

Tout transformer en données pour tout transformer par les données

Julien Lévy - Directeur du Centre IDEA (Innovation & entrepreneuriat) d'HEC - Directeur de la Chaire AXA Stratégie Digitale et Big Data.

On comprend aujourd'hui peu de choses à la révolution numérique si on ne voit pas qu'elle s'inscrit très précisément dans le prolongement de la science moderne.

Le projet scientifique de quantifier le monde

Depuis la Renaissance, la vocation de la science est de déchiffrer le monde... en le chiffrant. Le propre de la démarche scientifique est de transformer l'univers de notre perception – tout ce que nous voyons, nous entendons, nous touchons – en grandeurs qu'on peut mesurer et manipuler. Ces grandeurs sont le temps, la longueur, la masse, la température, la quantité de matière, etc. et ce que les physiciens appellent les « grandeurs dérivées » : superficie, volume, vitesse, pression, etc.

Comme l'a écrit Galilée en 1623 dans le *Saggiatore* :

« L'Univers [est un livre] écrit dans la langue mathématique et ses caractères sont des triangles, des cercles, et autres figures géométriques sans les moyens desquelles il est humainement impossible d'en comprendre un mot. Sans eux, nous errons vainement dans un univers obscur. »

C'est en transformant l'univers de notre perception en données objectives que la science arrive à en rendre compte et acquiert une puissance de prédiction sans égale.

Descartes, quatorze ans après la publication de l'ouvrage de Galilée, en tire les conclusions pratiques dans le *Discours de la méthode* (6ème partie) :

« Sitôt que j'ai eu acquis quelques notions générales touchant la physique (...) elles m'ont fait voir (...) qu'on en peut trouver une pratique, par laquelle, connoissant la force et les actions du feu, de l'eau, de l'air, des astres, des cieux, et de tous les autres corps qui nous environnent, aussi distinctement que nous connoissons les divers métiers de nos artisans, nous les pourrions employer en même façon à tous les usages auxquels ils sont propres, et ainsi nous rendre comme *maîtres et possesseurs de la nature*. »

Le XVIIème siècle européen a amorcé le grand tournant de la *quantification du monde*. Si rien dans ce projet n'est de nature à nous surprendre, tellement il nous semble naturel, il a pourtant un caractère inouï : jamais auparavant on n'avait tenté, ni même conçu rien de semblable. La révolution numérique, qui est une révolution technologique, s'inscrit précisément dans ce mouvement de quantification du monde.

La révolution numérique donne une nouvelle ampleur à la quantification du monde

Si l'informatique s'est toujours préoccupée de traiter des données, trois facteurs en ont profondément changé la nature : l'accroissement vertigineux des capacités de traitement (la fameuse « loi » de Moore), la chute considérable des coûts de traitement et la miniaturisation des équipements. C'est grâce à la conjonction de ces trois facteurs que l'informatique est devenue ubiquitaire : elle pénètre chaque aspect de notre vie personnelle, sociale et économique.

Internet, en permettant de connecter les ordinateurs les uns aux autres et en rendant accessibles les données diffusées dans le réseau, a insufflé une nouvelle ambition à cette dynamique. Il suffit de lire le *mission statement* de Google pour en prendre la mesure : "La mission de Google est d'organiser l'information mondiale et de la rendre universellement accessible et

utilisable." Autrement dit : faire en sorte que toutes les données à travers le monde soient accessibles en ligne, et les organiser d'une façon qui les rende utiles.

Cette vocation remarquablement ambitieuse – qui permet de comprendre la stratégie de Google – semble être presque modeste au vu des évolutions présentes. Aujourd'hui, il ne s'agit pas seulement d'*organiser* les données, mais d'en *produire sur tout, tout le temps, partout, et en flux continu*.

