



HAL
open science

Les biais algorithmiques, fatalité ou symptôme ?

Thibault de Swarte, Jean-Baptiste Janvier

► To cite this version:

Thibault de Swarte, Jean-Baptiste Janvier. Les biais algorithmiques, fatalité ou symptôme ?. Les biais de l'intelligence humaine et de l'intelligence artificielle, Institut Psychanalyse et Management, Nov 2021, Paris, France. <hal-03890486>

HAL Id: hal-03890486

<https://hal.science/hal-03890486v1>

Submitted on 8 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Les biais algorithmiques, fatalité ou symptôme ?

Thibault de Swarte, Maître de Conférences, IMT Atlantique

Jean-Baptiste Janvier, *Chief Data Scientist*, Société Générale

Résumé

La mesure des biais est essentielle afin de pouvoir modifier chaque fois que nécessaire les paramètres des algorithmes en fonction de valeurs humaines spécifiques : neutralité, loyauté, équité... On se focalise dans ce texte sur l'équité. Mieux on mesure les biais, plus la réflexion sur des choix équitables peut progresser. Le « rasoir d'Ockham » permet de formuler une définition simple du biais algorithmique comme symptôme, notamment pour ce qui concerne la part non modélisable d'une décision. Cette part correspond à des a priori, des stéréotypes, des cas exceptionnels, des relations de causes à effet non observables dans les observations passées, ou sur des événements futurs. Un biais est aussi un « phénomène subjectif qui traduit les états morbides et qui est lié aux troubles fonctionnels ou lésionnels qui le déterminent ». L'algorithmique peut alors permettre de révéler un symptôme humain resté jusque-là invisible ou mal connu. Le point crucial est ici la subjectivité. La machine en manque définitivement quand l'humain est parfois dépassé par elle. La question est donc d'articuler la mesure des biais et la réflexion éthique sur le machine learning plutôt que de les opposer. C'est en ce point précis que peuvent se rencontrer deux théories des biais en première analyse incompatibles : l'une issue de l'algorithmique et l'autre issue de la théorisation de l'inconscient humain et de la subjectivité de chacun.

Summary

The measurement of biases is essential in order to be able to modify the parameters of the algorithms whenever necessary according to specific human values: neutrality, fairness, equity, etc. In this text, we focus on equity. The better we measure the biases, the more the reflection on fair choices can progress. Ockham's razor" allows us to formulate a simple definition of algorithmic bias as a symptom, especially concerning the part of a decision that cannot be modelled. This part corresponds to a priori, stereotypes, exceptional cases, cause and effect relationships not observable in past observations, or on future events. A bias is also a "subjective phenomenon that reflects morbid states and is linked to the functional or lesional disorders that determine it". Algorithms can then reveal a human symptom that was previously invisible or poorly known. The crucial point here is subjectivity. The machine definitely lacks subjectivity when humans are sometimes overwhelmed by it. The question is therefore to articulate the measurement of bias and the ethical reflection on machine learning rather than to oppose them. It is at this precise point that two theories of bias, which are incompatible at first sight, can meet: one stemming from algorithmic theory and the other from theorizing the human unconscious and the subjectivity of each individual.

Mots clés

Intelligence artificielle, éthique, biais, modélisation, organisation, Lacan

Key words

Artificial intelligence, ethics, bias, modeling, organization, Lacan

Ce qui nous émeut... ce n'est pas que le monde ne parvienne pas à être entièrement juste (qui s'y attendrait parmi nous ?), c'est qu'il existe autour de nous des injustices manifestement réparables, que nous voulons éliminer ».

Amartya Sen (Prix Nobel d'Economie, 1998)

Introduction

Les biais algorithmiques seraient-ils une sorte de *fatum*, de fatalité à laquelle seraient condamnées nos sociétés et organisations de plus en plus complexes et numérisées ? Ne seraient-ils pas plutôt un symptôme des interrogations que suscite l'indispensable dialogue entre humains, dialogue qui a vocation à être nourri et enrichi par l'intelligence artificielle (IA) ?

