



Source : sekulicn

Résumé

- L'explosion des données est une réalité incontestable dont il convient de prendre toute la mesure : même si d'autres déluges de données ont pu être rencontrés dans le passé, l'ampleur du phénomène actuel est inédite.
- Outre son impact environnemental, cette évolution pose, directement ou indirectement, différents problèmes au sein de nos sociétés : économie de l'attention, bulles de filtres, surcharge informationnelle, nouvelles inégalités, etc.
- Pour prévenir le risque de submersion lié à cette lame de fond, des solutions technologiques existent mais un combat culturel en faveur de démarches de sobriété s'impose pour limiter l'accumulation obsessionnelle de données, véritable « syndrome de Diogène numérique ».

Ludovic Haye, sénateur

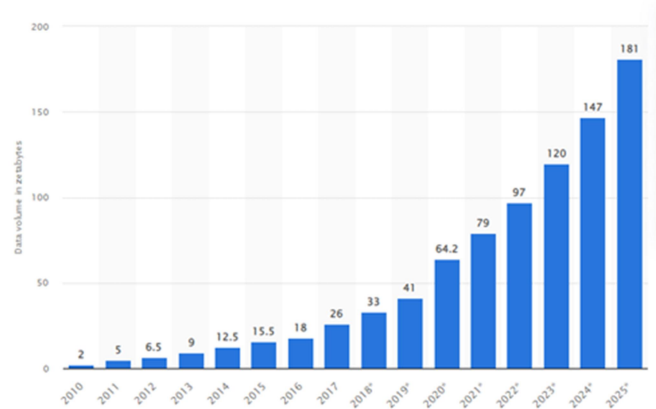
■ La longue histoire d'une question récurrente

La collecte, l'accumulation et le traitement des données furent des activités humaines bien avant l'apparition de moyens automatiques permettant d'effectuer ces opérations. Ces tâches présentes dès les premières civilisations urbaines – Mésopotamie ou Égypte – incombaient aux institutions comme l'Église et l'État. Comme le rappelle l'historien Robert Darnton « chaque âge a été, à sa manière, un âge de l'information »⁽¹⁾ : à cet égard, l'explosion des données n'est pas un phénomène nouveau et d'autres contextes ont pu engendrer des angoisses aujourd'hui oubliées⁽²⁾. Néanmoins, les précédents déluges de données paraissent bien modérés par rapport à l'ère actuelle de la civilisation numérique et du *Big Data*, marquée par les « 3 V » (volume, vitesse et variété des données).

■ Prendre la mesure de l'explosion des données

Le volume de données créées augmente en effet exponentiellement passant – au niveau mondial – de 2 zettaoctets en 2010 à 18 zettaoctets en 2016, puis, selon les prévisions, à 64 zettaoctets en 2020 et 181 zettaoctets en 2025⁽³⁾, ce qui pose notamment la question du stockage des données, question déjà traitée dans une note scientifique de l'Office⁽⁴⁾ et qui ne sera pas abordée dans la présente note. Les capacités humaines et technologiques pour traiter et classer les données ont du mal à suivre un tel rythme, en raison de coûts très élevés et de la concentration des moyens autour d'un nombre limité d'acteurs.

Évolution du volume de données (en zettaoctets)



Source : Statista

Selon Daniel Rosenberg, les données sont « une catégorie de faits et de principes définis, par accord, comme étant au-delà de l'argument »⁽⁵⁾ : les données ont donc un contenu variable, fonction d'un consensus social et historique. Par exemple, alors que les lettres étaient rarement sauvegardées dans le passé et n'étaient pas des données, les *e-mails* en sont de nos jours : ils sont conservés pour être scrutés par des algorithmes aujourd'hui et demain. De plus, alors qu'auparavant l'État était avec l'Église le principal collecteur de données, celles-ci sont aujourd'hui principalement récoltées par des entreprises privées à travers des sources multiples (Internet, capteurs physiques, caméras, satellites, systèmes de navigation, processus industriels, santé, activités

scientifiques, génomique...). Ces évolutions confirment la spécificité de l'explosion actuelle des données. Pour comprendre mais aussi visualiser et, le cas échéant, traiter les données, différentes typologies visent à les classer en fonction de critères divers, dont leur nature, leur structure ou, encore, leur usage⁽⁶⁾. Enfin, le fait que la quantité de données inexacts augmente fortement renforce la difficulté à traiter toutes les données disponibles.

■ Nouvelles opportunités offertes par le *Big Data*

L'engouement autour de l'analyse des données massives (*Big Data*) s'explique d'abord par les nouvelles opportunités qu'elle offre. Si les données isolées ont peu d'utilité, leur agrégation et leur traitement permettent d'en extraire des connaissances et de faire des prédictions. Alors qu'à l'aide des lois physiques dégagées par la recherche *a priori* de causes et de déterminants, les déductions permettent traditionnellement de faire des prédictions grâce à des modélisations, comme en météo, les technologies numériques recourent au *Big Data* à travers des approches inductives, en analysant les données massives afin d'en dégager des lois. On peut distinguer deux types d'analyse : l'application de modèles statistiques aux données massives et l'apprentissage automatique, dans lequel c'est la machine qui construit les modèles algorithmiques de prédiction, au cœur des recherches actuelles en intelligence artificielle ou IA⁽⁷⁾.

Les opportunités du *Big Data* sont nombreuses notamment sur un plan économique, comme en témoignent ces métaphores : nouvel or noir, pétrole du XXI^e siècle, révolution du *Big Data*, etc. Parmi les bénéfices escomptés pour les entreprises : l'amélioration de l'efficacité de l'entreprise et de la gestion des ressources humaines, l'optimisation de la relation client, par un marketing personnalisé et/ou prédictif, la monétisation des informations sur les clients, etc. Les domaines d'applications sont nombreux : la santé, les transports, les sciences, etc. On peut citer dans la santé des progrès en épidémiologie, en diagnostic et en précision des traitements⁽⁸⁾. Dans le transport, le *Big Data* peut servir à optimiser l'utilisation des différents moyens de transport et à personnaliser l'expérience du voyageur⁽⁹⁾.

Les données ont aussi conduit à l'émergence de nouveaux modèles d'affaires, mis en place notamment par les entreprises Google, Facebook, Amazon, Twitter, etc. Ces entreprises utilisent les données des utilisateurs pour le fonctionnement de leurs services, pour générer des revenus par la publicité ciblée⁽¹⁰⁾, souvent en revendant ces données sous la forme de profils mis aux enchères *via* des *data brokers*, et pour prévoir à l'aide d'algorithmes les intentions individuelles, ce que l'on va acheter ou ce que l'on va faire. L'explosion des données a conduit beaucoup d'entreprises, même en dehors du numérique, à ajouter le traitement de données à leurs activités traditionnelles⁽¹¹⁾ et à revisiter celles-ci en vue de les optimiser.

■ Un impact environnemental et sociétal préoccupant

Au-delà de ses bénéfices réels ou escomptés, le *Big Data* pose différents problèmes. Les outils numériques ne relèvent pas que du « virtuel » : leurs conséquences sont bien réelles. Ainsi, leur impact environnemental croît de 6% par an : ils causent au moins 3,5 % des émissions mondiales de gaz à effet de serre (GES) et 4,2 % de la consommation mondiale d'énergie primaire en 2019⁽¹²⁾. Un rapport de The Shift Project souligne d'ailleurs que 44 % des Français considèrent les ordinateurs et Internet comme une menace pour l'environnement.

