



**HAL**  
open science

## Illusions et vérités du big data

Valérie Charolles

► **To cite this version:**

Valérie Charolles. Illusions et vérités du big data. Le Debat , 2019, n°207, pp.132-140. 10.3917/deba.207.0132 . hal-04086062

**HAL Id: hal-04086062**

**<https://hal.science/hal-04086062v1>**

Submitted on 1 May 2023

**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.

# Illusions et vérités des big data

Valérie Charolles

*Preprint (version auteur)*

Pour citer cet article : Charolles, V. (2019). Illusions et vérités du big data. *Le Débat*, 207, 132-140. <https://doi.org/10.3917/deba.207.0132> [hal-02395333v1](https://hal.archives-ouvertes.fr/hal-02395333v1)

« Big data » : le terme est apparu il y a vingt ans pour décrire les données massives produites dans le monde, en particulier à travers les réseaux numériques et il marque incontestablement un changement d'échelle. Jamais nous n'avons eu autant d'informations disponibles sous une forme traitable au plan statistique. Les volumes en cause (plus de 900 000 exaoctets publiés en 2018 sur les seuls réseaux numériques) et leur forte croissance représentent un nouveau régime par rapport au premier grand moment de la statistique : celui de « l'avalanche de chiffres » qui a caractérisé la société postnapoléonienne<sup>1</sup> et la période d'ordonnement de la statistique publique qui l'a suivie, superbement décrite par Alain Desrosières<sup>2</sup>.

Le big data repose sur deux phénomènes concomitants. D'une part, le nouveau vecteur de communication que constitue le numérique est désormais alimenté par plus de la moitié de la population du globe et présente l'avantage d'être aisément lisible, puisque les informations y sont transmises sous forme de chiffres. Le second élément déterminant est la puissance de calcul atteinte par les processeurs informatiques qui permet l'usage de formes de statistiques plus complexes que celles issues du modèle standard gaussien.

Entre ceux qui y voient, souvent du côté des entreprises, un nouvel Eldorado pour leur développement et ceux qui y décèlent la concrétisation du cauchemar imaginé par Philippe K. Dick dans *Minority Report*, les avis sont en général très tranchés et il est

---

1. Selon l'ouvrage de référence de l'histoire des probabilités de Ian Hacking, *The Emergence of Probability*, Cambridge University Press, Mass., 1975.

2. Alain Desrosières, *La Politique des grands nombres. Histoire de la raison statistique*, La Découverte, 2000.

difficile de faire la part des choses et de comprendre ce qui se joue : simple changement d'échelle dans une société déjà depuis longtemps habituée à recourir de manière privilégiée à des observations sous une forme chiffrée, ou changement de nature plus profond ?

C'est en réalité à deux séries de questions que renvoient le big data et son usage : l'une est épistémologique et porte sur la solidité scientifique des résultats ainsi produits mais aussi, et plus profondément, sur le mode de savoir que nous pouvons rechercher ; l'autre est de nature ontologique et concerne les formes d'êtres que dessinent les informations massives et les types d'échanges, souvent marchands, et de société qui les sous-tendent. Ces deux questions se mêlent intimement. Nous les analyserons ici autour de six propositions.

### *Quand on passe de l'univers infini au système réfléchi*

Dans un précédent ouvrage<sup>3</sup>, j'ai analysé l'évolution représentée par l'écran (du cinéma, de l'ordinateur, puis des réseaux) et l'ai comparée à celle décrite dans *Du monde clos à l'univers infini* d'Alexandre Koyré lorsque, aux XVI<sup>e</sup> et XVII<sup>e</sup> siècles, l'apparition de la lentille astronomique a modifié la conception du système solaire, débouchant sur la science classique et ses lois éternelles et immuables, mathématiques universelles du cosmos.

Les découvertes intervenues depuis la fin du XIX siècle nous ont permis, avec le cinéma, de capturer une image en mouvement, ce que Léonard de Vinci et René Descartes avaient cherché ardemment à réaliser, puis, avec la télévision, de transporter ces images au cœur des foyers. Le remplacement de l'analogique par le numérique a ensuite rendu aisé le traitement et la circulation de toute forme d'information, au moment où les capacités de calcul des ordinateurs autorisaient la manipulation de données en très grand nombre. Enfin, depuis à peine trente ans (vingt ans pour la téléphonie mobile), l'apparition des réseaux a permis à une large part de la population du monde d'alimenter en boucle ces données.

