

MIT Technology Review 6/2025 p. 80

Rapport

Intelligence artificielle



L'expérience de l'aide sociale basée sur l'IA

Amsterdam voulait faire examiner les demandes d'aide sociale par une IA. Le projet n'a pas abouti. Il montre à quel point il est difficile de développer des algorithmes équitables.

Justin-Casimir Braun, Gabriel Geiger et Eileen Guo (traduction : Jo Schilling)

Le professeur de sport Hans de Zwart est un défenseur des droits numériques. Il est secrétaire général de Bits of Freedom, la principale ONG pour les droits numériques aux Pays-Bas, et raconte qu'il est presque tombé de sa chaise lorsqu'il a entendu parler du projet d'Amsterdam d'évaluer toutes les demandes d'aide sociale à l'aide d'un algorithme.

à l'aide d'un algorithme.

C'était en février 2023 et de Zwart travaillait depuis près de deux ans comme consultant informel pour la municipalité d'Amsterdam. Selon les documents de la ville, ce modèle d'IA spécial, appelé « Smart Check », était destiné à examiner les demandes des bénéficiaires potentiels de l'aide sociale et à identifier ceux qui avaient peut-être commis une erreur dans leur demande. « Il existe des problèmes très fondamentaux et insolubles » lorsqu'on applique cet algorithme « à de vraies personnes », explique-t-il.

Paul de Koning, consultant pour la ville, a déjà travaillé pour différentes autorités de l'État providence néerlandais et considérait ce système avec fierté. Il avait dirigé la phase pilote de Smart Check et était enthousiasmé par le potentiel du projet. Celui-ci devait rendre les procédures plus efficaces et éliminer les préjugés du système d'aide sociale d'Amsterdam.

« Il existe des problèmes très fondamentaux et insolubles lorsque cet algorithme est appliqué à de vraies personnes. »

Une équipe avait travaillé pendant des années sur Smart Check. La ville avait consulté des experts, effectué des tests pour vérifier les biais du système, pris des mesures de sécurité techniques et recueilli les commentaires des personnes qui seraient concernées par le programme, suivant ainsi plus ou moins toutes les recommandations du guide éthique pour l'IA. « J'avais un bon pressentiment », déclare M. de Koning.

De Zwart et de Koning incarnent, avec leurs opinions respectives, le débat mondial sur l'équité des algorithmes. Pour de Koning, Smart Check est une avancée technologique vers un système social plus équitable et plus transparent. Pour de Zwart, cela représente une menace considérable pour les droits des bénéficiaires de l'aide sociale.



Hans de Zwart milite avec Bits of Freedom pour les droits numériques aux Pays-Bas.

Photo : Hans de Zwart

Les avantages devraient être des services publics plus efficaces. Dans les systèmes sociaux en particulier, ils devraient permettre de récupérer des fonds qui échappent aux pouvoirs publics. Dans la pratique, nombre de ces systèmes ont été mal conçus dès le départ, ce qui a donné lieu à une série de scandales qui dure depuis plus de dix ans : aux États-Unis, des candidats non blancs ont été écartés des bases de données de candidatures, au Japon, des familles ont été injustement signalées dans le cadre d'enquêtes sur des cas de maltraitance infantile et, en Inde, des habitants à faibles revenus se sont vu refuser des aides alimentaires

En réponse, les législateurs, les bureaucrates et le secteur privé, d'Amsterdam à New York, tentent de développer des systèmes algorithmiques qui intègrent les principes d'une « IA responsable », une approche qui vise à orienter le développement de l'IA vers le bien de la société tout en minimisant les conséquences négatives.

