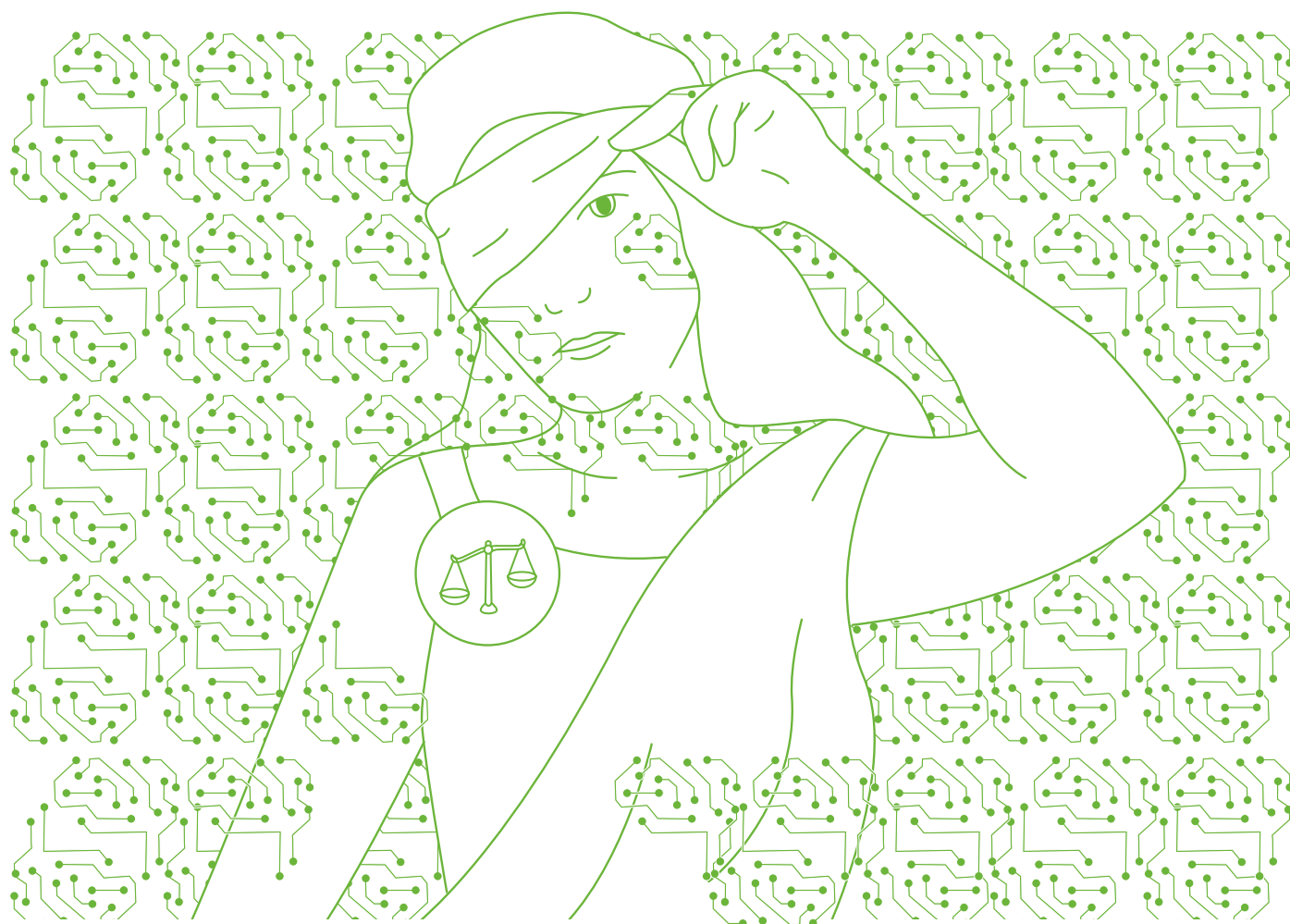


L'intelligence artificielle n'est pas plus discriminante que les données qu'on lui fournit

Entretien avec **Henrike Schlottmann**
Jan Kuhlen



L'utilisation de l'intelligence artificielle (IA) se répand dans toutes les branches pour la prise de décisions basées sur les algorithmes et généralise le risque de discrimination de personnes et de groupes. Henrike Schlottmann et Jan Kuhlen, du «Zentrum für digitalen Fortschritt» (centre du progrès numérique) nous expliquent ce qui peut être entrepris pour réduire les inégalités de traitement. Leur estimation s'articule autour de la thèse selon laquelle l'intelligence artificielle ne fait que refléter les inégalités sociales et la systématisation des erreurs facilite la correction des discriminations – dans la vie numérique et la vie analogique.