Le constat est aujourd'hui bien établi que les biais algorithmiques sont nombreux et augmentent au fur et à mesure que l'usage des algorithmes se répand. Ce type de biais doit par conséquent être quantifié afin de pouvoir être qualifié. La démarche de recherche complémentaire : du qualitatif vers le quantitatif ou de la réflexion vers la mesure est pour nous tout aussi légitime.

Ces deux étapes quantitatives puis qualitatives sont à nos yeux indissociables, tant il est vrai que c'est l'interprétation de résultats qui peut conduire à une herméneutique. Aux yeux d'Aristote, l'*hermeneutikè* était avant tout de la logique.

Les biais algorithmiques de l'intelligence artificielle ne sont pas une fatalité, mais un symptôme.

Plan

On rappellera tout d'abord le contexte de la recherche (partie 1). Dans « la formation de l'esprit scientifique », Bachelard (1970) indique qu'« il faut réfléchir pour mesurer et non pas mesurer pour réfléchir ». Avec les biais algorithmiques, il nous a semblé préférable d'inverser les termes de la proposition. On devra tenir les deux bouts de la chaîne et mesurer pour réfléchir (partie 2) puis réfléchir pour mesurer (partie 3).

1. Contexte de la recherche

Notre texte vise à dépasser une opposition assez fréquente et quelque peu stérile entre l'intelligence humaine, l'intelligence émotionnelle, l'intelligence artificielle et l'algorithmique. Ce chapitre repose sur un traitement statistique de biais algorithmiques et sur une interprétation de ce traitement.

D'un point de vue épistémologique, il s'agit d'un travail à l'articulation du traitement de données algorithmiques, d'une recherche d'amélioration des capacités de gestion des biais algorithmiques dans une organisation et d'une analyse plus spéculative des questions ci-dessus basée sur les travaux de Jacques Lacan.

1.1. Travaux antérieurs

Des travaux antérieurs sur « éthique et intelligence artificielle » ont été conduits. De nombreux domaines applicatifs de l'intelligence artificielle mettent ainsi en présence plusieurs acteurs qui doivent interagir, décider conjointement et coopérer (Vallée, Bonnet, de Swarte, 2018). Le terme utilisé, notamment en IA, est souvent celui d'agents. Cela permet d'effectuer un *distinguo* entre les agents artificiels, notamment les algorithmes et les agents humains réputés naturels.

1.2. Définitions

L'antonyme de biais est « droiture, simplicité ». Ce sont là les qualités qu'on attend d'un algorithme même si la simplicité peut entrer en contradiction avec la droiture et réciproquement. Un algorithme sera biaisé lorsque son résultat n'est pas neutre, loyal ou équitable.

Nous choisissons ici la définition figurée du symptôme donnée par Larousse « ce qui permet de deviner un état de fait à venir ou mal connu » plutôt que la définition littérale. Issue de la médecine et de la psychopathologie elle serait inappropriée s'agissant d'une intelligence artificielle. En revanche, l'algorithmique peut permettre de révéler un symptôme humain resté jusque-là invisible ou mal connu.

1.3. Méthodologie

Ce texte résulte de la collaboration entre un ingénieur *chief data scientist* (responsable scientifique des données) dans une grande organisation européenne et un enseignant-chercheur en sciences de gestion ayant publié différents travaux sur la psychanalyse lacanienne dans les organisations « high tech », notamment numériques.

La méthode utilisée est :

- mathématique pour ce qui concerne l'éthique et l'équité dans la mise en œuvre des applications de *Machine Learning*, en français « l'apprentissage machine ». L'anglicisme « Machine Learning » sera noté ici « ML ». Une étude de cas de mathématiques appliquées montrant que l'IA et les algorithmes peuvent permettre d'atteindre un certain niveau d'intégrité et d'éthique a été réalisée durant l'été 2020.