Ces évolutions reconfigurent l'espace-temps dans lequel nous évoluons : le lointain est maintenant très aisément accessible, faisant perdre à la distance établie en mètres sa valeur de repère absolu ; les nouveaux moyens d'observation qui en ont découlé sur le plan scientifique, les télescopes géants notamment, nous ont donné accès à des informations sur le passé de l'univers et conduit à constater que ce dernier avait une

---

3. Valérie Charolles, *Philosophie de l'écran. Dans le monde de la caveme*, Fayard, 2013.

histoire. Contrairement à ce que la science classique nous avait appris, il n'était donc ni éternel, ni immuable.

Dans le même temps, les actions des hommes en très grand nombre ont modifié la physionomie de la Terre ; l'échange marchand s'est développé à un niveau inégalé dans l'histoire, donnant au système économique un poids nouveau sur nos vies. En trente ans, la topographie des marchés financiers a été totalement modifiée : opérant à l'échelle mondiale et à des vitesses largement en dessous de la seconde, les acteurs des marchés financiers ont massivement tiré parti de la puissance de l'informatique et des réseaux pour automatiser le traitement des transactions, puis, depuis peu, la décision même de les réaliser.

Ces marchés, Internet ou encore les réseaux sociaux sont les témoins les plus évidents d'un mode de fonctionnement reposant sur des mécanismes d'échos et de miroitement. Une transaction sur les marchés financiers verra son effet démultiplié par les opérations dérivées auxquelles elle donnera lieu (couvertures contre les risques de change, de taux...) et ses conséquences le seront encore par les techniques de placements indexés (se concentrant par exemple sur le CAC 40). Le même type de fonctionnement est à l'œuvre sur Internet et les réseaux sociaux : plus un site est recherché, plus il apparaît tôt dans les résultats, ce qui renforcera encore sa position, tout comme il en va du nombre de connaissances qui est proposé sur les réseaux sociaux, induisant des phénomènes de grappes, des îlots se réfléchissant continûment.

Toutes ces évolutions, dont le big data est un élément, dessinent une nouvelle forme du monde, à savoir non plus un univers infini mais un système réfléchi<sup>4</sup>. Force est de reconnaître que nous rencontrons de grandes difficultés pour le comprendre et avoir prise sur lui. La séparation absolue opérée par Emmanuel Kant entre « le ciel étoilé » au-dessus de nous (que la science a vocation à décrire de façon mécanique) et « la loi morale » en chacun de nous (par principe déliée de toute forme de calcul) continue à constituer la toile de fond à laquelle nous nous référons. Mais celle-ci s'avère inadaptée pour rendre compte du fonctionnement du monde contemporain ; elle s'est d'ailleurs en pratique traduite par l'entrée en force du règne du calcul, scientifique mais aussi économique et commercial, pour traiter des comportements humains.

Nous n'avons pas encore déployé les outils conceptuels permettant de décrire et maîtriser un monde dans lequel il n'y a plus, d'un côté, l'homme et, de l'autre, la nature, mais un entrelacs. C'est ce que j'ai aussi appelé la fin de la naturalité, qui se traduira peut-être, si les scientifiques trouvent un consensus sur ce point, par la reconnaissance d'une nouvelle ère géologique, l'anthropocène. Cela appelle un nouveau discours de la méthode, permettant de ranger ce qui se produit autrement que sous la forme de la crise

---

4. Ces points sont développés aux chapitres I, II et VI de *Philosophie de l'écran*, *op. cit.*

(lorsque les faits limites se multiplient, c'est que notre modèle de compréhension du monde est à revoir, insuffisamment large pour embrasser le fonctionnement du réel), permettant surtout, grâce à cet effort de reconceptualisation, de trouver des prises légitimes pour faire évoluer la réalité et non pas être, en tant qu'humains, dépossédés par elle.