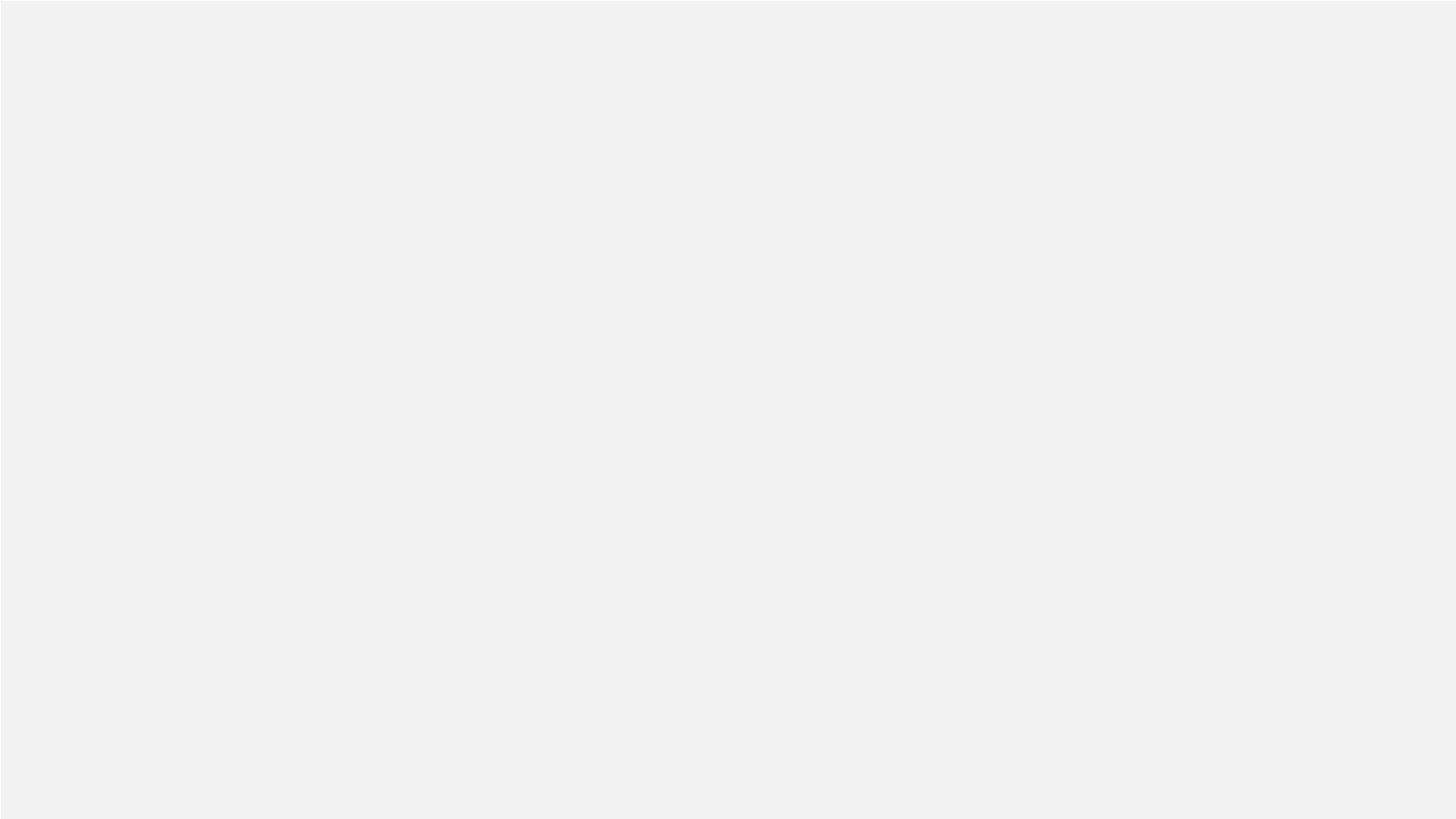
Pour l'Union européenne, le développement et l'introduction d'une IA éthique sont une priorité absolue, et Amsterdam pensait être sur la bonne voie. Les fonctionnaires des services sociaux étaient convaincus qu'ils pouvaient développer une technologie capable de prévenir la fraude tout en protégeant les droits des citoyens. Ils ont suivi les derniers exemples de bonnes pratiques et ont investi beaucoup de temps et d'argent dans un projet qui a finalement traité de réelles demandes d'aide sociale. Cependant, dans le cadre de leur projet pilote, ils ont constaté que le système qu'ils avaient développé n'était ni équitable ni efficace. Pourquoi ?

ERREUR DANS LE SYSTÈME

Lighthouse Reports, MIT Technology Review et le journal néerlandais *Trouw* ont eu accès au système afin de déterminer si les systèmes algorithmiques pouvaient tenir leurs promesses ambitieuses dans les meilleures conditions possibles. La ville a divulgué plusieurs versions de l'algorithme Smart Check ainsi que des données sur l'évaluation des demandes d'aide sociale réelles. Mais la réponse à cette question est loin d'être simple.

Il faut d'abord comprendre comment Amsterdam en est venue à vouloir lutter contre la fraude au système social à l'aide de l'IA. L'histoire remonte à quatre décennies : Albine Grumböck, mère célibataire divorcée de trois enfants, percevait depuis plusieurs années déjà des allocations sociales. En 1984, elle a appris qu'un de ses voisins, employé au service social local, surveillait secrètement sa vie. Il avait documenté les visites d'un ami masculin qui aurait théoriquement pu apporter des revenus non déclarés à la famille.

Ces observations ont conduit l'administration à supprimer l'aide sociale d'Albine Grumböck. Elle a contesté cette décision devant les tribunaux et a obtenu gain de cause.