- de type *grounded theory* en sciences sociales car ces méthodes comprennent la collecte et l'analyse simultanées des données, l'utilisation de cadres interprétatifs différents l'échantillonnage théorique et une volonté d'intégration de catégories hétérogènes dans une théorie homogène.

- à l'articulation du machine learning, du management et de la psychanalyse lacanienne pour la partie interprétative : quelle est la part subjective et non mesurable existant dans la relation entre le langage d'un sujet humain et « l'apprentissage machine » ?

1.4. Intelligence humaine et intelligence artificielle

On peut imaginer 4 situations simplifiées possibles, toutes basées sur l'hypothèse du développement de l'intelligence artificielle (IA) et de l'intelligence humaine (IH).

	Intelligence humaine + (IH)	Intelligence artificielle + (IA)
Opposition	En amplifiant les capacités de l'IH, l'IA générerait des « mondes » totalitaires (ex. crédit social chinois)	L'IA détruirait les capacités humaines de raisonnement et la richesse de la vie psychique
Articulation	Permet une aide à la décision humaine plus objective	Automatise des tâches répétitives

1.5. Introduction des biais algorithmiques dans le raisonnement

Un algorithme est biaisé lorsque son résultat n'est pas neutre, loyal ou **équitable**. Les biais algorithmiques peuvent conduire à des situations de discrimination.

On se focalise ci-dessous sur le point particulier de **l'équité algorithmique**

	Biais Intelligence humaine + (IH)	Biais Intelligence artificielle + (IA)
Inéquité	Un groupe d'experts non aidé par l'IA est moins objectif	Cette situation « perdant-perdant » est souvent évoquée dans les <i>media</i>
Équité	Nécessité d'une analyse mathématique des biais inévitables afin de promouvoir des algorithmes plus équitables	Un groupe d'experts aidé par l'IA est plus objectif

2. Mesurer pour réfléchir

D'un côté l'IA suscite des interrogations du fait de l'émergence de nouveaux types de biais, de l'autre elle donne une formidable opportunité d'élargir le champ du « manifestement réparable ». L'enjeu de ce texte est de tenter de dépasser une critique parfois mal documentée des algorithmes et de montrer comment on peut imaginer de corriger les biais algorithmiques en les mesurant afin de tendre vers plus d'éthique.

Le terrain d'investigation est celui d'une grande organisation européenne qui cherche à articuler au mieux les intelligences humaines et artificielle, et à promouvoir l'éthique et l'équité dans les prises de décisions.

2.1. L'IA : un concept flou et polysémique.

Le terme d'intelligence artificielle est polysémique. Il associe en effet deux concepts très larges : « intelligence » et « artificiel ». Le simple fait d'évoquer ce que pourrait être la « bêtise naturelle », opposé de l'IA suffit à mettre en évidence l'ambiguïté du concept d'IA.

Un gastéropode est peut-être naturellement « bête » par rapport à l'intelligence humaine mais du point de vue du biologiste, parler de bêtise serait inapproprié. On préférera donc dans ce texte se focaliser sur l'apprentissage machine, en Anglais le *Machine Learning*.

2.2. Principe fondamental du Machine Learning

Le terme d'Intelligence Artificielle recouvre plusieurs concepts. L'un des principaux est celui de Machine Learning (Minsky, 1963), puis beaucoup plus récemment (Bonaccorso, 2017). Le Machine Learning consiste à identifier automatiquement, via un algorithme c'est-à-dire une suite d'instructions logiques, un modèle à partir d'un ensemble d'observations en minimisant une fonction de coût, elle-même somme d'une mesure d'écart entre les prédictions et les observations et d'une mesure de complexité du modèle.

Ce problème se formalise sous la forme :

$$\text{Min (mesure écart [prédiction, observation] + mesure complexité du modèle)}$$

Le terme de limitation de la complexité du modèle est appelé « Régularisation » et vise à limiter la variance des résultats, pour améliorer la stabilité du modèle.

