

Zones grises du management et intelligence artificielle

Un regard psychanalytique sur les biais algorithmiques

Pour citer la référence

De SWARTE Th., JANVIER J. B. (2025). « Zones grises du management et intelligence artificielle. Un regard psychanalytique sur les biais algorithmiques », *Revue Psychanalyse & Management – Édition académique en Ligne* ISSN 2739-9656 - N° 02_2025, pp. 257-268

Thibault de SWARTE, Directeur adjoint LASCO Lab, IMT

Jean Baptiste JANVIER, Chief Data Scientist, grande banque européenne

Avertissement : l'essentiel de ce texte a été rédigé lors du colloque IP&M qui s'est tenu en Sorbonne en novembre 2021 mais n'avait pas été publié. On l'a actualisé au regard de plusieurs développements de l'IA depuis 2021 et d'un approfondissement théorique (Ockham vs Lacan)

Résumé : Le développement de l'IA s'accompagne de celui de nouvelles zones grises du management parmi lesquelles se trouvent les biais algorithmiques. La mesure de ces 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 la 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.

Mots clés : Intelligence artificielle, éthique, biais, organisation, Lacan, Ockham

Abstract: The development of AI has been accompanied by the emergence of new grey areas in management, including algorithmic bias. Measuring this bias is essential in order to be able to modify algorithm parameters whenever necessary in line with specific human values such as neutrality, fairness and equity. This article focuses on fairness. The better we measure bias, the more we can advance our thinking about equitable choices. Ockham's razor provides a simple definition of algorithmic bias as a symptom, particularly with regard to the non-modelled part of a decision. This part corresponds to preconceptions, stereotypes, exceptional cases, cause-and-effect relationships that cannot be observed in past observations, or 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 therefore reveal a human symptom that was previously invisible or poorly understood. The crucial point here is subjectivity. Machines definitely lack this when humans are sometimes overwhelmed by them. The question is therefore to articulate the measurement of bias and ethical reflection on machine learning rather than opposing them. It is on this precise point that two theories of bias, which at first glance appear incompatible, can meet: one derived from algorithmics and the other from the theorization of the human unconscious and the subjectivity of each individual.

Keywords: Artificial intelligence, ethics, bias, organization, Lacan, Ockham

Introduction

Zones grises du management et intelligence artificielle semblent en première analyse antinomiques dans la mesure où l'IA s'assigne comme projet de diminuer voire de supprimer les zones grises propres à l'intelligence humaine. Un regard psychanalytique sur les biais algorithmiques inverse les termes du problème : l'IA développe de nombreuses zones grises, de nombreux biais que seules l'intelligence humaine, l'éthique et la prise en compte de la subjectivité de chacun peuvent interpréter.

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. A cet égard les travaux de Lacan et d'Ockham sont non seulement compatibles mais leur rencontre s'avère féconde.

On rappellera tout d'abord, sur la base du cas des banques, dans quelle mesure on peut considérer que les biais algorithmiques constituent des zones grises du management. On précisera ensuite le contexte de la recherche. 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 puis réfléchir pour mesurer.

1. Zones grises du management et biais algorithmiques en banque

Les biais algorithmiques, particulièrement dans le secteur bancaire, représentent une zone grise managériale par excellence. Ils remettent en question le mythe de la neutralité technologique et placent les institutions financières face à des dilemmes éthiques, légaux et de gouvernance complexes.

Les Biais Algorithmiques en Banque : Une Zone Grise du Management

L'Opacité et le Mythe de la Neutralité Technologique

Le cœur de cette zone grise réside dans la nature même des systèmes d'Intelligence Artificielle (IA) appliqués à l'évaluation du risque de crédit ou à la personnalisation des offres. Ces systèmes, souvent des modèles d'apprentissage automatique (*Machine Learning*), sont entraînés sur des données historiques qui sont, par essence, le reflet des discriminations et préjugés sociétaux passés (O'Neil, C., *Weapons of Math Destruction*, 2016).

