

## Avant-propos

Cet ouvrage fait suite au livre *Du Big Data au Smart Data, au service d'un monde connecté* (ISTE Editions, Londres, 2015), titre auquel nous pourrions ajouter « sans latence », car la temporalité devient le maître mot : il s'agit d'agir plus vite, mieux que ses concurrents dans un environnement digitalisé, où les informations *via* Internet voyagent à la vitesse de la lumière. Jamais plus qu'aujourd'hui dans l'histoire de l'humanité la temporalité n'a représenté un « actif immatériel » avec une aussi forte valeur ajoutée (le *trading* haute fréquence opéré par les banques en est un exemple criant ; je vous invite à lire à ce propos le livre de Michael Lewis, *Flash Boys: A Wall Street Revolt*<sup>1</sup>, W. W. Norton, New York, 2014).

Force est de constater qu'une grande partie de nos décisions et, par capillarité, de nos actions (personnelles ou professionnelles) sont adossées au monde digital (qui mélange informations et algorithmes de traitement de ces informations). Essayez d'imaginer une journée sans votre ordinateur portable, votre smartphone ou autre tablette, vous verrez à quel point nous avons organisé nos vies autour de cette « intelligence digitale » qui nous rend bien des services, augmente notre autonomie, tout en accentuant notre dépendance, voire addiction à ces technologies (quel paradoxe !).

Ce « nouveau » monde structuré autour d'Internet impose aux entreprises de devoir décider et agir dans un environnement très concurrentiel, sur la base de

---

1. Cet ouvrage de Michael Lewis nous présente les tenants et les aboutissants du *trading* haute fréquence (HFT : *High Frequency Trading*) : son histoire, les moyens mis en œuvre, les enjeux...

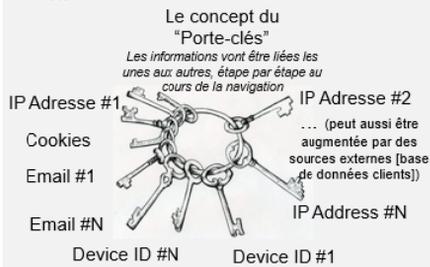
données complexes et dans une temporalité qui se chiffre en millisecondes (voire moins). C'est un monde où l'« expérience client » devient le maître mot ; notre exigence en tant que consommateurs (de tous types de biens, services ou contenus : messages, produits, offres, informations...) ne fait que croître. Nous exigeons d'être « traités » de façon pertinente, et ce même si nous naviguons dans ce monde digital de façon « anonyme » (sans avoir au préalable utilisé un compte personnel d'authentification), ce qui implique que d'autres mécanismes sont en œuvre pour permettre cette « traçabilité ». Qui a dit que « l'habit ne fait pas le moine » ? Je crains que dans ce monde digitalisé, les habits d'Internet que sont les traces (navigation, cookies, adresse IP, ce que nous téléchargeons...), que nous laissons de façon volontaire ou non, vont permettre la constitution d'une identité numérique à notre insu, que nous ne maîtrisons donc pas ou peu ! Ces informations seront rapprochées les unes des autres, liées entre elles au fur et à mesure de leur génération, selon un principe de « porte-clés » (voir figure AP.1, « Résolution d'identité »). Elles seront ensuite exploitées par les solutions de ciblage, segmentation, moteur de recommandations, etc., mises en place depuis une dizaine d'années, architecturées sur la base d'agents logiciels adossés à des moteurs de règles (moteur de recommandations), ceci afin de répondre à cette attente de « pertinence » du contact.

« Un client n'appartient pas à une entreprise, mais seulement le temps qu'il veut bien lui consacrer. » Durant ce fameux temps, qui devient le « graal » pour les entreprises qui dépensent des trésors d'imagination (mais aussi financiers) pour les attirer vers leurs canaux de contacts (Internet, centre d'appels, boutiques...), elles doivent être les plus « pertinentes » possible. Ces solutions en place (moteur de règles/recommandations) interagissent assez peu avec leur environnement (modèles prédéfinis à partir d'un nombre limité de variables descriptives de la situation), elles ne sont que peu auto-apprenantes (mise à jour des modèles à l'issue de traitements analytiques souvent très lourds) ; le tout ayant pour résultat que les mêmes causes (identifiées par quelques variables) provoquent les mêmes effets. Ces solutions ne tiennent pas ou peu compte des variations de contexte en temps réel (comment l'internaute est arrivé sur cette page, quel contenu lui a été proposé lors de sa dernière navigation, quelle est la nature de sa recherche...), du résultat des décisions et actions précédentes, et finalement ne permettent pas ou peu d'exploiter l'ensemble des données contextuelles (comportement de navigation, ce qui a déjà été proposé en matière de contenu, les actions résultantes, la géo-localisation...).