La numérisation du réel est un phénomène à la fois massif et en pleine accélération : on numérise les actualités (toute la presse et ses archives sont en ligne), le savoir humain et les arts (livres, peinture, musique, photographie, cinéma...), les relations sociales (réseaux sociaux, contenu généré par les utilisateurs...), notre environnement géographique (cartographie, géolocalisation...), physique (reconnaissance visuelle, sondes et senseurs), l'ensemble des comportements en ligne, notre corps et le fonctionnement de notre cerveau par le biais d'objets portables (*wearable*), la logistique, les différentes opérations de production et de distribution grâce à l'Internet des objets industriels... La masse des données produites en flux continu et intégrées dans un réseau est sans commune mesure avec le passé : une étude souvent citée d'IBM indique que 90% des données mondiales ont été produites ces deux dernières années et le cabinet IDC estime que la production de données est exponentielle et double tous les deux ans !

Le mot d'ordre de la révolution numérique

Les ordinateurs, les objets connectés, Internet sont un terrain d'innovation formidable, car la science ne progresse qu'avec la technologie, mais ils ne sont pourtant que le support de la révolution numérique. Le *mot d'ordre* de la révolution numérique peut être résumé ainsi : *Il faut tout transformer en données, pour traiter toutes les données, afin de tout transformer par les données*.

La production et la collecte de données ont donné lieu et donnent encore lieu à nombreuses innovations qui en accélèrent le mouvement. Mais les entreprises sont encore très loin du monde idéal de la gouvernance des données – l'univers merveilleux dans lequel l'entreprise exploite les promesses de l'exploitation massive des données. La plupart vous diront qu'elles ont déjà bien du mal à recenser et à regrouper les données qu'elles produisent ou auxquelles elles ont accès. Comme aiment à le dire les *data scientists* : « on passe 80% de notre temps à collecter les données, et les 20% restant à nous en plaindre. » La collecte et l'organisation des données sont aujourd'hui un défi pour les entreprises, mais c'est un défi organisationnel. L'enjeu technologique majeur s'est déplacé : il n'est plus dans la production et la collecte des données, mais dans leur *exploitation*, qui est le plus souvent frustré et fragmentaire.

Face à l'explosion des données produites, la capacité à n'en traiter qu'une infime partie est de plus en plus perçue comme un gaspillage coupable : ce sont autant de ressources inexploitées. Le big data s'est développé pour répondre à cette fracture entre la production de ces masses de données et leur traitement. Et l'intelligence artificielle est sortie de son long hiver – la série des échecs et des espoirs déçus des années 70 aux années 90 – parce qu'un saut qualitatif dans le traitement des données devenait nécessaire.

Les promesses ambiguës de l'intelligence artificielle

Au cours des années 2000, les innovations ont beaucoup porté sur les usages liés à l'ubiquité de la puissance informatique (plus puissante, plus petite, meilleur marché), ainsi que les enjeux de transparence (veiller / surveiller) dans la production des données. Ces questions restent majeures aujourd'hui. Mais on a constaté ces dernières années un essor considérable du nombre d'innovations issues de l'intelligence artificielle.

La renaissance de l'intelligence artificielle est principalement venue du courant de recherche qui travaille sur les *réseaux de neurones artificiels* (un terme métaphorique : les neurones en question sont des fonctions mathématiques et n'ont pas de réalité physique). Cette forme d'intelligence artificielle ne s'appuie pas sur des systèmes experts (un corpus de connaissances), mais sur le traitement statistique massif d'informations à travers un processus d'apprentissage itératif. Plus clairement : la machine traite les données et donne des réponses à partir d'un algorithme initial ; on lui indique quelles réponses sont vraies et fausses ; elle corrige son traitement pour aboutir aux bonnes réponses ; on lui donne de nouveaux jeux de données, elle se trompe à nouveau, et se corrige à nouveau en fonction des réponses qu'on lui donne... jusqu'à ce qu'elle ne se trompe plus ; on peut alors repartir dans un cycle où on lui demande des traitements de complexité supérieure. Les moteurs de recherche, les domaines très étendus de la reconnaissance visuelle et vocale ou bien encore les traductions automatiques fonctionnent *tous* sur ce modèle, de même qu'une grande partie des systèmes de réponse aux questions en langage naturel (les fameux *bots*).