### *Une science descriptive plutôt qu'une science des modèles*

Dans le cadre ainsi posé, concentrons désormais la focale sur le big data. Sur un plan scientifique, il se singularise par le fait que les « données » s'organisent d'elles-mêmes sans nécessiter d'hypothèse préalable. Cette caractéristique a conduit Chris Anderson à annoncer, dans un article retentissant de juin 2008, « la mort de la théorie » et le fait que « le déluge de données rend la méthode scientifique obsolète »<sup>5</sup>. Son constat est le suivant : avec suffisamment de points d'observations, les informations parlent d'elles-mêmes ; aucune analyse, causale ou sémantique, n'est nécessaire ; ceci permet de se passer de modèles *a priori* et des erreurs qu'ils contiennent, tout en obtenant des résultats d'une fidélité sans précédent par rapport à la réalité. On reviendra plus tard sur le problème de la fidélité, restons-en ici à la question épistémologique. C'est le modèle de la science poppérienne, faite d'hypothèses et de réfutations, que mettait ainsi en question Chris Anderson : tout comme la mécanique quantique, encore fondée sur un modèle, a montré en quoi la physique newtonienne était trop fruste, la science des données offrira une vision plus large que celle de la physique quantique, avec comme nouveau slogan : « Laisser les algorithmes statistiques trouver les configurations à l'œuvre quand la science n'y parvient pas. »

Cet article a fait beaucoup réagir<sup>6</sup>. Certaines de ses formulations sont à l'évidence exagérées, comme sa conclusion provocatrice : « Il est temps de se demander : qu'est-ce que la science a à apprendre de Google ? » Les exemples sont néanmoins multiples de la fécondité d'une démarche observationnelle pour faire émerger de nouveaux objets de connaissance, plus complexes que ceux qui étaient supposés (citons le développement de l'épigénétique après que le génome humain a été décodé). Que les informations parlent d'elles-mêmes repose sur l'usage de la statistique selon une forme différente de celle dominante jusqu'à présent et ne présuppose pas le recours au modèle gaussien avec ses observations qui s'organisent de façon harmonieuse autour de la moyenne (selon la fameuse courbe en cloche). Le big data mobilise des formes de calcul beaucoup plus complexes, issues de l'œuvre de Thomas Bayes (1702-1761), que la

---

5. Dans la revue *Wired* dont l'auteur était le rédacteur en chef (consultable en ligne).

6. Voir notamment dans le numéro de mai 2019 d'*Esprit* l'article de Jean-Pierre Dupuy « La nouvelle science des données ».

puissance des processeurs des nouveaux ordinateurs permet de manipuler : la classification et la segmentation des observations y sont réalisées par l'outil statistique lui-même selon les régularités qu'il observe entre les faits.

La recherche d'une description du monde qui n'enserme pas le réel *a priori* dans un modèle pour voir comment les informations vont y réagir n'est pas une mauvaise évolution pour la science. Si l'on prend le cas de l'économie, elle devrait permettre de se passer du postulat de rationalité qui modélise l'*homo œconomicus* comme un agent cherchant à maximiser son utilité, le transformant, selon les termes de Kenneth Arrow, en idiot rationnel. On pense aussi au modèle néoclassique de fonctionnement des marchés financiers selon lequel le mieux que l'on puisse faire est de suivre le marché et que ses acteurs utilisent massivement : cela induit des phénomènes autoréférentiels, d'échos et de miroitement, qui se traduisent inmanquablement par des bulles suivies de cracks, ce modèle n'en restant pas moins le standard enseigné et utilisé en économie et en gestion.

Bref, que l'information soit traitée dans une visée descriptive et non filtrée à l'avance par des modèles n'est en soi une évolution délétère ni pour la démarche scientifique ni pour la société, bien au contraire. Elle marque la fin d'une période, celle de la suprématie de la loi dite normale de Laplace-Gauss, qui a été celle des calculs effectués pour caractériser la moyenne et, dans les sciences humaines, l'homme moyen, avec toutes les simplifications et les dérives, notamment idéologiques, qui peuvent en résulter.

Il est à cet égard paradoxal que les nouveaux formats scientifiques soient massivement utilisés par les entreprises pour faire des propositions d'achat personnalisées : on n'est plus alors dans l'ordre de la description mais dans celui de la prescription, certes sans modélisation *a priori*, mais fondé sur une ontologie très fruste. Ce n'est pas sans rappeler l'histoire vécue par Einstein quant aux applications de ses théories en matière d'armement : les évolutions scientifiques ont différents types d'usages dont il est utile de choisir lesquels méritent d'être développés et lesquels doivent être encadrés, le maniement bayésien de l'information qu'est le big data en étant un nouveau témoignage.