### Comment puis-je augmenter ma connaissance client ?

En utilisant les données ouvertes comme :

- les cookies ;
- le device ID (qui en dit beaucoup sur les appareils matériels) ;
- l'adresse IP (très utilisée pour géolocaliser) ;
- ...



> 98 % des activités de navigation sur Internet sont opérées de façon anonyme

Ces données pourront ensuite être croisées avec une base de données clients, ce qui permettra d'augmenter la résolution de l'identité

**Figure AP.1.** Résolution d'identité

Ce besoin d'agir et de réagir en temps réel dans un environnement complexe est identifié depuis des années ; l'avènement du Big Data et des objets connectés n'a fait qu'augmenter la complexité de traitement de ces informations, les solutions et organisations (statisticiens, analystes décisionnels...) sont débordées par ce flux continu de données (Internet ne dort jamais). Pas ou peu de solutions sont proposées au travers des processus et outils analytiques historiques en possession des entreprises, qui sont trop lourds, trop complexes à faire évoluer et nécessitent des ressources pour leur mise en œuvre. Or ces ressources sont de plus en plus rares (sans doute un des futurs gros problèmes de ces prochaines années dans ce domaine : le manque d'experts en Business Intelligence et de statisticiens va se faire criant !). Les comportements d'achat et de consommation sont en évolution permanente (les plateformes collaboratives comme Uber, Airbnb, entre autres, ont « inventé » ce nouveau modèle économique), ce qui va *in fine* créer de nouveaux risques (pour ceux qui ne sauront pas s'adapter à ce monde en mouvement permanent), et des opportunités (pour qui saura exploiter ce nouvel Eldorado qu'est le Big Data).

L'intelligence artificielle (IA) est une des réponses (sans doute la plus prometteuse) à l'exploitation massive, auto-apprenante, autonome de ces Big

Data. Plus précisément, le Deep Learning (apprentissage profond, en couches), qui voit le jour dans les années 1980 avec l'avènement des réseaux de neurones, devient aujourd'hui la « clé de voûte » de cette nouvelle génération de solutions. Les progrès de la technologie, associés aux flux de données digitalisées, ont ouvert de nouveaux horizons pour cette discipline, ce qui ne trompe pas les grands acteurs technologiques de la Business Intelligence qui s'y sont engouffrés, comme la suite logique du Big Data. Les domaines d'application possibles de l'intelligence artificielle sont nombreux et variés, tels que : la robotique (les voitures connectées et autonomes), la domotique (la maison connectée), la santé (l'aide au diagnostic médical), la sécurité, les assistants personnels (qui deviendront des outils indispensables de notre quotidien), les systèmes experts, la reconnaissance d'images, de sons, de visages (et pourquoi pas l'analyse des émotions), le traitement du langage naturel, etc., mais aussi la gestion de la relation client (dans un mode qui pourrait anticiper nos attentes, voire les dépasser). Tous ces systèmes seront auto-apprenants, leurs connaissances ne feraient que croître avec le temps, ils pourraient échanger entre eux des savoirs.

Certains, comme Bill Gates, le fondateur de Microsoft, le serial entrepreneur Elon Musk ou Steve Wozniak, le cofondateur d'Apple, ou encore le scientifique Stephen Hawking, se sont émus de ce que l'intelligence artificielle pourrait changer dans notre société : le risque que l'humanité soit un jour contrôlée par des machines (ce qui ressemble fort au scénario de « Matrix »). Le but de ce livre n'est pas philosophique ou éthique, quoique ce débat soit intéressant (et nécessaire), les questions qu'il pose sont pertinentes. Ce que l'on peut constater dans l'histoire de l'humanité est que le développement technologique a toujours accompagné son évolution, et ce pour le « meilleur » et pour le « pire ». Je consacrerai donc mon ouvrage au positionnement (actuel et dans un futur proche) de l'intelligence artificielle dans le monde de la Business Intelligence, voire même en quoi l'intelligence artificielle pourrait remplacer (ou supplanter) la Business Intelligence que nous connaissons aujourd'hui ; ou comment les entreprises commencent à adopter des solutions architecturées autour de plateformes d'intelligence artificielle, et en quoi ces solutions vont permettre de « créer » des ponts entre la Business Intelligence « traditionnelle » et celle liée au Big Data.