Pour que l'apprentissage s'effectue, le programme a besoin de beaucoup de données. Moins il y a de données, moins la machine est intelligente, c'est-à-dire moins elle est capable de fournir les bonnes réponses à de nouveaux jeux de données. Non seulement cette forme d'intelligence artificielle est adaptée pour le traitement de masses considérables de données, mais elle *requiert* ces quantités pour devenir intelligente. Plus il y a de données, mieux c'est : l'intelligence artificielle est le complément nécessaire de la production continue et massive de données.

Un essai de prospective : quand les tendances s'amplifient

Bien qu'on encourt toujours le ridicule à vouloir anticiper l'avenir, sautons le pas pour imaginer où ces tendances nous conduisent.

Tout indique que la tendance à la production de données sur tout, tout le temps, par tous les moyens va se poursuivre, avec ses opportunités et ses risques. Cette nouvelle ressource ne pouvant rester inexploitée, on verra se multiplier les innovations en matière d'usage des nouvelles capacités de traitement des données et cela concerne et concernera des domaines aussi variés que la santé, le corps et le fonctionnement du cerveau, notre habitat, nos déplacements et nos loisirs, les relations sociales, le monde du travail, le monde économique : production, logistique, commerce, marketing, relation client, finance, etc.

Ce mouvement de création de nouveaux usages est porté par le développement de l'intelligence artificielle qui n'en est aujourd'hui qu'à ses premiers pas. On assistera à un accroissement considérable de la capacité à traiter des masses d'information pour fournir des connaissances, des prévisions (anticipation), des prescriptions (recommandations) et pour actionner des robots.

La place de l'homme, à cet horizon, est problématique. Beaucoup des inquiétudes viennent de la peur de ce qu'on appelle *l'intelligence artificielle forte*, c'est-à-dire la capacité qu'auraient les machines, en plus de réaliser des tâches cognitives complexes, à être conscientes d'elles-mêmes et à se fixer des objectifs : autrement dit à être un double de l'homme. Cette perspective n'est pas à exclure, mais les développements actuels de l'intelligence artificielle n'en prennent pas le chemin. On réussit de mieux en mieux à *tromper* les hommes dans l'échange avec les machines mais on ne fait que simuler un comportement « humain » en apprenant à la machine, par des outils statistiques, à analyser les signaux émis par les êtres humains et à émettre ensuite des signaux que ces derniers interpréteront à tort... comme étant humains. D'intelligence artificielle forte, on n'en voit pas aujourd'hui le début d'une réalisation.

En revanche, *l'intelligence artificielle faible*, qui consiste à développer des méthodes et des programmes pour résoudre des tâches cognitives complexes et spécifiques (comme la reconnaissance visuelle, le traitement du langage, etc.) connaît bien un développement sans précédent. Par exemple, des logiciels permettent d'identifier, sur photo, des mélanomes et de les distinguer des grains de beauté aussi bien, et bientôt mieux, que des dermatologues. L'analyse automatique de scanners du cerveau pour établir un diagnostic est aujourd'hui de plus en plus performante et le sera bientôt plus que celle des radiologues. Alors que le cerveau humain fonctionne de façon heuristique (par des raccourcis cognitifs), que la machine fonctionne de façon systématique en ayant appris de façon itérative par l'analyse de millions de cas. Autrement dit, si la machine ne démontre aujourd'hui ni créativité, ni intuition (intelligence artificielle forte)... on ne demande nullement à un diagnostic d'être créatif, mais d'être fiable.

Pour prendre l'exemple de l'étude du génome humain, on peut aujourd'hui faire séquencer son ADN pour 1000 euros, en cherchant les parties de son code génétique qui donnent une prédisposition à certaines maladies héréditaires (par exemple, le cancer du sein). Demain, pour le même prix, on séquencera les 2,3 milliards de paires de nucléotides qui constituent notre ADN, mais on ne l'analysera plus pour tester des hypothèses spécifiques. Les programmes d'intelligence artificielle analyseront et modéliseront l'ensemble de son ADN pour établir un profil qui permette d'adapter ses traitements (*médecine personnalisée*) : aucun être humain, fut-il un médecin hautement compétent, n'aura la capacité d'établir ce diagnostic et ces prescriptions.

