

L'éthique, grande oubliée des algorithmes ?

Du matin au soir, nous sommes confrontés à des algorithmes. Mais cette exposition n'est pas sans dangers. Leur influence sur nos opinions politiques, sur nos humeurs ou sur nos choix est prouvée. Loin d'être neutres, les algorithmes portent les jugements de valeur de leurs développeurs, et nous les font subir le plus souvent sans que nous en ayons conscience. Il devient alors nécessaire de questionner leur éthique, et de trouver des solutions aux biais que subissent leurs utilisateurs.

Quel est le métier de Facebook ? Ou celui de Twitter ? Plus généralement, quel est le métier d'un réseau social ? La réponse simpliste, mais néanmoins juste, est la suivante : sélectionner l'information qui vous sera présentée sur votre mur, afin de vous faire passer le plus de temps possible dessus. Derrière ce « fil d'actualités » chronophage se cache une sélection de contenus, publicitaires ou non, optimisée pour chaque utilisateur à grand renfort d'algorithmes. Grâce à eux, les réseaux sociaux déterminent ainsi ce qui sera le plus intéressant pour vous. Sans questionner l'utilité de ces sites, leur fonctionnement soulève néanmoins de lourdes questions éthiques. À commencer par celle-ci : tous les utilisateurs sont-ils conscients du poids des algorithmes sur leur perception de l'actualité, sur leurs opinions ? Et pour aller plus loin encore : quels impacts ces algorithmes ont-ils sur nos vies, sur nos décisions ?

Pour Christine Balagué, chercheuse à Télécom École de Management et membre de la [Cerna](#), « le sujet de la captation des données personnelles est connu, mais moins celui du traitement de ces données par les algorithmes ». Si les utilisateurs font plus attention à ce qu'ils partagent sur les réseaux sociaux, ils ne se demandent pas encore nécessairement comment fonctionne le service qu'ils utilisent. Et cette méconnaissance ne concerne pas que Facebook ou Twitter. Les algorithmes sont partout dans nos vies, présents dans nos applications mobiles et les services web que nous utilisons. Du matin au soir nous sommes confrontés à des choix, des suggestions, des informations qui sont traitées par des algorithmes : Netflix, Citymapper, Waze, Google, Uber, TripAdvisor, AirBnb...

« Ils gèrent nos vies » affirme Christine Balagué. « Un nombre fortement croissant d'articles de chercheurs issus de différents domaines est publié au sujet du pouvoir que les algorithmes ont sur les individus. » En 2015, Robert Epstein, chercheur à l'institut américain de recherche comportementale, montrait notamment comment un moteur de recherche pouvait jouer sur les résultats d'une élection. Son étude, menée sur plus de 4 000 personnes, a permis de déterminer que le classement des candidats dans les résultats de recherche influençait au moins 20 % des votants indécis. Autre exemple frappant : une recherche de 2012 conduite par Facebook sur 700 000 utilisateurs de son service a démontré que les personnes exposées précédemment à des publications

négatives postaient majoritairement du contenu négatif. De même, celles exposées précédemment à des publications positives postaient essentiellement du contenu positif. Ceci prouve que les algorithmes sont susceptibles de manipuler les émotions des individus sans qu'ils en soient conscients ni informés. Quelle place pour nos préférences personnelles dans un système d'algorithmes que nous ne connaissons même pas ?

Le côté obscur des algorithmes

Dans cette opacité réside l'un des principaux problèmes éthiques des algorithmes. Sur un moteur de recherche comme Google, deux utilisateurs effectuant la même requête n'auront pas le même résultat. L'explication avancée par le service est la personnalisation des réponses pour mieux répondre aux attentes de chacun des deux individus. Mais les mécanismes de sélection des résultats sont obscurs. Parmi les paramètres pris en compte pour déterminer quels sites seront affichés sur la page, plus d'une centaine concernent l'utilisateur effectuant la requête. Sous couvert de secret industriel, la nature exacte de ces paramètres personnels et la façon dont ils sont pris en compte par les algorithmes de Google sont inconnues. Difficile de savoir comment l'entreprise nous catégorise, détermine nos centres d'intérêt ou prédit nos comportements. Et une fois cette catégorisation faite, est-il même possible d'en sortir ? Comment rester maître de la perception que l'algorithme crée de nous ?

Du fait de cette opacité, il est même impossible de connaître les biais qui peuvent survenir du traitement des données. Pourtant, ils existent bel et bien, et s'en prémunir est un véritable enjeu de société. Un exemple de la non-équité des individus face aux algorithmes ? Des travaux menés par Grazia Cecere, économiste à Télécom École de Management, ont mis en évidence une discrimination entre hommes et femmes dans les algorithmes d'association de centres d'intérêt d'un grand réseau social. « En créant une publicité sur les STEM¹, nous avons constaté que le logiciel la distribuait préférentiellement aux hommes, alors même que les femmes montrent plus d'intérêt pour cette thématique » détaille Grazia Cecere. Loin du mythe de l'intelligence artificielle maléfique, l'origine de ce genre de biais est à chercher dans l'action humaine. Trop souvent oubliée, la présence de développeurs derrière chaque ligne de code doit être rappelée.

