



HAL
open science

Analyse des données et choix de société

Serge Abiteboul

► **To cite this version:**

Serge Abiteboul. Analyse des données et choix de société . Parole publique, 2016, Communiquer à l'ère du Big Data. Promesses et périls de l'action publique algorithmique. hal-01273439

HAL Id: hal-01273439

<https://inria.hal.science/hal-01273439>

Submitted on 12 Feb 2016

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.

Analyse des données et choix de société

Serge Abiteboul, Inria et ENS Cachan

<http://abiteboul.com>

« La faute aux algorithmes ! » Le programme informatique, de plus en plus au cœur de la gouvernance de notre société, est souvent mis au banc des accusés. Notre propos est ici d'expliquer qu'on est en droit d'attendre des algorithmes des transformations extrêmement positives de cette société, et qu'on est également en droit d'exiger d'eux qu'ils aient un comportement véritablement responsable. Nous allons en particulier nous intéresser aux algorithmes de traitement de données massives [1].

L'analyse de données massives, popularisée sous le nom de Big data, porte en elle d'incroyables promesses : améliorer la vie des gens, accélérer la découverte scientifique et l'innovation, ou apporter des changements positifs dans la société. De nouvelles techniques d'analyse de données permettent par exemple d'extraire des connaissances à partir des quantités de données de plus en plus importantes produites quotidiennement par chaque individu lorsqu'il utilise un téléphone intelligent qui le géo-localise, qu'il surfe sur internet et collecte malgré lui des cookies, ou utilise des réseaux sociaux.

Avec le Big data, les gouvernements, comme les quelques géants de l'internet qui peuvent avoir accès aux données et à cette technologie, disposent d'un pouvoir énorme. Il nous paraît donc nécessaire que la collecte et l'analyse des données deviennent « responsables ». Nous entendons par cela qu'il faut que la collecte et l'analyse de données deviennent équitables, transparentes et qu'elles favorisent la diversité.

Équité. L'équité, pour ce qui est de l'analyse de données, c'est avant tout l'absence de biais. Par exemple, il est injuste d'utiliser un ensemble de données sur la criminalité, dans lequel certains quartiers sont sous-représentés, ou de biaiser un calcul en utilisant les valeurs de certaines variables cachées, comme l'origine ethnique, reconstruites pour l'occasion. De tels biais peuvent refléter les intérêts commerciaux, les préférences politiques, religieuses, ou autres des concepteurs des algorithmes. Qu'ils soient intentionnels ou pas, ils peuvent être immoraux, voire illégaux.

Transparence. Les individus aimeraient connaître et également contrôler, les données qui les concernent, savoir à quoi elles servent, comprendre comment des décisions sont prises à partir d'elles. Une telle transparence est indispensable pour construire la confiance. Liée à l'absence de transparence, on trouve la dissymétrie, l'algorithme se comportant en un dispositif opaque accédant vos données les plus personnelles quand vous ignorez tout de ce qu'il en fait. Les fameuses « boîtes noires » de la Loi de renseignement sont un exemple particulièrement frappant de ce qu'il faut absolument éviter.

La diversité. Les algorithmes de recommandation (d'information, de pages Web, de films, etc.) font la part belle à la popularité au risque d'occulter une grande partie des choix possibles et d'accentuer encore plus des inégalités. Ils excellent de plus en plus à personnaliser les réponses au risque de nous enfermer dans des territoires intellectuels de plus en plus étroits.

Ces trois propriétés ne sont pas nées avec les algorithmes. Depuis des siècles, les démocraties s'essaient à s'améliorer pour garantir aux citoyens plus d'équité, de transparence, et ce en respectant la diversité. Les algorithmes ne sont pas, comme on l'entend parfois, par nature, injustes, et opaques ; ils n'encouragent pas, par nature, l'uniformité. Au contraire, bien utilisés, ils peuvent nous permettre de progresser dans ces trois directions.

Open data, open source, provenance. La transparence peut s'appuyer sur l'ouverture des données et des programmes. Si le fait que les données et le code sont en accès ouvert, ne garantit pas, par exemple, l'équité, cela en facilite la vérification. Evidemment, toutes les données ne peuvent être ouvertes (par ex, les données médicales) et tous les logiciels en accès libres (par ex, les systèmes commerciaux). Mais sans aller jusqu'à une ouverture complète, quand on publie des données, ou qu'on appuie des décisions sur des jeux de données, on devrait spécifier comment les données ont été acquises, et quels traitements ont été réalisés.

Vérification. Pour vérifier le comportement d'un programme, on peut soit en analyser le code (l'auditer), soit en analyser les effets. L'analyse de code est directement liée à la preuve de théorèmes en mathématiques ; c'est techniquement compliqué. L'analyse des effets s'apparente plutôt à l'étude par exemple statistique de phénomènes physiques ou biologiques comme le climat ou le cœur humain. On peut vérifier par exemple le code d'un logiciel de vote. On peut tester les réponses d'un moteur de recherche. La vérification et le test de logiciels ont fait d'énormes progrès, par exemple pour garantir leur fiabilité. Il reste beaucoup à faire pour garantir leur « responsabilité ».

Mais revenons à la transparence. Elle implique aussi d'être capable d'expliquer des décisions. Dans certains cas, cela peut être relativement simple. Par exemple, on peut rendre public un algorithme de calcul d'impôts ou de prestations sociales, comme le fait par exemple l'entreprise openfisca. On peut alors suivre pas à pas le processus qui conduit à une décision. Il est beaucoup plus complexe d'expliquer les millions d'opérations qui ont permis, par exemple, de décider qu'une rue doit être mise en sens unique en se basant sur des téra octets de données récoltés par un service de l'équipement. Dans les cas où la technique se révèle incapable d'expliquer, ne vaudrait-il pas mieux se passer de telles analyses ?

L'autonomie des algorithmes. Les algorithmes décident déjà pour nous dans des situations simples, comme les radars sur les routes. Cela pose problème dans des cas complexes : peut-on laisser à un algorithme, par exemple, la décision de remise de peine d'un condamné. Et si l'algorithme se contente de conseiller, un juge osera-t-il prendre le risque d'une remise en liberté contre l'avis de l'algorithme ? La question du partage des décisions administratives entre humains et algorithmes va se poser de plus en plus. Nous sommes encore loin de nommer un robot responsable d'un service administratif. Vraiment ?

L'opposabilité. Il doit être possible de s'opposer à une décision administrative, notamment d'un algorithme, qui même s'il est potentiellement infaillible, peut ne pas avoir toutes les informations, ne pas avoir été programmé pour tenir compte d'aspects humains pourtant absolument essentiel. Cette opposabilité des décisions administratives est mal prise en compte aujourd'hui. Elle est essentielle pour que les décisions des algorithmes soient acceptées par les administrés. Il faut donc une place pour des médiateurs, qui un jour seront peut-être eux-mêmes aussi... des programmes.

Les algorithmes peuvent être plus justes que des humains ; ils peuvent apporter plus de transparence dans les procédures administratives ; ils peuvent offrir un traitement plus personnalisé qui tiennent compte de la diversité des administrés. Ils peuvent tout ça. Mais ils peuvent aussi les contraires. C'est à nous de choisir. Il ne faut donc pas avoir peur des algorithmes mais nous battre pour construire avec eux une société plus juste, plus transparente, plus respectueuse des diversités.

[1] S. Abiteboul, J. Stoyanovich, Data, Responsibly, ACM Sigmod Blog, 2015.