Si la mise en œuvre prend des formes différentes selon les méthodes, l'idée reste toujours de limiter le nombre de coefficients du modèle ainsi que leurs valeurs.

Par exemple dans XGBoost (Chen, 2016), une méthode se basant sur des arbres de décisions, la mesure de complexité du modèle est une fonction de la taille des arbres et du poids de leurs coefficients. Dans les réseaux de neurone (LeCun, 2015), la limitation des poids (Nowlan, 1992) et la limitation des coefficients via méthode du « dropout » (Srivastava, 2014)

En faisant cela l'algorithme met en œuvre le principe du « Rasoir d'Ockham » (Jennings, 2007). Ce principe philosophique remonte aux travaux de Guillaume d'Ockham (XIV^e siècle) et vise à éliminer des explications improbables d'un phénomène.

2.3. Le rasoir d'Ockham : simplifier les problèmes complexes

Dès la fin des années 80, lors de la première vague de développement de l'IA. Blumer & al. (1987) ont posé le principe de la pertinence du rasoir d'Ockham pour un sujet tel que le nôtre. Vapnik (1995) a ensuite posé le principe de la minimisation de la taille d'un modèle en IA.

« *Le rasoir d'Ockham est le principe philosophique selon lequel il ne faut pas introduire inutilement des entités ou des hypothèses. Il est souvent interprété comme signifiant qu'il faut chercher l'explication la plus simple d'un phénomène¹* ».

L'espace dans lequel le modèle est cherché dépend de l'algorithme. La taille de cet espace grandit au fur et à mesure que des algorithmes plus performants sont proposés, permettant de générer des modèles plus complets, capable de prendre en compte des relations entre variables plus complexes.

2.4. La prise de décision : le modélisable et le non-modélisable

Une prise de décision se base sur un ensemble d'éléments hétérogènes : observations de cas similaires, connaissances de relations de causalité, d'événements futurs, etc...

Elle peut se décomposer entre une part modélisable, et une part non modélisable. La part modélisable correspond aux règles qui peuvent être définies à partir de la seule observation des décisions passées, sans a priori. La part non modélisable correspond à d'autres éléments, qui peuvent être factuels (connaissances de relations de causalité, d'événements futurs, connaissance de cas exceptionnels, etc...), ou non factuels (*a priori*, des stéréotypes, etc...).

Le Machine Learning vise à trouver une approximation de la part modélisable. Sur cette partie la machine permet d'avoir une référence objective sur laquelle s'appuyer. Le savoir humain peut venir ensuite en complément pour apporter des corrections prenant en compte des éléments non visibles dans les données.

2.5. Biais et équité : mesurer pour corriger

Le problème de minimisation exposé précédemment implique nécessairement la présence d'un biais. Ce biais est même nécessaire au bon fonctionnement du modèle pour éviter les instabilités des résultats, c'est le dilemme bias-variance (Geman, 1992).

¹ L'utilisation du rasoir d'Ockham en tant qu'outil d'inférence fait l'objet d'une certaine controverse, qui semble être centrée sur la signification de l'expression "explication simple" (Lazar, 2010 p.1).

Dans la conception d'un modèle, ces biais sont observés, et plusieurs méthodes existent pour les limiter et si besoin les corriger. Dans les entreprises, des organismes de gouvernance se mettent en place pour maîtriser les risques liés à l'utilisation des modèles.

Le concept d'équité algorithmique est encore mal défini. 21 définitions différentes ont pu être proposée (Narayanan, 2018).

Nous abordons ici le problème de l'équité comme le problème de l'identification d'un biais caché, qui seraient non observable lors de la mise en œuvre du problème de minimisation tel qu'il est posé précédemment, et qui ne serait pas souhaité.

Il s'agit d'abord d'un problème de mesure : une fois ce biais non voulu mesuré il peut alors être corrigé.