### *Corrélation et causalité : deux espaces de pensée*

À cet égard, il est décisif de ne pas se méprendre sur l'espace de pensée et de savoir que déploie la statistique, quelle que soit sa forme. Elle établit des corrélations et non des relations de causalité : elle permet de dire que des séries de faits sont en relation positive, négative, ou ne présentent pas de lien eu égard aux observations disponibles,

mais elle ne permet pas d'établir de lien de causalité ( $p$  implique  $q$ ) ; et il se peut qu'il y ait une causalité entre deux phénomènes même si l'on n'y voit pas de corrélation (si intervient une autre variable qui a caché la causalité).

Tout étudiant qui manie des statistiques apprend cela très tôt : la statistique délimite des relations, mais elle ne les explique pas (même si elle peut infirmer ou confirmer une hypothèse). Les statisticiens les plus intègres le revendiqueront, à l'instar de Benoît Mandelbrot<sup>7</sup> : l'explication de ce qui cause les faits est une boîte noire, à laquelle il n'est pas nécessaire d'avoir accès. Dans la plupart des cas, néanmoins, cette caractéristique est escamotée et l'on prend le parti de voir dans la corrélation une causalité (quel que soit  $x$  et quel que soit le temps  $t$ , si  $x$  a la caractéristique  $p$ , alors il aura la caractéristique  $q$ ).

C'est un travers d'autant plus dommageable que la corrélation peut ouvrir sur des formes de relations très intéressantes, fondées sur le « il existe » et non sur le « quel que soit », reposant donc sur la reconnaissance des singularités<sup>8</sup>. On peut, ce faisant, dépasser une vision déterministe ou réductionniste de la relation, bref, laisser du jeu pour le libre arbitre. Pour autant que l'on prenne la corrélation pour ce qu'elle est, la toile de fond à laquelle elle se rattache est profondément différente de celle qui sous-tend la science classique : le changement  $y$  a toute sa place puisqu'il n'est pas supposé de permanence de la relation. Cette caractéristique n'est pas sans lien avec l'intérêt renouvelé pour le cadre de pensée bouddhiste (impermanence, interdépendance, non-ego au sens où l'ego n'a pas de stabilité)<sup>9</sup>.

Pris au sérieux, l'espace de pensée de la corrélation<sup>10</sup> peut ainsi se révéler non seulement plus adapté pour décrire le fonctionnement d'un monde où les interactions d'acteurs en très grand nombre dominent et se réfléchissent les unes les autres, mais aussi pour laisser place à des modalités d'agir plus ouvertes, indépendantes de liens de causalités qui viendraient enchâsser l'homme sous la figure d'un automate dénué de temporalité. Sous cette forme, le savoir pourrait saisir dans le même temps les possibilités qu'offre le libre arbitre et celles de l'élaboration de connaissances situées. Cela n'interdit évidemment pas de continuer par ailleurs à rechercher des connaissances scientifiques à visée architectonique.

### *Le problème des données*

---

7. Avec Richard Hudson, *The (Mis)behavior of Markets. A Fractal View of Risk, Ruin and Reward*, Basic Books, 2004.

8. Nicolas Bouleau, « Sur la place du singulier dans la connaissance », *Communications*, 2015 / 1.

9. La trilogie *Matrix* en est une illustration.

10. Ou celui des systèmes complexes (voir, par exemple, en ligne, sur le site de l'Institut des systèmes complexes de Paris, Île-de-France, ISC-PIF).

Si le modèle scientifique sur lequel repose le big data est donc loin de mériter toutes les critiques dont il est souvent l'objet, le big data n'en présente pas moins, de par sa définition même, une faille intrinsèque : il dépend des informations qu'il pourra analyser et qu'il qualifie de « données ». C'est là sa principale faiblesse et elle s'analyse à un double niveau.