L'opacité de certains modèles (le problème de la « boîte noire ») amplifie cette problématique. En l'absence de traçabilité complète, il devient difficile pour les managers non-experts et même pour les régulateurs d'identifier précisément les variables et les corrélations qui mènent à une décision biaisée (Burrell, J., *How the machine 'thinks' : understanding opacity in machine learning algorithms*, 2016). Cette opacité managériale contraint les entreprises à arbitrer entre la performance prédictive du modèle et l'explicabilité de ses décisions, créant une tension directe entre l'efficacité opérationnelle et la responsabilité éthique.

Le Risque Systémique et le Dilemme Éthico-Légal

Le secteur bancaire est réglementé de manière stricte, et l'utilisation d'algorithmes discriminatoires expose les institutions à un **risque systémique** et réglementaire majeur.

Les algorithmes peuvent utiliser des variables *proxy*, des critères neutres en apparence (tels que le code postal ou le type de navigateur web) qui sont fortement corrélés à des motifs de discrimination prohibés (genre, origine, situation sociale). La Banque Centrale Européenne (BCE) et l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) identifient l'IA comme un facteur de risque systémique émergent, soulignant que les modèles déployés sans supervision humaine peuvent amplifier les inégalités d'accès au crédit et à l'assurance. La discrimination algorithmique est punie par la loi et engendre des sanctions

sévères ainsi qu'une dégradation de la réputation (Défenseur des Droits & CNIL, *Algorithmes : prévenir l'automatisation des discriminations*, 2020).

Cette situation force les managers à faire face à un dilemme éthico-légal : maximiser la performance financière en acceptant un risque de discrimination implicite, ou garantir l'équité au prix potentiel d'une performance amoindrie. Ce choix n'est plus seulement technique, mais se situe au niveau de la stratégie et des valeurs de l'organisation.

Le Déficit de Gouvernance et l'Impératif Managérial

La zone grise des biais algorithmiques se traduit par un flou sur les chaînes de responsabilité traditionnelles (qui est responsable de la décision : l'outil ou son concepteur ?).

La solution académique et institutionnelle réside dans l'établissement d'une Gouvernance de l'IA rigoureuse (Floridi, L. et al., *AI for the good of society : between control and freedom*, 2020). Cela implique une gestion proactive des risques, dépassant la simple conformité réglementaire :

- **Audit d'Équité (*Fairness Audit*)** : Mettre en place des tests réguliers pour évaluer l'impact du modèle sur différents groupes démographiques, en utilisant des métriques d'équité spécifiques (Dastin, J., *Amazon scraps secret AI recruiting tool that showed bias against women*, 2018 - bien que non bancaire, cet exemple est paradigmatique).
- **Explicabilité (XAI)** : Exiger des solutions techniques et managériales pour que les décisions algorithmiques puissent être comprises et justifiées, garantissant le droit des clients à une explication de la décision les concernant (ACPR, *Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier*, 2023).

Au total, la gestion des biais algorithmiques n'est plus une simple préoccupation technique. Elle est devenue un impératif de management stratégique qui requiert l'intégration de l'éthique, du droit et de la science des données au sein d'une politique de gouvernance claire, afin de sortir des zones grises et de préserver la confiance des clients et des régulateurs.

2. 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.

2.1. 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.

2.2. 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.

2.3. 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 » ?

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

Figure 1 Intelligence artificielle et intelligence humaine

2.4. 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	Le groupe d'experts A non aidé par l'IA est moins objectif	Cette situation « perdant-perdant » est souvent évoquée dans les <i>media</i>
Equité	Nécessité d'une analyse mathématique des biais inéquitables afin de promouvoir des algorithmes plus équitables	Le groupe d'experts B aidé par l'IA est plus objectif

Figure 2 Biais de l'IH et biais de l'IA

3. 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 anticiper sur l'évolution du cadre réglementaire. Elle réfléchit dès aujourd'hui à ce que pourraient être des algorithmes éthiques où la subjectivité des experts serait limitée par l'aide à la décision fournie par les algorithmes.