Sur l'usage des données massives s'est construite une économie de l'attention, dans laquelle les entreprises profitent des données pour capter de plus en plus finement l'attention des utilisateurs, les exposer à plus de publicité et, de façon circulaire, collecter encore plus d'informations. Ce « nouveau modèle capitaliste », que certains qualifient de « capitalisme cognitif »⁽¹³⁾, pousse les entreprises du numérique à enfermer, d'une part, l'internaute dans des bulles de filtres⁽¹⁴⁾, qui confirment ses points de vue, confortent ses croyances et les sujets qui semblent l'intéresser, et à inciter, d'autre part, le consommateur à toujours consommer davantage sous l'effet de cette « industrie de l'influence », ce qui par rétroaction aggrave la problématique environnementale⁽¹⁵⁾. Cette course à l'attention a aussi profondément transformé le journalisme. Aujourd'hui, l'intérêt des médias porte de plus en plus sur la vitesse de publication plutôt que sur la qualité de l'information⁽¹⁶⁾. D'un point de vue politique, le glissement vers une information pauvre et uniformisée qui cherche à attirer avant d'informer est une menace pour la démocratie. Celle-ci est également fragilisée par la surveillance de masse que permet le numérique, aggravée par les enjeux géopolitiques propres au *Big Data*, la domination des entreprises et des *data-centers* américains caractérisant la gouvernance des données dans un véritable capitalisme de surveillance⁽¹⁷⁾. Ce dernier peut se mettre au service d'un État, comme en témoigne le programme américain *Prism* qui, depuis 2007, permet aux États-Unis de récupérer dans le monde entier les milliards de données personnelles collectées par les entreprises. La Chine n'est pas en reste, n'hésitant pas à utiliser TikTok à des fins de surveillance, comme l'illustre l'exemple récent de l'espionnage de journalistes étrangers⁽¹⁸⁾.

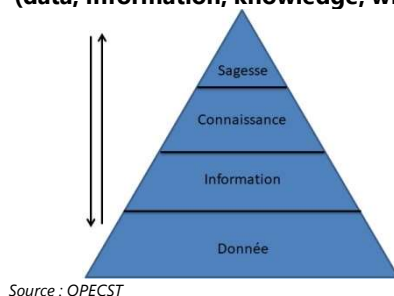
L'ère du *Big Data* accroît également les inégalités liées à la transformation numérique. Selon Eszter Hargittai, une seconde fracture numérique est ainsi apparue⁽¹⁹⁾ : si la première, liée à l'écart d'accès aux technologies numériques s'est réduite, la seconde, qui relève de la capacité à s'approprier les technologies numériques (en fonction du milieu social et du niveau d'éducation), se creuse. De plus, l'utilisation du *Big Data* peut renforcer des inégalités existantes du fait des biais des algorithmes dont les résultats reproduisent les biais des échantillons de données utilisés. Le

logiciel de prédiction de la récidive utilisé par la justice américaine, *Compas*, a ainsi tendance à surestimer le risque pour les personnes noires.

■ Le lien indirect entre explosion des données et surcharge informationnelle

Pour comprendre la relation entre explosion des données et surcharge informationnelle, il est nécessaire d'identifier les articulations entre les données, l'information, les connaissances et la sagesse. Ces concepts sont liés mais les frontières qui les séparent restent floues et appellent des précisions. Une donnée est une description élémentaire d'une réalité, qui, prise isolément, n'a pas ou quasiment pas de sens ; c'est un élément ou un ensemble d'éléments objectifs sur un fait. L'information est une série de données contextualisées dont on peut tirer une signification. En comprenant le sens de l'information puis en l'intégrant à notre système de savoirs et de valeurs elle acquiert le statut de connaissance. Enfin, des connaissances, on peut tirer la sagesse, c'est-à-dire des comportements conformes à une éthique. Albert Einstein écrivait ainsi que « *la connaissance s'acquiert par l'expérience, tout le reste n'est que de l'information* »⁽²⁰⁾ et T. S. Eliot se demandait « *où est la sagesse que nous avons perdue dans la connaissance ? Où est la connaissance que nous avons perdue dans l'information ?* »⁽²¹⁾. Les relations entre ces concepts peuvent être illustrées par l'exemple de la météorologie. Les mesures de température relevées chaque jour dans une station météo sont des données. Une courbe donnant l'évolution dans le temps de la température moyenne dans un lieu est une information. Le fait que la température sur Terre augmente en fonction de l'activité humaine est une connaissance. Moduler l'activité anthropique selon son impact sur le réchauffement climatique est de la sagesse. Le graphique ci-après modélise ces relations.

**La pyramide DIKW
(data, information, knowledge, wisdom)**



Source : OPECST

La volumétrie d'information (trop d'informations à traiter par un individu ou une organisation dans un temps imparti entraîne de la surcharge informationnelle), la surcharge cognitive (atteinte des limites des capacités cognitives des individus à traiter un volume d'information), enfin la surcharge de communication (une grande part de la surcharge d'information provient des nouveaux moyens de communication électronique : *e-mails*, forums de discussion en ligne, messageries instantanées, etc.) sont trois paramètres qui convergent vers la surcharge informationnelle, définie

comme un concept tridimensionnel par Michel Kalika et d'autres chercheurs⁽²²⁾.

La surcharge informationnelle ou *infobésité*

La surcharge informationnelle désigne l'excès d'informations qu'une personne ne peut traiter ou supporter sans nuire à elle-même. Apparue pour la première fois en 1962, le terme surcharge informationnelle est popularisé en 1970 par le futurologue américain Alvin Toffler. Le concept d'*infobésité* est quant à lui apparu en 1993 sous la plume de David Shenk, qui établit un parallèle symbolique entre surcharge informationnelle et obésité, le surplus d'informations qui étouffe nos processus intellectuels serait pour lui à rapprocher de l'excès de matières grasses qui caractérise l'obésité.

■ Difficultés liées à la surcharge informationnelle et mutations de nos structures cognitives

Comme le souligne un rapport du Centre d'analyse stratégique et de la Direction générale du travail, « *La "surinformation" (ou infobésité) est l'un des plus grands problèmes à résoudre par les organisations pour les dix prochaines années* »⁽²³⁾. La surcharge informationnelle est un concept récent qui décrit l'idée ancienne selon laquelle la capacité de traitement de l'information des êtres humains est limitée, à la fois en matière de quantité et de capacité de conservation d'informations sur une période donnée⁽²⁴⁾. Bien entendu, la réaction à cette surcharge informationnelle est inégale et dépend des individus car, comme le rappelle Michel Kalika, « *nous ne sommes pas tous égaux face à la surcharge informationnelle* »⁽²⁵⁾. Plusieurs causes peuvent expliquer celle-ci qui serait, pour chacun, une conséquence de l'explosion de la communication, à travers la quantité de messages reçus et la quantité d'applications et de canaux de communication utilisés⁽²⁶⁾. Les technologies de l'information et de la communication (TIC) ont permis d'augmenter significativement ces quantités. La surcharge informationnelle est aussi encouragée par les principes régissant nos comportements sociaux : addiction à la communication, angoisse de rater une information (*Fear of Missing Out* ou FOMO, syndrome pouvant engendrer des dépressions, particulièrement présent chez les jeunes mais qui se généralise), mutations et instabilités propres au monde du travail⁽²⁷⁾, *multitasking*, disparition des frontières entre sphère professionnelle et sphère privée, gratuité apparente des communications électroniques, etc. En particulier, les *e-mails* semblent être, selon Caroline Sauvajol-Rialland, l'un des principaux outils de la surcharge informationnelle et communicationnelle car « *le rythme des échanges est devenu trop rapide et incertain. La croissance du volume, exponentielle, rend de plus en plus difficiles la gestion et le traitement des courriels par les salariés* »⁽²⁸⁾. Cette *infobésité* entraîne de nombreuses conséquences, dont les principales sont : au plan individuel, stress, anxiété, dépression, réduction de la créativité, épuisement professionnel ; au niveau des

organisations, baisse de productivité, saturation, désorganisation ; et sur le plan sociétal, gaspillage d'énergie, empreinte environnementale, et carbone en particulier, considérable.

