

Gérer l'intelligence artificielle de manière responsable

Une partie de la société manifeste une profonde méfiance vis-à-vis de la numérisation. Il s'agit de prendre ce scepticisme au sérieux pour pouvoir exploiter les chances que présente l'intelligence artificielle pour la société. *Clemens Mader, Johann Čas, Anne Scherer*

Abrégé Dans une étude mandatée par la Fondation suisse d'évaluation des choix technologiques, une équipe interdisciplinaire de l'université de Zurich, du Laboratoire fédéral d'essai des matériaux et de recherches et de l'Académie autrichienne des sciences a examiné les champs d'application de l'intelligence artificielle (IA) du point de vue des chances et des risques. Qu'il s'agisse de publicité personnalisée sur Internet ou de (pré-)sélection de candidats dans les entreprises, nous sommes confrontés quotidiennement à l'IA, consciemment ou non, ce qui soulève des questions éthiques et juridiques. Cet article aborde deux domaines particulièrement pertinents du point de vue de la politique économique : le travail et la consommation. Les auteurs concluent qu'il faut gérer l'IA de façon responsable et intelligente, et émettent des recommandations à l'attention du monde politique, des formateurs, des entreprises et du grand public.

Nous sommes confrontés quotidiennement à l'intelligence artificielle (IA). Pourtant, nous réfléchissons peu à la manière d'en gérer les applications : nous n'en connaissons la plupart du temps pas le fonctionnement ni ne savons de quelle manière elle nous influence.

Qu'il s'agisse de recommandations concernant des produits de boutiques en ligne ou d'offres d'emploi, toutes nous paraissent à première vue insoupçonnables. Ce n'est qu'en y regardant de plus près que nous découvrons que ces recommandations apparemment pratiques cachent des algorithmes qui ne défendent probablement pas nos seuls intérêts et qui sont également susceptibles de se tromper.

Les problèmes fondamentaux des applications de l'IA sont évidents (voir *illustration*)¹.

¹ Voir à ce sujet Christen et al. (2020) ainsi que Christen (2019).

Les premières difficultés surgissent dès l'analyse et dans les données elles-mêmes. Des données inexactes ou un traitement insuffisant peuvent en effet entraîner des discriminations. C'est pourquoi on parle aussi d'un problème de biais. Certains groupes de personnes peuvent par exemple être désavantagés lors de candidatures. Les utilisateurs – ici les candidats à un emploi – ne remarquent pas ce biais, car ils ne connaissent pas le fonctionnement de l'algorithme.

Un autre problème surgit lorsqu'il s'agit d'algorithmes d'apprentissage qui évoluent constamment. Il peut alors arriver que même les développeurs de l'algorithme ne puissent plus reconstituer comment l'algorithme est arrivé à sa décision. Dans ce problème dit «de la boîte noire», la complexité des procédures nous empêche de comprendre clairement l'algorithme.

Il existe en outre un problème dit «d'équité». Les hypothèses normatives des développeurs peuvent en effet également influencer la mise au point d'algorithmes. Des valeurs et des intérêts spécifiques peuvent être privilégiés par rapport à d'autres. Contrairement aux problèmes de biais, ce sont ici les préjugés du développeur qui posent problème et non les biais de l'algorithme.

Ces difficultés révèlent les problématiques de fond des systèmes d'IA. Or, il arrive que des individus fassent davantage confiance à ces systèmes qu'à leurs congénères ou, inversement, qu'ils accordent davantage de crédit à d'autres personnes dans des cas où les algorithmes prendraient des décisions plus objectives. Cette situation attire notre attention sur le problème de la confiance, qu'il faut résoudre pour chaque application.

