Algorithmes et biais cognitifs

Gilbert ISOARD

Ex-consultant en systèmes expert et intelligence artificielle,
consultant et analyste en informatique, conseiller en stratégie et intelligence économique

Résumé : Les biais cognitifs, nombreux et aux causes diverses, sont des raccourcis de nos schémas de pensées et provoquent des associations mentales faussement logiques liées à nos perceptions et à un confort de fonctionnement cérébral simplifié. L'avènement de l'usage massif de données statistiques dans des cycles de traitements de type algorithmiques tout aussi volumineux, pose la question de leur fiabilité et de leurs usages en fin de cycle. De fait peu de gens sont capables de décoder leurs contenus et rendre intelligible les écueils sociétaux que cela peut engendrer. Les solutions sont multiformes et ne passeront pas par un principe de précaution au sein des compétitions de tous ordres, mais par l'information tout aussi proportionnelle, l'éducation des acteurs et la facilitation des pouvoirs publics aux meilleurs usages et aux bonnes pratiques, de l'école aux acteurs économiques et aux citoyens. Une société civile qui est de plus en plus orientée par les algorithmes, influant jusqu'à notre imaginaire, et nos pensées traduites en actes quotidiens doit aussi nous alerter sur les dérives potentielles, colorées et augmentées des biais cognitifs, en passant par une éducation et des usages critiques, vigilants, d'une société de l'information baignant dans des flux aux multiples influences.

Mots-clés : biais de perception, cycle de traitement, rationalité limitée

« L'entendement humain n'est pas une lumière froide, il est soumis à l'influence
de la volonté et des émotions, ce qui engendre un savoir fantaisiste :
l'homme croit de préférence ce qu'il désire être vrai. »
Francis Bacon (1561-1626) *in Novum Organum*

1 Algorithme, biais cognitifs et perceptions individuelles

« Al-Khwarizmi », mathématicien arabo-perse du IX^{e} siècle av. JC, ne se doutait pas de la portée future de son patronyme, déformé du grec (« arithmos » : nombre), puisqu'il nous donnera le terme : « Algorithme ». Un algorithme est une séquence finie et non ambiguë d'opérations destinée à la résolution d'un problème.[1] On dit aussi souvent qu'une recette de cuisine est un algorithme, de même une simple notice de

[1] Le terme, quelque peu réducteur, d'« algorithme » comprend déjà une forme de biais en tant que raccourci sémantique. La notion d'algorithme représente en effet un cycle de traitement informatisé plus complexe et sujet à de nombreuses potentialités de biais.

montage d'un meuble en kit, ou bien sûr en informatique, toute programmation en un langage codé.

Un biais cognitif en psychologie cognitive[2] est un schéma de pensée trompeur et faussement logique. Ces biais influencent nos choix, lorsqu'il faut gérer avec des algorithmes une quantité importante d'informations, notamment en un temps limité. Il se produit alors un dysfonctionnement appliqué sur les résultats qui seront potentiellement amplifiés par un « effet papillon », celui-ci octroie à de petites causes, de grands effets, notamment à caractère prédictif, perturbant la réalité et par la suite influençant nos perceptions.

Passons à la pratique : vous donnez la première réponse qui vous vient spontanément à l'esprit.

Imaginons une raquette de tennis avec sa boite de 4 balles, le tout représentant une valeur de 300 euros, or sachant que la raquette de tennis vaut 280 euros de plus que la boite de balles, Combien vaut une balle de tennis ?

En quelques instants… 85 à 90 % des personnes répondent spontanément à 5€ la balle (20/4).

Le piège se trouve dans le fait que la raquette vaut 280… de *plus* qu'une boîte de balles.

Nous aurons donc le montant d'une boîte de balle +280 € pour obtenir le coût de la raquette, soit 280 plus le prix d'une boite de balles [soit $(280+10)+(300-290) = 300$] ce qui fait que la balle de tennis vaut 2,5 € $(4 * 2,5)$.

Si vous aviez répondu mentalement 5€, il ne faut pas s'inquiéter. Plus de 80% des diplômés de l'enseignement supérieur testés font de même ; il s'agit du résultat d'un phénomène bien documenté que l'on appelle justement un biais cognitif. Un biais cognitif est une erreur de raisonnement mais il ne faut pas pour autant en déduire que toutes les erreurs constituent des biais (ce qui serait également un biais cognitif de raisonnement).