3.1. 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*.

3.2. Principes fondamentaux du Machine Learning

Le terme d'Intelligence Artificielle recouvre plusieurs concepts. L'un des principaux est celui de Machine Learning (Minsky, M. (1963), puis beaucoup plus récemment Bonaccorso, G. (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.

Il existe des méthodes mathématiques d'optimisation de l'équité algorithmique. Ce problème se formalise sous la forme :

Min (mesure écart [prédiction, observation] + mesure complexité du modèle)

La démarche générale se décompose en trois temps :

1) Préciser la donnée sensible à étudier. Pour savoir si un modèle est équitable ou pas, il faut tout d'abord savoir par rapport à quoi. Il est donc essentiel de préciser quelle donnée est susceptible d'être sensible afin d'étudier l'impact de celle-ci sur les résultats du modèle. Par exemple : l'âge, le sexe, les opinions politiques ou religieuses, etc...

2) Déterminer le groupe de référence par rapport auquel on veut étudier l'équité : Il s'agit du groupe privilégié qui est souvent majoritaire et favorisé par l'algorithme. L'objectif est de donc de mesurer l'écart entre ce groupe et les autres groupes qui seront appelés les groupes protégés. Par exemple : considérer le groupe des hommes comme étant un groupe de référence pour le comparer au groupe de femmes dans le cas d'un modèle d'octroi de crédit bancaire.

3) Déterminer la métrique d'équité à calculer :

Le bon choix de la métrique est essentiel pour bien mener l'étude. En effet, ce choix dépend du système décisionnel à auditer.

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.

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

Des réponses sont apportées 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 ».

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).

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.

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 au regard des décisions passées, la part non modélisable correspond à des connaissances comme 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.

4. 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 :

- Être identifiés
- Être 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.

4.1. Une courbe de Gauss inversée ?

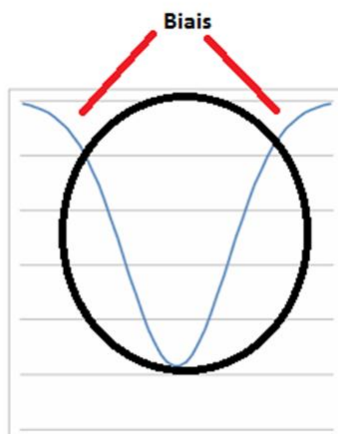


Figure 3 : les biais aux marges d'une courbe de Gauss inversée

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.

4.2. Lacan vs Ockham

L'étude des biais algorithmiques peut être éclairée par un cadre théorique paradoxal mais puissant, articulant la cybernétique et le stade du miroir de Jacques Lacan. Cette convergence met en lumière la formation et la persistance des biais en les percevant comme des "erreurs structurées" inhérentes à tout système, qu'il soit psychique ou technologique.

4.2.1. L'Inconscient comme Système Logique : Cybernétique et Biais

L'Inconscient Formalisé

Lacan s'est intéressé à la cybernétique (science du contrôle et de la communication) dans les années 1950 (Séminaire II, *Le Moi dans la théorie de Freud...*), y voyant un outil pour formaliser la structure de l'inconscient. L'inconscient n'est plus un lieu obscur, mais un réseau logique structuré par les lois du langage et de la théorie de l'information.

Le Biais comme Boucle Répétitive

Dans ce cadre, le concept de biais dans un système algorithmique trouve un écho dans la logique psychique :

- Biais Algorithmique : Une distorsion systématique due à des boucles de *feedback* ou à des données initiales.
- Biais Subjectif : Le symptôme ou la névrose obsessionnelle peuvent être conceptualisés comme des boucles logiques rigides et répétitives (compulsion de répétition). Ces boucles sont des "erreurs structurées" qui résistent à l'information nouvelle (Lacan, 1954-1955).