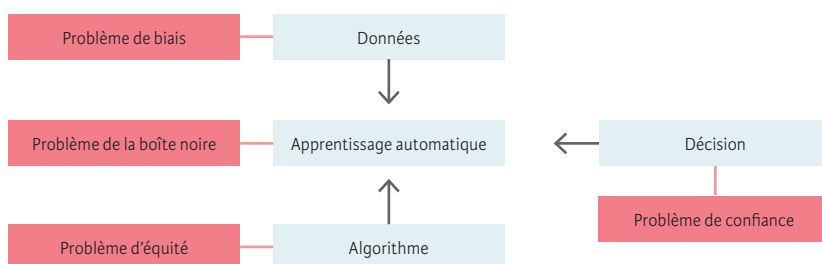
La peur de perdre son emploi

Dans une étude mandatée par la Fondation suisse d'évaluation des choix technologiques (TA-Swiss), une équipe interdisciplinaire a évalué les risques et les chances liés à l'utilisation de l'IA dans les domaines du travail, de la consommation, de la formation et de la recherche, des médias et de l'administration publique, et quels problèmes fondamentaux peuvent alors se poser. Cet article se limite au travail et à la consommation et s'arrête en particulier sur les questions de politique économique qui en résultent.

Le progrès technique s'est certes toujours accompagné de changements dans le monde du travail, mais l'IA constitue aujourd'hui un véritable saut dans l'inconnu sur ce point, car elle est capable de seconder l'homme dans certaines activités, voire de le remplacer dans des tâches réservées jusqu'ici à l'intelligence humaine. Il n'est donc pas surprenant que cet outil suscite de grandes inquiétudes quant à l'avenir du travail.

En 2013, les scientifiques d'Oxford Carl Benedikt Frey et Michael Osborne estimaient dans une étude que 47 % des métiers étaient

Quatre problèmes d'intelligence artificielle qui soulèvent des questions éthiques



CHRISTEN ET AL. (2020) / LA VIE ÉCONOMIQUE



Bon nombre de consommateurs ne savent pas qu'ils laissent des traces numériques – ni ce qu'il advient ensuite de ces données. Scène de vie dans les rues à Cuba.

menacés par la numérisation, relançant ainsi d'anciens débats². Plusieurs études ont par la suite tenté de confirmer ou d'infirmer ce potentiel de rationalisation. D'un côté, on rappelait des évolutions historiques selon lesquelles les vagues d'automatisation entraînaient à long terme une croissance de l'emploi³; de l'autre, on mettait en doute la légitimité de la comparaison, sous prétexte que le progrès actuel se distinguait par une qualité toute différente⁴.

Le monde du travail se métamorphose

Pour que la peur de perdre des emplois ne devienne pas un obstacle à l'innovation, il est nécessaire que la société régie l'introduction de l'IA et qu'elle favorise une répartition équitable des avantages. Ce point concerne en particulier les possibilités de trouver du travail et d'obtenir un revenu. Les entreprises responsables et les milieux politiques doivent veiller à ce que le remplacement du travail humain par des systèmes d'IA n'entraîne pas de chômage et de pertes de revenu. La nécessité est plus urgente que jamais dans le contexte de la crise économique déclenchée par la Covid-19. Des évolutions attendues à long terme en raison du recours à l'IA – pertes massives d'emplois et effondrement de certains secteurs – sont devenues réalité en quelques semaines. C'est donc au niveau politique qu'il s'agit de décider quelles sont

les mesures concrètes à prendre pour rétablir la confiance nécessaire à la reprise économique comme à une utilisation sans heurt des technologies liées à l'IA.

Les effets macroéconomiques possibles de la révolution numérique en général et de l'IA en particulier – notamment les possibilités toujours plus nombreuses de remplacer le travail humain – devraient donc servir de prétexte au lancement d'un débat sociétal sur les processus d'adaptation : faudrait-il par exemple réduire l'horaire normal de travail en cas de taux de chômage durablement élevé ? Comment conserver les normes sociales et celles régissant le droit du travail face à des conditions de travail toujours plus souples ? Comment empêcher les écarts salariaux de s'aggraver ? Et quelles offres de formation continue faut-il créer pour améliorer les chances de chacun tout en promouvant le développement et l'utilisation de l'IA ?