Une autre difficulté provient du mode de raisonnement utilisé pour traiter les données. Si les raisonnements par induction sont prometteurs et que, temporairement du moins, le raisonnement par déduction décline, ces deux raisonnements devraient cohabiter sans quoi le Big Data risque de conduire à la « mort de la théorie scientifique »⁽²⁹⁾. Il convient également de faire attention à la dérive des usages de statistiques : les volumes de plus en plus massifs de données multiplient les occasions de réaliser des traitements apportant parfois faussement l'impression de capturer la vérité. Selon la formule de deux chercheurs, « les chiffres sont comme les gens. Si on les torture assez, on peut leur faire dire n'importe quoi »⁽³⁰⁾. Ainsi que l'écrivait dans une sorte de mise en garde Alain Desrosières, « les données ne sont pas données »⁽³¹⁾, elles sont construites et n'ont pas de signification en elles-mêmes, elles ne décrivent qu'une partie de ce qui se passe en réalité et s'accompagnent de très nombreux biais.

En outre, nos structures cognitives elles-mêmes pourraient bien être modifiées par cette surcharge informationnelle à l'ère de l'explosion des données : plus qu'une attraction temporaire de l'attention, nous courons le risque d'une dissolution durable de nos capacités à nous souvenir, à nous concentrer et à traiter l'information, les enfants étant particulièrement exposés à ces évolutions⁽³²⁾. Nous serions les victimes d'une civilisation dans laquelle la santé mentale est menacée à partir de plus de 30 minutes consécutives d'exposition aux réseaux sociaux et aux écrans d'Internet, une civilisation dans laquelle la durée d'attention sur un sujet serait réduite à quelques secondes, faisant de chacun un poisson rouge enfermé dans le bocal de son écran, ainsi que l'affirme Bruno Patino dans un livre important, d'où la question posée par Nicholas Carr, « Google nous rend-il tous idiots »⁽³³⁾ ? Par ailleurs, plusieurs études neurologiques montrent les limites de nos capacités réelles au *multitasking*, soulignant de surcroît comment notre attention est de plus en plus facilement perturbée par des distractions extérieures⁽³⁴⁾.

■ Solutions technologiques et recommandations

À l'heure d'une sorte de « syndrome de Diogène numérique » où chaque personne, chaque organisation, stocke le plus de données possibles, les remèdes à cette « syllogomanie 2.0 » sont indispensables. La recherche de solutions technologiques pour prévenir et contenir l'explosion des données et la surcharge informationnelle doit se renforcer, notamment la recherche sur l'agrégation et la curation de

données. Celle-ci consiste à sélectionner, éditer et partager les seuls contenus appropriés⁽³⁵⁾ : en sélectionnant plus efficacement les données pertinentes, il serait possible de limiter le volume de données conservées. De même, plusieurs projets sur les bases de données cherchent à effacer progressivement certaines données grâce à une technologie de « pourriture contrôlée »⁽³⁶⁾. On peut aussi citer les « Personal Information Management Systems » (PIMS), technologies encore embryonnaires, qui pourraient permettre aux utilisateurs de gérer et de contrôler toutes leurs données personnelles. De manière générale, l'informatique appuyée sur le *Big Data* ouvre de grandes perspectives. Des avancées majeures sont ainsi attendues au cours des prochaines années pour améliorer les technologies numériques en termes de gestion et de traitement des données : progrès en matière d'architectures de calcul et de stockage, d'intégration et de curation des données, de modèles, de logiciels, d'algorithmes, d'intelligence artificielle, de dispositifs d'acquisition et de visualisation, ou encore de technologies quantiques⁽³⁷⁾.

Mais la sobriété numérique est d'abord un combat culturel. C'est pourquoi la formation aux enjeux du numérique est essentielle. Il ne s'agit pas tant d'apprendre à utiliser ces outils qu'à comprendre leurs conséquences et la législation qui les encadre. Par exemple, en dépit du règlement général sur la protection des données (RGPD) qui encadre le traitement des données personnelles au sein de l'Union européenne, seuls 6 % des Français refusent les cookies sur Internet⁽³⁸⁾. La législation sur la protection des données pourrait être renforcée ou, au moins, s'accompagner d'une plus grande pédagogie. Il conviendrait de promouvoir la sobriété numérique et de développer l'hygiène numérique à l'échelle individuelle et au niveau des organisations (entreprises et administrations), ce qui passerait par le nettoyage régulier de ses données (*clean up days*).

Enfin, l'ambition de la souveraineté numérique et de l'indépendance à l'égard des grandes entreprises du numérique requiert un débat public sur le monopole exercé par ces plateformes, le plus souvent américaines, ainsi que sur les algorithmes captant notre attention ou provoquant des bulles de filtre qui nous enferment. Cette ambition interroge la possibilité d'une gouvernance française et européenne des données et de leurs infrastructures. Plusieurs personnes auditionnées ont fait valoir que la solution ultime restait la déconnexion pure et simple⁽³⁹⁾, réflexe encore trop peu présent dans notre société.

Sites Internet de l'Office :

<http://www.assemblee-nationale.fr/commissions/opepst-index.asp>

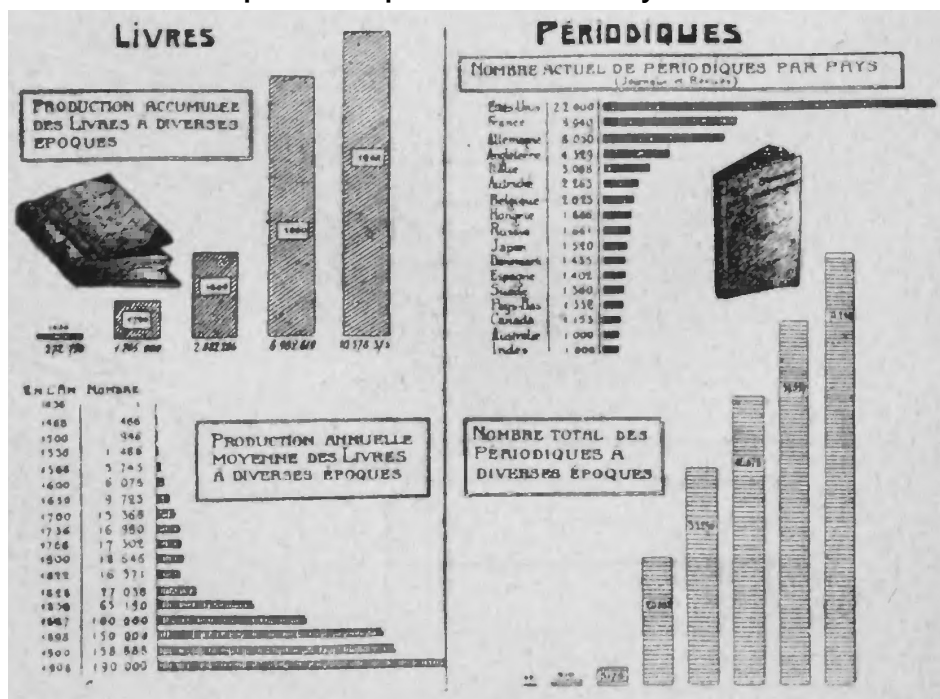
<http://www.senat.fr/opepst>

Références

¹ Selon Robert Darnton « Every age was an age of information, each in its own way ». Cf. « An early information society: News and the media in eighteenth-century Paris », *American Historical Review*, n° 1, vol. 105, 2000. Cf. <https://academic.oup.com/ahr/article/105/1/1/64445>

² Barnaby Rich, dès 1613, se plaignait de la prolifération de la littérature : « L'une des maladies de notre époque est la multiplicité des livres. Ils surchargent tellement le public que celui-ci est incapable de digérer l'abondance de matières oiseuses quotidiennement écloses et répandues dans le monde ». Derek de Solla Price a ainsi décrit un « effet Barnaby Rich » pour décrire la surcharge informationnelle. De même, Robert Burton, en 1621, s'exclamait : « Quelle surabondance de livres ! Qui peut les lire ? Nous aurons un vaste chaos dû aux livres, ils nous oppriment, la lecture fatigue nos yeux, nos doigts qui tournent les pages (...) ». En 1934, dans son *Traité de documentation*, Paul Otlet soulignait l'inflation documentaire, notant à propos des livres et documents que « leur masse énorme accumulée dans le passé s'accroît chaque jour, chaque heure, d'unités nouvelles en nombre déconcertant, parfois affolant (...). D'eux comme de l'eau tombée du ciel, on peut dire qu'ils peuvent provoquer l'inondation et le déluge ou s'épandre en irrigation bienfaisante » (cf. aussi son *Manuel de la bibliothèque publique*, 1922.)