Différentes méthodes ont été développées récemment, avec notamment des boîtes à outils regroupant plusieurs outils de mesure comme « AI Fairness 360 » (Bellamy, Rachel KE, et al, 2018) ou « Aequitas » (Saleiro, Pedro, et al, 2018).

3. Réfléchir pour mesurer

Après avoir mesuré pour réfléchir, il faut réfléchir pour mesurer. Ainsi, lorsqu'on mesure pour réfléchir, on traite des biais (en rouge ci-dessous) et on cherche à les minimiser afin que l'algorithme soit aussi proche que possible de la réalité. Les biais algorithmiques doivent alors :

- Etre identifiés
- Etre mesurés
- Faire l'objet de propositions de corrections
- Permettre l'émergence d'un algorithme plus réaliste et conforme à l'éthique, i.e. non discriminant pour des raisons qui n'auraient pas été identifiées auparavant.

3.1. Une courbe de Gauss inversée ?

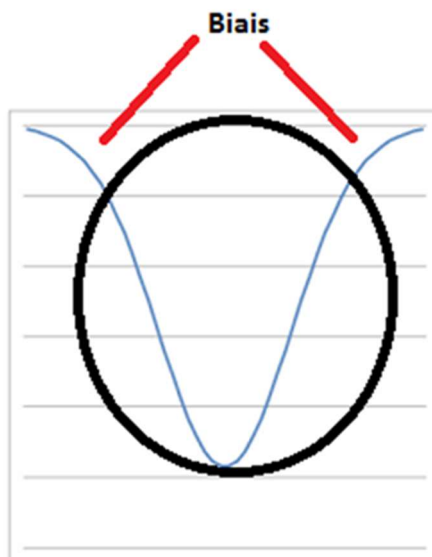


Figure 1 : les biais aux marges d'une courbe de Gauss inversée

Graphiquement, on aurait alors une courbe de Gauss inversée. Le statisticien s'intéresse généralement aux valeurs les plus fréquentes, au centre du schéma. Celui qui travaille sur les biais va étudier les limites à gauche et à droite de la valeur modale. Le cercle représente ce qui n'intéresse pas le lacanien, à savoir le fonctionnement statistiquement normal. On voit bien ci-dessus que les biais pointés en rouge sont susceptibles de bénéficier aussi bien d'une analyse statistique que d'une analyse lacanienne. Les deux épistémès semblent radicalement différentes ex ante. Mais ex post -après coup d'un point de vue lacanien- la focalisation de l'analyse sur les biais tire profit de deux éclairages différents mais au final complémentaires. On peut alors commencer à réfléchir pour mesurer.

3.2. Simplifier vs complexifier : Lacan et le rasoir d'Ockham

Pour Gault (2012), Lacan demeure fidèle au principe d'économie d'Ockham. Nous allons donc tenter de voir comment tenter de concilier la mesure préalable à toute réflexion, la réflexion sans laquelle la mesure est inutile et...D'être nous-mêmes mesurés dans la réflexion.

Lacan cherche en effet à formuler de manière concise une « formule » de l'inconscient et de la subjectivité humaine. Notons qu'il a toujours été intéressé par les avancées de la science : cybernétique dans les années 50, puis tentative de modéliser l'inconscient dans les années 70, suite à sa rencontre avec le groupe de mathématiciens « modernes » nommé Bourbaki.

On cherchera ci-dessous à dégager une convergence entre l'algorithmique et le miroir lacanien sur la question de l'analyse des biais.

3.3. Le réel et son image

Un algorithme n'est somme toute qu'une série de décisions binaires prises par une machine afin de tenter de représenter une réalité de manière dynamique. Aucun algorithme ne prétend englober toute la réalité. Réciproquement, aucune réalité complexe ne saurait être réduite à un algorithme. Le point de triangulation qui nous intéresse ici est celui de la représentation par un humain de la part de réalité qu'un algorithme lui livre. A cet égard, les biais algorithmiques sont une composante essentielle de l'analyse dans la mesure où ils indiquent la limite de ce que l'algorithme peut traiter et de ce qu'il ne peut pas traiter.