Pourquoi l'intelligence artificielle est-elle discriminante?

Jan Kuhlen: La notion de discrimination désigne l'égalité de traitement de choses ou personnes inégales ou l'inégalité de traitement de choses ou personnes égales – sans justification. La technologie n'est pas discriminante en soi, sauf si l'humain en tire des conclusions discriminantes ou la forme de manière à ce que les résultats entraînent automatiquement une discrimination. Il en va de même pour l'intelligence artificielle. Son utilisation peut entraîner une discrimination involontaire, en raison du manque de connaissance sur la manière avec laquelle un réseau neuronal artificiel en arrive à ses résultats. On peut constater un manque de prise de conscience des critères déterminants pour une évaluation des défis liés à l'intelligence artificielle.

Quelles formes de discrimination ont été identifiées lors de l'utilisation de l'intelligence artificielle?

Henrike Schlottmann: La discrimination peut avoir lieu sur quatre niveaux: elle dépend des données de formation, des algorithmes choisis, de l'équipe et des objectifs. C'est l'humain qui fixe des objectifs discriminants ou qui emploie des équipes contenant déjà des points aveugles. Dans le domaine de l'intelligence artificielle, la discrimination s'appuie en général sur les attributs protégés, comme le sexe, la couleur de la peau ou la religion. C'est en principe la même problématique que dans la vie réelle, puisque l'algorithme alimenté avec des données ne fait que refléter la déformation, également appelée biais. Celle-ci existe déjà dans la société.

Pouvez-vous nous donner un exemple?

Henrike Schlottmann: Nous avons l'exemple bien connu de l'algorithme de recrutement chez Amazon¹, qui avait des effets discriminatoires dans la procédure de candidature: il en arrivait à la conclusion que les candidatures de femmes devaient être évaluées de manière moins avantageuse. L'algorithme avait en effet été entraîné sur la base d'un ensemble de données déséquilibré - c'est-à-dire à partir de candidatures reçues par Amazon au cours des dix dernières années, dont la plupart étaient des candidatures d'hommes. Un modèle dont le système a conclu que les hommes étaient des candidats privilégiés. Autre exemple: le «Predictive Policing» – ou prévision policière, qui consiste à appliquer des décisions automatisées dans les poursuites pénales. C'est ainsi que des prévisions des taux de récidives ont été faites aux USA à partir d'un biais inhérent dans les données de formation – et que les non-blancs étaient systématiquement discriminés sur la base de leur couleur de peau².

Entendez-vous aussi par «discrimination rationnelle» l'optimisation de l'individu au détriment de la solidarité sociale, comme le fait le sociologue Stefan Selke?

Jan Kuhlen: Le «Life-Logging», c'est-à-dire l'enregistrement et l'archivage de grandes quan-

1: L'intelligence artificielle discrimine (encore): <https://www.zeit.de/arbeit/2018-10/bewerbungsroboter-kuenstliche-intelligenz-amazon-frauen-diskriminierung>

2: Predictive Policing: Sur les traces du crime du futur: <https://www.bpb.de/dialog/netzdebatte/238995/predictive-policing-dem-verbretchen-der-zukunft-auf-der-spur?type=galerie&show=image&i=238997>

tités de données de la vie quotidienne d'un individu, généralise la mesure et l'évaluation de l'humain. Cela pourrait également avoir pour conséquence une généralisation de la mesure de la société. Ce phénomène permet une comparaison à grande échelle entre les personnes. Cette mesure nous prive d'une part de vie privée, et c'est là un argument critique.

Henrike Schlottmann: Le «rationnel» implique toujours aussi une part de systématique. En matière de discrimination, cela signifie que le biais existe déjà dans la société et est reflété par l'algorithme. Mais c'est l'application qui systématise le tout. Ce n'est donc plus l'humain qui décide en fonction de la situation, mais le système qui agit de manière systématique. En même temps, la discrimination est plus visible et plus facile à corriger.

