



## GRUPE DE RÉFLEXION K2

# Les enjeux du Big Data

**Co-pilotes**

Messaoud Chibane et Kevin Dumoux

Shirine Benhenda, Gilbert Cette, Sonia Dahech, Aymar De la Mettrie,  
Franck DeCloquement, Franck Duval, Yara Furlan, Jean-Baptiste Harry,  
Thibaut Heckmann, Timothé Hervé, Maxence Jeunesse, Ano Kuhanathan,  
Yann Levy, Aurélie Luttrin, François Marchessaux, Benjamin Meisse,  
Ilan Ouanounou, Adil Redouani, Henri Michel Reynaud, Aurélie Sale,  
Jill-Jènn Vie, Jessica Vonderscher, Jun Zhou

# Avant-propos

Nos groupes de travail essaient de rendre compte d'une idée force du K2: la connaissance appartient à tout le monde. Malgré les raisons qui ont trop longtemps conduit le genre humain à se séparer, et à séparer les connaissances, nous pensons que les efforts de réflexion et de partage de professionnels issus de tous horizons dans une démarche pluridisciplinaire permettront de retrouver de l'unité dans la connaissance et donc une meilleure compréhension globale des choses. Telle est la démarche qu'ont entreprise les membres du Groupe K2 "Les enjeux du Big Data".

Il nous a donc semblé que cette tentative de penser le monde de façon globale nécessitait de modifier les perspectives habituelles de la production d'une réflexion commune.

Il nous est apparu que le point de vue de l'expert, auquel nous sommes habitués dans ce genre de rapport, ne devait plus être exclusif. Un besoin croissant du citoyen de connaître par lui-même se fait sentir en tout domaine. C'est la raison pour laquelle nos groupes de travail sont composés de spécialistes mais aussi de nombreuses personnalités qui ne disposent pas, a priori, de compétences particulières sur le sujet traité mais qui souhaitent y réfléchir.

Permettre au plus grand nombre de mieux appréhender des sujets d'actualité ou de fond nous paraît donc plus important qu'énumérer de savantes réflexions destinées à quelques spécialistes. C'est pourquoi chacun de nos rapports commence par une note

de synthèse écrite par les membres du groupe laquelle doit permettre de comprendre les grandes dynamiques du sujet. Elle est suivie de contributions (articles ou entretiens) confiées à des experts, français ou étrangers, mais celles-ci ont vocation à être accessibles au plus grand nombre.

## Parce que le monde impose une approche pluridisciplinaire.

L'animation d'un groupe de travail est enfin chose difficile. Cette tâche est assumée par un Pilote dont la mission est de conduire les énergies et faire en sorte qu'elles se réunissent sans pour autant écraser les différences. Son rôle est donc de relier les réflexions, les travaux des participants, trouver des équilibres et créer entre eux le lien indispensable à une œuvre commune. Que Messaoud Chibane, co-pilote de ce groupe sur les enjeux du Big Data, soit ici remercié pour ce rôle qu'il a parfaitement assumé.

Bonne lecture...

**Kevin Dumoux - Jean-Michel Icard - Général Jean-Pierre Meyer - Krys Pagani**

Co-créateurs du Cercle K2 et membres du Collège de la gérance

# Sommaire

Avant-propos .....	2	<b>11 - L'intelligence artificielle au service des Finances publiques</b> .....	76
Sommaire .....	4	par Franck Duval	
Avertissement .....	6	<b>12 - Propriété intellectuelle et Data : réflexions d'ordre général</b> .....	80
Composition du Groupe K2 .....	8	par Henri Michel Reynaud	
<b>01 - Note de synthèse - Le Big Data, une nouvelle révolution industrielle</b> .....	10	<b>13 - Le Big Data et les mutations du monde du travail</b> .....	90
<b>02 - Les organisations et l'enjeu de la Data : un tournant peu maîtrisé</b> .....	22	par Gilbert Cette	
par Aurélie Sale		<b>15 - Le chef d'entreprise au centre du jeu de la donnée</b> .....	96
<b>03 - Big Data et la conduite du changement : l'agilité avant tout</b> .....	28	par Benjamin Meisse	
par Adil Redouani		<b>14 - Les nouveaux métiers du digital : focus sur celui de Social Media Trader</b> .....	100
<b>04 - Pour une approche pragmatique du « Big Data »</b> .....	32	par Yara Furlan	
par Ilan Ouanounou		<b>16 - Data et Key Opinion Leader : un pivot de l'influence du nouveau monde ?</b> .....	104
<b>05 - La prospective et la donnée : « Foresight with Data »</b> .....	38	par Jun Zhou	
par Aymar De la Mettrie		<b>17 - Data visualisation &amp; Data Storytelling : quand les compétences techniques ne suffisent plus pour exploiter ses données</b> .....	114
<b>06 - Le Big Data au sein des Institutions financières sous le prisme de la gestion des risques</b> .....	44	par Yann Levy	
par Timothé Hervé		<b>18 - Big Data et cybersécurité</b> .....	120
<b>07 - Données bancaires : la transformation bancaire au cœur du marché unique de la donnée</b> .....	50	par Thibaut Heckmann	
par François Marchessaux		<b>19 - Biais, confidentialité, RGPD</b> .....	124
<b>08 - Quelques réflexions sur l'usage de la donnée dans le domaine de l'assurance</b> ..	60	par Jill-Jénn Vie	
par Maxence Jeunesse		<b>20 - Comment l'utilisation d'un matériel dédié au rendu graphique a permis la troisième vague de l'intelligence artificielle et ses applications concrètes ?</b> .....	128
<b>09 - Les données de santé et leurs enjeux</b> .....	64	par Jean-Baptiste Harry	
par Shirine Benhenda		<b>21 - La révolution de la data pourrait être à plusieurs vitesses : les dirigeants, salariés et citoyens doivent en être conscients</b> .....	134
<b>10 - Data et Justice : quelles données pour quels enjeux ?</b> .....	70	par Ano Kuhanathan	
par Jessica Vonderscher		<b>22 - La souveraineté numérique au fondement de notre performance nationale</b> ..	138
		par Franck DeCloquement et Aurélie Luttrin	
		<b>Bibliographie</b> .....	150

