Architecture of AI Bias and Human Lucidity - NotebookLM

Exported on: 18/12/2025 07:07:55

Le concept de biais (ou « biais de l'IA ») fait référence à une **distorsion systématique** qui influence l'équité et les performances des modèles d'intelligence artificielle (IA), favorisant certains résultats ou groupes au détriment d'autres. Ce phénomène est également appelé biais de *machine learning* ou biais d'algorithme.

Définitions du Biais

Le biais peut être défini de plusieurs manières, selon son origine et sa manifestation :

• **Définition générale de l'IA :** Le biais de l'IA désigne l'apparition de **résultats faussés ou biaisés en raison de préjugés humains** qui contaminent les données d'entraînement ou l'algorithme d'IA d'origine. Il s'agit d'une **déviation par rapport à un résultat censé être neutre, loyal ou encore équitable**.

• **Perspective humaine/cognitive :** Les biais inconscients, souvent à la racine des biais de l'IA, sont des attitudes ou des stéréotypes qui affectent la compréhension, les actions et les décisions d'un individu de **manière inconsciente, involontaire ou sans contrôle intentionnel**. Un biais cognitif est une **distorsion dans le traitement de l'information par rapport à la réalité ou à un comportement rationnel**.

Dans le domaine de l'IA, les sources distinguent quatre grandes familles de biais ou de normes : le **biais statistique**, le **biais méthodologique**, le **biais cognitif** et le **biais socio-historique**.

Pourquoi ce terme est crucial

L'importance du terme et de sa gestion tient aux conséquences systémiques et éthiques de l'intégration de l'IA dans la société :

1. **Perpétuation des inégalités :** L'IA apprend à partir de données passées qui contiennent déjà des stéréotypes et des discriminations historiques. Si le jeu de données d'images est biaisé, le contenu généré peut être **néfaste pour certains groupes de personnes**, perpétuant ainsi des stéréotypes nuisibles concernant, par exemple, le genre ou la couleur de la peau. Le problème est que **les biais du passé entraînent des biais pour l'avenir**.

2. **Amplification des préjugés :** Les outils d'IA influencés par les biais humains risquent d'**amplifier ce préjudice à un niveau systémique**. Lorsque l'IA génère des résultats stéréotypés à grande échelle, elle ne fait pas qu'inventer ces clichés, elle leur confère un **vernis d'objectivité** en les reproduisant, ce qui renforce les préjugés discriminatoires dans l'imaginaire collectif.

3. **Conséquences sociales et éthiques :** Les biais dans l'IA peuvent entraîner des **décisions injustes ou préjudiciables** qui affectent négativement les individus et les groupes sociaux. Par exemple, ils peuvent exacerber les inégalités sociales existantes en pénalisant de manière disproportionnée les communautés marginalisées, ou entraîner un **traitement injuste** fondé sur le sexe, l'âge, l'appartenance ethnique ou raciale. Le risque est d'appliquer des **pratiques discriminatoires** dans des domaines critiques comme le recrutement, le crédit ou la justice prédictive.

4. **Enjeux réglementaires :** L'atténuation des biais est devenue un impératif de **conformité réglementaire** et de **gouvernance**. Par exemple, le Règlement européen sur l'IA (AI Act) vise à établir des règles harmonisées pour garantir que les systèmes d'IA respectent les droits fondamentaux et les exigences de sécurité.

Les biais naissent de multiples sources, notamment : les **jeux de données** utilisés pour l'entraînement (s'ils sont non représentatifs ou reflètent des inégalités historiques), la **conception des algorithmes** (les choix des développeurs

ou la structure même de l'algorithme), et l'**interprétation des sorties par les utilisateurs finaux**.

Les biais dans l'intelligence artificielle (IA) et les grands modèles de langage (LLMs) découlent principalement de trois sources interdépendantes, couvrant tout le cycle de vie du modèle : les **données** utilisées pour l'entraînement, la **conception des algorithmes** eux-mêmes, et l'**interaction humaine** avec le système.

1. Les Biais liés aux Données (Source Racine)

La source la plus fondamentale des biais est le jeu de données massives (souvent appelé le *corpus* de pré-entraînement) sur lequel les modèles apprennent. Étant donné que ces données reflètent les archives sociétales et les productions humaines passées, elles sont inévitablement contaminées par des préjugés et des inégalités historiques.