Existe-t-il des applications qui permettent une discrimination particulièrement rationnelle?

Jan Kuhlen: Le risque de discrimination rationnelle est particulièrement important avec les données de santé collectées par des applis. Mais rien ne permet d'attester le caractère systématique pour la discrimination par l'intelligence artificielle. Dans ce domaine, le risque de discrimination existe principalement quand il s'agit d'évaluer des personnes – notamment avec les attributs protégés.

Quel est le lien entre le «Life-Logging» et l'utilisation de l'intelligence artificielle?

Henrike Schlottmann: Le «Life-Logging» ou Big Data ne constitue pas à lui seul une intelligence artificielle, mais il est une condition indispensable aux analyses intelligentes.

Jan Kuhlen: L'intelligence artificielle, quand on l'utilise ainsi, peut faire à partir de ces données des déductions et des généralisations injustifiées. Par exemple: une personne qui se lève à sept heures et qui court trois kilomètres tous les matins est plus performante.

Existe-t-il aussi des discriminations rationnelles dans la vie réelle?

Jan Kuhlen: Oui. Quand l'assurance m'envoie un questionnaire, elle évalue le risque à partir des données que je lui fournis. Avec le Big Data, les évaluations sont plus fines et peuvent être systématisées plus facilement à l'aide de l'intelligence artificielle, ce qui permet d'obtenir de meilleurs résultats: elles permettent de prendre en compte la situation de risque personnelle. L'intelligence artificielle se répand massivement sur tout le système. Et nous avons jusqu'à présent trop peu pris conscience de la discrimination rationnelle.

Dans un papier de positionnement³, vous demandez une réglementation contraignante pour l'utilisation de l'IA afin d'éviter les abus.

Henrike Schlottmann: Nous demandons un conseil de déontologie qui garantisse le respect par tous de certaines règles et directives.

Comment cela permettrait-il de lutter contre les discriminations?

Jan Kuhlen: Il faut s'attaquer aux discriminations sur les quatre niveaux que nous avons cités: l'équipe doit être sélectionnée de manière à éviter les points aveugles, les données doivent être réparties en données d'apprentissage et en données d'inspection et, en cas de doute, contrôlées à l'aide d'analyses boîte noire – une méthode indirecte qui produit par exemple des modèles explicatifs a posteriori grâce à la visualisation. Même si la mécanique d'une boîte noire algorithmique, donc d'un système fermé, est inconnue, on peut en examiner le comportement, c'est-à-dire les entrées et sorties.

Henrike Schlottmann: Les systèmes doivent être aisément compréhensibles et explicables: on doit pouvoir comprendre comment les décisions sont prises. Les algorithmes peuvent par exemple faire l'objet d'un audit; les données seraient alors examinées pour déceler un

3: L'influence de l'intelligence artificielle sur la liberté, la justice et la solidarité: <https://d-64.org/wp-content/uploads/2018/11/D64-Grundwerte-KI.pdf>

biais. Les instruments nécessaires dépendent cependant toujours de la définition personnelle de l'équité et sont encore loin d'être optimaux.

Faut-il toujours argumenter avec des causalités pour justifier des décisions de manière plausible?

Jan Kuhlen: Pour les décisions officielles auxquelles nous ne pouvons échapper, nous sommes en droit d'obtenir des argumentations transparentes et compréhensibles. Ces décisions ne devraient pas être prises sur la base d'une probabilité. L'intelligence artificielle se base pourtant toujours sur des probabilités. La prudence est donc de mise avec l'IA. Nous ne pouvons cependant pas attendre d'une technologie qu'une entrée produise toujours une sortie totalement logique. On ne le demande pas non plus aux êtres vivants – même pas aux humains. La société va devoir s'habituer au fait que la technologie ne fournit pas toujours des résultats 100% logiques en laboratoire, mais qu'elle est quand même globalement beaucoup plus précise que l'humain.

Quels domaines d'utilisation de l'IA sont légitimes?