Dans sa célèbre représentation du miroir, évoquée dès 1936 à Marienbad, Lacan montre que l'image d'un vase de fleur est toujours différente du réel dudit vase. L'œil à droite du schéma ci-dessous (fig.2) voit un vase qui est reflété par un miroir concave à gauche. Cette concavité renvoie une image inversée du vase dont la base devient le sommet.

Le rasoir d'Ockham s'attache avant tout au col du vase. Plus il est étroit, moins les biais algorithmiques sont nombreux. Inversement, tenter de construire un algorithme de la partie la plus large du vase serait bien trop complexe pour le traitement de données d'aujourd'hui et probablement de demain. D'où la nécessité d'un dialogue humain autour des calculs de l'algorithme visant à laisser à ce dernier son rôle d'outil d'aide à la décision.

L'intervention humaine permet alors de limiter la concavité du miroir et d'approcher au mieux le réel de telle sorte que $i(a)$ tende vers une valeur la plus faible possible. Si l'image (algorithmique) correspond « presque » au réel, alors les biais humains et algorithmiques se corrigent mutuellement et on tend vers une représentation aussi équilibrée et fidèle que possible.

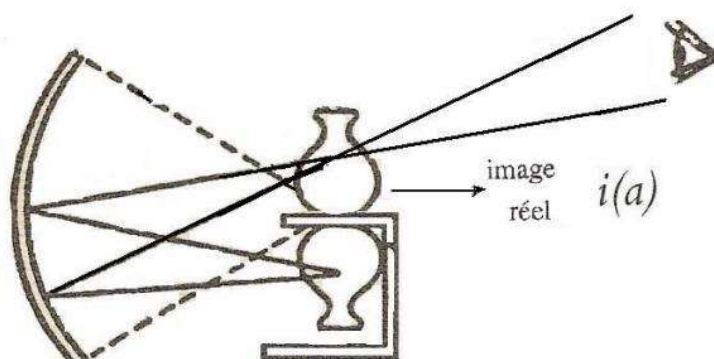


Figure 2

L'image et le réel d'après Lacan

Pour ce qui concerne la question de l'interprétation, sur la base des travaux d'Ockham, l'explication simple est celle qui doit être privilégiée car elle est la plus significative. Pour Lacan au contraire, la signification est construite à la suite d'un travail complexe. Revenons à ce stade de la réflexion aux algorithmes et à la question de ce qui est modélisable ou non.

3.4. Le modélisable et le non-modélisable

L'espace dans lequel le modèle est cherché dépend de l'algorithme. La taille de cet espace grandit au fur et à mesure que des algorithmes plus performants sont proposés, permettant de générer des modèles plus ambitieux et plus complets.

Une prise de décision peut alors se décomposer entre une part modélisable, et une part non modélisable. La part modélisable correspond aux règles qui peuvent être définies *ex ante* au regard de décisions passées. La part non modélisable correspond à des connaissances comme des *a priori* sur des cas exceptionnels, des relations de cause à effet non observables dans les observations passées, ou à des événements futurs.

On pourrait ici citer Freud : « celui qui ne connaît pas son passé est condamné à le répéter ». Mais on traite dans ce chapitre d'algorithmes qui doivent éviter d'être biaisés par des évènements ou des occurrences passées.

Lacan n'a jamais prétendu être un expert en futurologie. En revanche, il a travaillé sur les cas exceptionnels et sur des relations causales qui quoique non observables à l'œil nu sont interprétables. On voit alors que c'est autour de l'interprétation que se noue l'analyse des biais. L'IA et l'algorithmique ont vocation à repousser sans cesse l'espace laissé aux biais. Mais, plus celles-là se développent, plus ceux-ci (les biais) se développent aussi, en tout cas des biais inattendus, peu ou pas prévisibles.