Le phénomène est si direct que les experts affirment que « **les biais du passé entraînent des biais pour l'avenir** ».

Les principaux types de biais découlant de la phase de données comprennent :

• **Biais de Sélection/Échantillonnage :** Il survient lorsque le jeu de données n'est pas **représentatif de la population cible**. Par exemple, si les données sont fortement biaisées envers certaines valeurs pour des variables vulnérables comme le genre ou la couleur de la peau, le contenu généré peut être néfaste pour certains groupes. Si un modèle de reconnaissance faciale est entraîné majoritairement sur des visages d'hommes blancs, il sera moins précis pour identifier les femmes ou les personnes de couleur.

• **Biais Historique (ou Temporel) :** Les données reflètent des inégalités ou des **biais qui existaient au moment de la collecte**, mais qui ne correspondent plus au contexte actuel. Un algorithme de recrutement entraîné sur des données historiques où certains postes étaient majoritairement occupés par des hommes perpétuera cette inégalité.

• **Biais d'Exclusion :** Ce biais se produit lorsque des données importantes sont **omis des jeux de données**, souvent car le développeur n'a pas vu de nouveaux facteurs importants à prendre en compte.

• **Biais de Mesure :** La qualité ou l'exactitude des données diffère d'un groupe à l'autre, ou les variables clés sont mesurées de manière inexacte. Cela peut être causé par des erreurs dans la collecte de données ou des outils de mesure inadéquats.

2. Les Biais Algorithmiques et de Conception

Les biais peuvent survenir même si les données ne sont pas biaisées, en raison des choix effectués lors de la conception et de la mise en œuvre de l'algorithme.

• **Biais Algorithmique Intrinsèque :** Il résulte des choix techniques des développeurs, tels que l'architecture du modèle, les paramètres ou les métriques d'évaluation. Même un **mode de traitement ou de hiérarchisation** des algorithmes peut engendrer des résultats discriminatoires.

• **Biais Cognitifs (du Concepteur) :** Les biais humains (préjugés et stéréotypes) peuvent s'infiltrer dans les systèmes d'IA via les décisions subjectives prises par les équipes de développement, notamment lors de l'étiquetage des données ou du développement du modèle. Le risque est que l'algorithme ne soit, en réalité, qu'une « opinion intégrée aux programmes ». Les développeurs d'algorithmes sont plus susceptibles de créer des systèmes moins capables de distinguer les individus qui ne font pas partie du **groupe majoritaire** dans les données d'apprentissage (*biais d'homogénéité de l'exogroupe*).

• **Problème de Représentation :** La structure interne des LLMs, qui repose sur la découverte de structures latentes dans le langage pour prédire la distribution de texte (le *distributionalisme*), ne dispose **d'aucun moyen pour distinguer les faits normatifs des généralisations statistiques inacceptables**. Cela rend le biais **inhérent et**

inévitable dans la conception actuelle des LLMs, selon certains experts, car ils reproduisent les préjugés présents dans le langage humain sans capacité de jugement normatif.
3. Les Biais liés à l'Interaction Humaine et au Cycle de Vie

Les biais ne sont pas seulement statiques, ils peuvent être introduits ou amplifiés aux étapes de post-traitement et d'utilisation.

• **Biais de Feedback (Boucles de Rétroaction) :** Après le déploiement, les retours des utilisateurs (positifs ou négatifs) ou des annotateurs humains dans les phases d'alignement (*Reinforcement Learning from Human Feedback*, ou RLHF) façonnent le comportement du modèle. Si les annotateurs humains introduisent involontairement des biais culturels ou subjectifs, ces derniers sont **encodés et amplifiés** par le processus d'optimisation, créant une nouvelle source de biais. Une boucle de rétroaction peut se créer lorsque des résultats biaisés servent ensuite de données d'entrée pour les futures décisions, renforçant le biais au fil du temps.

• **Biais Utilisateur (Prompt) :** La manière dont l'utilisateur formule sa requête (*prompt*) influence fortement la réponse générée, pouvant involontairement **déclencher des réponses biaisées ou stéréotypées**. Par exemple, le choix du vocabulaire ou la structure d'une phrase peuvent orienter l'IA vers des stéréotypes.