Jan Kuhlen: Il existe déjà de nombreuses technologies basées sur des composants d'apprentissage automatisés et qui fonctionnent bien: les traductions, les analyses vidéo ou la reconnaissance vocale. Mais la traçabilité, c'est-à-dire la manière dont un algorithme obtient de bons résultats de traduction, n'intéresse que peu de monde. Ce qui compte, c'est que la phrase soit bien traduite. De même, la plupart des utilisateurs de la reconnaissance vocale – par exemple les saisies vocales converties en texte – ne se préoccupent pas de savoir pourquoi tel mot est reconnu.

Mais pourquoi ne peut-on pas argumenter que les technologies qui ne sont pas logiques devraient être complètement interdites du fait qu'elles peuvent toujours être discriminatoires de manière injustifiée?

Jan Kuhlen: Si on appliquait cette question aux humains, on devrait également remettre en question les décisions des humains, car elles peuvent toujours être discriminantes de manière injustifiée. L'intelligence artificielle recèle d'énormes potentiels. Mais comme nous l'avons dit, dans certains domaines, il faut contrôler son utilisation, par exemple quand elle est utilisée pour des décisions officielles.

La mise en place d'un conseil de déontologie apparaît illusoire, étant donné que les principes normatifs ne s'appliquent pas dans tous les pays, comme l'a montré le jeu en ligne «Moral Machine» du Massachusetts Institute of Technology (MIT)⁴: les participants étaient confrontés à des situations d'accident de voitures autonomes avec différents usagers de la route et devaient choisir le dommage le plus acceptable. Résultat: les principes éthiques varient fortement selon les pays.

Jan Kuhlen: Certaines règles ou normes se sont imposées dans le monde entier. Ce n'est pas exclu dans le domaine de l'intelligence artificielle. En Allemagne par exemple, l'interdiction des armes autonomes est déjà demandée dans l'accord de coalition entre la CDU, la CSU et le SPD⁵. Il est évident que cela ne peut fonctionner que si tous les pays y adhèrent, comme pour l'arme nucléaire. De la même manière, il faudrait que tous les pays prennent conscience de la nécessité d'informer que les assistants virtuels et les avatars ne sont pas des humains, mais des machines. Il faut également définir dans quels cas les décisions doivent être explicables et compréhensibles. Bien sûr, plus on essaie d'as-

4: Moral Machine: une voiture autonome peut-elle agir de manière éthique? <https://www.dw.com/de/moral-machine-kann-ein-selbstfahrendes-auto-ethisch-handeln/a-46045294-0>

5: Accord de coalition entre la CDU, la CSU et le SPD: <https://www.bundesregierung.de/breg-de/themen/koalitionsvertrag-zwischen-cdu-csu-und-spd-195906>

socier de pays et de cultures, plus il est difficile de trouver un code de valeurs éthiques commun.

Que se passera-t-il si nous n'y parvenons pas?

Henrike Schlottmann: On peut imaginer ce que serait le monde si les USA ou la Chine fixaient leurs propres règles. Il suffit pour cela de regarder le Social Credit System. L'Empire du Milieu tente actuellement de mettre en place une technique de surveillance capable d'identifier immédiatement chaque individu dans une très grande foule. Ceux qui se comportent correctement sont récompensés. Les mauvais comportements sont sanctionnés. Une question se pose: est-ce le monde que nous voulons? Je trouve que nous devrions en profiter pour débattre certaines règles au niveau international, car l'intelligence artificielle est devenue un thème mondial. Ses applications sont disponibles tout autour du globe, et l'internet permet à tous les humains d'accéder à ce marché.

Elles favorisent en outre une utilisation solidaire des données. Qu'entendez-vous par là?

Jan Kuhlen: Il s'agit de tirer une plus-value intelligente de ces données. Cela signifie que les données non personnelles devraient être mises à disposition de la société de manière solidaire afin de permettre l'émergence d'innovations sociales et axées sur le bien commun. Les généralisations de l'intelligence artificielle peuvent ainsi bénéficier au service public. De nombreuses idées échouent aujourd'hui en raison du manque de données ou de l'absence de préparation des données, qui prend beaucoup de temps et de ressources. Ainsi, l'économie allemande a actuellement du mal à recueillir des données de qualité suffisante. En même temps, les géants américains tels que Facebook ou Google en savent beaucoup sur le comportement des consommateurs allemands. L'économie allemande dispose aussi de données, mais elles ne sont pas partagées et sont trop spécifiques.