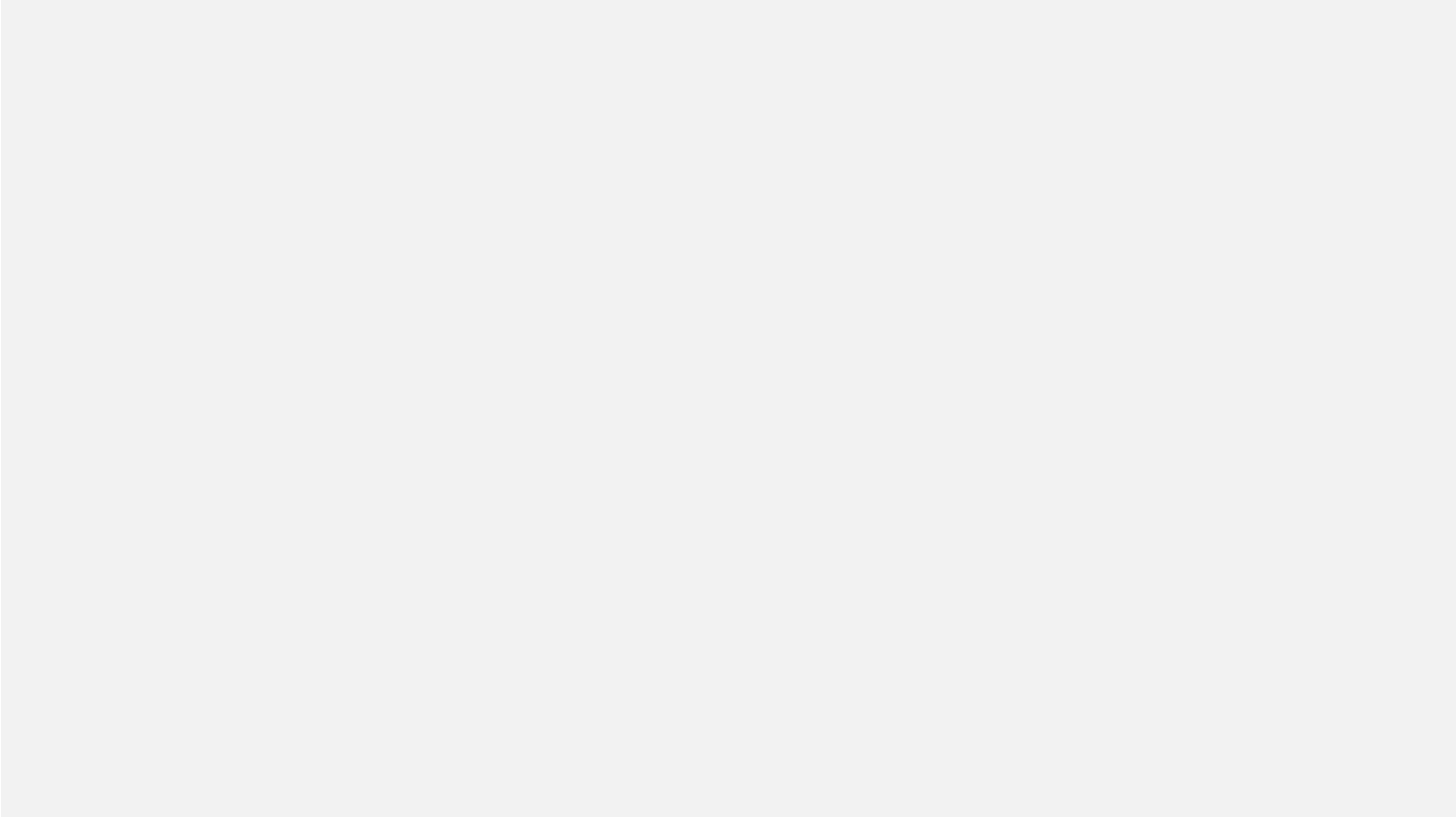
Albine Grumböck, dont les prestations sociales avaient été supprimées dans les années 1980, apprend que le tribunal lui a accordé une aide provisoire.

Photo : Rob Bogaerts / Nationaal Archief

Malgré ce scandale, les services sociaux néerlandais ont continué à employer des enquêteurs – appelés « compteurs de brosses à dents » – et les ont autorisés à bouleverser la vie des gens. Cela a contribué à créer un climat de méfiance, explique Marc van Hoof, avocat qui aide depuis des décennies les bénéficiaires de l'aide sociale néerlandaise à s'y retrouver dans le système : « Le gouvernement ne fait pas confiance à ses citoyens, et les citoyens ne font pas confiance au gouvernement. »

Harry Bodaar, fonctionnaire de longue date, a observé la politique sociale des Pays-Bas pendant cette période, d'abord en tant que travailleur social, puis en tant qu'enquêteur chargé des fraudes et aujourd'hui en tant que conseiller en politique sociale de la ville. Les 30 dernières années lui ont montré que « le système tient grâce à des élastiques et des agrafes », dit-il. « Et quand on se trouve au bas de l'échelle, on est le premier à passer à travers les mailles du filet. »

Un meilleur système pour les bénéficiaires a été un facteur de motivation important pour le développement de Smart Check en 201U, ajoute-t-il. « Nous voulions uniquement contrôler de manière équitable les personnes qui nous semblaient devoir l'être », explique M. Bo-daar, contrairement à la politique menée jusqu'en 2007 par l'administration, qui consistait à effectuer des visites à domicile chez chaque demandeur.



Si les demandeurs résident par exemple dans le quartier Kolenkitbuurt, connu pour sa forte proportion d'habitants issus de l'immigration, cela entraîne déjà des préjugés, tant chez les personnes que chez les algorithmes.