# Avertissement

**Le Cercle K2 n'entend donner ni approbation ni improbation aux opinions émises dans les rapports : ces opinions doivent être considérées comme propres à leurs auteurs.**



# Composition du Groupe K2



**Kevin Dumoux**  
Co-créateur du Cercle K2,  
Conseil en Stratégie,  
Transformations digitales  
et M&A  
*co-pilote du Groupe*



**Messaoud Chibane**  
Directeur du MSC  
Finance & Big Data,  
NEOMA Business School,  
Lauréat du Trophée K2  
"Finances" 2018  
*co-pilote du Groupe*



**Shirine Benhenda**  
Experte en Biologie  
moléculaire,  
données OMICS, PhD



**Sonia Dahech**  
Directrice CRM, Trafic  
et Data omnicanal,  
BUT



**Franck DeCloquement**  
Expert en intelligence  
stratégique  
Enseignant à l'IRIS et l'IHEDN  
Spécialiste Cyber



**Franck Duval**  
Administrateur des  
Finances publiques,  
Directeur adjoint du pôle  
gestion fiscale, DDFiP 92



**Yara Furlan**  
Trader Social Media,  
Publicis Media



**Jean-Baptiste Harry**  
HPC & AI Solution Architect  
& pre sales EMEA,  
NEC Europe



**Timothé Hervé**  
Risk Manager,  
Banque de France



**Yann Levy**  
Data Analyst,  
Expert BI



**Aurélie Luttrin**  
Présidente,  
Eokosmo



**François Marchessaux**  
Senior Partner,  
Franz Partners  
Conseil en Stratégie  
& Management



**Aurélie Sale**  
Coach Agile,  
Renault Digital



**Jun Zhou**  
Entrepreneur, Lecturer &  
Consultant in Chinese  
Social Media, MSIE,  
HEC Paris

## 01

## Note de synthèse

# Le Big Data, une nouvelle révolution industrielle

Messaoud Chibane, Kevin Dumoux

Shirine Benhenda, Sonia Dahech, Franck DeCloquement, Franck Duval,  
Yara Furlan, Jean-Baptiste Harry, Timothé Hervé, Yann Levy,  
Aurélie Luttrin, François Marchessaux, Aurélie Sale, Jun Zhou

Au cours des trois dernières décennies, la quantité de données auxquelles les entreprises et les individus ont accès a explosé. La datasphère globale, c'est-à-dire l'ensemble des données disponibles sous forme digitale dans le monde, atteindra 80 zettabytes en 2022 et doublera d'ici 2025 (IDC, 2021). Pour décrire ce phénomène, le concept de Big Data (données massives) a vu le jour. Il permet notamment de mettre en lumière la capacité pour les entreprises de développer des outils plus précis d'aide à la décision (Big Data Analytics), d'améliorer leur performance organisationnelle et d'en tirer un avantage compétitif.

des entreprises européennes. Ainsi, selon une estimation du WISE Institute, le Big Data peut contribuer au PIB de l'Union européenne à hauteur de 1.9 % sur un horizon de 5 ans, soit approximativement l'équivalent d'une année de croissance de PIB.

Bien sûr, la seule présence des données ne peut à elle seule générer tout son potentiel économique car celui-ci doit être exploité par les techniques les plus avancées de l'intelligence artificielle (IA). Selon une étude du cabinet Mc Kinsey, l'IA pourrait contribuer à 13 trillions de dollars d'activité économique soit l'équivalent de 1.2 % de PIB additionnel

**Si le Big Data est devenu un facteur majeur de développement économique, la question est donc, qui seront les pays dominants dans les décennies à venir et où se situeront l'Europe et la France sur l'échiquier.**

Certaines entreprises comme Netflix, Amazon, Google, articulent l'ensemble de leur stratégie sur l'exploitation du Big Data. Selon (Cavanillas, 2016) l'industrie des nouvelles technologies et des télécommunications contribue pour 5 % du PIB de l'Europe et, dans ce cadre, les solutions Big Data (et de l'open data) peuvent participer à améliorer la compétitivité

par année. Cet apport dans la richesse globale se ferait à travers une diversité de facteurs tels qu'une plus grande automatisation du travail, une accélération du rythme de l'innovation et un accroissement de productivité et de compétitivité. Si le Big Data est devenu un facteur majeur de développement économique, la question est de déterminer qui seront

les pays dominants dans les décennies à venir et où se situeront l'Europe et la France sur l'échiquier.

### Pourquoi le Big Data révolutionne-t-il le fonctionnement des entreprises ?

Selon (Philippon, 2019) l'association du Big Data et des technologies de l'IA peut être envisagée sous trois angles :

1. Elles constituent un actif incorporel pour l'entreprise.
2. Elles révolutionnent les outils permettant d'accroître sa part de marché :

Exemples: Amazon et Netflix utilisent des outils de Big Data pour améliorer les recommandations de produits à leurs clients.

3. Elles apportent des outils de prévision (et donc d'aide à la prise de décision) beaucoup plus puissants. Les outils de Big Data peuvent par exemple être utilisés pour adapter les politiques de prix en fonction de la démographie des consommateurs (âge, genre, lieu de vie, etc.) (Cavallo, 2017).

En exploitant les avantages du Big Data dans leurs organisation, les entreprises augmenteraient leur productivité de 5 % et leur profitabilité de 6 %. Elles diminueraient le coût d'acquisition de nouveaux clients de 47 % et augmenteraient la valeur de leur firme de 8 % (McAfee, 2012).