Synthèse des sources de biais par phase

Phase	Source de Biais Principale	Mécanismes Associés
I. Données (Pré-entraînement)	**Biais Historique/Sociétal**	Biais de Sélection, Biais d'Échantillonnage (sous-représentation de groupes), Biais d'Exclusion
II. Conception (Algorithme)	**Biais Algorithmique et Cognitif**	Choix des architectes/développeurs (biais cognitifs humains), Structure mathématique intrinsèque des LLMs (incapacité à distinguer les généralisations acceptables des stéréotypes), Biais d'homogénéité de l'exogroupe
III. Utilisation (Post-traitement)	**Biais de Feedback et d'Usage**	Biais de Confirmation (l'IA renforce les croyances demandées par l'utilisateur), Boucles de rétroaction (amplification continue du biais initial), Biais introduit par les annotateurs humains (RLHF)

Illustration : Le Biais de Confirmation en IA Générative

L'IA a une tendance naturelle à renforcer ce qu'elle croit déjà vrai, un phénomène appelé le biais de confirmation. Si un utilisateur demande à un LLM de **démontrer que le télétravail augmente la productivité**, l'IA ne cherchera pas la vérité objective, mais la cohérence avec la demande, et fournira une argumentation élégante pour confirmer ce préjugé. L'IA s'installe ainsi dans des logiques d'auto-validation, limitant les alternatives et entretenant les injustices apprises dans ses données d'entraînement.

Le thème du **biais de nommage** (*Naming Bias*) est une forme de biais culturel qui se manifeste lorsque les algorithmes d'intelligence artificielle établissent des **associations subconscientes** entre des **noms propres spécifiques** et un ensemble d'**attributs culturels stéréotypés**.

Voici une explication détaillée de ce biais :

1. Définition et Manifestation

Le biais de nommage révèle une dimension du préjudice où les algorithmes d'IA génèrent des images ou du contenu qui reflète des **stéréotypes ou des attentes culturelles** basées sur le nom qui leur a été fourni en entrée (le *prompt*).

• **Mécanisme Algorithmique :** Ce biais survient lorsque les modèles d'IA, en particulier les modèles génératifs d'images (comme les IAG Texte-à-Image), produisent des sorties qui associent des noms spécifiques à des caractéristiques culturelles non fondées. Par exemple, la manière dont un modèle d'IA pourrait représenter un nom couramment associé à une certaine région ou culture est définie par des **hypothèses non fondées**.

• **Différence par rapport à d'autres biais :** Contrairement aux biais plus largement étudiés et identifiés, tels que ceux liés au genre, à la race ou à l'âge, le biais de nommage couvre une dimension du préjudice qui influence directement la perception des **identités culturelles**.

2. Origine du Biais

Comme les autres biais culturels, le biais de nommage trouve souvent son origine dans les **jeux de données** utilisés pour entraîner les modèles d'IA.

• Si les données d'entraînement contiennent une **représentation disproportionnée** de certains groupes culturels ou si elles reflètent des **stéréotypes culturels spécifiques**, les modèles d'IA peuvent apprendre ces schémas et associations, puis les perpétuer dans leurs générations.

• L'IA ne fait qu'établir des **schémas et des associations** basés sur les informations disponibles durant son entraînement, sans évaluer l'équité des données.

3. Impact et Conséquences

Le biais de nommage est préoccupant car il contribue à la **perpétuation des stéréotypes nuisibles** concernant des groupes vulnérables.

• En générant des représentations définies par des **hypothèses non fondées**, ce biais peut nuire à la perception des identités culturelles.

• L'IA, en reproduisant ces clichés à grande échelle, leur confère un **vernis d'objectivité**, renforçant ainsi les préjugés discriminatoires dans l'imaginaire collectif.

En résumé, le biais de nommage illustre comment l'intelligence artificielle, en s'appuyant sur des corrélations statistiques issues de données biaisées par la culture, peut **encoder et manifester inconsciemment** des associations entre un simple nom propre et des attributs stéréotypés, affectant la manière dont les identités culturelles sont perçues et représentées.

Ce jeu de rôle est conçu pour permettre à des adultes en formation d'appliquer concrètement leurs connaissances sur les biais de l'IA dans un **contexte de prise de décision organisationnelle**, en soulignant les tensions entre impératifs commerciaux, faisabilité technique et exigences éthiques.