Photo : mauritius images / Peter Ptschelinzew / Alamy

Cependant, les tentatives du gouvernement néerlandais pour détecter les fraudes à l'aide de l'IA se sont parfois retournées contre lui : en 2019, il a été révélé que le gouvernement avait utilisé un algorithme pour établir des profils de risque liés aux allocations familiales. Le scandale qui a suivi a révélé que près de 35 000 parents, principalement des migrants ou des enfants de migrants, avaient été accusés à tort de fraude pendant six ans. En conséquence, l'ensemble du gouvernement a finalement dû démissionner en 2021.

À Rotterdam, Lighthouse Reports 2023 a enquêté sur un système de détection de la fraude sociale qui était biaisé à l'égard des femmes, des parents, des personnes dont la langue maternelle n'était pas l'allemand et d'autres groupes vulnérables. Ce système a dû être abandonné. D'autres villes, comme Amsterdam et Leyde, utilisaient un système appelé « Fraud Scorecard » (fiche de fraude), introduit il y a plus de 20 ans, qui se basait sur le niveau d'éducation, le quartier, la parentalité et le sexe comme facteurs de risque approximatifs pour évaluer les demandeurs d'aide sociale ; ce programme a également été abandonné.

Mais les Pays-Bas ne sont pas un cas isolé. Aux États-Unis, selon le Benefits Tech Advocacy Hub, il y a eu au moins onze cas où des États ont utilisé des algorithmes pour attribuer des prestations sociales, souvent avec des résultats inquiétants. Et en France

, des militants poursuivent l'organisme national de protection sociale en justice pour un algorithme qui, selon eux, discrimine les demandeurs à faibles revenus et les personnes handicapées.

UNE IA RESPONSABLE

Ces scandales, ainsi qu'une prise de conscience croissante de la discrimination raciale dans les systèmes algorithmiques, ont contribué à mettre l'accent sur une IA responsable. Selon Jiahao Chen, consultant en IA éthique, « c'est devenu un terme générique qui implique que nous devons réfléchir non seulement à l'éthique, mais aussi à l'équité ».

« Nous prenons de plus en plus conscience que nous avons besoin de transparence, de protection des données, de sécurité et de sûreté. »

Cette approche repose sur une formule bien connue : livres blancs, cadres élaborés par des groupes de réflexion et des organismes internationaux, et secteur lucratif du conseil. Ainsi, en 2019, l'Organisation de coopération et de développement économiques a publié ses *principes pour l'intelligence artificielle*, qui servent de lignes directrices pour le développement d'une « IA digne de confiance » : des systèmes explicables, la consultation des groupes d'intérêt publics et la réalisation d'audits.

Mais l'héritage de décennies d'erreurs algorithmiques est difficile à effacer. Alors que les Pays-Bas s'efforcent d'introduire des réformes basées sur une IA responsable à l'échelle nationale, l'ONG néerlandaise Algorithm Audit est arrivée à la conclusion que cette technologie ne devrait être utilisée que dans des conditions strictement définies. Dans le même temps, Amnesty International, des militants pour les droits numériques tels que de Zwart et certains bénéficiaires de l'aide sociale affirment que le secteur public ne devrait pas utiliser l'IA pour prendre des décisions qui affectent la vie des gens.

Amsterdam espérait avoir trouvé le juste équilibre. « Nous avons tiré les leçons du passé », explique M. Bodaar. Et cette fois-ci, la ville voulait mettre en place un système qui « montre aux habitants d'Amsterdam que nous agissons dans leur intérêt et que nous sommes équitables ».

Lorsqu'un habitant d'Amsterdam demande des prestations sociales, des agents chargés du dossier vérifient si la demande présente des irrégularités. Si elle semble suspecte, elle est transmise au service d'enquête de la ville. Cela peut entraîner un refus, une demande de rectification ou une réduction des prestations, même rétroactive.

Les fonctionnaires disposent de pouvoirs étendus : ils peuvent demander des documents bancaires, convoquer les bénéficiaires de prestations à la mairie et effectuer des visites à domicile sans préavis. Pendant la durée de l'enquête, les paiements urgents peuvent être retardés. Et souvent, selon Bodaar, dans plus de la moitié des cas suspects, la

ville ne trouve aucune preuve de comportement fautif. Dans ces cas-là, la ville a « injustement harcelé des gens », estime M. Bodaar.

Le système Smart Check devait remplacer les agents chargés de transmettre les dossiers au service d'enquête. L'algorithme utilise des caractéristiques personnelles pour identifier les demandes les plus susceptibles de contenir des erreurs importantes.

Si tout se passe bien, écrit la ville dans sa documentation interne, le système signalerait moins de demandeurs d'aide sociale pour enquête et identifierait en même temps plus de cas contenant des erreurs. Dans un document, la ville prévoit que le modèle permettrait d'épargner à 125 Amstellodamois une procédure et d'économiser 2,4 millions d'euros par an.

De Koning, qui devait diriger le projet après son lancement, était optimiste, car la ville avait adopté une approche scientifique. L'objectif était de « voir si cela fonctionne » plutôt que de partir du principe que « cela doit fonctionner et que nous allons poursuivre dans cette voie quoi qu'il arrive ».