Perception de l'explosion des données il y a un siècle



Source : Paul Otlet, *Manuel de la bibliothèque publique*, 1922.

Plus récemment, Emmanuel Didier constate la récurrence des discours sur l'explosion de données : « il y a une très belle chose qui frappe quand on est historien de la statistique, c'est que, depuis au moins 1850, à chaque fois qu'il y a une révolution technologique, s'installe un discours selon lequel nous sommes submergés par les données » (cf. ses propos recueillis par Edouard Mien et Alizé Papp, « L'afflux massif de données est-il nouveau ? », *Regards croisés sur l'économie*, 2018, n°23). Si les explosions précédentes de données ne nous dérangent plus et que la surabondance de livres n'est plus perçue comme un déluge de données c'est notamment parce que les outils de gestion de l'information ont permis de canaliser ces surcharges informationnelles, ainsi que l'ont expliqué Pascal Griset et Bruno Strasser lors de leurs auditions, citant l'exemple des encyclopédies ou du memex (memory extender) imaginé par Vannevar Bush en 1945 et préfigurant les systèmes informatiques de gestion de bases de données.

³ Les données de Statista se basent notamment sur un rapport d'IDC, qui, outre les données numériques créées ou répliquées, fournit aussi d'autres prévisions, sur la répartition d'entreprises utilisant les données, sur les bénéfices possibles de ces données ou, encore, sur le stockage de données. Il convient d'observer qu'une part réduite des données numériques créées est conservée : ainsi seules 2 % des données produites en 2020 ont été sauvegardées en 2021. Cf. leur site régulièrement actualisé <https://www.statista.com/> et, en particulier ce graphique relative à la croissance des données <https://www.statista.com/statistics/871513/worldwide-data-created/> <https://fr.statista.com> ainsi que le rapport de David Reinsel, John Gantz et John Rydning, *The Digitization of the World*, 2018, lien : <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

⁴ L'Office a en effet déjà abordé la question du stockage de données, ses enjeux actuels ainsi que la solution du stockage sous la forme d'ADN, dans la Note scientifique de Ludovic Hays « Le stockage de données sous la forme d'ADN », n° 29, décembre

2022, disponible aux liens suivants : sur le site du Sénat www.senat.fr/rap/r21-285/r21-2851.pdf ou sur celui de l'Assemblée nationale https://www2.assemblee-nationale.fr/content/download/451536/4391889/version/1/file/OPECST_2021_0063_Note_stockage_ADN.pdf

⁵ Traduction de « a category of facts and principles that were, by agreement, beyond argument » cf. le livre de Daniel Rosenberg, *Raw Data is an Oxymoron*, 2013.

⁶ En effet, on peut tout d'abord catégoriser les données en fonction de leur **contenu**, soit une donnée quantitative (un nombre), soit une donnée qualitative, elle-même séparée en deux catégories : donnée nominale (ne pouvant être classée, comme la couleur des yeux) ou ordinale (qui peut être triée, comme la taille des vêtements). Il est aussi possible de classer les données en fonction de leur **format**, les données structurées (soumises à un format fixe dépendant d'un modèle) s'opposant aux données non structurées (images, audios ou vidéos par exemple dont la forme n'est pas connue à l'avance), même s'il existe des données semi-structurées, intermédiaires entre données structurées et non structurées (elles n'ont pas de structure formelle mais possèdent tout de même des propriétés d'organisation, souvent grâce aux métadonnées, qui sont des « descriptions attachées à des données ou à des documents, dans le but de faciliter leur classement et l'accès à leur contenu informationnel. Elles peuvent être utilisées pour nommer, décrire, cataloguer et indiquer la propriété intellectuelle ou le droit d'auteur », cf. Mokrane Bouzeghoub et Rémy Mosseri (dir.), *Les Big Data à découvert*, CNRS éditions, 2017). On distingue ainsi les lacs de données (data lakes), qui peuvent accueillir tout type de données, notamment les non structurées, alors que les entrepôts de données (data warehouses) réunissent des données structurées, transformées et nettoyées. Il est également possible de classer les données en fonction de leur **fréquence d'utilisation**, entre données chaudes, utilisées très fréquemment (et qui nécessitent donc des supports de stockage accessibles rapidement ou immédiatement), et données froides, auxquelles on n'accède que rarement, voire jamais. Enfin, on peut catégoriser les données en fonction de leurs **sources**, les données first party sont ainsi des données directement récupérées auprès des utilisateurs ou clients de l'entreprise qui collecte les données, les données second party des données achetées auprès de l'entreprise qui les a collectées, les données third party des données achetées à une entreprise, mais qui ne les a pas collectées elle-même. Les opérations de classification des données sont très importantes, elles permettent une gestion et une protection plus efficace de celles-ci.

⁷ L'Office a, dès 2017, approfondi le sujet des technologies d'intelligence artificielle et de leurs enjeux dans le rapport n° 464 (2016-2017) de M. Claude de Ganay député, et Mme Dominique Gillot, sénatrice, « Pour une intelligence artificielle maîtrisée, utile et démystifiée ». Cf. les liens disponibles sur le site du Sénat <http://www.senat.fr/notice-rapport/2016/r16-464-1-notice.html> ainsi que sur le site de l'Assemblée nationale https://www.assemblee-nationale.fr/dyn/14/dossiers/intelligence_artificielle_maitrisee_utilite

⁸ Dans son livre sur le Big Data, Pierre Delort donne l'exemple de Google Flu Trends qui avec les centres pour le contrôle et la prévention des maladies (CDC) examinait entre 2008 et 2015, la corrélation entre les recherches Google comportant des termes en lien avec la grippe et le nombre de personnes contaminées par la grippe : ses résultats convaincants étaient annoncés deux semaines avant la détection de l'épidémie. Cet exemple d'une opportunité offerte par les données dans le domaine de la santé présente cependant des limites puisque l'on mesure aussi les réactions de la population à une médiatisation de la grippe ou à des rumeurs : faire des recherches sur une maladie ne signifie pas être malade. Cf. Pierre Delort, *Le Big Data*, PUF, 2018. Plus récemment, la pandémie de COVID a invité à une réflexion sur les facteurs d'analyse pertinents et la qualité des indicateurs. Ces exemples invitent à la prudence dans l'interprétation des données. De manière plus générale, l'usage des données en santé doit se faire à la lumière de principes scientifiques et éthiques, comme le rappelle l'Office parlementaire espagnol sur la science et la technologie, qui a ainsi fait récemment un point sur les relations entre l'intelligence artificielle et la santé dans un court rapport paru en novembre 2022, cf. https://www.oficinac.es/sites/default/files/informes/20221114_Report%20C%20IA_0.pdf

⁹ Un travail du Parliamentary Office of Science and Technology (POST) du Royaume-Uni explique comment les données sont aujourd'hui utilisées dans le transport et les perspectives qu'elles ouvrent, en particulier les possibilités d'optimisation des transports et la future démocratisation des voitures connectées. Il développe aussi les futurs enjeux comme l'accès aux données pour le public, le regroupement des données entre tous les acteurs du transport ou la gouvernance des données. Cf. la note d'août 2014 du POST, « Big and Open Data in Transport » : <https://post.parliament.uk/research-briefings/post-pn-472/>

¹⁰ Ainsi que l'expliquent Serge Abiteboul et Valérie Peugeot, ces sites proposent souvent des services « gratuits », dont la gratuité est illusoire : « l'utilisateur fournit en échange, non pas de l'argent sonnante et trébuchante, mais son attention et de l'info-monnaie, des informations sur lui-même ». Cette réalité s'accompagne d'un « phénomène monopolistique ou oligopolistique à l'échelle planétaire » : un petit nombre de grandes entreprises accaparent la plus grande masse des données personnelles, ce qui leur donne un avantage commercial, nuit à l'innovation, accentue la relation toujours plus asymétrique entre les utilisateurs et ces entreprises, fragilise la protection des données personnelles et pose le problème de la mobilisation de ces infrastructures informationnelles à des fins de surveillance. Cf. Serge Abiteboul et Valérie Peugeot, *Terra Data. Qu'allons-nous faire des données numériques ?*, Le Pommier, 2017.