Il existe deux types d'intelligence artificielle : l'intelligence artificielle forte et l'intelligence artificielle faible.

L'intelligence artificielle forte fait référence à une machine capable de produire un comportement intelligent<sup>2</sup> et aussi d'éprouver une impression d'une réelle conscience de soi, de vrais sentiments. Dans ce monde, la machine serait capable de comprendre ce qu'elle fait (et donc les conséquences de ses actions). Le fait est que l'intelligence est issue de la biologie du cerveau, sur la base d'un processus d'apprentissage et de raisonnement (donc matériel et basé sur une logique « algorithmique »). Sur cette base, les scientifiques ne voient pas *a priori* de limites à pouvoir réaliser un jour une intelligence avec une certaine conscience, qui pourrait avoir des émotions, à partir d'une machine (ou d'un élément matériel équivalent). Ce sujet suscite de nombreux débats. S'il n'y a pas à l'heure actuelle d'ordinateurs ou de robots aussi intelligents que l'homme, ce n'est pas dû à un problème de matériel, mais de conception. D'après cela, nous pouvons considérer qu'il n'y a pas de limite fonctionnelle. Pour pouvoir déterminer si une machine peut être considérée comme ayant une intelligence artificielle forte, il faut que celle-ci ait réussi à passer le test de Turing<sup>3</sup>.

L'intelligence artificielle faible consiste à mettre en œuvre des systèmes de plus en plus autonomes, auto-apprenants, avec des algorithmes capables de résoudre des problèmes d'une certaine classe. Mais, dans ce cas, la machine fait comme si elle était intelligente, il s'agit plus d'une « simulation » d'intelligence humaine sur la base d'un apprentissage (supervisé ou non). On peut apprendre à des machines à reconnaître des sons, des images, sur la base de données représentant le type d'apprentissage attendu (par exemple reconnaître une voiture dans un lot d'images) : c'est l'apprentissage supervisé. Ou alors, laisser la machine découvrir par elle-même les éléments qu'elle analyse, jusqu'à les nommer. Pour reprendre l'exemple de l'image

---

2. Terme défini dans le chapitre 1 : « Qu'est-ce que l'intelligence ? ».

3. « Le test a été inspiré d'un jeu d'imitation dans lequel un homme et une femme vont dans des pièces séparées et les invités tentent de discuter avec les deux protagonistes en écrivant des questions et en lisant les réponses qui leur sont renvoyées. Dans ce jeu, l'homme et la femme essaient de convaincre les invités qu'ils sont tous deux des femmes. À l'origine, Alan Turing a imaginé ce test pour répondre à sa question existentielle : "Une machine peut-elle penser ?", en donnant une interprétation plus concrète de sa question. Une idée intéressante de sa proposition de test est que les réponses doivent être données dans des intervalles de temps définis. Il imagine que cela est nécessaire pour que l'observateur ne puisse pas établir une conclusion qui soit fondée sur le fait qu'un ordinateur puisse répondre plus rapidement qu'un humain, surtout sur des questions de mathématiques. » (source : Wikipédia).

d'une voiture, la machine analyse les images qui lui sont proposées, et de proche en proche (grâce au Deep Learning, ou apprentissage profond *via* des réseaux de neurones) va d'elle-même associer le concept de « voiture » aux images analysées ; lorsqu'une des images associées aura été labélisée comme une voiture, elle saura le « verbaliser » : c'est l'apprentissage non supervisé.

# Introduction

## I.1. L'ère du « tout » numérique a sonné, pouvons-nous y échapper ?

Rien n'est moins sûr. Lorsque l'on voit la vitesse exponentielle à laquelle nous nous approprions les technologies numériques, c'est sans doute sans équivalent dans l'histoire de l'humanité (chaque nouvelle génération accélère ce mouvement). Son impact sur nos sociétés est/sera au moins aussi fort que l'invention de l'imprimerie par Gutenberg en 1450 (qui a permis au livre d'exister, reliant ainsi les connaissances, les cultures, les idées, les savoirs...). Aurions-nous évolué aussi rapidement sans l'imprimerie ? On ne peut que constater cette numérisation frénétique de notre monde, où le réel et le virtuel se confondent, créant un modèle d'« assimilation digitale » plus ou moins consciente dont Internet est le support. À ses débuts (*grosso modo* avant l'an 2000 et l'avènement des blogs<sup>1</sup>, des réseaux sociaux...), Internet était un médium aux mains des entreprises et des acteurs institutionnels, qui étaient (de fait) les « maîtres » du contenu. Le rôle de l'internaute se résumait à consommer de façon assez passive ces contenus, aucune ou peu d'interactions étaient possibles entre l'internaute et Internet.