Les consommateurs peu critiques

Alors que le recours à l'IA s'accompagne souvent de réserves dans le monde du travail, il est déjà très avancé dans le domaine de la consommation. Il y a longtemps que cet outil est utilisé pour identifier les segments attractifs de la clientèle, établir des profils clients détaillés ou personnaliser les offres. Pour économiser des coûts et améliorer leur efficacité, des géants technologiques comme

Google, Amazon, Facebook et Apple élaborent et mettent constamment en œuvre de meilleurs systèmes d'IA à l'interface avec la clientèle. Les consommateurs sont ainsi confrontés quotidiennement à l'IA – plus ou moins consciemment. Google Maps leur montre par exemple le plus court chemin pour rentrer chez eux, le système de recommandations d'Amazon leur propose d'acheter d'autres produits, Facebook Newsfeed trie les contenus et l'assistant vocal d'Apple Siri interagit couramment avec eux.

Les avantages de ces systèmes d'IA sont manifestes : d'un côté, les entreprises profitent des processus d'automatisation, par exemple en matière de planification ou de placement de la publicité, et économisent des coûts, par exemple en recourant à des assistants basés sur l'IA et à des « robots conversationnels » (« chatbots ») toujours plus humanisés pour répondre aux questions les plus fréquentes de la clientèle ; d'un autre côté, elles sont en mesure d'améliorer leur efficacité dans la mesure où elles s'adressent de façon ciblée aux consommateurs et proposent des offres taillées sur mesure.

Les consommateurs profitent eux aussi de ces informations personnalisées : les contenus et les offres jugés de moindre intérêt sont occultés et les décisions de consommation et d'achat largement étayées – mais aussi influencées – par cette présélection. À une époque où l'information est devenue pléthorique, beaucoup de clients apprécient cette personnalisation. Mais les consommateurs devraient également être conscients des éventuels problèmes de biais et d'équité décrits précédemment et donc ne pas se fier aveuglément aux systèmes d'IA.

Le nombre croissant d'empreintes électroniques laissées derrière soi donne aujourd'hui lieu à une personnalisation dynamique des profils clients sans qu'il soit la plupart du temps nécessaire de demander leurs préférences aux consommateurs. Quelques centaines de « j'aime » sur Facebook suffisent pour évaluer avec beaucoup de précision des caractéristiques personnelles sensibles⁵. C'est justement parce que ces systèmes d'IA fonctionnent sans la participation active des consommateurs qu'ils restent une boîte noire opaque. Il devient donc toujours plus difficile pour les consommateurs d'imaginer les inférences que l'IA tire de leurs données (« inferred data ») et même de savoir où ils ont affaire à de l'IA. Ainsi, selon une étude américaine de 2015, 62 % des utilisateurs de Facebook ne savaient pas que le service Newsfeed est personnalisé par IA⁶.

² Voir Frey et Osborne (2013).

³ Voir Autor (2015).

⁴ Voir Lovergine et Pelleri (2018).

⁵ Youyou, Kosinski et Stillwell (2015).

⁶ Eslami et al. (2015).

Les auteurs de l'étude mandatée par TA-Swiss recommandent donc aux entreprises qui utilisent des systèmes d'IA dans le domaine de la consommation de rendre transparent leur recours à l'IA et de communiquer simplement les données personnelles qui entrent dans le système et les inférences qui en sont tirées.

Le gagnant rafle la mise

La plupart des systèmes d'IA secondent aujourd'hui les consommateurs dans des domaines encore strictement définis. Les spécialistes s'attendent cependant à ce qu'un petit nombre d'auxiliaires d'IA comme Assistant (Google), Siri (Apple) ou Alexa (Amazon) s'imposent pour aider les consommateurs dans des pans entiers de leur vie. Au lieu d'assistants spécialisés dans les questions de santé ou dans la coordination des calendriers, un seul assistant pourrait à l'avenir secondar un consommateur dans une multitude de tâches, ce qui paraît à première vue avantageux.

Mais comme plus ces systèmes disposent de données d'un utilisateur, plus ils deviennent précis, des coûts élevés pourraient à long terme survenir pour les consommateurs qui changent de fournisseur. Les entreprises affronteront donc par ricochet des entraves croissantes à l'accès au marché si elles ne disposent pas de données d'utilisateurs pour alimenter leurs propres systèmes d'IA. Les données d'utilisateurs deviennent ainsi de plus en plus le carburant de l'ère numérique. Pour échapper à une telle capitalisation des données, il est nécessaire de trouver des solutions durables.