Un biais cognitif est une erreur de raisonnement qui provient d'heuristiques[3] simplifiant la réalité.

Tous, nous possédons ces stratégies communes de raisonnement (les heuristiques) nous permettant de faire face à notre environnement. Ces stratégies sont imparfaites. C'est en fait cette insuffisance que l'on appelle un biais cognitif. Ces biais sont présents à différents niveaux chez tous les humains, ce qui est d'ailleurs rassurant, mais devrait-on alors faire confiance à des humains « experts dans leur domaine » et sujets à des biais cognitifs ?

[2] L'étude des biais cognitifs, ou des déviations du jugement par rapport à un jugement rationnel, est en effet au cœur des travaux en psychologie cognitive. Le terme « biais » fait référence au fait que l'erreur commise est souvent systématique et ses déterminants sont à rechercher dans les mécanismes mis en jeu dans le traitement cognitif appliqué à une situation donnée (v. https://fr.wikipedia.org/wiki/Biais_cognitif).

[3] Heuristiques : L'heuristique ou euristique (du grec ancien, « je trouve ») est définie en résolvant des problèmes à partir de connaissances incomplètes. Ce type d'analyse permet d'aboutir, en un temps limité à des conclusions vraisemblables. *Celles-ci peuvent toutefois s'écarter de la solution optimale.*

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Le modèle le plus simple, étudié en sciences cognitives quant à la prise de décision, est celui des théories à processus inspiré d'une logique de simplification duale. En effet, l'esprit humain comporte deux grands types de processus : les processus, disons par exemple de type X qui sont rapides, peu coûteux en énergie de concentration et intuitifs, et les processus de type Y qui eux sont lents, coûteux en énergie liée à la concentration et qui demandent une réflexion analytique plus approfondie. Une réponse intuitive à une question… combien coûte une balle de tennis ? est le résultat d'un processus de type X, tandis que l'exercice d'en faire la démonstration sera plutôt analytique et de type Y.

Se lever, prendre un petit déjeuner et se brosser les dents avant d'aller travailler sont plutôt des décisions prises rapidement et intuitivement (Type X). En clair, il est tout simplement impossible que chacune de nos décisions soit le résultat d'un processus analytique (Type Y), parce que la capacité cognitive d'un individu est toujours liée au concept d'une rationalité limitée[4]… proposé par Herbert Simon dès 1955, à cause de nos limites biologiques face à un environnement complexe.

Pendant toute son évolution, l'Humain a donc développé des stratégies devant lui permettre malgré tout de décider pour agir : Les heuristiques.

Pour conclure cette première partie nous signalons en Mars 2020 un rapport sur le sujet du contrôle des biais, par l'Institut Montaigne[5] et une parution de Paris-Telecom en Février 2019, qui indique : « *Les biais cognitifs sont une distorsion de la manière dont l'information est traitée par rapport à un comportement rationnel ou à la réalité.* »[6]

2 Biais cognitifs et « cycle de traitement »

Humains, les biais cognitifs sont présents dans tous les domaines d'activités, et nous appliquerons notre raccourci sémantique et cognitif, plus spécifiquement à l'impact potentiel des traitements dits d'intelligence artificielle, et notamment les processus d'apprentissage par le biais d'algorithmes spécifiques[7]. En effet, ces processus sont susceptibles d'influencer très fortement la vie des individus à travers le monde. Les algorithmes modifient la manière dont nous vivons, nous apprenons et nous travaillons, ils modifient nos loisirs, nos relations interpersonnelles, notre façon de nous informer et de raisonner et même la façon dont nous sommes diagnostiqués ou soignés…

Ce qui nous perturbe est principalement l'effet « boîte noire », c'est-à-dire une forme de cécité sociétale de compréhension, tant le nombre de spécialistes de ces algorithmes,

[4] Théorie de la rationalité limitée : la rationalité procédurale de la décision, le modèle I/M/C – Herbert Simon | SI & Management (sietmanagement.fr), La rationalité limitée est l'idée selon laquelle la capacité de décision d'un individu est altérée par un ensemble de contraintes comme le manque d'information, des biais cognitifs ou encore le manque de temps. Dans cette optique, les décideurs ont tendance à choisir des solutions satisfaisantes plutôt qu'optimales.