C'est en ce point précis que peuvent se rencontrer deux théories des biais en première analyse incompatibles : l'une issue de l'algorithmique et l'autre issue de la théorisation mathématique de l'inconscient.

Conclusion

Le biais algorithmique : une fatalité ?

Les biais algorithmiques ne sont pas une fatalité. Les algorithmes sont de puissants outils d'aide à la décision qui certes demandent en amont une modélisation parfois très poussée de problèmes plus ou moins complexes. Mais en aval, les algorithmes sont objectifs, rationnels et dénués de sentiments, contrairement aux humains.

La mesure des biais est ainsi essentielle afin de pouvoir modifier chaque fois que nécessaire les algorithmes en fonction des valeurs humaines choisies *ex ante* : neutralité, loyauté, équité... On s'est dans ce chapitre focalisé sur l'équité algorithmique. Mieux on mesure les biais, plus la réflexion sur des choix équitables peut progresser. On comprend bien qu'à l'inverse, la non-mesure des biais serait susceptible d'entraîner des décisions inéquitables sans qu'il soit possible d'avoir connaissance des dites inéquités puisqu'elles resteraient invisibles ou inconscientes.

De quoi le biais algorithmique est-il le symptôme ?

Les biais algorithmiques nous semblent ainsi être avant tout des symptômes. Ils permettent de « deviner un état de fait à venir ou mal connu » à plusieurs niveaux.

Du point de vue des principes fondamentaux du *machine learning*, il s'agit via des méthodes de mathématiques appliquées, d'optimiser l'équité algorithmique sur la base de : *Min (mesure écart [prédiction, observation] + mesure complexité du modèle)*.

Le « rasoir d'Ockham » permet de formuler une définition simple du **biais algorithmique comme symptôme**, notamment pour ce qui concerne la part non modélisable d'une décision. Cette part correspond à des a priori, des stéréotypes, des cas exceptionnels, des relations de causes à effet non observables dans les observations passées, ou sur des événements futurs. Il s'agit bien là « d'états de faits à venir ou mal connus ».

De manière certes plus spéculative, la définition littérale d'un biais s'avère très utile : c'est un « phénomène **subjectif** qui traduit les états morbides et qui est lié aux troubles fonctionnels ou lésionnels qui le déterminent ». Issue de la médecine et de la psychopathologie, une telle définition est bien entendu inappropriée s'agissant d'une intelligence *artificielle* (cf. partie 1 : mesurer pour réfléchir ». En revanche, l'algorithmique peut permettre de révéler un symptôme humain resté jusque-là invisible, mal connu ou inconscient. Le point crucial est ici la **subjectivité**. La machine en manque définitivement quand l'humain est parfois dépassé par elle.

Traiter et dépasser le symptôme : mesurer **et** réfléchir

La question est donc **d'articuler la mesure et la réflexion** plutôt que de les opposer. C'est en ce point précis que peuvent se rencontrer deux théories des biais en première analyse incompatibles : l'une issue de l'algorithmique et l'autre issue de la théorisation de l'inconscient.

Suivant la pensée d'Aristote, Paul Ricœur formule ainsi la définition de l'interprétation : « Dire quelque chose de quelque chose, c'est, au sens complet et fort du mot, interpréter. »

Dans cet article, nous avons cherché à montrer que « quelque chose », à savoir le biais algorithmique devait être identifié, mesuré, analysé et interprété d'une part. D'autre part, une « théorie du biais », notamment la psychanalyse lacanienne, montre que tout est interprétable, y compris le silence, le rire ou un soupir. Il s'agit bien de « dire quelque chose de quelque chose ». C'est à la fois une mise en abyme dans le champ « littéraire » et une suite de décisions et/ou de biais tendant vers l'infini d'un point de vue mathématique.

Éléments de bibliographie (ouvrages consultés).

Arnaud, G., & Vidaillet, B. (2018). Clinical and critical: The Lacanian contribution to management and organization studies. *Organization*, 25(1), 69-97.