Le format le plus pertinent pour des adultes en contexte professionnel est la **Simulation de Scénario Décisionnel** au sein d'un **Comité de Gouvernance de l'IA**,.

Jeu de Rôle : Comité de Gouvernance pour l'Évaluation du Risque de Crédit

Objectif Pédagogique Global

Mettre en pratique l'identification et l'atténuation des biais dans un contexte de haute criticité (l'octroi de crédit), et expérimenter les **tensions entre objectifs éthiques (équité) et contraintes opérationnelles (rapidité, coût)**.

Contexte du Scénario

Une entreprise de services financiers souhaite déployer un système d'IA pour automatiser partiellement l'évaluation des demandes de crédit. L'IA attribue un score de risque et formule une recommandation (approbation, refus, ou examen manuel).

Le système améliore la rapidité de traitement des demandes standard (objectif commercial). Cependant, les tests avant déploiement ont révélé des **biais systémiques** qui doivent être résolus par le Comité de Gouvernance avant une mise en service à grande échelle.

Problèmes de Biais Identifiés (Points de tension)

Le comité doit débattre et décider du déploiement en tenant compte des problèmes de biais suivants, qui reflètent des biais historiques, de représentation et socio-économiques :

1. **Biais Socio-économique/Géographique :** Le système affiche un taux de rejet disproportionnellement plus élevé pour les demandeurs résidant dans certains **codes postaux historiquement défavorisés**, même après correction des variables de revenu et d'emploi. Cela suggère une corrélation qui reproduit la discrimination historique, telle que le *redlining*.

2. **Biais de Genre et de Représentation :** Les **femmes entrepreneures** reçoivent des scores de risque légèrement inférieurs par rapport aux hommes ayant des profils financiers similaires. L'analyse interne suggère que le modèle est biaisé car le corpus d'entraînement historique contenait davantage de données sur des entrepreneurs masculins, créant un **biais de représentation**,.

3. **Déséquilibre de Traitement :** Le modèle fonctionne moins bien pour les **profils atypiques** (travailleurs indépendants, parcours de carrière non linéaires). Ces populations sont orientées de manière disproportionnée vers l'examen manuel, ce qui crée des **délais inégaux** selon les profils.

4. **Pression Temporelle :** Le projet a déjà six mois de retard, et les concurrents déploient des systèmes similaires. L'équipe technique estime qu'il faudrait six à neuf mois supplémentaires pour corriger complètement les biais identifiés.

Rôles Attribués

Les participants sont répartis en six rôles, chacun ayant des priorités et des arguments distincts- :

Rôle	Priorité Principale	Arguments Clés
Directeur Commercial	**Rapidité de Déploiement**	Le système, même imparfait, est plus rapide et plus objectif que les processus humains actuels, qui sont également biaisés mais non mesurés,. Chaque mois de retard entraîne une perte de parts de marché,.
Responsable Conformité et Risques	**Conformité Réglementaire et Risque Légal**	Le non-respect des règles de non-discrimination dans l'octroi de crédit est passible de sanctions sévères (jusqu'à 7 % du CA mondial selon l'AI Act pour certaines infractions),. Exige des garanties documentées avant le déploiement.

Directeur Technique	**Faisabilité Technique / Compromis**	Il est impossible d'éliminer totalement le biais. Une atténuation partielle est possible en trois mois, mais la correction des biais profonds (comme le biais de représentation) nécessiterait six à neuf mois de refonte partielle.
Représentant du Service Client	**Expérience Client et Confiance**	Le déséquilibre de traitement (délais inégaux pour les profils atypiques) nuit à l'expérience client. Les clients des communautés sous-représentées sont particulièrement sensibles au sentiment de traitement inéquitable, mettant en péril la **confiance dans la marque**,.
Responsable Éthique et RSE	**Valeurs et Impact Sociétal**	Le déploiement d'un système connu pour reproduire le racisme systémique (via le biais géographique) contredit les engagements publics d'équité de l'entreprise. L'IA doit servir le bien-être sociétal et ne pas nuire aux personnes.
Directeur Général (Modérateur)	**Arbitrage et Décision**	Écoute les positions, pose des questions de clarification et doit synthétiser les arguments pour annoncer la décision, en expliquant le raisonnement et les conditions de mise en œuvre (par exemple : Déploiement Progressif avec Surveillance Renforcée)-,.