L'algorithme se comporte de manière biaisée lorsque sa logique est trop fermée et répétitive, incapable d'intégrer l'aléa. Le but de la psychanalyse, comme de l'audit algorithmique, est de "déboucler" cette logique.

4.2.2. Le Stade du Miroir : Matrice du Biais Imaginaire

Le Moi : Une Fiction Aliénante

Le stade du miroir (Lacan, 1949) est, par excellence, le lieu de la constitution d'un biais fondamental : le Moi.

- L'Illusion d'Unité : L'enfant s'identifie à son image spéculaire, totale et unifiée (Moi Idéal), alors que son corps réel est encore immature et fragmenté.
- Le Biais de Méconnaissance : Le Moi ainsi constitué est un leurre et une fiction aliénante (Lacan, 1966). Ce biais imaginaire est la source de toutes les illusions de contrôle, de maîtrise et de rationalité parfaite. Le Moi filtre et réinterprète la réalité pour maintenir son illusion d'unité, d'où la résistance aux prises de conscience.

Résonance avec les Biais de l'IA

Cette analyse résonne avec l'étude des biais dans l'IA :

1. Le Biais comme Filiation : Les algorithmes, qui modélisent la décision ou le "sujet" (client, emprunteur), sont structurés par des catégories humaines et des données historiques (biais culturels, sociaux).
2. Le Rasoir d'Ockham : L'économie lacanienne (recherche de la structure minimale et rigoureuse de l'inconscient) peut être vue comme une application radicale du Rasoir d'Ockham au domaine psychique. Ce principe d'économie de pensée (*Entia non sunt multiplicanda praeter necessitatem*) invite à déconstruire les entités illusives. En IA, cela se traduit par le refus de l'entité illusoire de la neutralité algorithmique et la recherche de la structure *minimale et nécessaire* du modèle, quitte à en accepter la complexité formelle (mathèmes, formules de la sexuation).

Le travail lacanien rappelle ainsi que tout système (psychique ou algorithmique) construit sa "réalité" à partir d'un point de vue fondamentalement biaisé (l'image du miroir ou le jeu de données initial).

4.2.3. Simplicité vs. Complexité : La Fidélité Paradoxale de Lacan à Ockham

Principe	Rasoir d'Ockham	Psychanalyse Lacanienne	Convergence
Objectif	Favoriser l'hypothèse la plus simple et parcimonieuse.	Décrire la structure minimale et nécessaire du sujet.	Recherche d'une rigueur formelle extrême .
Méthode	Réduction des entités théoriques.	Introduction de concepts complexes (RSI, objet a , $\$$ barré).	Minimalisme formel malgré la densité conceptuelle .
Leurre	Rejeter les entités métaphysiques inutiles.	Refuser l'illusion d'un Moi unifié et d'un Autre complet (L'Autre, $\$$ barré).	Refus des simplifications psychologiques faciles (Gault, 2012).

Figure 4 Lacan vs Ockham : vers une convergence ?

La démarche de Lacan, bien que complexe en apparence, est une recherche de concision et de rigueur formelle extrêmes (via la logique, la topologie et l'intérêt pour des groupes comme Bourbaki). Il s'agit d'une fidélité paradoxale au principe d'économie : utiliser la complexité pour identifier et déconstruire les entités illusoire (comme le Moi autonome ou la neutralité de l'algorithme) et atteindre la formulation la plus économe et nécessaire de la structure du sujet.

En somme, la tension est entre la complexité des concepts (RSI, objet a) nécessaires pour échapper aux simplifications psychologiques faciles, et la rigueur minimaliste (mathèmes, formules) qui cherche la formulation la plus économe et nécessaire pour décrire la structure de l'inconscient. On cherchera ci-dessous à dégager une convergence entre l'algorithmique et le miroir lacanien sur la question de l'analyse des biais au travers de l'analyse du réel et de son image.