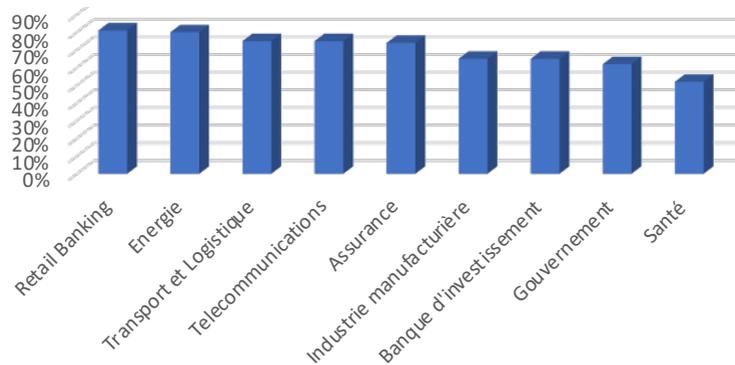


Figure 1: Taux d'adoption du Big Data par industrie au Royaume-Uni en 2020  
(Source: Statista 2020)

De plus, le Big Data constitue un nouveau vecteur d'innovation pour les entreprises. Tout d'abord, il permet d'optimiser la rentabilisation de leurs actifs, de personnaliser leurs services ou de lancer de nouveaux produits (Trabucchi, 2019).

Certaines industries sont déjà bien avancées dans leur transition vers l'adoption du Big Data. À titre d'exemple, la Figure 1 montre le taux d'adoption du Big Data par industrie en 2020 au Royaume-Uni. On peut constater que la banque de détail est en tête avec 80 %, suivie par l'industrie de l'énergie et les transports. La santé arrive en dernière position avec 52 % d'adoption seulement.

Par ailleurs, l'adoption du Big Data permet de nourrir, au cœur des écosystèmes, les processus d'innovation collaborative directe et indirecte, passive ou intentionnelle entre les entreprises, leurs fournisseurs, et les différentes institutions (Bresciani, 2021).

Si l'on ausculte l'interaction entre le Big Data et l'innovation du point de vue de la recherche scientifique, on assiste à un vrai changement de paradigme. En effet, la démarche scientifique classique consiste, en partant de conjectures, à émettre des hypothèses et à les vérifier à partir de données échantillonnées. Au contraire, avec le Big Data, il est courant

de collecter un grand nombre de données et d'extraire leurs sens et leurs corrélations *a posteriori*.

D'autre part, la variété immense des données de natures différentes nécessite la mise en place de projets pluridisciplinaires et l'intervention d'experts de domaines parfois radicalement différents, ce qui requiert un nouveau mode d'organisation du monde académique. Aujourd'hui, il existe peu de revues scientifiques de haut niveau qui soit véritablement pluridisciplinaire. Pourtant, les pays qui seront à même de mettre en place ce nouveau type d'environnement académique, permettant au Big Data d'exprimer toute son potentiel, seront très probablement aux premières places mondiales de l'innovation.

Enfin, le Big Data est en passe de révolutionner le monde de l'éducation. En effet, grâce aux développements des cours online comme Uдеми, OpenClassrooms ou Coursera, un volume important d'information est produit. La digitalisation des échanges entre professeurs et élèves permet aux premiers d'accéder aux travaux de ces derniers, de contrôler leur performance et de leur renvoyer du feedback quasiment en temps réel. Cette immédiateté dans l'interaction pédagogique couplée au volume important de données produites permet de construire

des modèles d'apprentissage qui peuvent s'adapter aux préférences et compétences des étudiants afin de construire un système éducatif plus efficace et plus juste.

Il semble donc clair que le Big Data, couplé aux technologies de l'IA, doivent occuper une place majeure dans la stratégie économique et d'innovation des entreprises et des grandes puissances. Il est légitime, nécessaire et urgent de se demander comment se positionne la France et plus généralement l'Union Européenne (UE) dans ce nouvel ordre mondial.

## **Le Big Data, couplé aux technologies de l'IA, doit occuper une place majeure dans la stratégie économique et d'innovation des entreprises et des grandes puissances.**

Quand on pense aux super puissances de la data, les deux pays qui viennent à l'esprit sont les Etats-Unis et la Chine, notamment grâce à leurs entreprises iconiques dans le domaine, respectivement les GAFAM et Alibaba, Tencent et Huawei. Cependant, si l'ère du Big Data peut redessiner un nouvel ordre mondial dans un avenir proche, comme le pétrole l'a fait au 20ème siècle, cette transformation se fera de manière différente

car ce nouvel or digital n'est pas produit uniquement dans ces deux pays. Pour classer les pays selon un nouveau critère de GDP (Gross Data Product), l'étude de la Harvard Business Review (Chakravorti Bhaskar, 2019) classe les pays selon un nouveau jeu de critères :

- Volume : la quantité totale de données consommées à haut débit par un pays, comme approximation des données brutes générées ;
- Usage : nombre d'utilisateurs actifs d'internet comme approximation de la diversité des utilisateurs, de leur com-

portements et de leurs besoins;

- Accessibilité: c'est-à-dire le degré auquel les institutions du pays donnent accès aux différentes applications, aux chercheurs en IA et aux innovateurs dans leur généralité ;
- Complexité: le volume des données haut débit consommée par utilisateur, qui permet de mesurer la complexité et la sophistication de l'activité digi-

tales d'un pays. Cette dernière mesure est comparable au concept de PIB par habitant qui permet de comparer les niveaux relatifs de prospérité et de développement des pays.