Les algorithmes sont avant tout là pour proposer des services, le plus souvent commerciaux. Ils s'inscrivent donc dans le cadre d'une stratégie d'entreprise qu'ils refléteront inévitablement pour répondre aux attentes économiques de celle-ci. « Les data scientists travaillant sur un projet visent l'optimisation de leurs algorithmes sans nécessairement réfléchir aux questions éthiques des choix qui sont effectués par ces programmes » pointe Christine Balagué. En outre, les humains ont des perceptions de la société qu'ils intègrent plus ou moins consciemment dans les logiciels qu'ils développent. Le jugement de valeur d'un algorithme est alors bien souvent un jugement de valeur porté par ses créateurs. Dans le cas des travaux de Grazia Cecere, celle-ci explique le biais mis en évidence de façon simple : « L'algorithme apprend ce qu'on lui demande d'apprendre, et réplique les stéréotypes s'ils ne sont pas triés. »

¹ sciences, technologie, éducation, mathématiques

Un exemple emblématique de ce phénomène concerne l'imagerie médicale. Un algorithme classifiant une cellule comme malade ou non devra être configuré pour faire un arbitrage entre le nombre de faux positifs et le nombre de faux négatifs. En effet : les développeurs doivent décider à quel point il est tolérable d'avoir des tests positifs sur des personnes saines pour ne pas passer à côté de personnes malades dont les tests reviendraient négatifs. Pour les médecins, il est préférable d'avoir des faux positifs plutôt que des faux négatifs. Pour les scientifiques qui développent l'algorithme en revanche, il est préférable d'avoir des faux négatifs plutôt que des faux positifs, car les connaissances scientifiques sont cumulatives. Selon les valeurs défendues par les développeurs, ceux-ci privilégieront l'une ou l'autre des professions.

La transparence ? Oui, mais pas que !

Pour lutter contre ces biais, l'une des propositions est de rendre plus transparents les algorithmes. Depuis octobre 2016, la Loi pour une République numérique proposée par Axelle Lemaire, ancienne secrétaire d'État chargée du numérique, impose la transparence de tous les algorithmes publics. C'est grâce à cela notamment que le code du site web Admission post-bac (APB) a été communiqué au public. Les entreprises jouent également de plus en plus cette carte. Depuis le 17 mai 2017, Twitter propose ainsi à ses utilisateurs de connaître les centres d'intérêt qui leur sont associés. Mais malgré ces bonnes intentions, la transparence peine à être suffisante pour garantir la dimension éthique. D'abord, l'intelligibilité des codes est encore souvent négligée : les algorithmes sont parfois livrés sous des formats qui ne facilitent pas leur lecture et leur compréhension, même pour des professionnels. Ensuite, la transparence peut être factice. Dans le cas de Twitter, « rien n'est communiqué sur la manière dont les centres d'intérêts sont attribués » constate Christine Balagué.

Centres d'intérêt de Twitter

Voici certains des centres d'intérêt qui vous correspondent, sur la base de votre profil et de votre activité. Vous pouvez les ajuster si quelques choses ne semble pas convenir.

- Action et aventure
- Actualités et informations générales sur la santé
- Actualités et informations générales sur les films
- Actualités et informations générales sur les voyages
- Actualités technologiques
- Affaires et finances
- Animation
- Actualités et informations générales sur l'éducation
- Actualités et informations générales sur les affaires
- Actualités et informations générales sur les livres
- Actualités financières
- Actualités scientifiques
- Affaires et actualités
- Alternative
- Biologie

Quelles publications de cet utilisateur l'ont amené à être classé dans « Action et aventure », une thématique à la signification très large ? Quelle est la pondération effectuée par les algorithmes de Twitter entre « Actualités scientifiques » et « Affaires et finances » pour afficher des contenus dans le fil d'actualité de l'utilisateur ?

Pour aller plus loin, il faut évaluer le degré de transparence des algorithmes. C'est le sens de l'initiative [TransAlgo](#), lancée par Axelle Lemaire également, et pilotée par l'Inria. « Il s'agit d'une plateforme mesurant la transparence, en regardant quelles données sont utilisées, quelles données sont produites, à quel point le code est ouvert... » explique Christine Balagué, membre du conseil scientifique de TransAlgo. La plateforme est la première du genre en Europe, et fait de la France une nation avancée dans la réflexion sur la transparence. Dans la même veine, DataIA est un institut de convergence sur les données initié sur le plateau de Saclay pour une durée de 10 ans. Ce programme unique est interdisciplinaire et comprend des recherches sur les algorithmes en intelligence artificielle, leur transparence et les enjeux éthiques.

En réunissant des équipes scientifiques pluridisciplinaires, l'objectif est d'étudier les mécanismes de développement des algorithmes. Les sciences humaines ont beaucoup à apporter sur l'analyse des valeurs et des décisions qui se cachent derrière l'élaboration des codes. « Il devient de plus en plus nécessaire de décortiquer les méthodes algorithmiques, de faire du retro-engineering, de mesurer leurs potentiels biais et discriminations, de les rendre plus transparents » insiste Christine Balagué. « Plus largement, des recherches ethnographiques sur les développeurs sont nécessaires, en s'immergeant dans leurs intentions, et en étudiant l'assemblage socio-technique des algorithmes. » À mesure que les services numériques prennent de plus en plus de place dans nos vies, il est en effet primordial d'arriver à cerner les risques que les algorithmes font encourir à leurs utilisateurs.

Une commission publique dédiée à l'éthique du numérique

Depuis 2009, les grands acteurs français de la recherche et de l'innovation sur le numérique sont regroupés au sein d'Allistene (Alliance des sciences et technologies du numérique). En 2012, cette alliance décidait de créer une commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique : la Cerna. Sur la base de travaux multidisciplinaires, rassemblant des expertises et des contributions de l'ensemble des acteurs du numérique aussi bien au niveau national qu'international, la Cerna questionne le numérique sur ses aspects éthiques. En prenant comme objets d'étude des thématiques aussi variées que l'environnement, la santé ou la robotique ou les nanotechnologies, elle souhaite sensibiliser et éclairer les concepteurs de technologies aux questions éthiques. Ses rapports sont téléchargeables sur son site web.