[5] Institut Montaigne ; « Algorithmes : contrôle des biais S.V.P. » Rapport Mars 2020. https://www.institutmontaigne.org/publications/algorithmes-controle-des-biais-svp

[6] Bertail, Patrice , *et al.* (2019) « Algorithmes : Biais, Discrimination et Équité. » Fondation Abeona et Paris Telecom) https://www.telecom-paris.fr/wp-content-EvDsK19/uploads/2019/02/Algorithmes-Biais-discrimination-equite.pdf.

[7] Il s'agit d'algorithmes qui analysent un ensemble de données afin de déduire des règles, avec des étapes dites d'entraînement, constituant des connaissances permettant d'analyser de nouvelles situations.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

rapporté à la population est extrêmement faible. Les algorithmes sont faussement habillés des atouts de la neutralité et de l'objectivité. Mais ils sont intégrés dans nos sociétés et en prolongent les défauts, voire insidieusement, en renforçent les inégalités structurelles. Il s'agit d'une tendance récurrente, pour l'usage des techniques issues de la science en général, de plus en plus pertinente au fil du temps. Toute technologie appliquée à l'échelle de la vie en société, avec en corollaire son temps social d'usage, est accélérée inexorablement, augmentant le niveau d'incompréhension et de cécité qui augmente d'autant en diminuant la confiance du corps social, et engendre des phénomènes de rejets, au pire, et au mieux d'incompréhension totale pour la plupart des personnes.

Les algorithmes sont très fortement dépendants des données et en ayant acquis une certaine forme d'autonomie, les biais, sont associés à une vraie « fausse rigueur », à l'insu des services qui en sont attendus et de ceux qui les représentent et/ou des organismes d'États les utilisant massivement. Ceci sans aucune remise en cause de leur niveau réel de pertinence ou d'équité, sur l'ensemble des processus digitaux qui nous concernent et des données à partir desquelles ils fonctionnent. Ils induisent diverses recommandations ou des décisions[8] directement préétablies, pesant sur les individus. Celles-ci sont mécaniquement soumises à divers « biais cognitifs », pour un « cycle de traitement » plus complet. Or ces biais, déjà présents dans les données, qui sont « retenus en mémoire(s) » par l'apprentissage automatisé de type *« machine learning »*[9], sont potentiellement amplifiés par les algorithmes de renforcement de l'apprentissage machine qui utilisent ces mêmes données.

Un cycle de traitement, avec des « biais cognitifs » potentiels à chaque étape, ne s'arrête donc certainement pas aux données, ni aux traitements algorithmiques proprement dits, mais devrait s'appliquer aux divers domaines ci-dessous :

- Qualification des sources de données / diversité, exhaustivité, fiabilité
- Acquisition des données / périmètre et collecte
- Mise en forme des données et formatage des données / traitement envisagé
- Tests de cohérence sur un échantillon de données / validation
- Programmation algorithmique / règles de validation
- Traitement algorithmique (Biais culturels associés)
- Interprétation des résultats / Règles de cohérence (nombreux paramètres)

[8] Divers travaux ont ainsi identifié des biais dans de multiples domaines : jugement de probabilité et prise de décision (Allais, Maurice (1953). « Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'école Américaine ». Econometrica. vol. 21, n° 4, p. 503–546), raisonnement déductif et inductif (Frederick, S. (2005). « *Cognitive reflection and decision making* ». Journal of Economic Perspectives. vol. 19, n° 4, p. 25–42 ; Kahneman, D. & Slovic, P. & Tversky, A. (1982), « *Judgement under uncertainty : Heuristics and biases* », Cambridge University Press), ou encore relations sociales (Drozda-Senkowska, Ewa (1999). Psychologie sociale expérimentale, collection Cursus, Armand Colin).

[9] Hamilton, Evan ; 2017, *"Benchmarking Four Approaches to Fairness-Aware Machine Learning."* Haverford College Computer Science Spring 2017.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

- Mise en forme des résultats /Type de validation des sorties (écrits, tableaux, schémas, etc.)
- Communication des résultats / Canaux, forme et quantité, qualité d'interprétations.
- Médiation avec des tiers (journalistes ou experts d'une problématique sociétale, langage, croyances, intérêts…)
- Analyse et revue de perception catégorielle et sociétale de la façon dont sont perçus les résultats
- Choix des mesures correctives, adaptées par occurrence, suivies de leur mise en application.