Que proposez-vous?

Jan Kuhlen: On pourrait par exemple créer des plateformes de données intelligentes qui entraîneraient ou échangeraient les données. Il faudrait alors des normes communes permettant aux différents acteurs de charger leurs données sur une plateforme et de les mettre à disposition des autres, gratuitement ou non.

Henrike Schlottmann: Il faut bien sûr veiller à ce que les données soient anonymes, afin d'éviter que certaines données soient associées à certaines personnes. Et qui souhaiterait que ses informations personnelles soient librement accessibles? Il n'est pas simple d'anonymiser des données, car il ne suffit pas de supprimer le nom. Il faut aussi empêcher toute identification par triangulation à partir d'autres attributs, comme la rue ou la formation.

Préconisez-vous également une utilisation solidaire des données de santé?

Jan Kuhlen: De mon point de vue, la collecte et le partage de données de santé sont problématiques, car ils touchent des connaissances hautement privées. Il s'agit ni plus ni moins que de l'état de santé d'une personne. Ce n'est pas pour rien que les médecins sont soumis au secret médical. Il est reconnu socialement que ces connaissances ne doivent pas être partagées. Il ne s'agit pas, de toute façon, de partager des données personnelles, mais des données anonymisées. Il nous paraît également important que les systèmes de données soient protégés contre les attaques, sinon nous aurions un problème avec l'auto-détermination. Tout le monde a besoin de la même protection de sa santé, qu'il fasse du jogging à sept heures du matin ou pas. Les caisses d'assurance maladie obligatoire n'ont pas le droit de faire de distinctions algorithmiques, que ce soit pour l'éventail des prestations ou les coûts.

La solidarité est un concept humain. Qu'est-ce que cela signifie pour l'évolution des systèmes d'IA?

Jan Kuhlen: La technique est toujours neutre, mais elle peut bien sûr aussi être utilisée à des fins de solidarité. Si par exemple l'IA permet de mieux gérer les réseaux d'électricité et de réduire ainsi la consommation de courant, tout le monde en profite. Toutes les utilisations axées sur le bien commun peuvent et doivent être encouragées.

Henrike Schlottmann: C'est une question de volonté politique. Une chose est claire: si mon objectif n'est pas solidaire, le résultat ne le sera pas non plus. Si j'utilise par exemple un algorithme pour déterminer l'activité physique ou le poids d'une personne et qu'en conséquence je lui propose un tarif d'assurance maladie moins avantageux, cela n'est pas solidaire. L'IA ne peut donc être solidaire que si les personnes qui utilisent l'algorithme ont aussi des objectifs solidaires.

Pour conclure: quelle peut être selon vous l'utilité sociale de l'IA?

Jan Kuhlen: Il faut résolument combattre l'image négative de l'intelligence artificielle. Sa grande prouesse est d'automatiser les processus et de réduire les travaux simples. Car cela libère de la place pour les nouvelles missions et la créativité. Nous devons être conscients que la révolution numérique est une révolution industrielle.

Henrike Schlottmann: Je souhaite revenir sur la discrimination, car l'IA a également des avantages dans ce domaine: les algorithmes mettent en évidence une certaine discrimination telle qu'elle est vécue dans la société. La systématisation permet de la combattre activement.

Henrike Schlottmann et **Jan Kuhlen** dirigent le groupe de travail Intelligence artificielle au D64 – Zentrum für digitalen Fortschritt, une association à but non lucratif consacrée à la politique numérique. Henrike Schlottmann détient un MSci de mathématiques du Unique College London. Elle a travaillé plusieurs années dans une entreprise de conseil en management international, où elle a suivi des projets dans le domaine de la numérisation et de l'innovation. Elle est actuellement Co-CEO de ProjectTogether Social Entrepreneurs. Jan Kuhlen est juriste et sociologue de formation. Avec son cabinet KUHLEN Partnerschaft von Rechtsanwälten mbB, il conseille des start-ups, des agences et des moyennes entreprises pour toutes leurs questions de droit économique. Jan Kuhlen est membre de la commission d'enquête sur l'intelligence artificielle du Bundestag allemand et du conseil d'experts Intelligence artificielle.