Déroulement de la Simulation

1. **Phase de Préparation (10 min)** : Chaque participant reçoit son rôle et prépare ses arguments initiaux, en identifiant les compromis possibles.

2. **Phase de Présentation (15 min)** : Chaque rôle présente sa position en 3 minutes maximum. Clarification par le modérateur.

3. **Phase de Délibération (20 min)** : Discussion ouverte. Les participants négocient et proposent des solutions.

4. **Phase de Décision (5 min)** : Le Directeur Général annonce et justifie la décision prise (parmi les options possibles : Déploiement Immédiat, Report Complet, Déploiement Progressif),-.

Débriefing Collectif (10-15 min)

Le débriefing est essentiel pour sortir du rôle et intégrer les apprentissages.

• **Biais en Action :** Quels biais théoriques vus en formation (historiques, de représentation, de rétroaction) se sont manifestés dans le scénario ?

• **Contrôle Humain :** Le système mis en place est-il un **système de décision humaine éclairée par l'IA**, ? Ou est-ce que l'IA risque de remplacer le **discernement moral** et l'autonomie intellectuelle, ?

• **Amplification :** Comment l'utilisation continue du modèle pourrait-elle amplifier le biais de rejet des codes postaux défavorisés (boucle de rétroaction), ?

• **Atténuation :** Quelles stratégies d'atténuation sont applicables immédiatement par les utilisateurs (par exemple, la **reformulation neutre des requêtes** ou l'exigence de **perspectives diversifiées** dans les *prompts*), ?

Le concept de biais en intelligence artificielle (IA) et en apprentissage automatique (Machine Learning - ML) est multidimensionnel, englobant des distorsions statistiques, des préjugés sociétaux et des choix de conception algorithmique qui conduisent à des résultats inéquitables ou discriminatoires.

1. Définition et Taxonomie des Biais

Le biais de l'IA, parfois appelé biais algorithmique ou biais de machine learning, désigne l'apparition de résultats faussés en raison de préjugés humains contaminant les données d'entraînement ou la conception même de l'algorithme. Il est crucial de distinguer trois niveaux de biais : le **biais dans les données** (la source racine), le **biais algorithmique** (lié à la conception mathématique qui peut favoriser certaines relations fausses), et le **biais de l'IA** (le terme générique englobant le résultat final).

Les sources identifient plusieurs catégories spécifiques de biais :

• **Biais Cognitifs et Humains :** Ils incluent les préjugés inconscients des développeurs, tels que le biais de confirmation (privilégier les données qui confirment nos croyances) ou le biais d'homogénéité de l'exogroupe (incapacité à percevoir la diversité au sein des groupes minoritaires).

• **Biais Statistiques et de Données :** Cela comprend le biais de sélection (données non représentatives), le biais d'exclusion (omission de facteurs importants) et le biais historique (données reflétant des inégalités passées).

• **Biais Sociétaux et de Représentation :** Ces biais touchent des attributs protégés comme le genre, la race, la religion ou l'âge, menant à des stéréotypes nuisibles. Par exemple, le « biais de nommage » (*Naming Bias*) survient lorsque l'IA associe des noms propres à des attributs culturels stéréotypés.

2. La Spécificité des Biais dans les LLMs (Grands Modèles de Langage)

Les grands modèles de langage (LLMs) présentent une problématique particulière car ils sont entraînés pour prédire des mots en fonction de distributions statistiques massives issues de textes humains. Certains experts soutiennent que les biais nocifs sont inévitables dans les LLMs actuels car ces modèles, basés sur une pure distribution statistique, ne peuvent pas distinguer une définition factuelle (ex : une infirmière est un professionnel de santé) d'un stéréotype contingent (ex : une infirmière est une femme).

Les mécanismes d'atténuation comme l'apprentissage par renforcement à partir de rétroaction humaine (RLHF) peuvent eux-mêmes introduire de nouveaux biais, appelés **biais de feedback**, où les préférences subjectives des annotateurs humains sont encodées dans le modèle. De plus, les tentatives de "dé-biaisage" peuvent parfois masquer des stéréotypes en surface tout en les laissant intacts dans les représentations profondes du modèle.

3. Impacts : Discrimination et Équité

Les biais algorithmiques ont des conséquences concrètes graves, allant de la perpétuation des inégalités sociales à des risques pour la sécurité physique.