Les « Algorithmes », pour évoquer un cycle de traitement plus complet, symbolisent également un raccourci lié à la façon de développer une pensée ou bien une forme de raisonnement. Les données historiques utilisées par les algorithmes (vous aurez compris un cycle de traitement), peuvent notamment être biaisées et défavoriser des personnes en fonction de leur âge, de leur origine sociale ou ethnique, de leur genre, ou de leurs revenus d'existence. Or ces biais, déjà présents dans les données, sont « amplifiés » par l'apprentissage des algorithmes de renforcement qui utilisent ces mêmes données. Ces données reflètent finalement nos biais humains en « miroir », alors que celles des biais d'amplifications seront des données faussées et ce sont celles-ci qui doivent pouvoir être combattues et rectifiées.

Avec la technologie 5G et l'Internet des objets connectés, de plus en plus de données fines de consommation seront collectées (domotique, type de flux et consommations, applications mobiles de santé, de vacances, de loisirs, de mobilité, etc.). Les algorithmes, avec leurs biais, auront une capacité à détecter, à normaliser mais pourront aussi amplifier des discriminations ou proposer des conclusions pouvant être erronées.

Un algorithme peut multiplier et amplifier des biais sociétaux et avoir des conséquences en profondeur sur les individus. Il n'est donc pas souhaitable de limiter l'analyse aux seuls biais techniques : la réflexion doit intégrer pleinement le risque de transmission des biais de sociétés, qui sont par nature multiformes. Bien heureusement, par diverses contingences précédentes, d'éducation et d'informations critiques partagées, le sujet évolue.

Comment peut-on et doit-on utiliser, dans notre société apprenante et évolutive, des « algorithmes » de manière raisonnée et juste, en adéquation avec l'éthique nécessaire qui devrait prévaloir démocratiquement à leur usage ?

3 Biais cognitifs, algorithmes et société

Nous retiendrons que :

- Les algorithmes au sein d'un cycle de traitement, sont programmés par des personnes avec leurs propres biais cognitifs, conscients ou inconscients.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

- Il y a près de 250 biais cognitifs répertoriés[10]
- Dans le langage de la technique, un biais sera l'écart-moyen entre sa prédiction et la valeur normée.
- Un fort biais signifie que l'algorithme manque de relations pertinentes entre données d'entrée et de sortie pour effectuer des prédictions et des propositions correctes.
- Les « biais techniques » sont pris en compte car ils viennent directement diminuer la performance.
- L'algorithme peut être très abouti techniquement tout en étant « biaisé » d'un point de vue social. Ces biais sont en fait la reproduction *via* l'algorithme, de biais qui sont déjà présents au sein de la société.
- Ces biais relèveront du débat public lorsqu'ils désavantagent un groupe social spécifique sur des critères de discrimination.

Bien évidemment, un algorithme qui reconnaît un peu moins souvent les variétés d'un végétal plutôt qu'un autre sur des images sera biaisé techniquement, mais le préjudice sociétal en matière d'injustice sera quasi-inexistant. La distinction entre biais technique et biais de société est importante car ceux-ci n'attirent pas la même attention de la part des développeurs. Les biais techniques réduisent la performance de l'algorithme, et entravent la réalisation de son objectif. Réduire les biais techniques représentera un coût, mais également souvent un bénéfice clair pour le développeur. A contrario, suivre les biais de société permet à l'algorithme[11] d'être également statistiquement plus performant…

Nos différents biais, car il s'agit de nous tous, sont répertoriés par grandes catégories :

- Biais sensori-moteurs (illusions liées aux sens et à la motricité)
- Biais attentionnels ou biais d'attention (problèmes d'attention)
- Biais mnésiques (en rapport avec la mémoire)
- Biais de jugement (déformation de la capacité de juger)
- Biais de raisonnement (paradoxes dans le raisonnement)
- Biais liés à la personnalité (en rapport avec la culture, la langue, l'influence sociale…) etc.

[10] Associations Libres : https://associationslibres.wordpress.com/2016/10/14/petit-guide-exhaustif-des-biais-cognitifs.