Cette approche a attiré des techniciens tels que Loek Berkers. Dans un café situé derrière l'hôtel de ville d'Amsterdam, Berkers se souvient à quel point il a été impressionné lors de son premier contact avec le système : « C'était très innovant, surtout pour un projet au sein de l'administration municipale », dit-il.

Smart Check utilisait un algorithme appelé « Explainable Boosting Machine », qui vise à aider les gens à comprendre comment les modèles d'IA font leurs prédictions. Le modèle prenait en compte 15 caractéristiques, notamment si les demandeurs avaient déjà sollicité ou reçu des prestations sociales, le montant de leur patrimoine et le nombre d'adresses déclarées. Le modèle attribuait ensuite à chaque personne une note de risque. Les facteurs démographiques susceptibles d'entraîner des distorsions, tels que le sexe, la nationalité ou l'âge, ont été délibérément omis. Le système évitait également les facteurs « proxy », tels que les codes postaux, qui, bien qu'ils ne semblent pas sensibles à première vue, peuvent être statistiquement associés à un groupe ethnique particulier.

LES RÉSERVES

Le modèle a été entraîné à partir d'un ensemble de données provenant de 3 400 demandes antérieures. Cependant, l'utilisation d'anciennes enquêtes pose d'emblée des problèmes potentiels, selon Sennay Ghebreab, directeur scientifique du Civic AI Lab (CAIL) de l'université d'Amsterdam. Le problème lié à l'utilisation de données historiques pour créer des modèles est que « nous nous retrouvons finalement avec des biais historiques ». Si, par exemple, les travailleurs sociaux ont commis davantage d'erreurs dans le passé avec un certain groupe ethnique, le modèle pourrait en déduire à tort que ce groupe ethnique commet plus souvent des fraudes.



L'équipe de rédacteurs a obtenu de la ville d'Amsterdam l'accès aux modèles Smart Check et a vérifié leur partialité à l'aide du personnage fictif de Susan.

Graphique : Technology Review

La ville a décidé de revoir son système en profondeur. Cependant, la définition des préjugés et la signification de l'équité pour un algorithme font l'objet de vifs débats. Les autorités d'Amsterdam ont choisi une définition de l'équité qui met l'accent sur la répartition équitable de la charge que représentent les enquêtes *illégitimes* entre les différents groupes de population. En d'autres termes, elles espéraient que cette approche garantirait que les demandeurs d'aide sociale issus de milieux différents subiraient dans une même mesure la charge que représentent les enquêtes injustifiées.

Lors du développement de Smart Check, Amsterdam a notamment consulté le délégué interne à la protection des données de la ville et la commission de protection des données d'Amsterdam, ainsi que des organisations privées telles que la société de conseil Deloitte. Tous ont donné leur accord au projet.

toute l'équipe a reçu le feu vert. Un groupe important n'était toutefois pas impliqué : un comité consultatif de 15 personnes composé de bénéficiaires de l'aide sociale, de représentants d'intérêts et d'autres acteurs non gouvernementaux. Tout comme le militant pour les droits numériques de Zwart, le comité était profondément préoccupé par les conséquences que le système pourrait avoir pour les personnes en situation de précarité.

« Nous nous sommes méfiés du système dès le début. »

La municipalité a présenté son projet pour la première fois au conseil participatif à l'automne 2021. Anke van der Vliet, aujourd'hui âgée de plus de 70 ans, est membre de longue date du conseil. « Nous nous sommes méfiés du système dès le début », explique cette femme qui s'engage sans relâche depuis des décennies en faveur des bénéficiaires de l'aide sociale de la ville. « Nous voulions savoir si cela était avantageux ou désavantageux pour nous. »

Deux autres réunions n'ont pas non plus réussi à les convaincre. Leur rétroaction a certes conduit à des modifications importantes, telles que la réduction du nombre de variables initialement prises en compte par la ville pour calculer le score d'un demandeur. Ou encore la suppression de l'âge des personnes. Néanmoins, après six mois, le conseil de participation a complètement cessé sa collaboration avec la ville. La justification donnée en mars 2022 : « Le conseil estime qu'une telle expérience porte atteinte aux droits fondamentaux des citoyens et doit être abandonnée. » Étant donné que seulement environ 3 % des demandes d'aide sociale sont frauduleuses, l'utilisation de l'algorithme est « disproportionnée », poursuit le communiqué. Non seulement cette technologie peut commettre des erreurs, mais celles-ci sont encore plus difficiles à corriger, et elle permet aux fonctionnaires de « se cacher derrière des murs numériques », explique Henk Kroon, de l'association caritative d'Amsterdam, un syndicat fondé dans les années 1970.

IL EST TEMPS DE TESTER

Malgré ces objections, la ville a décidé de tester le modèle Smart Check dans la pratique, et dès les premières utilisations en mai 2022, les faiblesses de l'algorithme en matière de biais sont apparues : la première version du modèle avait par exemple tendance à marquer à tort les demandeurs non néerlandais. Et les personnes de nationalité non occidentale avaient presque deux fois plus de chances d'être signalées. Le modèle envoyait également les hommes vers une enquête avec une probabilité 14 % plus élevée.