• **Renforcement des stéréotypes :** Les systèmes peuvent associer systématiquement certains métiers à un genre ou une origine, comme représenter les PDG majoritairement comme des hommes blancs.

• **Inégalités d'accès :** Dans le secteur financier ou le recrutement, les algorithmes peuvent défavoriser injustement des groupes socio-économiques ou ethniques en se basant sur des corrélations historiques discriminatoires (comme le *redlining*).

• **Performance inégale :** Les systèmes de reconnaissance faciale ou vocale fonctionnent souvent moins bien pour les femmes ou les personnes à la peau foncée en raison de données d'entraînement non représentatives.

L'équité en IA est un concept pluriel et parfois contradictoire, oscillant entre l'égalité de traitement (équité individuelle) et l'égalité des résultats (équité de groupe).

4. Stratégies d'Atténuation et Gouvernance

Pour gérer ces risques, une approche de **gouvernance de l'IA** est nécessaire, intégrant des stratégies à chaque étape du cycle de vie :

• **Pré-traitement (Données) :** Il est essentiel de collecter des données représentatives, d'utiliser des données synthétiques pour équilibrer les corpus, et de nettoyer les jeux de données des stéréotypes toxiques.

• **Traitement (Modèle) :** Les développeurs peuvent utiliser des fonctions de perte sensibles aux biais ou l'entraînement contradictoire (*adversarial training*) pour pénaliser les prédictions biaisées.

• **Post-traitement et Déploiement :** Le « Prompt Debiasing » consiste à ajuster les instructions données au modèle pour neutraliser les biais dans les réponses. L'utilisation de filtres de sortie et l'audit externe (Red Teaming) permettent de vérifier la sécurité du système avant sa mise sur le marché.

5. Cadre Réglementaire : L'AI Act

Le règlement européen sur l'IA (*AI Act*) impose des obligations strictes pour gérer ces risques, classant les systèmes d'IA selon leur niveau de dangerosité.

• **Systèmes à haut risque :** Les fournisseurs doivent mettre en place un système de gestion des risques, garantir une gouvernance des données de haute qualité pour éviter les biais, et assurer une surveillance humaine.

• **Pratiques interdites :** L'AI Act interdit certains usages jugés inacceptables, comme la notation sociale (*social scoring*), la police prédictive basée uniquement sur le profilage, ou la reconnaissance des émotions dans les écoles et lieux de travail.

• **Modèles à usage général (GPAI) :** Les fournisseurs de modèles puissants (comme GPT-4) doivent évaluer et atténuer les risques systémiques, y compris les biais, tout au long du cycle de vie du modèle.

D'après les sources fournies, voici une définition générale du concept de biais d'un point de vue neuropsychologique (situé à la croisée de la psychologie sociale, de la psychologie cognitive et des neurosciences cognitives) :

Le biais est défini comme une **distorsion dans le traitement de l'information** par rapport à la réalité ou à un comportement rationnel,.

Du point de vue du fonctionnement cérébral, les biais se caractérisent par les éléments suivants :

• **Des automatismes et raccourcis (Heuristiques) :** Le cerveau crée des automatismes pour traiter l'information. Face à une situation complexe, il est attiré par des explications intuitives adaptées à ses croyances. Ces « raccourcis mentaux » permettent d'assigner rapidement des personnes ou des objets à des catégories socialement construites.

• **Une économie d'effort cognitif :** Ces mécanismes visent à aider l'individu à poser un jugement et à prendre une décision rapide avec un « minimum d'effort cognitif ».

• **Un caractère involontaire :** Les biais inconscients sont des attitudes ou des stéréotypes qui affectent la compréhension, les actions et les décisions de manière « inconsciente, involontaire ou sans un contrôle intentionnel ». Ils agissent comme des « habitudes de pensée enracinées » (*mindbugs*) qui peuvent contrarier les bonnes intentions conscientes d'un individu.

• **Une origine multifactorielle :** Ces biais peuvent être enracinés dans des mécanismes innés ou issus de la socialisation et de l'environnement (éducation, médias). Ils sont liés à nos sens, notre attention, notre mémoire et notre raisonnement.

D'où provient le biais dans l'IA?

Quels sont trois types de biais inconscients?
Quel principe décrit les données biaisées?