Peut-être que le critère le plus difficile à appréhender pour classer les différents pays dans ce nouvel ordre digital est celui de l'accessibilité: nous avons le problème de la protection des données qui peut en apparence être en conflit avec celui d'une certaine ouverture des données, nécessaire pour alimenter de façon efficiente les outils d'IA. Parmi les grandes économies, La Chine présente un avantage concurrentiel à court terme puisque, par la taille même de sa popula-

## **Les trois premières puissances dans l'ordre mondial du Big and Open Data sont les Etats-Unis, Le Royaume-Uni et la Chine suivi par la Suisse, la Corée du Sud et en sixième position la France.**

tion, elle génère une quantité importante de données sur laquelle les institutions gouvernementales ont un droit de préhension tout en limitant l'accès de ces données aux autres pays.

Si l'on ne considère que les critères de l'accessibilité et de la complexité (Figure 1), on constate que les Etats-Unis dominent le classement alors que la Chine

figure parmi les mauvais élèves sur ces deux critères. Sans surprise, l'Union Européenne montre un bon niveau d'accessibilité mais son score de complexité reste faible comparé à d'autres pays européens non UE tels que le Royaume Uni ou la Suisse. À l'exception de la Suède, on constate une certaine homogénéité dans le classement des différents pays de l'UE avec une France qui domine le critère d'accessibilité.

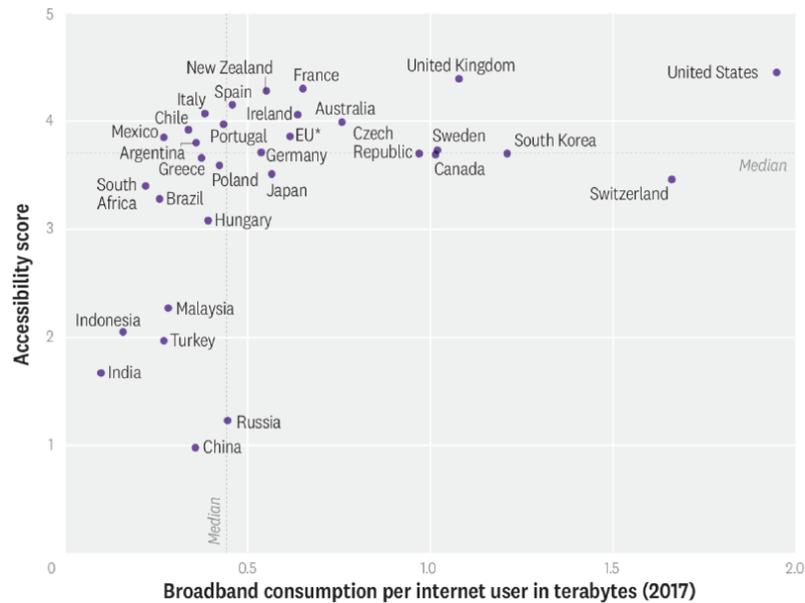
Lorsque les 4 critères sont pris en considération, les trois premières puissances dans l'ordre mondial du Big and Open Data sont les Etats-Unis, le Royaume-Uni et la Chine suivi par la Suisse, la Corée du

Sud et en sixième position la France.

Il semble donc au vu de ces éléments que la France soit bien positionnée pour profiter de l'expansion du Big Data dans les années à venir. Cette analyse est confirmée par la progression de la France au sein du Global Innovation Index (GII 2021). Elle devance la Chine et talonne l'Allemagne (en 10ème position). Cette

amélioration est à mettre en perspective avec le mouvement de la French-tech. Pour rappel, le nombre de Licornes françaises est passé, de 2018 à 2021, de 4 à 26. Il faut cependant relativiser ces données positives notamment au regard du positionnement moyen de la France (18ème) sur le Network Readiness Index. Cet index mesure entre autres le niveau fondamental de l'infrastructure ICT (Information and Communication Technology), la sophistication des entreprises

et des individus dans leur utilisation de ces technologies, ainsi que le niveau des investissements publics dans leurs infrastructures technologiques. De même, dans un contexte plus général, l'OCDE mesure l'intensité de recherche et de développement dans les nouvelles technologies des différents pays (le niveau de dépenses R&D en pourcentage du PIB). Sur cette échelle, la France alloue seulement 2.5 % de son PIB contre 4.9 % pour Israël et 4.6 % pour la Corée du Sud.



\*The EU data point contains 12 EU countries and almost 81% of the EU population.  
Source: Analysis of Eurostat, Cisco, ITU, Global Open Data Index, Open Government Partnership, and CNIL data by The Digital Planet Initiative at The Fletcher School, Tufts University; and Mastercard

HBR

Figure 2: Classement International sur les critères d'accessibilité des données et de la consommation haut débit par utilisateur (Source HBR)

Enfin un futur positionnement avantageux de la France dans le classement du Big Data ne pourra se matérialiser que si sa politique de protection des données est à la hauteur des enjeux économiques et stratégiques. Aujourd'hui, celle-ci est classée 14ème, derrière des pays attendus comme les Etats-Unis (1er), le Royaume-Uni (2ème) mais aussi des

### Le risque d'une stratégie seulement basée sur le critère économique

L'adoption généralisée du Big Data dans les économies occidentales favorisera les grandes compagnies au détriment des plus petites. En effet, d'après (Bessen, 2017), les nouvelles technologies propulsent la productivité des grandes entreprises ainsi que leurs marges opéra-

### La France possède un certain nombre d'atouts pour atteindre une place dominante dans le nouvel ordre mondial induit par le Big Data.

pays plus surprenants tels que l'Arabie Saoudite (3ème), l'Estonie, l'Espagne ou encore la Russie.