- Bachelard G. (1970) *La formation de l'esprit scientifique*, p.213, Librairie Philosophique J. Vrin,
- Bellamy, Rachel KE, et al. "AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. 2018
- Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information processing letters*, 24(6), 377-380.
- Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.
- Chen, T., & Guestrin, C. (2016) Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- Clavurier Vincent, « Réel, symbolique, imaginaire : du repère au nœud », *Essaim*, 2010/2 (n° 25), p. 83-96. DOI : 10.3917/ess.025.0083. URL : <https://www.cairn.info/revue-essaim-2010-2-page-83.htm>
- Crozier, M., & Friedberg, E. (1977). L'acteur et le système.
- De Swarte Thibault, (2012) Sciences de gestion et psychanalyse : Lacan, l'imaginaire et l'organisation high tech. Gestion et management. Université de Rennes 1, 2012. [tel-00784764](tel:00784764)
- Green, C. F. (1989). Business ethics in banking. *Journal of Business Ethics*, 8(8), 631-634.
- De Clerck, F. (2009). Ethical banking. In *Ethical prospects* (pp. 209-227). Springer, Dordrecht.
- De Larosière, J. (2011). L'aléa moral. *Commentaire*, (2), 389-396.
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural computation*, 4(1), 1-58.
- Jennings B.K (2007) On the Nature of Science, *Physics in Canada*, 63(7).
- Lacan J. (1974-1975) *Séminaire XXII : R.S.I.* . Ce séminaire n'est pas publié au Seuil à ce jour. <http://staferla.free.fr/S22/S22%20R.S.I.pdf>
- Lazar, N. (2010). Ockham's razor. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), 243-246. <https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.75>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Léger Claude, « Une leçon de maïeutique lacanienne », *Champ lacanien*, 2004/1 (N° 1), p. 115-125. DOI : 10.3917/chla.001.0115. URL : <https://www.cairn.info/revue-champ-lacanien-2004-1-page-115.htm>
- Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29.
- Loevenbruck, H. (2011). *Le rasoir d'Ockham*. Flammarion.
- Mahesh, B. (2020). Machine Learning Algorithms-A Review. *International Journal of Science and Research (IJSR).[Internet]*, 9, 381-386.

Minsky, M. (1963). Steps toward artificial intelligence. In Feigenbaum, E. and Feldman, J., editors, *Computers and Thought*, pages 406–450. McGraw-Hill, New York

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Narayanan, A. (2018) Translation tutorial: 21 fairness definitions and their politics. In *Conference on Fairness, Accountability, and Transparency*.

Nowlan S. and Hinton G. (1992). Simplifying neural networks by soft weight-sharing. *NeuralComputation*, 4(4).

Saleiro, Pedro, et al. (2018) *Aequitas: A bias and fairness audit toolkit*.

Stoianoff S., « Lacan logicien », *Le Portique* [En ligne], 35 | 2015, document 2, mis en ligne le 10 mars 2016, consulté le 25 juillet 2021. URL : <http://journals.openedition.org/leportique/2827> ; DOI : <https://doi.org/10.4000/leportique.2827>

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Machine Learning Res.* 15, 1929–1958.

Vallée T., Bonnet G. , de Swarte T. Modélisation de valeurs humaines : le cas des vertus dans les jeux hédoniques. *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle*, Lavoisier, 2018, {10.3166/RIA.32.519-546}. {hal-02087704}

Valls Martínez, M. D. C., Cruz Rambaud, S., & Parra Oller, I. Me. (2020). Sustainable and conventional banking in Europe. *PloSéthiue one*, 15(2), e0229420.

Vapnik V.N. (1995) *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995

Sites web consultés

<https://www.universalis.fr/encyclopedie/interpretation/1-de-l-hermeneutique-d-aristote-aux-sciences-humaines/>

Serafino MALAGUARNERA, 15 septembre 2006,

<http://lacan.e-monsite.com/pages/categorie-cachee/l-angoisse-cernee-par-le-schema-optique.html>