La révolution numérique menace donc autant des emplois de service, comme le personnel de centres d'appels, que des emplois hautement qualifiés. Cette menace est nécessairement croissante : la révolution numérique fournit sans cesse plus de données et les outils d'intelligence artificielle, sans cesse plus perfectionnés, accompliront de mieux en mieux les tâches complexes qui étaient réservées aux êtres humains. L'enjeu est moins aujourd'hui celui d'un Golem, d'une machine dotée de conscience, que la multiplication de programmes spécialisés sans cesse plus performants parce qu'ils n'affrontent pas de limites biologiques.

Les voitures autonomes ont aujourd'hui un taux d'accidents pratiquement nul (plus de 90% des accidents automobiles sont dus à des erreurs humaines selon les études américaines), les avions autonomes feront chuter... le nombre d'accidents, les médecines automatiques, le nombre d'erreurs médicales, et peut-être demain, le recrutement du personnel, les actions militaires, etc. seront plus rapides, plus sûrs et moins coûteuses grâce à l'intelligence artificielle.

En 1962, J.C.R. Licklider et Welden Clark ont publié un article prémonitoire sur l'intelligence artificielle, en assurant que l'avenir n'était pas à la substitution, mais à la complémentarité :

« Les hommes fixeront les buts, formuleront des hypothèses, détermineront des critères et exécuteront les évaluations. Les ordinateurs feront le travail routinier, qui doit être fait pour préparer les idées et les décisions propres à la pensée technique et scientifique. (...) Aujourd'hui, nous pensons que l'homme et la machine se complètent l'un l'autre et que la capacité intellectuelle de cette symbiose homme-machine sera beaucoup plus puissante que celle de chacun pris séparément.¹ »

Dans cet article, les auteurs identifiaient six domaines de supériorité humaine, deux domaines partagés et deux où dominerait la machine : d'où l'intérêt de la complémentarité. Mais si on reprenait aujourd'hui ces dix domaines, la supériorité des machines est aujourd'hui démontrée dans six, voire sept d'entre eux, réduisant ainsi sensiblement le domaine de la suprématie humaine et la nécessité de cette complémentarité.

¹ J. C. R. Licklider and Welden Clark (August 1962). "On-Line Man-Computer Communication", *AIEE-IRE '62* (Spring), pp. 114-115.

Le pire n'est jamais sûr

La révolution numérique s'inscrit dans la continuité de la révolution scientifique de quantification du monde. Que se passe-t-il quand on numérise tout, à tout propos, constamment et à flux continu ? On cherche à organiser et à exploiter ces données pour mieux agir sur le monde. L'intelligence artificielle est une étape de ce processus. Ses domaines d'application sont immenses et les innovations en matière d'usage sont directement corrélées aux progrès rapides de ce qui est aujourd'hui une science en formation. Mais ces progrès soulèvent une interrogation, qui est celle de la place de l'homme dans un monde dominé par les données. Le projet humaniste qui fait de l'homme le référent de toutes choses, et donc le principal bénéficiaire de la science qui le rend « comme maître et possesseur de la nature », est remis en cause dans ses fondements.

Mais l'innovation digitale est par nature ambivalente. Un monde de transparence est un monde de surveillance comme on le constate aujourd'hui dans l'usage que le gouvernement chinois fait des outils d'intelligence artificielle. Mais l'intelligence artificielle représente aussi un champ d'innovations considérables pour améliorer notre vie, notre santé, l'univers dans lequel nous vivons, en même temps qu'il conduit à réévaluer progressivement ce qu'être humain veut dire, et sa place entre les autres Êtres vivants et les machines.