Dans le cadre de la formation modèle, la ville a également collecté des données sur les personnes que ses agents administratifs avaient sélectionnées pour un contrôle : contrairement à Smart Check, ils avaient tendance à classer les Néerlandais et les femmes dans la mauvaise catégorie.

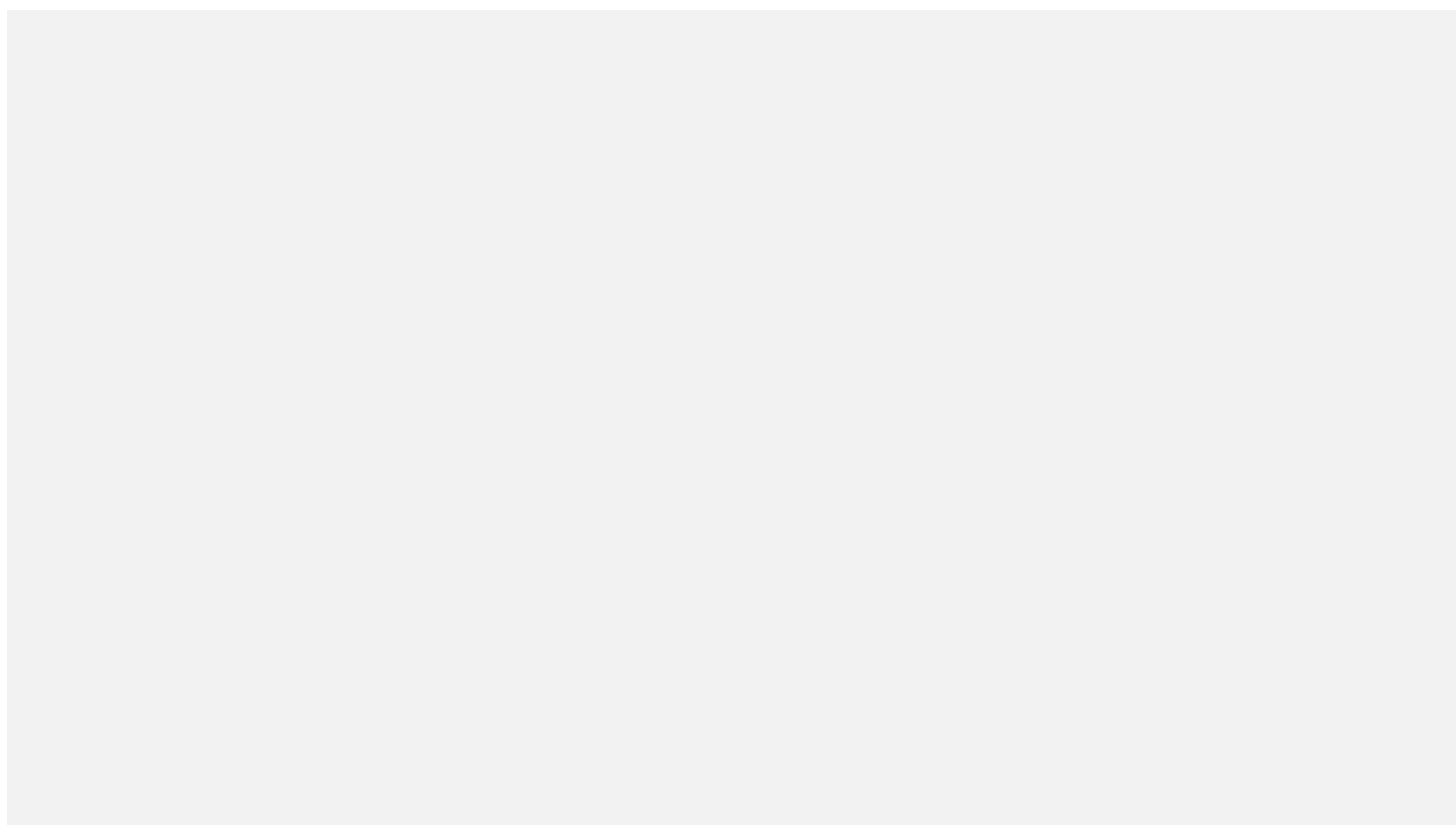


Photo : Chantal Jahchan

Afin de corriger ce biais dans Smart Check, l'équipe a procédé à une « repondération des données d'entraînement » (Training-Data Reweighting) : les demandes non occidentales comportant des erreurs importantes ont reçu moins de poids dans les données, tandis que les demandeurs de nationalité occidentale ont reçu plus de poids. Après la pondération du modèle, les ressortissants néerlandais et non néerlandais ont été marqués à tort avec la même fréquence. De Koning estime que c'est une bonne chose : « Comme c'était équitable, nous avons pu poursuivre le processus. »

Comme le modèle semblait également être 20 % plus performant pour identifier les demandes nécessitant un contrôle

