



Résumé

Les constats et enjeux

Malgré le risque de biais dans certains cas, les algorithmes sont, à bien des égards, un progrès en matière de lutte contre les discriminations. Les hommes et les femmes portent régulièrement, bien

que parfois inconsciemment, des jugements biaisés. Ils sont inconstants dans leurs décisions. **Utiliser un algorithme revient à formaliser des règles applicables à tous, à en mesurer les résultats et donc à se donner les moyens d'assurer l'absence de biais.**

Les biais algorithmiques conduisant à des discriminations sont rarement dus à un code erroné de l'algorithme. Les données, incomplètes, de mauvaise qualité, ou reflétant les biais présents dans la société, sont bien plus souvent à l'origine de ces biais. Le combat des biais algorithmiques est donc avant tout un combat contre des discriminations déjà existantes au quotidien. L'enjeu n'est pas seulement de produire des algorithmes équitables mais aussi de réduire les discriminations dans la société.

Ce combat est difficile à plusieurs égards. Tout d'abord, **définir ce que serait un algorithme sans biais est complexe.** Certains biais sont volontaires, comme le fait de promouvoir les boursiers dans le cursus scolaire. D'autres biais sont inconscients, et conduisent à discriminer certains groupes.

Un algorithme traitant les individus de manière équitable s'approche d'un algorithme sans biais indésirable, sans pour autant le garantir. **Le détour par l'équité ne clôt pourtant pas le débat, car l'équité d'un algorithme peut prendre des formes multiples.** L'appréciation de ce qui est juste comporte une dimension culturelle et morale. L'attitude équitable ne sera pas la même selon qu'on parle d'un algorithme d'analyse de radiographie des poumons ou de recommandation de publicités politiques. L'équité totale entre individus et l'équité totale entre groupes sont mathématiquement incompatibles. Il y aura toujours des choix à faire, des choix de société, des choix politiques.

Ensuite, **corriger un algorithme pour le rendre équitable, c'est souvent réduire sa performance.** Lorsqu'on développe un algorithme, on choisit une ou plusieurs métriques qui permettent de l'optimiser et d'évaluer s'il remplit bien sa tâche. Ces métriques définissent sa performance. Ajouter une contrainte, c'est limiter la capacité d'optimiser l'algorithme vis-à-vis de son critère de performance initial. Il est toujours plus difficile de poursuivre plusieurs buts à la fois

plutôt qu'un seul. Il sera donc complexe et coûteux pour de nombreux acteurs de produire des algorithmes performants et équitables.

Enfin, **lutter contre les biais des algorithmes consiste à réaliser une synthèse entre la protection des citoyens contre les discriminations et la possibilité d'expérimenter, cruciale dans l'économie numérique.** Restreindre fortement l'usage des algorithmes sur la suspicion qu'ils pourraient avoir des biais, c'est se priver de nouveaux outils pouvant objectiver nos décisions, c'est brider l'industrie française du numérique et subir à long terme une domination technologique américaine et chinoise. Laisser faire, c'est ignorer le potentiel de destruction de telles innovations sur notre tissu social.

Les recommandations

Face à ces enjeux, il faut être clair : nous ne recommandons ni une loi contre les biais des algorithmes commune à tous les secteurs d'activités, ni un contrôle systématique par l'État de l'absence de biais dans les algorithmes.

Il existe d'ores et déjà de nombreux textes s'attaquant aux discriminations. Ceux-ci s'appliquent au monde physique comme au monde numérique et sont de nature à limiter le risque de biais. Compte tenu du faible recul dont nous disposons, une loi spécifique aux biais risquerait de bloquer toute innovation sans même résoudre le problème de fond.

Le règlement général sur la protection des données (RGPD) a montré que l'usage des données personnelles est bien trop répandu pour qu'une autorité administrative puisse en contrôler l'intégralité avant leur utilisation. Nous pensons qu'il en sera de même pour les algorithmes et qu'il est illusoire d'attendre de l'État qu'il contrôle chacun des algorithmes pour s'assurer de leur caractère éthique avant leur utilisation.

Nous nous sommes attachés à formuler des recommandations les plus réalistes possible afin de permettre un développement rapide des nouvelles technologies dans un cadre respectueux de nos modes de vie et de nos valeurs.

Tester la présence de biais dans les algorithmes comme l'on teste les effets secondaires des médicaments.

À l'instar des nouveaux médicaments, il est difficile de comprendre le fonctionnement de tous les algorithmes, notamment ceux utilisant l'intelligence artificielle. Par ailleurs, comprendre leur fonctionnement ne garantit pas qu'il n'aura pas de biais algorithmiques. C'est *in fine* par le test de l'absence de biais qu'il est possible de créer la confiance dans le caractère équitable des algorithmes.

Tester l'équité d'un algorithme a un coût et nécessite des données de test qui incluent spécifiquement certaines données sensibles (genre, origine sociale). Les développeurs et acheteurs d'algorithmes devront intégrer cette contrainte, et recourir à des tests fonctionnels ou de performance pour s'assurer de l'absence de biais. Dans certains cas où la création de ces bases de données est difficile ou problématique, l'État pourrait la prendre en charge.