Le premier point d'achoppement est que nous disposons d'informations non pas sur tout ce qui se produit, mais uniquement sur certaines de ses parties : les « données » ne peuvent ainsi que fournir une vision biaisée de la réalité si elles prétendent l'embrasser dans sa totalité. Les « traces » que nous laissons sur Internet, comme les appelle Pierre-Antoine Chardel<sup>11</sup>, ne sauraient, en particulier, constituer des images fidèles, et ce quand bien même nos achats et rentrées d'argent donnent à notre banque des informations très précises. Ce n'est pas une ontologie qu'elles dessinent. C'est uniquement dans des contextes très particuliers – explorations scientifiques visant une partie ciblée du ciel par exemple – que l'on peut estimer avoir collectionné toutes les informations que nous sommes en capacité d'avoir sur un problème et donc de caractériser, provisoirement, une entité.

La seconde faiblesse du big data provient du fait qu'il n'existe pas, à proprement parler, de « données » : ce sont d'informations qu'il est question, en l'occurrence intégrées dans des ordinateurs sous forme d'input, mais aucune d'entre elles n'est « donnée » par une expérience qui se ferait sans médiation aucune. Bernard Stiegler les appelle des « produits », Bruno Latour parle d'« obtenues » ; je préfère, pour ma part, le terme de « construits<sup>12</sup> ». Le processus de construction des informations à l'œuvre dans le big data se joue sur plusieurs strates.

Il est d'abord propre au big data : si le traitement statistique bayésien ne suppose pas le choix de variables *a priori*, il n'en passe pas moins par un certain nombre d'étapes qui font intervenir des choix sur ce qu'il sera pertinent de retenir : pour donner un exemple, comprendre le champ à explorer et les « données » disponibles, préparer les « données » (en supprimant certains points aberrants), explorer les « données » pour obtenir des résultats, évaluer ces résultats pour déployer enfin l'algorithme obtenu<sup>13</sup>. L'algorithme, s'il pourra après fonctionner de façon automatique, ne se construit pas lui-même. Il est

---

11. Pierre-Antoine Chardel, avec Olaf Avenati (sous la dir. de), *Datalogie. Formes et imaginaires du numérique*, Loco, 2016.

12. Cf. mon livre *Et si les chiffres ne disaient pas toute la vérité ?*, Fayard, 2008. On pourra aussi se référer à *Quand le monde s'est fait nombre* d'Olivier Rey, Stock, 2016.

13. Méthode CRISP DM largement utilisée.



en outre dépendant, comme le montre Éric Guichard<sup>14</sup>, de ce que le codage informatique reconnaît comme entités ayant une signification, là encore selon des modes qui ne sont pas entièrement automatisés et laissent en particulier la part belle aux signes issus des langues des pays développés.

À un niveau plus général, enfin, l'évolution du monde et de nos connaissances doivent nous conduire à reconnaître qu'il n'est pas d'observation qui puisse être faite sur des données parfaitement brutes : c'est le sens de la théorie quantique au plan scientifique, mais c'est aussi la marque de notre état de civilisation que de nous mettre face à des faits résultant de diverses interactions. Pour en donner une illustration, les cours de bourses qui constituent, parfois avec les informations produites par les analystes financiers, les observations dont se nourrissent les algorithmes de marché, s'ils apparaissent comme des chiffres, ne sont pas des données abstraites. C'est le résultat de multiples informations reposant *in fine* sur des bilans comptables, eux-mêmes établis à partir de normes dont l'orientation peut être mise en question.

Faire parler les informations, sur le modèle du big data, suppose ainsi dans le même temps de s'interroger sur les formes de langage qui servent à construire ces informations et sur leurs biais éventuels, renforçant ainsi l'intérêt de mener par ailleurs des analyses et recherches conceptuelles.

### *Tout ce que le big data ne fera pas, quoi qu'il prétende*

Les forces et faiblesses épistémologiques du big data permettent de dissiper certaines illusions sur ce qu'il pourra apporter, quoi qu'en disent les entreprises qui en ont fait leur spécialité<sup>15</sup>.

Il ne sera pas en mesure de prévoir le futur lointain puisqu'il est seulement capable d'établir des corrélations entre des informations. Ces corrélations ne sont pas déterministes mais ont une valeur de prédiction, et elles peuvent être contredites par une information ou des configurations nouvelles. De même, il ne pourra pas embrasser notre conscience et prévoir nos pensées. Certes, les formes mathématiques utilisées sont en capacité d'intégrer de nombreuses observations et de faire droit à l'évolution des relations qu'elles entretiennent ; mais, d'une part, comme y insiste Jean Dhombres, l'intelligence artificielle ne traite que des nombres calculables, soit une goutte d'eau par

---

14. Éric Guichard, « La philosophie des techniques revue à l'aune de l'internet et du numérique », *Le Numérique, enjeux et débats*, Éditions universitaires de Dijon, 2017. Je le remercie pour sa relecture aussi professionnelle que bienveillante de cet article.