[11] Borrajo, L., Cao, R. 2018. Nonparametric Mean Estimation for Big-But-Biased Data. In the Mathematics of the Uncertain, Studies in Systems, Decision and Control, Springer : Cham, Switzerland, 142, 55-65.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Mais plus intéressantes à mettre en lumière sont, les diverses raisons associées aux biais cognitifs, pour leur influence sur un cycle de traitement et ceci par rapport à 4 grandes thématiques liées aux processus cognitifs :

Tout d'abord, le rapport entre « information » et « saturation ». Nous remarquons préférentiellement ce qui a déjà sensibilisé notre mémoire, ou ce qui est souvent répété. Les choses déviantes de la norme sont plus marquantes que celles qui ne le sont pas. Ainsi, lorsqu'un élément change, nous le remarquons plus facilement. Nous sommes de ce fait plus fortement influencés par ce qui confirme nos propres convictions.

Le rapport entre « mémorisation » et « rétention sélective » est par ailleurs essentiel. Selon le contexte, ou l'expérience sensible, nous conservons sélectivement les souvenirs. Nous réduisons par commodité les événements à leurs éléments clés et négligeons des spécificités pour nous souvenir plus facilement des généralités. Ainsi, nous modifions et renforçons certains souvenirs, *a posteriori*.

Le lien entre « action » et « réaction » participe également à la formation de biais cognitifs : Nous préférons simplicité et complétude face aux options complexes et ambiguës. Nous préservons ainsi notre position au sein d'un groupe, afin de conserver des opportunités d'actions/réduction du risque. Nous préférons, même à tort, terminer ce qui nous a déjà coûté en temps et en énergie, ou encore, nous privilégions ce qui est immédiat et qui a du sens pour nous. Enfin, pour agir, nous préférons avoir une influence et sentir que ce que nous faisons est important.

Le rapport entre « sens » et « non-sens » joue un rôle dans la constitution de biais, et ce, en particulier, pour les raisons suivantes : Nous projetons notre état d'esprit et nos suppositions sur le passé et pour le futur. Nous simplifions les probabilités et les nombres pour qu'ils soient plus faciles à appréhender. Nous accordons par ailleurs naturellement du crédit aux choses et aux gens que nous apprécions…Enfin, nous extrapolons sur la base de stéréotypes, de généralités, et d'antécédents et nous conscientisons des histoires et des motifs, y compris à partir de données éparses.

Voilà en quelques grandes thématiques, résumées, la matrice filtrée des biais cognitifs, qui oriente nos vies… La toile de fond de cette matrice étant de permettre aux heuristiques, valorisant notre préférence à « réduire notre risque » et pour le définir plus exactement, notre perception du risque.

Alors que peut-on faire pour diminuer les impacts potentiels des biais ?

Comme pratiquement en toute chose, la solution vient de l'équilibre par la multiplicité des points de vue, et du dialogue entre les sciences, les techniques et les expertises. Assurer l'éducation, la formation technique et s'assurer de la diversité des équipes de conception et de déploiement des algorithmes, pourraient être des voies porteuses de plus d'équilibre et assurer la limitation des biais. Toutefois, il est assez clair et réaliste que ces évolutions sociétales sont et seront malheureusement lentes et difficiles à développer sur plusieurs plans, car toute notre économie est cannibalisée par la concurrence (entre États, services, entreprises, individus…) et donc avec des évolutions technologiques de plus en plus rapides, sans être suffisamment assimilées par la société.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

- Tout en sachant qu'être capable de définir ce que serait un algorithme sans biais, est également relativement complexe. Certains biais sont d'ailleurs volontaires, comme le fait de promouvoir des boursiers dans un cursus scolaire, d'autres biais sont inconscients, et conduisent à discriminer certains groupes.
- Un algorithme traitant les décisions impactant des individus de manière totalement équitable s'approche d'un algorithme sans biais indésirable, sans pour autant avoir le pouvoir de le garantir.
- L'équité d'un algorithme peut prendre des formes multiples.
- L'appréciation de ce qui est juste comporte une grande dimension culturelle et morale, par nature différente selon les continents et souvent entre pays ou même provinces voisines.
- Les domaines d'application sont tous différents : Les présupposés « justes » ne seront pas les mêmes selon que l'on évoque un algorithme d'analyse d'une radiographie ou bien de recommandation de publicités sur les réseaux sociaux.
- Chaque typologie de cas mérite une éducation préalable et des informations pour tous les acteurs impliqués.

Et cela prend une durée significative et structurellement inadaptée à nos temps d'actions économiques impulsés par les acteurs eux-mêmes.