En conclusion, la France possède un certain nombre d'atouts pour atteindre une place dominante dans le nouvel ordre mondial induit par le Big Data. Mais ce seul objectif économique ne peut se faire sans prendre le risque de créer un grand nombre d'inégalités à la fois entre les petites et les grandes entreprises mais aussi entre les individus qui ne seront pas tous équipés du même bagage culturel et éducationnel pour « surfer sur la vague » du Big Data.

tionnelles. L'ère du Big Data pourrait donc causer la quasi-disparition des entreprises de petite taille qui n'auraient pas la capacité de générer suffisamment de données pour bénéficier des économies d'échelle qui profitent à leurs concurrents. De plus, puisque les grandes entreprises génèrent plus de données que leurs concurrentes, plus modestes, les institutions financières, elles-mêmes disposeront de plus d'informations sur les grandes entreprises et leurs proposeront donc des opportunités de financement plus avantageuses (Begenau, 2018).

Même si l'ère du Big Data générera un grand nombre d'opportunités économiques et créera beaucoup d'emplois dans la technologie pour les data scientists, data analysts de demain, elle automatisera un grand nombre de tâches et provoquera probablement l'abandon d'un nombre important de postes aujourd'hui occupés par des êtres humains. Ces personnes, qui ne possèdent pas les compétences appropriées aux traite-

Enfin, l'ère du Big Data va accentuer le fossé entre les pays développés et les pays en voie de développement. En effet, ces derniers risquent de devenir de simples fournisseurs de données qui seront exploités par la suite par les entreprises dominantes des pays développés.

On peut ainsi voir que l'impact du Big Data sur tous les secteurs du monde s'apparente à une véritable révolution

## L'impact du Big Data sur tous les secteurs du monde s'apparente à une véritable révolution industrielle et est en passe de changer les rapports de force économique au sein des sociétés mais aussi entre pays.

ment des données Big Data, auront des difficultés à maintenir leur employabilité. Nous pourrions donc assister à une substitution « Machine contre Humain ». Nous devons donc anticiper cette possibilité et nous préparer à ce que l'innovation Big Data favorise le consommateur en partie mais cause des externalités négatives pour le travailleur non formés à ces nouvelles technologies.

industrielle et est en passe de changer les rapports de force économique au sein des sociétés mais aussi entre pays. L'ensemble de ses ramifications est tellement vaste qu'il faudrait plusieurs ouvrages pour toutes les traiter. Nous allons ici nous concentrer sur les aspects du Big Data qui nous semblent essentiels pour bien appréhender ses conséquences à court et moyen termes. Nous centrerons notre analyse sur les thèmes suivants :

### 1. Le Big Data, ou comment les entreprises doivent opérer leur mutation pour profiter de cette révolution industrielle

Le monde économique a pris conscience du volume extraordinaire des données disponibles et pressent son potentiel en termes de nouvelles opportunités de création de valeur et d'amélioration des processus de décision et d'exécution. Mais de la prise de conscience à l'exploitation optimale des données des pas de géants restent à faire. Quels sont les changements à opérer par les entreprises pour adopter le Big Data dans leur business model ?

### 2. Le Big Data et son impact sur le secteur financier

Contrairement aux idées reçues, ce n'est pas l'industrie des télécommunications ou le secteur de la technologie qui sont le plus en pointe dans l'adoption du Big Data mais bien le secteur financier avec au premier rang la banque de détail, alors que la banque d'investissement est en septième position. Les sources de données exploitées par les services financiers sont multiples, s'étendant des profils et comportements clients de la banque de détail, aux flux d'investissements, aux prix disponibles sur les marchés financiers en passant par les analyses fournies

par les experts financiers et les agences de notation.

L'industrie financière est forcée d'opérer une véritable mutation en raison de l'expansion des Fintech et de leur business model plus agile qui offre des services plus larges et moins chers à leurs consommateurs. Parmi les plus connues, on peut citer Quonto, Lydia, ou encore Finexkap, Fintech spécialisée dans l'affacturage qui utilise des algorithmes de Big Data pour évaluer le risque client avec une précision accrue. Dans la mesure où le secteur financier irrigue l'économie, il est crucial de comprendre comment le Big Data va impacter la gestion des risques, le financement des entreprises et des particuliers et le secteur de l'assurance

### 3. Le Big Data et les services publics

Le Big Data impacte tous les pans de l'économie et de la société. Comment envisager un secteur privé performant dans son utilisation des données sans l'existence d'un service public qui adopte lui-même une approche sophistiquée dans la collecte, l'organisation et la valorisation de celles-ci.

Prenons par exemple le secteur de la santé où une quantité importante de données patients est stockée dans les systèmes informatiques des hôpitaux

publics, cliniques privées, médecins généralistes et spécialistes sans que ces systèmes communiquent nécessairement entre eux. Une des conséquences de cette absence de cohérence est le coût financier supporté par le système de santé du fait de la multiplication d'exams médicaux, lesquels pourraient être parfois évités si toutes les données médicales des patients étaient centralisées. Un autre problème est aussi l'éventuelle « perte de chances » subie par le patient, due à ce manque de communication. Ces données pourraient être mis à profit afin de concevoir des algorithmes de prédiction qui permettraient d'obtenir des signaux faibles de présence de maladies graves qui seraient traitées en amont avant que l'état de santé des patients ne se dégrade.

Un parallèle peut aussi être établi avec le domaine de la Justice ou celui de notre infrastructure fiscale où le manque de cohérence entre les différentes bases de données publiques conduisent à des pertes d'efficacité, des erreurs de procédure ou même des fraudes alors qu'une exploitation optimale du Big Data pourrait résoudre ces problèmes et donc améliorer le service rendu aux individus et la situation financière des Etats.

#### 4. Le Big Data et la refonte du monde du travail

On voit qu'une telle mutation engendra à moyen et long termes un changement profond des économies occidentales. Qu'en est-t-il du marché du travail ?