4.3. 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.

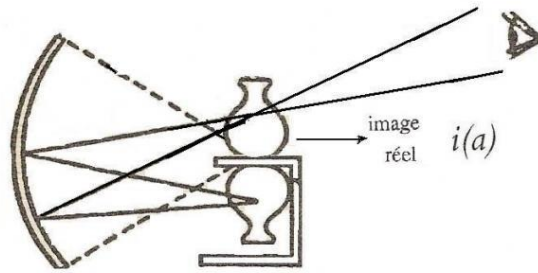


Figure 5 : L'image et le réel d'après Lacan

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.

4.4. 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 événements 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

Nous étions partis d'une interrogation sur le fait que la gestion des biais algorithmiques n'est pas une simple préoccupation technique. Elle est devenue un impératif de management stratégique qui requiert l'intégration de l'éthique, du droit, de la psychanalyse et de la science des données au sein d'une politique de gouvernance claire, afin de sortir des zones grises et de préserver la confiance des clients et des régulateurs.

Le biais algorithmique : une fatalité ?

Les biais algorithmiques ne sont pas une fatalité. Les algorithmes sont de puissants outils d'aide à la décision. En aval, ils sont objectifs, rationnels et dénués de sentiments, contrairement aux humains. La mesure des biais est essentielle pour modifier les algorithmes selon les valeurs humaines choisies ex ante : neutralité, loyauté, équité. Mieux on mesure les biais, plus la réflexion sur des choix équitables progresse. À l'inverse, la non-mesure laisserait les inéquités invisibles ou inconscientes.

De quoi le biais algorithmique est-il le symptôme ?

Les biais algorithmiques nous semblent avant tout être des symptômes. Le « rasoir d'Ockham » formule une définition simple du biais comme symptôme de la part non modélisable d'une décision. Cette part correspond aux a priori, stéréotypes, cas exceptionnels, ou relations de causes à effet non observables dans les données passées. De manière plus spéculative, la définition littérale d'un biais – un « phénomène subjectif qui traduit les états morbides » – s'avère pertinente. Bien qu'inappropriée pour l'IA, l'algorithmique peut révéler un symptôme humain resté invisible ou inconscient. La machine manque de subjectivité quand l'humain en est parfois dépassé.

Traiter et dépasser le symptôme : mesurer ET réfléchir

La question est d'articuler la mesure et la réflexion. Deux théories des biais, l'une algorithmique, l'autre issue de la théorisation de l'inconscient, se rencontrent en ce point. Suivant Paul Ricœur, l'interprétation consiste à « Dire quelque chose de quelque chose ». Nous avons montré que le biais algorithmique devait être identifié, mesuré, analysé et interprété. La psychanalyse lacanienne souligne que tout est interprétable. Le champ algorithmique devient ainsi un espace d'interprétation riche, à la croisée des mathématiques et des sciences humaines.

L'IA quantique au défi de l'inconscient

Afin de tendre vers une gouvernance aussi éthique que possible, notons pour une recherche future que l'interprétation des biais est sur le point d'être radicalement complexifiée par l'émergence de l'Intelligence Artificielle Quantique (IAQ). Si l'IA classique nécessitait de déchiffrer la "boîte noire", l'IAQ pourrait créer des modèles d'une opacité et d'une puissance de calcul sans précédent, rendant l'identification et la mesure des biais beaucoup plus ardues. Ce saut technologique nous oblige à anticiper dès aujourd'hui la nécessité d'une méta-gouvernance : comment interpréter, contrôler et aligner sur nos valeurs humaines des systèmes dont la complexité formelle défie potentiellement toute compréhension humaine conventionnelle car ils ne savent pas tenir compte de l'existence de l'inconscient humain ?

Bibliographie (réf. ouvrages consultés en 2021).