Ce modèle de communication est aussi connu sous le nom de « Web 1.0 », en opposition au Web 2.0, qui voit son essor avec la mise en place au début des années 2000 des blogs (contraction de *web logs*) qui ont permis un

---

1. Un blog, anglicisme de site web – ou une partie d'un site web –, est utilisé pour la publication périodique et régulière d'articles, généralement succincts, et rendant compte d'une actualité autour d'un sujet donné ou d'une profession.

nouveau mode d'expression et de partage sur le Web. Cette évolution se caractérise principalement par le rôle contributif des internautes. C'est le courant *open source*<sup>2</sup> (logiciel libre) qui est à l'origine de cette « révolution ». L'*open source* est une notion créée en 1982 par Richard Stallman<sup>3</sup>, qui stipule que les utilisateurs d'un logiciel sont libres de l'utiliser, de le faire évoluer et de le publier (voire de le distribuer) de façon gratuite ou non (toute entreprise ou particulier est donc en droit de le commercialiser). Les blogs ont été une des premières étapes (sans doute la plus marquante) de l'émergence du Web 2.0, rapidement suivies par d'autres, comme des outils permettant de publier des contenus numériques à travers les wikis, mais aussi le partage de photos, de vidéos et finalement les réseaux sociaux tels que Facebook, Twitter... (voir figure I.1) qui ont définitivement et de façon irréversible changé notre approche au monde et son réseau planétaire qu'est Internet.

Ces évolutions ont induit par capillarité des changements de comportement (de consommation de biens ou de services) grâce à cette énorme place de marché qu'est devenu Internet, où « tout » (et son contraire) est accessible en quelques clics. Les entreprises ont dû intégrer ces changements dans leur approche de la relation client (plus de volatilité du client, moins de fidélité à une marque, plus de facilité de comparaison... la concurrence fait rage sur le Web). La relation client est en pleine mutation, le client n'est plus considéré comme « appartenant » (au sens marketing du terme) à une entreprise, mais seul le temps que l'internaute est prêt à lui consacrer (*via* les différents canaux de contact) est finalement exploitable par l'entreprise. Ce temps est donc devenu précieux pour les entreprises, qui doivent être en mesure de l'exploiter dans la temporalité de l'internaute, et non plus dans celle de l'entreprise (au travers de politiques de contact « traditionnelles ») afin d'optimiser cette relation.

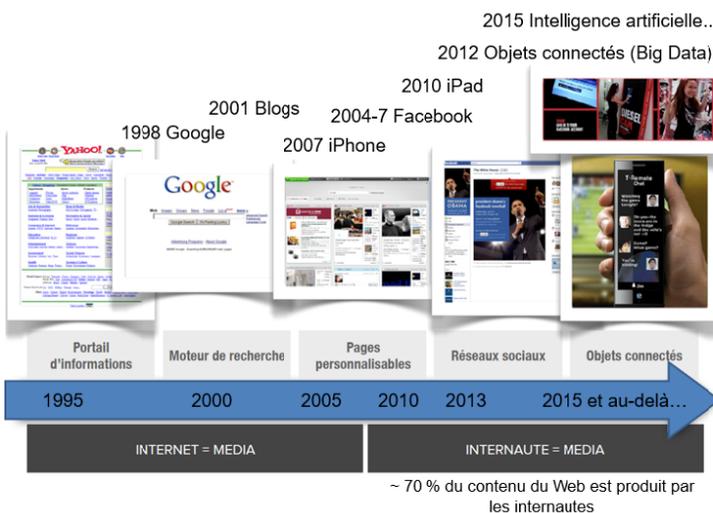
Les entreprises l'ont bien compris, et celles qui savent/sauront s'adapter le plus vite pour occuper cet « écosystème » (nous devrions dire « cyberspace ») auront un avenir plus assuré que les autres (bienvenue au

---

2. *Open source*, ou « code source ouvert », s'applique aux logiciels dont la licence respecte des critères précisément établis par l'*Open Source Initiative*, c'est-à-dire les possibilités de libre redistribution, d'accès au code source et de création de travaux dérivés. Mis à la disposition du grand public, ce code source est généralement le résultat d'une collaboration entre programmeurs.