¹¹ En effet, plusieurs métiers autour du Big Data se sont développés avec l'arrivée de cette explosion de données. Ces métiers s'articulent autour de différentes étapes nécessaires pour travailler avec le Big Data : la collecte des données ; le nettoyage des données ; l'analyse des données ; enfin, la présentation des résultats. En voici quelques exemples : Data engineer, Data

steward, Data architect, Data scientist, Data analyst, Data visualisation, Data protection officer, etc. Ces métiers sont de plus en plus demandés, avec des formations qui commencent à émerger. Serge Abiteboul propose avec d'autres des pistes pour former les futurs « data scientists » et pour créer de nouvelles filières de formation (cf. Serge Abiteboul et al., « L'émergence d'une nouvelle filière de formation : "data scientist" », INRIA Saclay, CNAM-Paris, CEDRIC laboratory, 2014).

¹² Le chiffrage des émissions de GES dues au numérique repose sur le travail de Hugues Ferreboeuf (rapport « Pour une sobriété numérique », The Shift Project, 2018, mis à jour en 2021 : <https://theshiftproject.org/wp-content/uploads/2018/11/Rapport-final-v8-WEB.pdf>), l'évaluation de la consommation mondiale d'énergie primaire a été faite par Frédéric Bordage (cf. « Empreinte environnementale du numérique mondial », GreenIT.fr, 2019 : www.greenit.fr/empreinte-environnementale-du-numerique-mondial/?msclkid=6b823725ba7211ec92840260f951396f). Ces données sont reprises dans le rapport commun de l'Ademe et de l'Arcep, « Évaluation de l'impact environnemental du numérique en France et analyse prospective », 2022 : https://www.arcep.fr/uploads/tx_gspublication/etude-numerique-environnement-ademe-arcep-note-synthese_janv2022.pdf. Le cycle de vie des données est constitué de quatre étapes qui consomment des ressources naturelles et émettent des gaz à effet de serre (GES). Il faut tout d'abord extraire les métaux pour fabriquer les objets connectés, capteurs, smartphones, serveurs, etc. Il faut fabriquer ces différents équipements. La donnée est alors transportée, traitée, stockée et assemblée à d'autres données. Enfin, les ressources sont consommées lors de la phase d'usage de ces dispositifs. (cf. l'ouvrage du groupe ECOINFO coordonné par Françoise Berthoud, Les impacts écologiques des technologies de l'information et de la communication, EDP Sciences, 2012). 44 % de l'empreinte carbone du numérique à l'échelle mondiale résulte de la fabrication des terminaux, centres informatiques et des réseaux, et 56 % de leur utilisation. À l'échelle nationale, le cabinet Citizing a estimé dans une étude que le numérique constitue 2 % du total des émissions de GES en 2019. L'étude démontre que les terminaux (smartphones, ordinateurs portables, etc.) sont à l'origine d'une très grande part des impacts environnementaux du numérique (81 %). En particulier, la fabrication et la distribution de ces terminaux engendrent 86 % de leurs émissions totales et sont donc responsables de 70 % de l'empreinte carbone totale du numérique en France. Ce chiffre est bien supérieur aux 44 % au niveau mondial en raison de l'importation d'équipements numériques de pays d'Asie, où l'intensité carbone de l'électricité est bien plus importante qu'en France (cf. le rapport d'information de Guillaume Chevrollier et Jean-Michel Houllégatte au nom de la commission de l'aménagement du territoire et du développement durable du Sénat, « Pour une transition numérique écologique », n° 555, 2019-2020 : www.senat.fr/rap/r19-555/r19-555.html). Il faudrait ajouter à ces impacts environnementaux le faible recyclage des objets connectés, risques d'une nouvelle source massive de déchets électroniques (74 millions de tonnes par an d'ici à 2030), avec 5,3 milliards de téléphones portables jetés pour la seule année 2022 (cf. les rapports 2022 du forum international sur les déchets d'équipements électriques et électroniques et du Global e-waste monitor ainsi que celui plus ancien de Thomas Graedel (dir.), « Recycling rates of metals, a status report », UNEP, International Resource Panel, 2011 : www.unep.org/resources/report/recycling-rates-metals-status-report?msclkid=eff8f215ba7211ec88247081146be314). En outre, comme l'a souligné Serge Abiteboul lors de son audition, on a tendance à focaliser l'attention sur l'impact environnemental des seuls data centers alors qu'il faudrait ne pas oublier la quantité de données numériques en circulation elle-même, les vidéos, surtout celles en haute ou très haute définition, étant un aspect important du problème.

¹³ L'expression « nouveau modèle capitaliste » se retrouve dans un article d'Alizé Papp « L'infobésité, une épidémie à l'âge des nouvelles technologies de l'information et de la communication » (Regards croisés sur l'économie, n° 23, 2018, cf. https://www.cairn.info/article.php?ID_ARTICLE=RCE_023_0105). Et c'est surtout Yann Moulier-Boutang qui a théorisé la notion de capitalisme cognitif, troisième stade du capitalisme après sa naissance mercantiliste et sa forme industrielle (cf. Le capitalisme cognitif : la nouvelle grande transformation, Éditions Amsterdam, 2007) D'autres auteurs ont écrit sur cette économie de l'attention, comme Yves Citton dans le livre dont il a dirigé la rédaction (cf. L'économie de l'attention, nouvel horizon du capitalisme ?, La Découverte, 2014.). Dès 1969, Herbert Simon soulignait : « Dans un monde riche en information, l'abondance d'information entraîne la pénurie d'une autre ressource : la rareté devient ce que consomme l'information. Ce que l'information consomme est assez évident : c'est l'attention de ses receveurs. Donc une abondance d'information crée une rareté de l'attention et le besoin de répartir efficacement cette attention parmi la surabondance des sources d'informations qui peuvent la consommer » (cf. Herbert Alexander Simon, The science of the artificial, MIT Press, 1969, traduction de Jean-Louis Le Moigne).

¹⁴ Cf. Eli Pariser, à l'origine de l'expression, The Filter Bubble : What the Internet Is Hiding from You, Penguin Press, 2011.

¹⁵ Cf. Hervé Le Crosnier, « De l'information à l'influence », revue Documentation et bibliothèques, numéro spécial « Quelle société de l'information ? Pour quelles bibliothèques/services d'archives ? », volume 64, n° 4, 2018.

¹⁶ Julia Cagé, Nicolas Hervé et Marie-Luce Viaud montrent ainsi que 64 % de l'information publiée en ligne correspond à du simple copié-collé. Dans leur étude, ils n'ont pas seulement regardé la quantité d'informations copiées, mais aussi la vitesse de diffusion de ces informations d'un site à un autre. En effet, pour les reprises pures et simples d'informations, « la moitié des événements se propagent en moins de 25 minutes ; un quart en seulement... 230 secondes, et 10 % en seulement 4 secondes » (cf. Julia Cagé, Nicolas Hervé et Marie-Luce Viaud « L'information à tout prix », Questions de communication, n° 32, 2017.). La problématique n'est pas totalement nouvelle puisque l'Agence Havas a systématisé l'usage des pigeons voyageurs et a été en pointe dans celui du télégraphe. Toutefois, d'après Alizé Papp, ce rythme effréné est accéléré par la civilisation numérique qui a pour conséquence une uniformisation des contenus et une baisse de qualité du bien offert. De

plus, cette évolution « faciliterait l'expansion de l'infobésité en habituant les individus à lire des contenus uniformisés et peu digérés, c'est-à-dire une forme de malbouffe informationnelle » (cf. son article « L'infobésité, une épidémie à l'âge des nouvelles technologies de l'information et de la communication », *Regards croisés sur l'économie*, n° 23, 2018).