4 Biais et futurs possibles, vers une incapacité sociétale ?

Tout, autour de nos perspectives digitales, nous démontre que les biais sont les nouveaux « virus cognitifs » encryptés et pour l'instant relativement et silencieusement acceptés. Nous sommes également les témoins et savons que le progrès technologique, dont la course en avant est décuplée par l'économie des marchés, va s'accélérer inexorablement dans les mois et les années à venir. Ceci du fait des applications sectorielles et transversales qui sont dépendantes des algorithmes, pour traiter des masses de données et orienter, voire modifier, en tout cas influencer de nombreuses décisions de notre vie quotidienne.

Nous sommes liés et ligotés :

Par les capacités citoyennes de corrections dans leur mise en œuvre

De nombreux citoyens du monde seront encore plus distancés et disqualifiés qu'à ce jour, et il faudrait prendre des mesures de protections à leurs égards, dès aujourd'hui et pour longtemps. Ceci sans vouloir oublier ou nier les services rendus par les algorithmes, puisque nous sommes bien dans la société du « Et-en-même-temps… », tout étant présent et juxtaposé en pleine cohabitation jusqu'à l'incohérence. Les algorithmes nécessitent une vigilance citoyenne d'autant plus accrue à leurs sujets, qu'ils sont actifs dans tous les domaines de nos vies, mais nous pouvons ressentir d'ores et déjà qu'ils risquent de gagner la bataille de la vitesse, plus rapide que les mesures lentes ne servant pas les intérêts latents de ceux pouvant les diminuer.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Par les difficultés d'entente Internationale dont celles européennes

Nos propos restent à visée nationale et des mesures possibles devraient aussi pouvoir concilier, d'une part le retard pris par l'Europe en matière de politiques numériques actives sur les territoires, ainsi qu'une déjà tardive régulation contre les biais algorithmiques qui ne devrait pas pouvoir s'aggraver davantage, alors qu'il y a d'autres fronts sanitaires ou économiques directs, en apparence relative, plus urgents. Et, d'autre part, un risque de déstabilisation sociale se profile, même en ne considérant que la part dite « algorithmique ».

Par la multiplication des applications sur les marchés et l'incapacité à les contrôler

Il est vrai également que le cadre juridique en matière digitale a proposé déjà de nombreuses solutions, toutefois, la puissance publique, ayant déjà et dès aujourd'hui de grandes difficultés à faire appliquer le règlement européen sur la protection des données, serait bien plus difficilement en mesure de contrôler des algorithmes, et leurs biais associés, au sein de dizaines de milliers de cycles de traitement, sur des secteurs très divers et dont la progression sur les marchés, est toujours croissante.

Ces évolutions ont une vitesse sans précédent qui réduit notre rapport au temps. Par cette accélération croissante, on risque fort de déstabiliser les individus, et de laisser de côté une grande part de la population, dans la compréhension de ces mécanismes dont les algorithmes ne sont pas uniquement le sujet : Les sciences humaines et toutes les sciences dites « douces » sont nécessaires pour appréhender ce qu'impliquent les sciences dites « dures », avec une dangereuse dissymétrie, du rapport qu'elles ont toutes deux, par rapport à leur temps d'action.

5 En conclusion : Comment minimiser l'emprise des biais cognitifs appliquée aux algorithmes ?

5.1 L'urgence d'une ouverture importante vers les acteurs et les citoyens

Bien que de nombreuses recherches soient encore nécessaires à une meilleure compréhension de la cognition humaine et leurs applications correctives, nous croyons que les quelques grandes notions très rapidement esquissées ici offrent dès maintenant, et de façon relativement compréhensible pour tous, quelques inquiétudes et quelques possibles rééquilibrages à venir permettant de détecter et lutter contre les diverses sortes de biais cognitifs, au moins ceux les plus sensibles.