Une hyper-digitalisation des échanges créera forcément de nouveaux métiers et une nouvelle façon d'appréhender les échanges humains. Mais ces nouvelles opportunités seront-elles l'apanage des classes les plus favorisées ou « ruisselleront-elles » sur l'ensemble de la société ?

#### 5. Le Big Data et les risques encourus : cybersécurité, dépendance matérielle et souveraineté numérique

Si toutes les révolutions industrielles s'accompagnent de gains de productivité, elles exposent aussi les économies qui en bénéficient à un certain nombre de risques. Dans le cas du Big Data, la performance des entreprises et des Etats en termes de cybersécurité est essentielle pour garantir la viabilité de cette mutation. De plus, ce nouvel « or noir » peut être transférable plus facilement que son ancêtre physique : se pose la question de la souveraineté des nations quant aux données qu'elles possèdent. Quelles sont les politiques nationales et internationales à mettre en place afin de protéger

les prérogatives des Etats quant à l'utilisation et/ou le partage de leurs données ? Pour la France en particulier, c'est une question stratégique qu'il est nécessaire de traiter en priorité.

Enfin, on oublie souvent que cette révolution digitale repose sur la performance du hardware utilisé pour stocker les données et exécuter les algorithmes d'intelligence artificielle qui les exploitent. Qui sont les constructeurs majeurs de ces composants matériels et comment éviter une dépendance trop importante vis-à-vis de ces acteurs ?

## 02

## Les organisations et l'enjeu de la Data: un tournant peu maîtrisé



### Aurélie Sale

Coach Agile,  
Renault Digital

Les enjeux de la data, la gouvernance des données, le pilotage, la sécurisation... tous ces sujets sont connus et étudiés par l'ensemble des organisations. Mais sont-ils vraiment pris en compte ?

Tout comme le volume mondial des données, les offres de formation et d'emploi se sont multipliées ces dernières années. Pour les entreprises l'enjeu n'est donc pas les ressources. La prise de conscience du Data Driven (approche consistant à prendre des décisions stratégiques sur la base d'analyses de données) est également présente dans la majorité des organisations françaises et européennes. Elles savent que l'enjeu est fort. Tout le monde en parle, tout le monde en fait. Avoir un « Pôle Data Scientist » ou une équipe d'expert en data est aujourd'hui monnaie courante. Les entreprises de services numériques (ESN) se ruent sur les profils pouvant être qualifiés de data scientist, les offres de conseil se multiplient et les projets abondent. L'objectif est clair : la data va aider l'orga-

nisation d'aujourd'hui à devenir celle de demain. Identifier les nouveaux services et les usages, définir les futurs profils de consommateurs, se transformer de vendeur de biens pour devenir une « Tech Company » ... Les enjeux business sont connus et l'intérêt des organisations à prendre le tournant de la data n'est plus à prouver. Mais y arrivent-elles ?

Avant de lancer un projet d'envergure ou d'orienter l'ensemble de l'organisation vers la culture du data driven, les entreprises ont, pour la plupart, le réflexe de réaliser des Proof of Concept (POC / Preuve de concept). Cette bonne pratique permet de vérifier la valeur délivrée par un projet, une équipe. Malheureusement les POC en lien avec la gestion des données sont rarement transformés en projets pérennes. Les POC sont lancés avec précipitation car les ressources – techniques et humaines – sont disponibles, mais l'écosystème, nécessaire à l'aboutissement d'un projet permettant d'utiliser la data pour développer le bu-

business d'une organisation bâtie sur des principes tout autres que le « data driven », est très souvent oublié. Hormis celles construites à partir de ce modèle, les entreprises souhaitant prendre le tournant de la data et bénéficier de l'ensemble des avantages doivent impacter leur organisation et éclaircir certaines zones d'ombre.

Dans une organisation, quand il est fait référence à la data, le sujet n'est pas simplement la donnée et son traitement. Il s'agit en réalité d'une culture. La culture de la data, et comment les données impactent le business, le modèle financier et la gouvernance de l'organisation. Pour prendre ce virage culturel, il faut prendre celui des nouvelles méthodes de travail telle que l'Agilité d'entreprise. Pour que la data science apporte de la valeur à une organisation, celle-ci doit pouvoir prendre en compte des changements rapidement et efficacement. Être résilient, répondre rapidement à un besoin et identifier les améliorations, sont les enjeux sous-jacents à celui de la data. Pour y répondre, deux changements structurants sont à mettre en place au sein d'une organisation :

- Développer la transparence et le droit à l'erreur : Il est encore rare de voir une organisation où les indicateurs sont transmis de la source aux décideurs.

Chaque donnée est analysée et revue par de multiples strates avant de servir à une prise de décision stratégique. Or, chaque strate à le syndrome de « l'indicateur pastèque ». Très répandu dans les grandes organisations, ce syndrome permet à chaque strate d'interpréter une donnée pour afficher une situation positive, là où la réalité opérationnelle n'est pas au rendez-vous attendu – afficher un indicateur vert alors que le cœur est rouge. Par exemple, une donnée sur la date de livraison de la première étape d'un projet est manipulée et interprétée de façon à laisser penser que le projet global sera livré en temps voulu, alors que la donnée seule montre un retard de X semaines. L'erreur, ou la non-adéquation entre une donnée et les attendus, sont vues comme une situation à sanctionner au lieu d'être vues comme une leçon à tirer. La priorité d'analyse des données est donc généralement biaisée par la mise en avant des nouvelles opportunités à détecter, au détriment des améliorations à apporter au système existant.

- Optimiser la collaboration entre les différentes directions : rapprocher la direction métier de la direction informatique n'est pas un nouvel enjeu. Il a été mis en exergue par les méthodes Lean et Agile dès les années 1990. Malgré cela les organisations rencontrent tou-

jours des difficultés, et les silos entre directions persistent. Les directions métier se rapprochent des directions informatiques pour optimiser la gestion des projets et accélérer le Time to Market mais les modèles de gouvernance et de gestion budgétaire peinent

données. La data science fait intervenir un ensemble important de domaines, mathématiques, programmation, intelligence artificielle, analyse prédictive... Et il faut tout autant de collaboration entre l'ensemble des directions d'une organisation pour en tirer de la valeur.