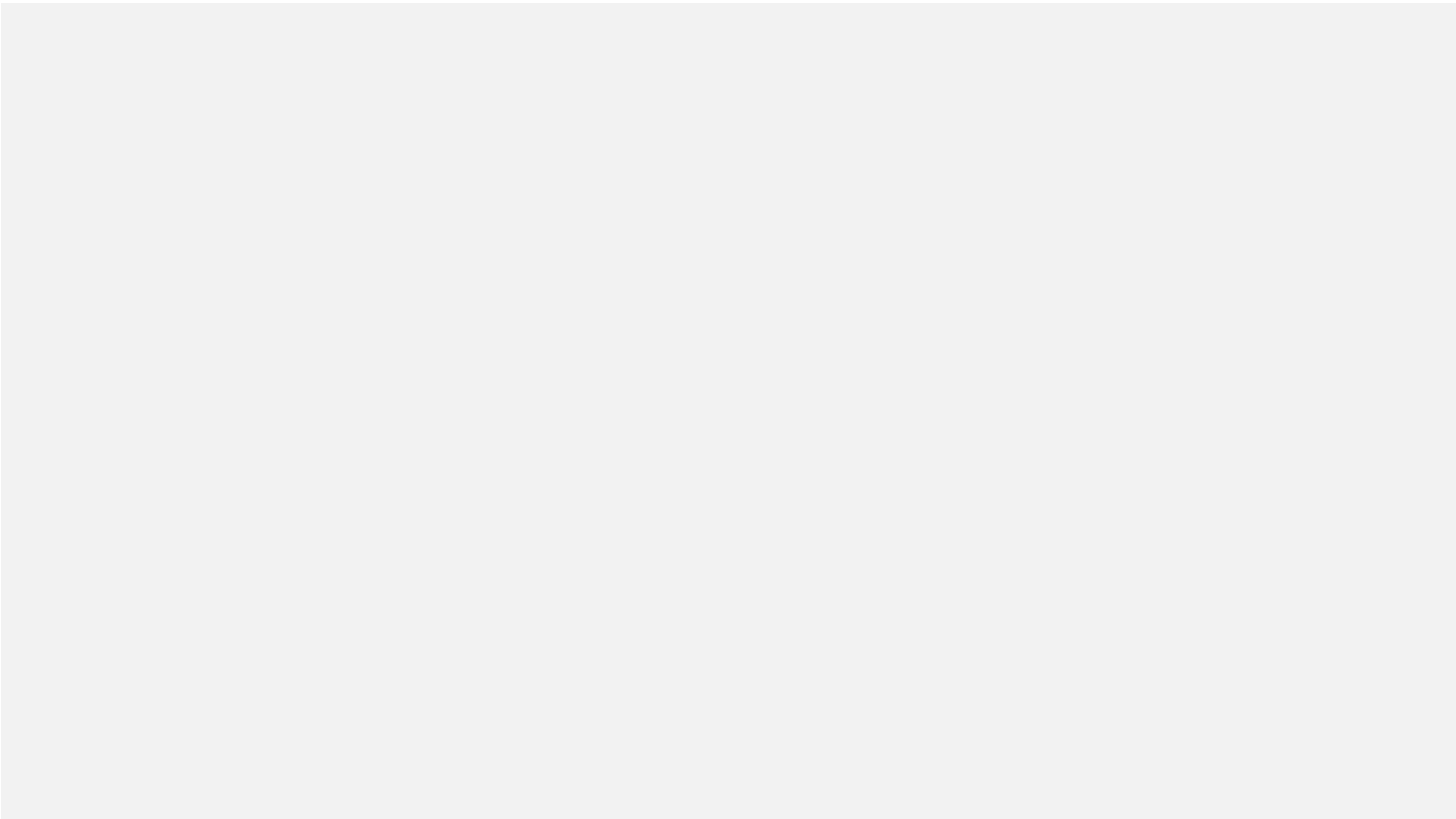
que les humains, la ville a enregistré Smart Check au printemps 2023 auprès de l'« Algo-rhythm Register », une initiative de transparence gérée par le gouvernement qui informe les citoyens sur les algorithmes d'apprentissage automatique utilisés par l'État.

Malgré toutes ces améliorations, M. De Zwart reste sceptique. Selon lui, la ville a utilisé tous les outils de la boîte à outils « Responsable AI ». « Ils ont fait tout ce que recommande le monde de l'IA responsable. » Mais en fin de compte, la question est de savoir s'il est légitime d'utiliser des données sur le comportement passé pour « évaluer le comportement futur de ses citoyens ».

En mars 2023, la ville a lancé le projet pilote avec 1600 demandes d'aide sociale réelles. Chaque demande a été examinée simultanément par une personne. Deux critères clés ont été surveillés : le système était-il capable d'évaluer les demandes sans préjugés ? Et Smart Check était-il réellement plus efficace et plus équitable que les humains pour détecter la fraude sociale ?

Le modèle a échoué sur ces deux points : le nombre de contrôles des demandes a augmenté au lieu de diminuer. Et il n'était pas plus efficace qu'un être humain pour identifier les personnes qui nécessitaient réellement un contrôle supplémentaire. De plus, le projet pilote en conditions réelles a révélé de nouvelles distorsions. Cette fois-ci, ce sont les citoyens néerlandais et les femmes qui ont été incorrectement signalés, et non plus les ressortissants étrangers et les hommes, ainsi que les demandeurs d'aide sociale avec enfants.