3. « Richard Matthew Stallman (né à Manhattan, le 16 mars 1953), connu aussi sous les initiales rms (en minuscules), est un programmeur et militant du logiciel libre. Initiateur du mouvement du logiciel libre. » (source : Wikipédia).

« darwinisme » numérique, où le *time to market* et la capacité à agir vite et bien deviennent les maîtres mots). L'avènement du smartphone (assez récent – l'iPhone est commercialisé en 2007 – même s'il nous semble avoir toujours été là, tant son usage s'est révélé pertinent), qui permet l'accès à Internet en tout lieu et à tout moment, ainsi que celui d'autres supports comme les tablettes n'ont fait qu'amplifier ce mouvement et contribuer à cette « assimilation numérique » qui, depuis, ne cesse de s'étendre. Les prochaines étapes de cette évolution seront sans aucun doute liées aux objets connectés et à tous les services qui vont leur être rattachés, dans des domaines comme la santé, le transport, la domotique, etc., mais aussi le monde de la réalité augmentée, *via* des métadonnées (textes, images, sons...) qui auront pour objectif d'« enrichir » en temps réel notre vision du monde. Un monde où le réel et le virtuel vont se confondre au point de ne faire plus qu'un (un cyberspace). Il y a fort à parier (et la tendance est déjà en cours) que nos smartphones vont évoluer vers des *smart* assistants, qui embarqueront des logiciels/algorithmes d'intelligence artificielle, qui apprendront de façon continue sur nous (nos mouvements, actions, préférences, achats, réseaux sociaux, etc.) pour nous aider à mieux gérer notre temps, nos actions... par anticipation (le mot est lâché, et sera sans doute très structurant). Ceci va accentuer le besoin de sécurisation de nos données personnelles, afin de ne pas passer de Big Data à Big Brother !



**Figure I.1.** *L'assimilation numérique*

## I.2. Comment sommes-nous en train de passer d'une « inconscience » à une « conscience » numérique ?

L'inconscience numérique pourrait se traduire par : « J'agis dans le monde digital sans me préoccuper de l'exploitation de mes données par des tiers. » Cette situation tient au fait qu'avant la numérisation de notre monde (*grosso modo*, avant Internet) nous n'étions pas ou peu confrontés à cette problématique. Nos données personnelles (numérisées) étaient disponibles principalement pour les administrations, les institutions (comme les banques ou assurances...), voire pour certaines entreprises à qui nous avons donné notre approbation (le fameux *optin*) pour une exploitation commerciale, le tout étant protégé par la loi, dont le bras armé était la CNIL<sup>4</sup> (Commission nationale de l'informatique et des libertés). Bref, nous nous sentions en « sécurité », et le Web 1.0 n'avait pas fondamentalement changé cette impression (les données, traces que nous laissions sur Internet, n'étaient pas ou que peu exploitées, faute de solutions [technologiques], et/ou à des coûts trop importants au regard de la valeur qui en était attendue).

Mais voilà, les choses n'en sont pas restées là ! La numérisation de notre monde s'est accélérée au début des années 2000, puis l'avènement des réseaux sociaux a mis à mal cette « croyance » que nos données personnelles nous appartenaient et que nous étions protégés contre leur utilisation (intéressez-vous aux conditions d'utilisation des applications que vous téléchargez tous les jours, vous pourriez être surpris !). La collecte de données personnelles est très souvent liée à une offre de services non payante (si vous ne payez pas pour un produit ou un service, c'est que finalement le produit c'est « vous »). Le temps de l'angélisme est maintenant révolu, nous savons que nos données personnelles font l'objet d'analyses de tous types, et les technologies liées au Big Data (principalement Hadoop) ont rendu possible l'analyse massive de ces données. Quelques scandales d'État (comme le projet PRISM<sup>5</sup> qui a défrayé la chronique et agité la « webosphère ») n'ont fait que confirmer cette crainte !