Avec comme grandes lignes de forces :

Ne pas recourir à la puissance publique pour se défausser de notre responsabilité individuelle. Ne pas produire de lois portant sur les biais des algorithmes ni un contrôle des algorithmes par l'État, ce serait trop long, trop complexe et avec des intérêts contradictoires, mais avec l'aide de la puissance publique, agir localement au niveau des territoires, pour améliorer la connaissance citoyenne des risques et opportunités de l'IA et des algorithmes avec leurs biais, pour prévenir les biais en développant de bonnes pratiques, en développant notamment des formations pour tous ceux qui produisent ou utilisent des algorithmes (chartes, diversité des équipes)

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

et en formant les techniciens et ingénieurs aux risques de biais. L'obligation de tester les algorithmes avant utilisation en s'inspirant des études sur les risques (militaire, santé, nucléaire, spatial…) devra en outre être systématisée. Cette approche implique d'adopter une démarche d'équité active autorisant l'usage de variables sensibles dans le strict but de mesurer les biais et d'évaluer les algorithmes. Mais aussi d'auditer et encadrer les algorithmes à fort impact pour limiter leurs risques (Santé, mobilités, finances, police…) et, être bien plus exigeant pour les algorithmes les plus sensibles (Droits fondamentaux, sécurité physique ou psychologique, conservation de l'accès à des services essentiels). Il sera nécessaire à cette fin, de mettre à disposition des bases de données de tests, publiques, pour permettre aux entreprises d'évaluer les biais relatifs à leur méthodologie. Enfin, il faudra soutenir l'émergence de labels pour renforcer la confiance du citoyen dans les usages critiques et accélérer la diffusion des algorithmes les plus bénéfiques.

5.2 Avec une responsabilité citoyenne majeure mais encore bien peu éduquée dans et par les institutions

Éduquer, former et accompagner sont les maîtres mots de ces évolutions numériques et algorithmiques. Former les personnes à comprendre est une des clés. Les personnes sont les mieux placées pour défendre leurs propres intérêts, droits d'usages et d'impacts, mais encore faut-il qu'on leur donne tous les éléments de compréhension. Ce n'est qu'à ces conditions qu'un meilleur dialogue, préventif et prospectif, entre les experts et les citoyens, peut s'instaurer, et que nous pourrons ensemble faire sens, et société… Or tout cela reste, in fine, du domaine de l'orientation institutionnelle à visées budgétaire et éducative, donc politique, avec une réelle dimension territoriale, que seuls les citoyens avec leurs élus peuvent impulser, avec des actions pour prévenir des risques bien plus systémiques de déstabilisation sociétale, tout en utilisant aussi, sans les nier, les formidables potentialités des algorithmes qui nous entourent dans de nombreux domaines.

Les évolutions de nos systèmes de formation au cœur de nombreuses problématiques devraient ainsi tendre vers plus de transversalité, pluridisciplinaires par nature et d'analogies fécondes, pour nous aider à appréhender notre monde en transformation. Cette problématique, comme d'autres, reste toujours une question de volonté politique, de compétences, de contexte et de dosage dans la régulation, qui demeure à faire évoluer en permanence.

Bibliographie

Andler D. ; Fagot-Largeault, A. et Saint-Sernin, B. « Processus cognitifs », Philosophie des sciences I, Paris, Gallimard/Folio (Essais), 2002 (p. 226-407), 672 p.

Barenboim, E. Tian, J. et Pearl J. 2014. Recovering from Selection Bias in Causal and Statistical Inference. Proceedings of the 28th National Conference on Artificial Intelligence (AAAI), CE Brodley and P. Stone, eds., AAAI Press, Menlo Park, CA p. 2410-2416

Bolukbasi T., Chang K-W., Zou J., Saligrama V., Kalai, A. 2016. Man is to Computer Programmer as Woman is to Homemaker ? Debiasing Word Embeddings, 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Lebraty, Jean-Fabrice ; Pastorelli-Negre, Ivan, 2018, « Biais cognitifs : quel statut dans la prise de décision assistée ? » https://fdocuments.fr/document/biais-cognitifs-quel-statut-dans-la-prise-de-decision-planifiee-dactivites.html

Megan, P., and Ball. P. 2014. Big Data, Selection Bias, and the Statistical Patterns of Mortality in Conflict. SAIS Review of International Affairs, 34(1), 9-20.

Pasquale, F. 2015. The Black Box Society : The Secret Algorithms That Control Money and Information. Cambridge, MA : Harvard University Press.

Steiner ; Pierre, « Introduction au cognitivisme et sciences cognitives », Labyrinthe, mis en ligne le 25 juin 2008 (http://journals.openedition.org/labyrinthe/754)

Tiberghien Guy (dir.), Dictionnaire des sciences cognitives, Paris, Armand Colin, 2002.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928