¹⁷ Sur ce sujet, Amaël Cattaruzza développe dans son livre l'idée que l'espace numérique est un espace « (géo)politique » dans lequel il faut s'inquiéter de la domination des firmes américaines. Bien que certains États essaient de développer leurs propres datacenters pour rétablir leur souveraineté numérique, c'est-à-dire maintenir les données sur leurs territoires, la France a connu l'échec du projet Andromède en 2011 (cf. Amaël Cattaruzza, *Géopolitique des données numériques, Pouvoir et conflits à l'heure du Big Data, Le cavalier Bleu*, 2019). À propos de la notion de « capitalisme de surveillance », cf. les références suivantes : Mark Andrejevic, « Surveillance in the Digital Enclosure », *The Communication Review*, vol.10, n° 4, 2007 ; John Bellamy Foster et Robert McChesney, « Surveillance Capitalisme. Monopoly-Finance Capital, the Military-Industrial Complex, and the Digital Age », *Monthly review*, vol.66, n° 3, 2014 ; Viktor Mayer-Schönberger et Thomas Ramge, *Reinventing Capitalism in the Age of Big Data*, Basic Books, 2018 ; Virginia Eubanks, *Automating Inequality: How high-tech Tools Profile, Police and Punish the Poor*, St. Martin's Press, 2018 ; Shoshana Zuboff, *L'âge du capitalisme de surveillance*, Zulma, 2020 ; Cédric Durand, *Techno-féodalisme. Critique de l'économie numérique*, La Découverte, 2020 ; Charleyne Biondi, *Dé-coder. Une contre-histoire du numérique*, Bouquins, 2022.

¹⁸ Cf. par exemple https://www.lemonde.fr/pixels/article/2022/12/23/l-entreprise-chinoise-bytedance-propretaire-de-tiktok-admet-avoir-espionne-des-journalistes_6155484_4408996.html

¹⁹ Cf. Eszter Hargittai, « Second-Level Digital Divide? Differences in People's Online Skills? », *First Monday*, n°7 (4), 2002 ; Eszter Hargittai et Paul Di Maggio, « The New Digital Inequality? Social Stratification among Internet Users? », intervention au congrès annuel de l'American Sociological Association, Chicago, 2002 ; le rapport du Conseil national du numérique, « Citoyens d'une société numérique : pour une nouvelle politique d'inclusion », 2013 ; ou, enfin, Fabien Granjon et al., *Inégalités numériques. Clivages sociaux et modes d'appropriation des TIC*, Hermès/Lavoisier, 2008. Selon Francis Jauréguiberry, il existerait des « nouveaux pauvres des télécommunications » qui ne peuvent pas échapper à l'obligation de répondre immédiatement à des urgences et doivent vivre dans l'interpellation continue sans pouvoir se déconnecter.

²⁰ Cf. Caroline Sauvajol-Riolland, *Infobésité, Comprendre et maîtriser la déferlante d'informations*, Vuibert, 2013.

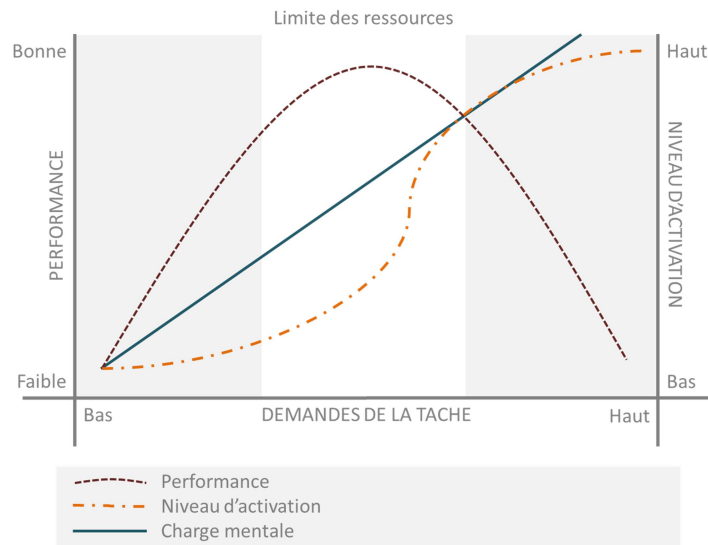
²¹ Traduction de « Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information ? » dans T. S. Eliot, *The Rock*, 1934. Tous les spécialistes ne sont pas d'accord pour séparer de cette manière les notions d'information et de donnée, cependant de nombreux auteurs et entreprises les distinguent ainsi. Pour certains d'entre eux, c'est justement le stockage de données qui n'a pas de sens, d'abord parce que trop de données rend l'identification de celles qui sont pertinentes difficile, ensuite parce que les données n'ont pas de signification en elles-mêmes, à la différence des informations (cf. Thomas H. Davenport et Laurence Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1998).

²² Cf. par exemple Éric Campoy, Michel Kalika et Henri Isaac « Surcharge informationnelle, urgence et TIC. L'effet temporel des technologies de l'information », *Management & Avenir*, vol.3, n° 13, 2007.

²³ Cf. le rapport commun du Centre d'analyse stratégique et de la Direction générale du travail, coordonné par Tristan Klein, « L'impact des TIC sur les conditions de travail », 2012, disponible au lien suivant : http://archives.strategie.gouv.fr/cas/system/files/raptic_web_light_final28022012.pdf

²⁴ La mémoire de travail, forme de mémoire courante à court terme qui permet de stocker et manipuler des informations pendant une courte durée en vue de les utiliser pour accomplir une tâche, ne peut gérer qu'un nombre limité d'informations à la fois, autour de sept plus ou moins deux (cf. l'un des articles les plus cités en psychologie, George A. Miller « The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information », *Psychological Review*, vol. 63, n° 2, 1956), depuis une dizaine d'années, les recherches ont ramené ce nombre d'objets à trois ou quatre (cf. Jeanne Farrington, « From the Research: Myths Worth Dispelling. Seven Plus or Minus Two », *Performance Improvement Quarterly*, vol. 23, n° 4, 2011). Parmi les études qui se sont intéressées aux performances des individus en fonction de la charge mentale imposée, on peut citer la métaanalyse de Mark S. Young (cf. Mark S. Young et al., « State of science: mental workload in ergonomics », *Ergonomics*, vol. 58, n° 1, 2015), sur laquelle Rémi Mounier s'est basé pour produire la figure ci-après. On observe notamment qu'autour d'un niveau maximum de performance pour une charge mentale précise, la productivité décroît rapidement au-dessus et en dessous. Cependant, les caractéristiques de cette courbe dépendent bien évidemment aussi de chaque individu et de la tâche assignée.

Performance optimale selon la charge mentale



Source : blog de Rémi Mounier, Octopus Ergonomie, article sur « La surcharge cognitive (ou mentale) », <https://octopus-ergonomie.com/blog-surcharge-mentale-cognitive-15>

²⁵ Formule employée par Michel Kalika lors de son audition.

²⁶ Pour certains auteurs, comme David Shenk, Caroline Sauvajol-Rialland ou Michel Kalika, il ne s'agit pas uniquement d'une surcharge informationnelle, mais d'une surcharge communicationnelle, qui à son tour renforce le sentiment de surcharge informationnelle (cf. David Shenk, *Data Smog : Surviving the Information Glut*, Harper & Collins, 1997, et Caroline Sauvajol-Rialland, *Infobésité, comprendre et maîtriser la déferlante d'informations*, Vuibert, 2013). En effet, alors que 90 % des salariés disent recevoir trop de courriels inutiles et 74 % des managers déclarent souffrir de surinformation dans le contexte d'un sentiment d'urgence généralisé et de compression temporelle, le problème se pose pour les individus mais aussi pour les organisations. Selon IBM, les grandes entreprises seraient incapables d'interpréter jusqu'à 90 % des données qu'elles génèrent, ce que l'on nomme les Dark Data. Certains critiquent toutefois l'idée d'un lien entre Big Data et infobésité, à l'instar de Marie-Anne Chabin selon laquelle « l'infobésité peut exister sans le Big Data ; et le Big Data ne provoque pas nécessairement l'infobésité ». Selon elle, le Big Data est une réalité objective, née des progrès de la technologie, alors que l'infobésité est une notion subjective, le problème se situant chez l'utilisateur et non dans l'existence des données elles-mêmes, cf. son article « Infobésité et Big Data : ne pas confondre ! », *Revue Documentaliste - Sciences de l'information*, Vol 51, 2014.