15. Cf. rapport de l'institut Aspen de 2010 (consultable en ligne).

rapport aux non-calculables ; d'autre part, la plasticité de notre cerveau et les propriétés émergentes de notre conscience, le fait qu'elle se transforme continuellement selon des modalités non prévisibles rendent cette entreprise illusoire, à moins de se concentrer sur des faits éminemment simples (réagir face à un danger grave, par exemple) qui ne constituent qu'une toute petite partie de notre ontologie.

Pour le dire de façon plus générale, le big data ne sera pas en capacité de prendre notre place en tant qu'êtres dotés de multiples facultés, d'une conscience aux propriétés émergentes, singuliers et inscrits dans un espace de vie collectif qui n'est pas que la somme des parties le constituant. Si ce que certains appellent la période de la singularité (celle où les machines prendront le pouvoir sur les hommes) n'aura pas lieu, c'est fondamentalement parce que les machines, en l'occurrence ici les algorithmes, même apprenants, sont liés à ce que nous y initialisons (ils ne sont pas des entités surgies de nulle part sans intervention humaine). C'est aussi parce qu'ils ne sont pas adaptés à la survie dans notre écosystème, dépendant notamment de façon décisive de l'électricité, ce qui les rend éminemment fragiles.

En revanche, le big data fait déjà très bien l'office d'un conseiller commercial personnalisé dont il est très difficile de se défaire ou encore d'un mouchard à disposition des services de sécurité et des banques, et ce sont là des questions qu'il faut aborder de front.

### *Comment nous préserver des mauvais usages du big data ?*

Le recours au big data s'est développé à une vitesse extrêmement rapide, pouvant donner l'impression d'une fatalité, d'une nouvelle féodalité<sup>16</sup>, à laquelle il serait impossible de se soustraire. Il y a pourtant des moyens de contrer ce qu'Antoinette Rouvroy et Thomas Berns qualifient de gouvernementalité algorithmique<sup>17</sup> : la remettre à sa place au plan épistémologique, rentrer dans la machine pour débusquer ses boîtes noires et ses biais, mais aussi se préserver des mauvais usages du big data sur un plan individuel et collectif. Le problème se pose à cet égard en termes différents selon les trois grandes configurations dans lesquelles apparaît le big data : à des fins scientifiques, de sécurité et de défense, ou, enfin, commerciales.

La manipulation des informations en grand nombre dans le domaine scientifique est le cas le moins difficile, même lorsqu'il s'agit de recherche médicale, pour autant que les informations produites ne soient pas ensuite transmises à des acteurs commerciaux

---

16. Alain Supiot, *La Gouvernance par les nombres*, Fayard, 2015.

17. « Gouvernementalité algorithmique et perspectives d'émancipation », *Réseaux*, 2013 / 1.

(compagnies d'assurance, banques) qui pourraient en profiter pour évincer certains clients ou adapter leur politique de prix. Logiquement, le serment d'Hippocrate et la déontologie médicale devraient prémunir contre ce type de dérives, à moins que les États ne modifient les règles en vigueur.

La situation est plus incertaine s'agissant de l'usage du big data par la puissance publique à des fins de sécurité et de défense. Accès aux informations de téléphonie mobile, de navigation sur Internet, de services de messagerie, de vidéosurveillance, informations captées par satellite : depuis les attentats du 11 Septembre, le principe selon lequel les correspondances privées (poste, téléphone) étaient en tant que telles protégées par la loi a subi de multiples brèches ; la NSA et, vraisemblablement, nos services de renseignement nationaux, ont accès à nos communications, même si les services nationaux ne disposent pas nécessairement des outils permettant de les traiter de façon fine.