## Deux zones d'ombre majeures mettent en difficulté les organisations dans le traitement des données: La sécurité et la protection des données.

à suivre le rythme. Prenons l'exemple d'un projet de développement d'une application web : la direction métier impliquée va être sollicitée pour participer activement au projet, pour allouer des ressources à temps plein au projet afin d'exposer son besoin et prendre des décisions rapidement, mais le budget et la responsabilité du développement de l'application reste à la main de la direction informatique. La vision business et l'optimisation budgétaire ne peuvent alors pas être alignés. Les données apportent ici une troisième dimension qui complexifie le modèle. Une donnée n'est pas propre à une direction, elle servira l'ensemble des directions ou presque. Il ne s'agit donc plus d'aligner une direction métier et une direction informatique autour d'un projet, mais d'aligner l'ensemble des directions autour des

En plus de ces impacts systémiques, deux zones d'ombre majeures mettent en difficulté les organisations dans le traitement des données : La sécurité et la protection des données. La sécurité des données et des systèmes est un sujet stratégique pour les organisations. Même si elles rencontrent encore des difficultés et que l'évolution constante des menaces rendent ce sujet sensible, les organisations disposent généralement des outils leur permettant de s'orienter dans la bonne direction. En revanche, le suivi du Règlement Général sur la Protection des Données (RGPD) est une ligne directrice difficile à suivre. La data science permet d'identifier des nouvelles tendances, des nouveaux usages, qu'il faut à leur tour analyser. Or, le lien entre ces nouveaux traitements et le RGPD n'est pas toujours fait, ou difficile à faire. Comment

être certain qu'une nouvelle donnée à disposition respecte le RGPD? D'autant plus lorsque les organisations sont silotées et ne font pas le lien entre la direction métier qui demande une analyse et la direction juridique qui peut émettre une recommandation sur le caractère personnel des données traitées. Avec la multiplication des projets liés au traitement des données il sera donc de moins en moins rare de voir des organisations se lancer dans des analyses ne respectant

pas, souvent sans s'en rendre compte, les règles de protection des données à caractère personnel.

L'étape de la prise de conscience sur l'apport majeur de la donnée étant passée, les organisations doivent maintenant répondre à ces enjeux de transformation systémique et assurer le lien entre leurs enjeux internes et l'écosystème réglementaire.

## 03

## Big Data et la conduite du changement: l'agilité avant tout



### Adil Redouani

Pilotage de projets de transformation,  
FrenchOps

« L'apprentissage et l'innovation vont de pair.  
L'arrogance du succès est de penser que ce que  
vous avez fait hier sera suffisant pour demain »

William Pollard

Le changement a longtemps été considéré comme un processus de passage d'un état stable à un autre état stable. Une construction linéaire et déterministe dans un environnement réputé prévisible.

Le pilotage des projets au sens large y compris les projets IT se sont inspirés de cette philosophie organisationnelle. On parle des méthodes traditionnelles de cycles en V, apparues dans le secteur de l'industrie bien avant les projets informatiques; en témoignent les appellations des activités de type maîtrise d'œuvre, maîtrise d'ouvrage, urbaniste ... qu'on retrouve encore aujourd'hui au sein des DSI des grands groupes français.

Avec la digitalisation des 20 dernières années et l'évolution fulgurante des technologies, nous avons assisté à un transfert des processus de production jadis protégés par les murs de l'entreprise vers un espace numérique sans frontières.

L'organisation traditionnelle des entreprises est mise à mal par une concurrence mondiale, native de ce monde digital. Ces nouvelles entreprises ont une approche nouvelle, elles ne considèrent plus le changement comme un processus temporaire entre deux états stables, le changement est dorénavant lui-même un état stable dans le sens où il est permanent. Il doit être accueilli comme une opportunité de progrès au fil du temps, pour être toujours au plus près des besoins des clients ou usagers.

Les règles du marché changent, les barrières à l'entrée ne sont plus aussi protectrices pour les acteurs historiques. La culture Projet est remplacée par la culture Produit, guidée par l'usage, le retour d'expérience et l'amélioration continue. Les consommateurs profitent de ce mouvement de marché et n'hésitent plus à aller vers les produits et les services qui leur apportent régulière-

ment le plus de valeur. La survie des entreprises traditionnelles, y compris celles qui bénéficiaient d'un soutien souverain, dépend ainsi de cette transition organisationnelle. La transformation vers l'agilité est en marche.

### En quoi l'agilité révolutionne les méthodes traditionnelles de gestion des projets ?

L'approche de gestion des projets agiles est une approche empirique. L'apprentissage et l'innovation étant à la base de la culture agile, elles représentent les piliers de toutes les méthodologies qui en découlent. On passe d'un processus de production logicielle en cascade, jalonnée par des livrables documentaires intermédiaires vers un processus de développement itératif qui implique le client du début à la fin du projet, avec des livraisons logicielles intermédiaires.

## Agilité et data, un lien éternel formé bien avant le digital auquel les Directions des Systèmes d'information (DSI) doivent s'adapter.

On passe d'une gestion de projets en silo, où le client exprime son besoin en début de projet et réceptionne la solution en fin de projet; un besoin réputé immuable, sous peine d'avenants contractuels; vers

une gestion de cycles courts permettant, au client de revoir ses besoins après chaque cycle; et d'en modifier le périmètre à sa convenance.

A la différence des méthodes traditionnelles de gestion des projets, la variable d'ajustement dans un projet agile est le périmètre et non plus le coût ni le délai.