---

4. La Commission nationale de l'informatique et des libertés (CNIL) est chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'Homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

5. PRISM, également appelé US-984XN1, est un programme américain de surveillance électronique par la collecte de renseignements à partir d'Internet et d'autres fournisseurs de

La conscience numérique serait donc : « Je comprends que mes actions dans le monde digital, dans lequel j'opère, peuvent faire l'objet d'analyses au travers de mes données. » Cette prise de conscience ne doit pas de transformer en défiance vis-à-vis d'Internet et de ces grands acteurs que sont les GAFA (Google, Amazon, Facebook et Apple, on pourrait y ajouter Microsoft), ou autres, mais nous permettre d'agir sur Internet en toute connaissance de cause (« Le jeu en vaut-il la chandelle ? »), de comprendre et de savoir (lorsqu'Internet est « gratuit », ce qu'il faut comprendre, c'est que le « produit » devient l'internaute) que nous sommes devenus la cible de toutes les attentions. La vraie question maintenant serait de savoir si nous pouvons y échapper. Sans doute pas ! Mais le renforcement des législations (avril 2016) concernant le stockage et l'exploitation des données personnelles est à l'ordre du jour (l'Europe a voté le RGPD<sup>6</sup> [pour Règlement général sur la protection des données ou, en anglais, *General Data Protection Regulation*, GDPR<sup>7</sup>]) et a pour objectif de redonner aux citoyens le contrôle de leurs données personnelles.

S'agissant des garanties à mettre en place, l'article 32 énumère certaines mesures susceptibles d'être utilisées par les sociétés, telles que :

- le recours à la pseudonymisation ;
- le chiffrement des données ;
- l'adoption de moyens permettant de garantir la confidentialité ;
- l'intégrité, la disponibilité et la résilience des systèmes ;
- l'adoption de moyens permettant de rétablir la disponibilité et l'accès aux données personnelles en cas d'incident technique ou physique ;
- la vérification régulière des mesures.

---

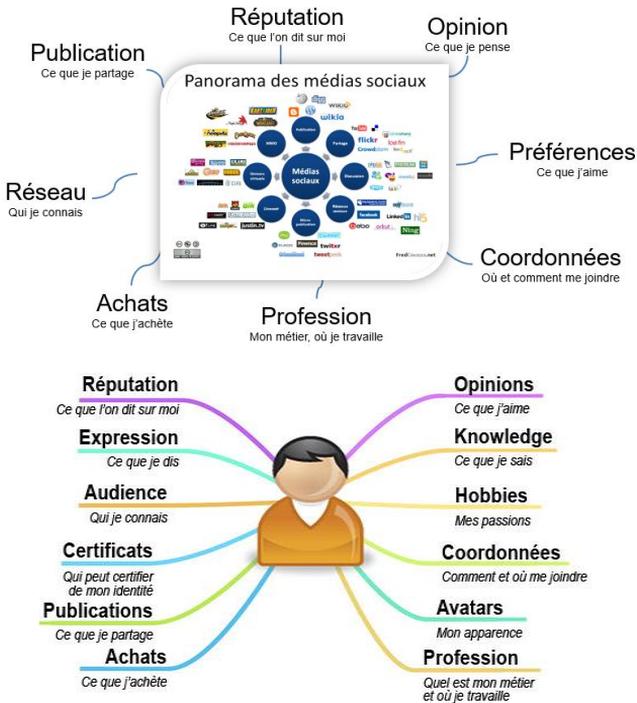
services électroniques. Ce programme classé, relevant de la National Security Agency (NSA), prévoit le ciblage de personnes vivant hors des États-Unis.

6. Disponible à l'adresse : [http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf).

7. « Le Règlement général sur la protection des données (GDPR) constitue le nouveau texte de référence, au niveau européen, en matière de protection des données à caractère personnel. Il renforce et unifie la protection des données pour les individus au sein de l'Union européenne. [...] Le règlement s'appliquera aux entreprises établies en dehors de l'Union européenne qui traitent les données relatives aux activités des organisations de l'UE. Les sociétés non européennes seront également soumises au règlement si elles ciblent les résidents de l'UE par le profilage ou proposent des biens et services à des résidents européens. » (source : Wikipédia).