Promouvoir une équité active, plutôt que d'espérer l'équité en ne mesurant pas la diversité

Pour lutter contre les discriminations, la France a longtemps fait le choix de ne reconnaître que des citoyens, égaux en droit, plutôt que des individus divers. En ce qui concerne les algorithmes, cette approche n'est pas pertinente. Un algorithme peut introduire des biais contre les femmes, même si l'on a explicitement exclu le genre des variables utilisées — il est en effet facile de déduire cette information à partir d'autres informations, comme le fait d'acheter en ligne des soins pour femme. Pour lutter contre les discriminations, il faut donc pouvoir les détecter.

Cela suppose de passer d'une approche qui attend l'équité par l'ignorance de la diversité à une équité active, c'est-à-dire accepter que l'équité d'un algorithme ne s'obtient pas en excluant toutes les variables protégées comme le sexe, l'âge ou encore la religion mais au contraire en les incluant et en testant l'indépendance du résultat vis-à-vis de ces variables. Pour ce faire, il est nécessaire de disposer de ces informations protégées. Mais si ces informations sont protégées, c'est justement parce qu'elles peuvent être source de discrimination. La collecte et l'utilisation de ces données doivent donc être strictement encadrées! Afin d'éviter les dérives, une telle collecte serait limitée aux tests de la présence de biais et elle serait restreinte à un échantillon des utilisateurs concernés. Par ailleurs, une telle approche devrait faire l'objet d'une analyse d'impact déclarée à la CNIL de manière préalable. Enfin, la nature des algorithmes testés devra justifier la collecte de telles données.

Être plus exigeant pour les algorithmes ayant un fort impact sur les personnes (droits fondamentaux, sécurité, accès aux services essentiels)

La sensibilité d'un algorithme vis-à-vis de la société dépend certes de son secteur d'activité mais surtout de son impact potentiel sur les citoyens. Celui-ci est fort dès lors que l'algorithme peut restreindre l'accès à des services essentiels comme un compte bancaire ou la recherche d'un emploi, mettre en danger la sécurité (santé, police), ou bafouer des droits fondamentaux. Ces domaines font déjà l'objet d'obligations fortes en matière de discrimination. Lorsqu'un algorithme y est introduit, cela ne peut être au prix d'une diminution des exigences.

Pour ces algorithmes, nous recommandons un cadre ad hoc prévoyant des obligations de transparence en ce qui concerne les données utilisées et les objectifs fixés à l'algorithme ainsi qu'un droit de recours contre la décision prise. La création d'un tel cadre ne

nécessite pas une nouvelle loi sur les biais des algorithmes mais plutôt la mise en œuvre de bonnes pratiques dans les entreprises et administrations, l'usage de dispositifs juridiques existants et l'ajout au cas par cas de dispositions dans des législations sectorielles.

Assurer la diversité des équipes de conception et de déploiement des algorithmes

Les algorithmes transforment les business models des entreprises. Pour cette raison, il importe que les responsables et utilisateurs soient de plus en plus impliqués dans leur conception.

Définir le comportement équitable de l'algorithme permet d'en modifier profondément les impacts économiques et sociétaux. Il est plus que jamais nécessaire d'intégrer une diversité de perspectives dans la prise de ce type de décision qui ne peut être le seul fait d'experts techniques. Au-delà de la diversité professionnelle, il est désormais clair que des équipes socialement diverses sont mieux armées pour prévenir les biais, pour éviter de reproduire des discriminations.

Intégrer une diversité de profils, de compétences, d'expériences, d'âges, de genres dans les équipes de conception, de production et de pilotage des algorithmes doit devenir une norme pour prévenir les biais algorithmiques.

Et pour aller plus loin...

Au-delà de ces quatre recommandations principales, nous sommes convaincus qu'un important travail demeure nécessaire en matière de formation. Cela doit concerner les chercheurs et développeurs bien sûr, notamment pour les biais algorithmiques, mais également les dirigeants et citoyens sur le cadre plus général de l'intelligence artificielle, afin que chacun puisse s'appropriier les opportunités et les dangers de cette technologie.

La vigilance sera également renforcée dans les organisations qui décideront de mettre en œuvre des chartes et bonnes pratiques. Il faut encourager ces initiatives qui, nous avons pu le constater lors de nos entretiens, génèrent une prise de conscience collective sur les dangers des biais algorithmiques, en complément des mesures techniques et opérationnelles.

Enfin, la vigilance doit être extérieure aux organisations. Il semble utile, dans le cas d'algorithmes à fort impact, de renforcer les exigences. Il faut d'abord soutenir l'émergence de labels qui garantissent la qualité des données utilisées et de l'organisation qui développe l'algorithme, l'existence de procédures de contrôle ou encore l'auditabilité de l'algorithme. Les acteurs économiques auront besoin de telles garanties pour s'emparer des algorithmes, et tirer pleinement parti de la révolution qu'ils représentent. Pour les algorithmes à fort impact, une capacité d'audit et de contrôle de certaines exigences pourrait être confiée à une tierce partie ou à l'État.