²⁷ Selon Caroline Sauvajol-Rialland, les TIC induisent plusieurs mutations qui augmentent la surcharge informationnelle, tout d'abord un travail de plus en plus abstrait et interactif, un travail transparent et contrôlé par le management, un travail dont le rythme s'accélère, imposant une gestion du temps de plus en plus difficile, un travail qui passe de plus en plus par la médiation des e-mails sans que les TIC n'aient pour autant diminué le temps de communication en face à face, et, parfois, un travail à distance qui soulève des problématiques de gestion de l'espace. Enfin, plus globalement, elles changent le travail en permanence et impliquent sans cesse des compétences nouvelles.

²⁸ Caroline Sauvajol-Rialland souligne cependant que les e-mails facilitent aussi le travail et que la vitesse de décision est globalement plus rapide grâce à la messagerie, « les effets contraignants du mail sont davantage la résultante d'une utilisation erronée et/ou abusive des fonctionnalités de l'outil que de l'outil lui-même » (cf. son livre *op.cit.*). Plusieurs faits peuvent expliquer cette mauvaise utilisation des e-mails, par exemple, le temps de réponse imposé, ainsi qu'un risque de redondance avec des mails en cascade.

²⁹ Cf. Martin Kersten et Lefteris Sidirourgos « A Database System with Amnesia », *CIDR*, 2017 : <https://www.cidrdb.org/cidr2017/papers/p58-kersten-cidr17.pdf>

³⁰ D'après Chris Anderson, la remise en cause des méthodes scientifiques traditionnelles par les données massives pourrait conduire à la fin de la théorie, c'est-à-dire à la fin de la science telle que nous la connaissons. Cf. Chris Anderson, « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », juin 2008 : <https://www.wired.com/2008/06/pb-theory/>

³¹ Avec cette formule, Alain Desrosières montre qu'en général les données ne sont pas automatiquement accessibles, il faut les récupérer, ce qui par conséquent révèle un choix dans les données récupérées, dans les modèles statistiques utilisés, etc. Des biais peuvent ainsi apparaître dans ces opérations de construction. Cf. Alain Desrosières, *La politique des grands nombres, Histoire de la raison statistique*, La Découverte, 2010.

³² Outre la surexposition des enfants aux écrans qui les rend dépendants aux outils numériques dès leur plus jeune âge et les effets - spécifiques sur eux - de la surcharge informationnelle qu'ils subissent, les impacts cognitifs et physiques existent, surtout quand les plus petits sont privés d'interactions avec leurs parents, ceux-ci étant eux-mêmes trop absorbés par leurs

écrans. Ces problèmes ont été soulevés il y a une dizaine d'années par le psychiatre Serge Tisseron, le philosophe Bernard Stiegler ou l'anthropologue Pascal Plantard, leurs enjeux conduisant l'Académie des sciences, puis l'Académie nationale de médecine et l'Académie des technologies à se saisir du sujet, qui fait l'objet d'enquêtes empiriques approfondies très préoccupantes, à l'instar des travaux de Marie Claude Bossière. Cf. en particulier les références suivantes : Serge Tisseron et Bernard Stiegler, *Faut-il interdire les écrans aux enfants ?*, Mordicus, 2009 ; Pascal Plantard, « Numérique et éducation : encore un coup de tablette magique ? », *Administration et Éducation*, n° 146, 2015, « Le collège et les pratiques numériques des adolescents », *Les Cahiers Pédagogiques*, n° 520, 2015, avec Jonathan Bernard et Sophie Jehel « Tablette, smartphone, console, télé, ordi... Faut-il les interdire aux enfants ? », *revue de l'Inserm*, 2019 ; Jean-François Bach, Olivier Houdé, Pierre Léna et Serge Tisseron, *L'enfant et les écrans, avis de l'Académie des sciences*, Le Pommier, 2013 ; « L'enfant, l'adolescent, la famille et les écrans », *appel commun de l'Académie des sciences, de l'Académie nationale de médecine et de l'Académie des technologies*, 2019, https://www.academie-sciences.fr/pdf/rapport/appel_090419.pdf ; Marie-Claude Bossière, « Le pédopsychiatre et la toxicité de l'omniprésence des écrans », *nouvelle revue de l'enfance et de l'Adolescence*, vol. 2, n° 1, 2020, avec Daniel Marcelli et Anne-Lise Ducanda, « L'exposition précoce et excessive aux écrans (EPEE) : un nouveau syndrome », *revue Devenir*, vol. 32, n° 2, 2020.

³³ Cf. Bruno Patino, *La civilisation du poisson rouge : Petit traité sur le marché de l'attention*, Grasset, 2019, et Nicholas Carr, « *Is Google Making Us Stupid ? What the Internet is doing to our brains* », *The Atlantic*, n° 7, 2008, disponible au lien suivant : <https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/> Bruno Patino rappelle dans son ouvrage que les ingénieurs de Google ont réussi à calculer la durée maximale de l'attention du poisson rouge qui tourne dans son bocal et semble redécouvrir le monde à chaque tour, soit huit secondes et que ces mêmes ingénieurs ont évalué la durée d'attention de la génération des millenials, celle qui a grandi avec les écrans connectés, soit neuf secondes. C'est pourquoi nous sommes, selon lui, devenus des « poissons rouges, enfermés dans le bocal de nos écrans, soumis au manège de nos alertes et de nos messages instantanés, totalement dépendants de technologies dangereuses car favorisant la solitude et la dépression ». Une étude fameuse, citée par Bruno Patino, évalue à 30 minutes le temps maximum d'exposition aux réseaux sociaux et aux écrans d'Internet au-delà duquel apparaît une menace pour la santé mentale, cf. Melissa G. Hunt, Rachel Marx, Courtney Lipson et Jordyn Young, « *No More FOMO: Limiting Social Media Decreases Loneliness and Depression* », *Journal of Social and Clinical Psychology*, December vol. 37, n° 10, 2018.

³⁴ En effet, Diego Mac-Auliffe Cabello écrit dans sa thèse de doctorat : « l'analyse des performances comportementales montre clairement que l'exécution simultanée des deux tâches réduit les performances globales. Nos résultats en iEEG indiquent que les deux tâches nécessitent des ressources cognitives communes – c'est-à-dire l'activité simultanée des mêmes populations neuronales dans les régions précitées - qui ne peuvent être allouées en même temps, de sorte que le système exécutif force les deux tâches à être réalisées l'une après l'autre plutôt qu'en même temps, avec pour effet immédiat une diminution de la vitesse de réponse et une augmentation du nombre d'erreurs ». (cf. *Neural mechanisms underlying external distraction by unexpected environmental stimuli or by a secondary task: an intracranial EEG investigation*, Université de Lyon, 2020). De même, Jean-Philippe Lachaux, en étudiant la réaction du cerveau aux distractions, conclut sur l'idée de plusieurs types d'« attentions » : l'attention volontaire (quand on choisit de donner notre attention), l'attention réflexe (exemple : face à une voiture qui menace de vous renverser) et l'attention basée sur l'émotion, en lien avec notre système de récompense (exemple : lorsqu'on reçoit un sms ou la notification d'un « like » sur Facebook). La croyance commune a tendance à surévaluer la place de l'attention volontaire, en réalité si l'on est bien toujours attentif ou presque, la question c'est plutôt à quoi.