C'est pourquoi les auteurs de l'étude mandatée par TA-Swiss recommandent d'examiner de quelle manière une mise en œuvre de la «portabilité des données» dans les systèmes d'IA serait possible. Le client qui prend un autre fournisseur pourrait alors emporter ses données, ce qui faciliterait le changement de prestataire. Il s'agirait d'examiner dans

quelle mesure les inférences élaborées par un système d'IA («inferred data») devraient être elles aussi portables, comme les données brutes.

Sensibiliser les esprits

Le débat sociétal sur la perte d'emplois ainsi que l'exigence de transparence et de portabilité des données montrent l'importance que le grand public soit informé et impliqué activement dans les décisions concernant l'utilisation et la gestion de l'IA.

D'amples mesures de formation et de recherche sont nécessaires afin que les citoyens aient davantage de contrôle sur leurs données. Le système éducatif doit par exemple préparer toutes les classes d'âge à manier l'IA. À cet effet, il s'agit d'étudier quelles compétences spécifiques doivent être enseignées pour que les élèves acquièrent une compréhension générale des capacités et des limites des systèmes d'IA. Les connaissances correspondantes doivent être intégrées dans les manuels didactiques et mises à disposition des enseignants et des élèves sur les plateformes existantes.

Un exemple de la manière dont le maniement de l'IA peut être enseigné dans les écoles est ce que l'on appelle la «pensée computationnelle». Cette capacité sensibilise les élèves aux problématiques complexes et au fonctionnement de base des applications d'IA et leur permet d'en faire un usage responsable. De telles capacités offrent en outre les prérequis qui permettent de s'imposer dans le monde du travail actuel et de profiter du recours à l'IA en tant que travailleur.

Outre le système éducatif, la recherche est également tenue de s'engager. Elle doit travailler à des solutions fiables d'applications de l'IA. Il s'agit également de promouvoir des compétences non-techniques comme la durabilité et l'éthique dans les hautes écoles, en particulier dans les disciplines techniques consacrées à l'élaboration de nouveaux systèmes d'IA. On s'assurera ainsi que les

approches normatives qui entrent aussi dans l'élaboration d'applications de l'IA s'inspirent de principes responsables et éthiques.

Il s'agit de s'attaquer aux problèmes fondamentaux pour permettre aux avantages de l'IA de l'emporter. Un pas important dans cette direction consiste à mener un dialogue communautaire et documenté sur les chances et les risques de l'IA, dialogue qu'il faut absolument encourager.



Clemens Mader

Collaborateur scientifique, division Technologie et société, Laboratoire fédéral d'essai des matériaux et de recherches (Empa), Saint-Gall



Johann Čas

Chercheur principal, Institut d'évaluation des choix technologiques, Académie autrichienne des sciences, Vienne



Anne Scherer

Professeure assistante de marketing quantitatif, université de Zurich

Bibliographie

Autor D. H. (2015). «Why are there still so many jobs? The history and future of workplace automation». *Journal of Economic Perspectives*, 29(3), 3–30.
Christen M., Mader C., Cas J. et al. (2020). «Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz». *TA-Swiss Publikationsreihe* (éd.), TA 72/2020, Zurich: vdf.

Christen M. (2019). «Peut-on programmer le sens de l'éthique?». *La Vie économique*, 12/2019, 13–15.
Eslami M., Rickman A., Vaccaro K. et al. (2015). ««I always assumed that I wasn't really that close to [her]»: Reasoning about invisible algorithms in the news feed». *CHI 2015*, ACM Press.

Frey C. B. et Osborne M. A. (2013). «The future of employment: how susceptible are jobs to computerisation?». *Technological Forecasting and Social Change*, 114, 254–280.
Lovergine S. et Pellero A. (2018). «This time it might be different: Analysis of the impact of digitalization on the labour market». *European Scientific Journal*, 14(36), 68.

Youyou W., Kosinski M. et Stillwell D. (2015). «Computer-based personality judgments are more accurate than those made by humans». *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.