### I.3. Les traces que nous laissons sur Internet (volontairement ou non) vont constituer notre identité numérique

L'identité numérique doit être comprise comme une identité virtuelle qui rassemble l'ensemble des informations (donc des données) que l'on trouve sur Internet nous concernant (voir figure I.2). Comme dans le monde réel, cette identité est en évolution permanente, elle représente différents éléments de notre personnalité et de la façon dont nous sommes perçus. Elle se décompose en deux types : l'identité déclarative, il s'agit des données que nous (ou un tiers) saisissons volontairement (réseaux sociaux, blogs, etc.), et l'identité d'activité (téléchargement, navigation, cookies, etc.). Chaque nouvelle connexion, navigation sur Internet va enrichir ce patrimoine informationnel nous concernant, dont nous ne sommes pas les dépositaires ! Et c'est là où le bât blesse : nous avons, de fait, « délégué » à un tiers (que sont les moteurs de recherche) la gestion de cette identité qui, *in fine*, deviendra notre e-réputation.



**Figure I.2.** Les traces que nous laissons sur Internet (volontairement ou non) vont constituer notre identité numérique

Un des usages liés à l'identité numérique est le *name googling* (recherche sur Internet à partir d'un nom), largement utilisé (et pas que par de futurs employeurs) pour savoir « qui est qui » et pour se faire une première impression (et, comme chacun sait, « nous n'avons jamais une seconde chance de faire une bonne première impression »). C'est une méthode qui, en quelques clics, permet de vérifier le profil d'une personne (son CV, son réseau professionnel, ce que l'on dit de lui, etc.), d'évaluer son influence sur le Web (présence sur certains forums...), bref, de se faire une idée plus « précise » sur cette personne.

#### **I.4. La numérisation de notre monde se poursuit et les objets connectés en sont la prochaine étape (l'Internet des objets)**

Alors qu'Internet ne s'étend pas au-delà du monde virtuel, l'Internet des objets va permettre la connexion entre le monde réel et le monde virtuel, *via* des échanges d'informations et de données provenant de « capteurs » présents dans le monde réel. L'Internet des objets est considéré comme la prochaine évolution du Web, sous le nom « Web 3.0 »<sup>8</sup>, le Web des objets, là

---

8. « L'Internet des objets (ou "IdO", en anglais "*Internet of Things*" ou "IoT") représente l'extension d'Internet à des choses et à des lieux du monde physique. Alors qu'Internet ne se prolonge habituellement pas au-delà du monde électronique, l'Internet des objets connectés représente les échanges d'informations et de données provenant de dispositifs présents dans le monde réel vers le réseau Internet. Considéré comme la troisième évolution d'Internet, baptisé "Web 3.0" (parfois perçu comme la généralisation du Web des objets, mais aussi comme celle du Web sémantique) qui fait suite à l'ère du Web social, l'Internet des objets revêt un caractère universel pour désigner des objets connectés aux usages variés, dans le domaine de la e-santé, de la domotique ou du *quantified self*. L'Internet des objets est en partie responsable d'un accroissement exponentiel du volume de données généré sur le réseau, à l'origine du Big Data (ou "mégadonnées" en français). Selon une équipe de l'ETH de Zurich, du fait des smartphones puis du nombre croissant d'objets connectés, en dix ans (2015-2025), 150 milliards d'objets devraient se connecter entre eux, avec Internet et avec plusieurs milliards de personnes. L'information issue de ce Big Data devra de plus en plus être filtrée par des algorithmes complexes, ce qui fait craindre une moindre protection des données personnelles, une information des personnes et de la société de moins en moins autodéterminée, notamment en cas d'appropriation exclusive de filtres numériques par des entités (gouvernementales ou privées) qui pourraient alors manipuler les décisions. L'ETH plaide donc pour des systèmes d'information ouverts et transparents, fiables et contrôlés par l'utilisateur. » (source : Wikipédia).

où le Web 2.0<sup>9</sup> était adossé à une approche sociale (blogs, réseaux sociaux, etc.). Les objets connectés vont accentuer (de façon exponentielle) le volume des données échangées, disponibles sur Internet (certaines études parlent d'un facteur 40 d'ici à 2020) (voir figure I.3). Le Big Data et l'intelligence artificielle vont se « nourrir », ils vont permettre la mise en place de nouveaux services dans des domaines d'application aussi divers que la domotique, la santé, le transport, etc.