³⁵ Pour Marc Rougier, fondateur et président de la plateforme de curation Scoop.it France, la curation est « la conjonction de trois axes qui consiste à sélectionner, organiser ou éditer puis partager du contenu existant ». Ces nouvelles technologies « permettent de trier l'information surabondante et de se concentrer sur ce qui est pertinent » selon Steven Rosenbaum, ce qui constitue une réponse à l'infobésité. Cependant, cette technologie a aussi ses détracteurs, qui lui reprochent de « favoriser le pillage du Web (par le Web) par l'abus de copier/coller » (cf. Caroline Sauvajol-Riolland, *Infobésité, comprendre et maîtriser la déferlante d'informations*, Vuibert, 2013.). Cependant, ces technologies de curation de contenu sont en pleine explosion, avec de nombreux outils et plateformes qui existent ou sont en développement, aussi bien pour des individus comme Pearltrees, que pour des entreprises comme Knowledge Plaza.

³⁶ Cf. Martin Kersten et Lefteris Sidirourgos « *A Database System with Amnesia* », CIDR, 2017 : <https://www.cidrdb.org/cidr2017/papers/p58-kersten-cidr17.pdf>

³⁷ En effet, selon la célèbre maxime de Rolf Landauer, l'information est de nature physique. Un support quantique de l'information doit donc permettre la transmission et le traitement quantique de cette information, avec des règles du jeu différentes de celles connues classiquement autour de nouvelles méthodes de calcul ou de cryptographie pouvant être plus efficaces pour gérer d'immenses quantités de données. Cette perspective attire l'intérêt de nombreux mathématiciens, physiciens, informaticiens et biologistes, en dépit de barrières technologiques et de compréhension. En 2019, Cédric Villani, alors président de l'OPECST, a présenté une série de notes scientifiques sur ces technologies (notes n° 13 « *Les technologies quantiques, introduction et enjeux* », n° 15 « *Technologies quantiques : l'ordinateur quantique* », n° 16 « *Technologies quantiques : la programmation quantique* », n° 18 « *Technologies quantiques : cryptographies quantiques et post-quantiques* ») disponibles aux liens suivants : <http://www.senat.fr/opecest/notes.html> ou <https://www2.assemblee->

[nationale.fr/15/les-delegations-comite-et-office-parlementaire/office-parlementaire-d-evaluation-des-choix-scientifiques-et-technologiques/\(block\)/48190](https://www.nationale.fr/15/les-delegations-comite-et-office-parlementaire/office-parlementaire-d-evaluation-des-choix-scientifiques-et-technologiques/(block)/48190)

³⁸ Selon une étude publiée en 2022 par NordVPN, seuls 6 % des Français refusent tous les cookies, la grande majorité des internautes les acceptant pour gagner du temps, y compris en sachant majoritairement qu'il faudrait s'en méfier (cf. <https://www.clubic.com/pro/legislation-loi-internet/donnees-personnelles/actualite-424824-vie-privee-seuls-6-des-francais-refusent-tous-les-cookies.html> et <https://siecledigital.fr/2022/05/31/46-pc-internautes-france-acceptent-tous-les-cookies/>). Un sondage de l'IFOP pour la Cnil, un peu moins récent, aboutissait au chiffre d'environ 22 % de Français qui refusent les cookies lorsque les sites web leur demandent leur accord, cf. Romain Bendavid et Delphine Poët, « Les Français et la réglementation en matière de cookies », étude Ifop pour la CNIL, 2019, disponible au lien suivant : https://www.cnil.fr/sites/default/files/atoms/files/les_francais_et_la_reglementation_en_matiere_de_cookies_-_sondage_ifop_pour_la_cnil_-_decembre_2019_.pdf

³⁹ Plusieurs personnes auditionnées, spécialistes des questions numériques, ont ainsi souligné cette perspective d'un recours à la déconnexion lors de leurs auditions et Ioana Manolescu a même rappelé pour sa part que « les enfants de Steve Jobs ont été éduqués sans ordinateur ». L'article L 2242-17 du code du travail, issu de la loi n° 2016-1088 du 8 août 2016 relative au travail, à la modernisation du dialogue social et à la sécurisation des parcours professionnels, a consacré le droit à la déconnexion. La question d'éventuelles obligations en matière de nettoyage des données reste ouverte tout comme reste en suspens le sujet des activités pour lesquelles il serait pertinent de réduire ou d'empêcher la production de données.

Personnes consultées

Institutions

Académie des sciences

- M. Serge Abiteboul, chercheur à l'Institut national de recherche en sciences et technologies du numérique (Inria), ancien professeur de la Chaire d'informatique du Collège de France, membre de l'Académie des sciences et du Collège de l'Autorité de régulation des communications électroniques, des postes et de la distribution de la presse, ancien membre du Conseil national du numérique, ancien président du Conseil scientifique de la Société d'Informatique de France

Académie des technologies

- M. Gérard Roucairol, président honoraire de l'Académie des technologies, président du pôle numérique de l'Académie, ancien directeur scientifique du groupe Bull, ancien professeur à l'École Normale Supérieure ainsi qu'à l'Université d'Orsay
- M. Claude Le Pape-Gardeux, membre de l'Académie des technologies, membre du Conseil scientifique de l'Institut Mines-Télécom, ancien membre du Conseil scientifique du CNRS, ancien cadre dirigeant d'entreprise

Gouvernement

- Mme Laure Lucchesi, directrice d'Etalab, direction interministérielle du numérique, services du Premier ministre
- Mme Ioana Manolescu, directrice de recherche à l'Institut national de recherche en sciences et technologies du numérique (Inria) et à l'École Polytechnique, directrice scientifique du Lab IA, Etalab
- M. Paul-Antoine Chevalier, responsable du pôle exploitation de données du Lab IA, Etalab

Chercheurs

Centre national de la recherche scientifique (CNRS)

- M. Pierre-Éric Mounier-Kuhn, chercheur au CNRS, historien de l'informatique

Universités

- Mme Anastasia Ailamaki, informaticienne, ingénieure et professeure titulaire à l'École Polytechnique Fédérale de Lausanne (EPFL), cofondatrice de la startup Raw Labs, spécialisée dans les logiciels de gestion de données
- M. Mokrane Bouzeghoub, professeur émérite à l'université de Versailles, coordinateur d'actions scientifiques sur la gestion des données au sein de la mission à l'interdisciplinarité du CNRS (MITI)

-
- M. Pascal Griset, professeur d'histoire contemporaine à l'Université Paris-Sorbonne, historien de l'innovation, président du comité pour l'histoire de l'Institut national de la santé et de la recherche médicale (Inserm)
 - M. Francis Jaureguiberry, professeur de sociologie à l'Université de Pau, membre du laboratoire Passages, coordinateur du projet « DEconnexion Volontaire des Technologies de l'Information et de la Communication » (DEVOTIC)
 - M. Michel Kalika, professeur émérite de l'Institut d'administration des entreprises de Lyon (IAE), président du Business Science Institute et de l'Université Jean-Moulin-Lyon-II, ancien directeur à l'École de management (EM) Strasbourg, ancien professeur à l'Université Paris Dauphine
 - Mme Caroline Sauvajol-Riolland, maîtresse de conférences à l'Institut d'études politiques de Paris (Sciences Po), à l'Institut international du commerce et du développement (ICD) et à l'Université catholique de Louvain
 - Mme Valérie Schafer, professeure d'histoire européenne contemporaine à l'Université du Luxembourg, historienne de l'informatique
 - M. Bruno Strasser, professeur à l'Université de Genève et à Yale, directeur du Bioscope, Laboratoire des sciences de la vie et des sciences biomédicales de l'Université de Genève

Entreprises

Orange

- Mme Valérie Peugeot, chargée de la prospective chez Orange, maîtresse de conférences à l'Institut d'études politiques de Paris (Sciences Po), membre du Collège de la Commission nationale de l'informatique et des libertés (CNIL) et ancienne vice-présidente du Conseil national du numérique (CNNum)

Shopify

- Mme Alizé Papp, *data scientist*

Dataiku

- M. Léo Deyfus-Schmidt, vice-président de la recherche
- M. Rémi Meunier, vice-président secteur public
- M. Vivien Tran Thien, directeur intelligence artificielle

Associations

Shift project

- M. Hugues Ferreboeuf, directeur de Virtus Management, polytechnicien, ingénieur du corps des Mines, responsable des questions de sobriété numérique au Shift project