior parte uomini.