Figure I.3. Nombre d'objets connectés par personne d'ici 2020

9. « L'expression "Web 2.0" désigne l'ensemble des techniques, des fonctionnalités et des usages qui ont suivi la forme originelle du Web, www ou *World Wide Web*, caractérisée par plus de simplicité et d'interactivité (sociabilité). Elle concerne en particulier les interfaces et les échanges permettant aux internautes ayant peu de connaissances techniques de s'approprier de nouvelles fonctionnalités du Web. Les internautes peuvent d'une part contribuer à l'échange d'informations et interagir (partager, échanger, etc.) de façon simple, à la fois au niveau du contenu et de la structure des pages, et d'autre part entre eux, créant notamment le Web social. L'internaute devient, grâce aux outils mis à sa disposition, une personne active sur la toile. Le Web 2.0 est donc l'évolution du Web vers l'interactivité, à travers une complexification interne de la technologie, mais permettant plus de simplicité d'utilisation, les connaissances techniques et informatiques n'étant pas indispensables pour les utilisateurs. L'expression "Web 2.0", utilisée en 2003 par Dale Dougherty, un cadre de la société O'Reilly Media, et diffusée par Tim O'Reilly en 2004, puis consolidée en 2005 avec l'exposé de position "*What Is Web 2.0*", s'est imposée à partir de 2007. » (source : Wikipédia).

Ces informations, ces traces que nous laissons volontairement ou pas sur Internet sont une partie importante de ce qui est maintenant appelé le « Big Data »<sup>10</sup>. Elles en disent beaucoup sur nous (et plus encore avec l'avènement des objets connectés), elles font/feront l'objet d'analyses de plus en plus précises, elles sont/seront la matière première d'une nouvelle forme d'intelligence numérique (l'intelligence artificielle<sup>11</sup>).

L'objectif de ce livre est de prendre un peu de hauteur par rapport à ce phénomène, voir en quoi il va modifier notre approche analytique (principalement au sein de l'entreprise, autour de la connaissance client), pour la rendre plus dynamique, plus réactive, plus apprenante, avec comme conséquence moins d'« humain » pour plus de « machine ». Ce mouvement est déjà initié : nous sommes passés lors de la dernière décennie au CRM<sup>12</sup> (*Customer Relationship Management*), où il s'agissait d'avoir une vision 360° du client, avec une interconnexion des canaux web et centre d'appel. C'était un temps où la référence client était le foyer (principalement identifié au travers de l'adresse postale et de sa constitution en matière de membres [adultes, enfants, seniors...]). Les évolutions technologiques, telles que le smartphone et les réseaux sociaux, ont changé la donne : on ne contacte plus un lieu (le foyer), mais une personne (en mouvement), on passe d'une approche 360°, à une approche que nous pourrions appeler « 37,2° » (la

---

10. Le Big Data désigne des ensembles de données qui deviennent tellement volumineux, avec des formats si divers et une vélocité si importante, qu'il devient impossible de les traiter avec des outils classiques de gestion de base de données ou de gestion de l'information.

11. L'intelligence artificielle (IA) peut se définir comme la capacité qu'aurait une machine à exécuter des fonctions normalement associées à l'intelligence humaine : compréhension, raisonnement, dialogue, adaptation, apprentissage, etc.

12. « La gestion de la relation client (GRC), ou gestion des relations avec les clients, en anglais *Customer Relationship Management* (CRM), est l'ensemble des outils et techniques destinés à capter, traiter, analyser les informations relatives aux clients et aux prospects, dans le but de les fidéliser en leur offrant le meilleur service. En ce qui concerne les applications informatiques, il s'agit des progiciels qui permettent de traiter directement avec le client, que ce soit sur le plan de la vente, du marketing ou du service, et que l'on regroupe souvent sous le terme de « *front-office* », ceci par opposition aux outils de *back-office* que sont les progiciels de gestion intégrés. » (source : Wikipédia).

température moyenne du corps humain). La « personnalisation » est née, elle va tirer un nouveau modèle de relation client, qui sera basé sur la capture et l'analyse de toutes les formes d'interactions avec le client : le CXM<sup>13</sup> (*Customer Experience Management*), avec comme finalité d'en optimiser l'expérience. Le CXM dépasse le CRM par bien des aspects, ces points sont développés dans le chapitre 4 (« Les usages de l'intelligence artificielle »).

---

13. Au début des années 1990, le CRM (*Customer Relationship Management*) s'est concentré sur comment capter, stocker et traiter les données clients. Maintenant, le CXM est une approche qui intègre l'ensemble des processus et organisations afin d'offrir un service individuel en plaçant les attentes du client au cœur des préoccupations de l'entreprise. Il devient alors impératif d'associer toutes les équipes de l'entreprise et pas seulement celles dédiées à la relation client.