- 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,
- Banque de France Autorité de Contrôle Prudenciel et de Résolution, <https://acpr.banque-france.fr/>
- 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.
- 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}](#)
- 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* (p. 209-227). Springer, Dordrecht.
- De Larosière, J. (2011). L'aléa moral. *Commentaire*, (2), 389-396.
- 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>
- 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.
- 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>
- 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

Références bibliographiques additionnelles (réf. Consultées en 2025)

Monographies et Articles Scientifiques

- Burrell, J. (2016). « How the machine ‘thinks’: Understanding opacity in machine learning algorithms ». *Big Data & Society*, vol. 3, n° 1, p. 1-12. DOI : 10.1177/2053951715622512.
- Défenseur des Droits & Cnil. (2020). *Algorithmes : prévenir l'automatisation des discriminations*. Paris : Rapport conjoint.
- Evans, Dylan. (1996). *Dictionnaire d'introduction à la psychanalyse lacanienne*. Paris : L'Harmattan. ISBN 978-2738445195.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). « AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations ». *Minds and Machines*, vol. 28, n° 4, p. 689–707. DOI : 10.1007/s11023-018-9482-5.
- Gault, C. (2012). « Lacan et le rasoir d'Ockham ». *Le Carnet Psy*, n° 167, p. 1-2.
- Lacan, Jacques. (1966). *Écrits*. Paris : Seuil.

- Lacan, Jacques. (1978). *Le Séminaire, Livre II : Le Moi dans la théorie de Freud et dans la technique de la psychanalyse*. Texte établi par Jacques-Alain Miller. Paris : Seuil. (Édition originale du séminaire : 1954-1955).
- Lacan, Jacques. (1991). *Le Séminaire, Livre XVII : L'Envers de la psychanalyse*. Texte établi par Jacques-Alain Miller. Paris : Seuil. (Édition originale du séminaire : 1970).
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York : Crown Publishing Group.

Ressources Électroniques (Sites Web et Articles en Ligne)

- Acpr. (2023). *Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier*. [En ligne]. Paris : Autorité de Contrôle Prudentiel et de Résolution. [Consulté le : 3 Novembre 2025]. Disponible sur : https://acpr.banque-france.fr/sites/default/files/medias/documents/2023_gouvernance_algorithmes_ia.pdf
- Culture Banque. (2025). *IA générative et métiers bancaires*. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.culturebanque.com>
- Dastin, J. (2018). « Amazon scraps secret AI recruiting tool that showed bias against women ». *Reuters*, 10 octobre 2018. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- Défenseur des Droits & Cnil. (2020). *Algorithmes : prévenir l'automatisation des discriminations*. [En ligne]. Paris : Rapport conjoint. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.defenseurdesdroits.fr/fr/publications/rapports/algorithmes-prevenir-lautomatisation-des-discriminations>
- Fédération Bancaire Française. (2025). *Étude interbranches sur l'IA générative*. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.fbf.fr>
- Malaguarrera, Serafino. (2006). *De l'herméneutique d'Aristote aux sciences humaines*. [En ligne]. *Universalis.fr*, 15 septembre 2006. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.universalis.fr/encyclopedie/interpretation/1-de-l-hermeneutique-d-aristote-aux-sciences-humaines/>
- Malaguarrera, Serafino. (2006). *L'angoisse cernée par le schéma optique*. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <http://lacan.e-monsite.com/pages/categorie-cachee/l-angoisse-cernee-par-le-schema-optique.html>
- Observatoire des Métiers de la Banque. (2025). *Communiqué de presse sur l'IA générative, avril 2025*. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <https://www.observatoire-metiers-banque.fr>
- Malaguarrera, Serafino. (2006). *L'angoisse cernée par le schéma optique*. [En ligne]. [Consulté le : 3 Novembre 2025]. Disponible sur : <http://lacan.e-monsite.com/pages/categorie-cachee/l-angoisse-cernee-par-le-schema-optique.html>