La ville était dans l'impasse et, fin novembre 2023, Rutger Groot Wassink, conseiller municipal chargé des affaires sociales, a pris place dans son fauteuil à la mairie d'Amsterdam, a jeté un coup d'œil à sa tablette et s'est adressé à ses collègues du conseil : « J'ai décidé de mettre fin au projet pilote. » Sa justification : « J'aurais trouvé très difficile de justifier un projet pilote qui a montré que l'algorithme présentait d'énormes distorsions. » La ville avait ainsi testé une approche innovante pour minimiser les risques en matière de détection des fraudes, constaté qu'elle ne tenait pas ses promesses et l'avait rejetée à temps.



Rutger Groot Wassink était conseiller municipal chargé des affaires sociales à Amsterdam lorsque Smart Check était en phase pilote. C'est lui qui a mis fin au projet.

Photo : Lumen Photo / VISUM

En mars, un an et demi après la fin du projet pilote, M. Bodaar a déclaré qu'il n'était peut-être tout simplement pas encore temps d'utiliser l'intelligence artificielle à cette fin. « Niente, zero, nada. Nous ne le ferons plus », a-t-il déclaré à propos de l'utilisation de l'IA pour évaluer les demandeurs d'aide sociale. « Mais nous continuons à y réfléchir : qu'avons-nous appris exactement ? »

C'est une question à laquelle la conseillère municipale IJmker réfléchit également. Lors des réunions du conseil municipal, elle a cité Smart Check comme exemple de ce qu'il *ne* faut pas faire. Elle s'est certes réjouie que les employés municipaux aient soigneusement rédigé leurs « nombreux procès-verbaux », mais elle craignait que le processus ne relègue certaines des questions les plus importantes au rang de « valeurs philosophiques » et « politiques ».

».

« Niente, zero, nada. Nous ne le ferons plus. Mais nous réfléchissons à la question suivante : qu'avons-nous appris exactement ? »

Des questions telles que « Quelle est notre position sur le profilage ? », « Qu'est-ce qui est justifié selon nous ? » ou même « Qu'est-ce que le préjugé ? ». Selon elle, ces questions « ne peuvent pas être simplement cochées dans une case ».

Maintenant que le projet pilote a été suspendu, elle craint que l'administration municipale ne veuille aller trop vite. « Je pense que beaucoup se disent : c'est fini, au revoir, fin de l'histoire », dit-elle. C'est « du gâchis », ajoute-t-elle, « car les gens y ont travaillé pendant des années ».

En abandonnant ce modèle, la ville revient à une procédure analogique qui, selon sa propre analyse, désavantage les femmes et les ressortissants néerlandais. Jiahao Chen, consultant en IA éthique, pose la question suivante : « Pourquoi plaçons-nous les systèmes d'IA à un niveau supérieur à celui des acteurs humains ? » Il n'y a aucune tentative de corriger systématiquement les préjugés des agents administratifs humains. Amsterdam a promis de rédiger un rapport sur les préjugés humains dans le processus d'aide sociale, mais la date a été reportée à plusieurs reprises.

« Dans la pratique, l'éthique revient à accepter que rien n'est parfait », explique Chen. « Il existe un principe fondamental, à savoir *la non-discrimination*, auquel nous pouvons tous adhérer, mais cet exemple montre certaines des complexités liées à la mise en œuvre de ce principe. »

Il est peut-être temps de repenser plus fondamentalement la manière dont l'équité doit être définie – et par qui. Au-delà des définitions mathématiques, certains chercheurs soutiennent que les personnes les plus touchées par les programmes en question devraient avoir davantage leur mot à dire. « De tels systèmes ne fonctionnent que si les gens les acceptent », explique Elissa Redmiles, professeure adjointe d'informatique à l'université de Georgetown, qui s'est penchée sur l'équité algorithmique.

Quelle que soit la forme que prendra ce processus, ce sont là des questions auxquelles tout gouvernement devra se pencher de toute urgence dans un avenir de plus en plus marqué par l'IA.

Et comme le fait valoir M. de Zwart, même les fonctionnaires bien intentionnés qui mettent en place des systèmes tels que Smart Check dans des villes comme Amsterdam seront condamnés à réapprendre sans cesse les mêmes leçons – ou à les ignorer – si ces questions fondamentales ne sont pas abordées.

« Nous nous laissons séduire par des solutions technologiques qui ne répondent pas aux vrais problèmes », dit-il. « Est-ce vraiment ce que nous voulons ? Pourquoi la municipalité ne développe-t-elle pas un algorithme qui recherche les personnes qui ne demandent pas d'aide sociale alors qu'elles y ont droit ? »

Cet article est le fruit d'une collaboration entre MIT Technology Review, Lighthouse Reports et Trouw, avec le soutien du Pulitzer Center.



Écrire une lettre à la rédaction



Télécharger l'article au format PDF



Partager sur Facebook



Partager sur X

Contact

Mentions légales

Politique de confidentialité

Conditions d'utilisation

Données médias

Résilier des contrats