Cette évolution nous rappelle que les principes de liberté et de protection de la vie privée sont entendus de façon variable dans le temps, y compris par les démocraties quand elles se sentent menacées. L'article 12 de la Déclaration universelle des droits de l'homme est pourtant clair<sup>18</sup>. Son respect appelle une attention soutenue tant les risques de dérive vers une société de surveillance discrète et néanmoins généralisée sont nombreux. La séparation des pouvoirs exécutifs, législatifs et judiciaires, ainsi que l'existence de contre-pouvoirs comme ceux de la presse représentent sur ce point des remparts que la loi ne doit pas venir effriter.

Le problème porte, enfin, sur l'usage marchand des informations personnelles. Les dérives sont multiples à cet égard et posent des questions ontologiques, éthiques et politiques. Certes, le Règlement général sur la protection des données (RGPD) et l'existence de la Commission nationale informatique et libertés viennent tempérer certaines tendances, mais force est de constater que les actions menées demeurent timides et les règlements surtout appliqués par les acteurs de bonne foi. Les publicités ciblées que les grandes plateformes délivrent, sur la base du *data profiling*, les relances incessantes des sites visités, l'adaptation du prix des produits en fonction des recherches déjà effectuées : tout cela est aux antipodes de l'idée de consommateur libre et éclairé et constitue, à travers la personnalisation effectuée, un changement de paradigme par rapport à l'ère de la publicité de masse, transformant par ailleurs les messageries en centres dont il faut constamment trier les déchets. Un cran supplémentaire est franchi quand les fichiers de profils sont cédés d'entreprise à entreprise, faisant de nos

---

18. « Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes. »

ontologies marchandes parcellaires des produits à vendre dans tous les domaines (y compris les campagnes électorales). Le secteur bancaire n'est pas en reste, grand utilisateur du big data, qui dispose sur ses clients d'informations détaillées et sensibles qu'il ne se prive pas d'utiliser pour proposer ou refuser un service et ajuster son prix.

Choisir des opérateurs qui n'entrent pas dans ces logiques (Qwant plutôt que Google, par exemple) est l'une des solutions ouvertes et il est de plus en plus d'utilisateurs éclairés qui se tournent vers elles. Il n'en demeure pas moins que la réponse par le droit, au niveau international, n'est pas à la hauteur des enjeux et témoigne d'un certain rapport de force. Pourquoi avoir choisi, avec le RGPD, de nous faire consentir au traitement automatique de nos données site par site et non pas en bloc, ce qui nous aurait ainsi permis de naviguer en refusant ou en acceptant ces usages une fois pour toutes (quitte à changer dans certains cas de profil) ? Pourquoi ne pas avoir contraint les entreprises d'Internet (mais aussi la puissance publique) à la transparence sur les algorithmes qu'elles utilisent pour nous classer, ranger et nous faire (ou non) des propositions ? Pourquoi ne pas avoir tout bonnement interdit la cession des profils utilisateurs de la même manière que nous prohibons la vente d'organes ? Pourquoi n'est-il pas proposé aux clients des banques de refuser le traitement de leurs données à des fins autres que de statistiques globales ? Pourquoi avoir accepté les transactions financières à deux décimales après la virgule, ce qui a permis le développement du *trading* automatisé à haute fréquence dont l'intérêt social reste à prouver ?

Toutes ces évolutions ne sont pas combattues avec les armes suffisantes au plan national et international. Un protocole de l'ONU garantissant, au-delà des textes en cours de déploiement, le droit au respect des données personnelles ou un élargissement de la Déclaration universelle des droits de l'homme seraient une première réponse. Certes, certains acteurs économiques en souffriraient, mais les sommes ainsi économisées pourraient peut-être servir à mieux rémunérer les salariés qui produisent les biens et services finalement vendus. Ce serait aussi l'occasion de marquer que la société ne vaut pas uniquement en tant qu'elle est marchande, mais aussi au nom d'autres aspects de la vie.

C'est ainsi un triple programme sur le big data qui se dessine : scientifique, pour comprendre ce qu'il permet de connaître effectivement et la place qu'il laisse aux modèles mathématiques mais aussi conceptuels ; intellectuel, pour éclairer sur les biais que ne peuvent manquer de contenir des informations parcellaires sur les êtres que nous sommes ; politique, enfin, sur la transparence que la cité est en droit d'exiger de la sphère marchande et sur les principes de respect de la vie privée auxquels cette dernière n'a pas de raison particulière d'être soustraite.