### Agilité et data, un lien éternel formé bien avant le digital auquel les Directions des Systèmes d'information (DSI) doivent s'adapter

L'exploitation de la donnée n'a pas attendu le digital ni le Cloud pour être agile. Le besoin d'exploitation agile des données dans l'entreprise existe depuis l'informatisation des processus de gestion. Ce mode de production et de consommation itératif a toujours cohabité avec le processus traditionnel de construction des solutions data.

Les tableurs font partis des technologies qui ont apportées cette souplesse depuis les années 70; si les directions IT ont historiquement conservé la production des

outils d'aide à la décision via les processus traditionnels de gestion des projets, les utilisateurs de leur côté se sont vus dotés d'outils puissants comme Excel, entre autres, qui leur ont permis de construire d'une manière agile et autonome les solutions dont ils ont besoin opérationnellement.

C'est ainsi qu'on peut voir, par exemple dans les salles des marchés, des traders avec des développeurs de proximité qui produisent des tableurs aussi complexes que nécessaires. Une organisation qui peut être qualifiée d'ancêtre aux plateformes de développement agiles.

Fort heureusement, la technologie a évolué sans pour autant détrôner les tableurs historiques. Les concepts de la Business Intelligence sont nés et avec eux le besoin d'une gouvernance centralisée et robuste de la donnée tout en gardant son exploitation aussi agile qu'un tableur. On parle de gisement de données sous la responsabilité des directions informatiques sur lequel les utilisateurs peuvent créer leurs propres outils opérationnels de visualisation et d'exploitation des données.

En même temps, la transformation agile des DSI a permis le transfert de la production de ces outils périphériques vers la DSI avec une relation non plus de client

/ fournisseur mais de partenariat.

La digitalisation, le cloud, les réseaux sociaux et les objets connectés ont généré une masse importante de données dont l'exploitation a été rendue possible avec les technologies de Big Data. Les projets de développement autour de ces technologies sont rarement des projets traditionnels. L'agilité devient une évidence.

A mesure que les gisements de données deviennent trop massifs, que les cas d'usage de l'intelligence artificielle se concrétisent et que les éditeurs offrent des solutions Cloud accessibles aux utilisateurs, on assiste de plus en plus à un retour au développement de proximité avec les problématiques classiques associées. Pour y remédier, de nouveaux concepts voient le jour comme celui de « Data Mesh », « Data as a product » ou « Data Product », où la data devient un produit à part entière avec des services à la demande en y associant autonomie et robustesse, ceci grâce à une architecture orientée services. Une technologie chasse l'autre, la DSI doit encore une fois se réinventer, l'agilité est une des clés ...

## 04

## Pour une approche pragmatique du « Big Data »



### Ilan Ouanounou

Customer goods & services executive

Le pétrole est connu depuis l'antiquité, mais il aura fallu attendre 1859, soit 30 siècles plus tard, pour que les techniques de forage inventées par Edwin Drake permettent son usage à l'échelle industrielle.

Qui n'a jamais entendu l'expression: « Data is the new oil! ». De fait, des zettabytes de données sont produites (ou jaillissent) chaque jour dans tous les secteurs économiques et l'acceptation générale nous amène à imaginer ces données comme une nouvelle ressource économique.

Chacun est en effet conscient de la richesse sous-jacente à l'exploitation de ces données, mais peu d'acteurs économiques en font réellement l'usage de façon massive et industrielle. Encore moins peuvent attester d'un lien direct entre l'usage de la donnée et la performance du business. Il y a certes Amazon, Meta ou Google qui ont fait de la « data » le centre de leur business model et l'essentiel de leurs profits. Ces GAFAM

avaient d'ailleurs conçu cette étape dès leurs origines. Cependant la plupart des entreprises plus traditionnelles, y compris les plus digitalisées d'entre elles, peinent encore à saisir l'opportunité de l'exploitation de la data à grande échelle.

### Quelles sont les raisons de cette inertie?

La donnée n'a aucune valeur en soi. C'est son usage qui en a. Or l'exploitation de la donnée nécessite un effort et des ressources importantes: collecter, organiser, structurer, croiser, modéliser, analyser, mettre en forme, traduire en « insights », convertir en actions, mesurer les résultats, corriger...et ainsi de suite. Du fait de l'intensité de cet effort, le retour sur investissement est rarement évident. Par ailleurs, ce sujet, éminemment technique, est souvent confié aux équipes IT, souvent éloignées des préoccupations quotidiennes du business, et finit par se noyer dans la liste des priorités. Combien de projets titanesques de création de

« datalakes », c'est à dire de nettoyage complet et organisé de la donnée ont échoué après de lourdes dépenses. La donnée est répartie dans de nombreux systèmes informatiques (flux financiers, CRM, etc...) qui n'ont pas été conçus pour communiquer entre eux. La création de ponts entre ces systèmes est complexe.

une liste limitée de « cas d'usage ». C'est à dire une série de questions concrètes à résoudre, directement liées à la performance du business, qui font appel à des séries de données facilement accessibles, et dont le résultat est mesurable. C'est en se limitant à ces quelques cas que les ressources marketing et IT peuvent être

## Il s'agit de poser le sujet de la data autrement, comme on poserait n'importe quelle autre question business: quel objectif l'entreprise poursuit-elle?

Il s'agit donc de poser le sujet de la data autrement, comme on poserait n'importe quelle autre question business: quel objectif l'entreprise poursuit-elle? Comment la data peut-elle contribuer à l'atteinte de cet objectif? Les entreprises poursuivent des objectifs simples: augmenter le chiffre d'affaires et les profits, satisfaire et fidéliser ses clients, engager et développer ses employés, optimiser ses coûts, contribuer positivement à la société.

### Comment la data peut-elle alors concrètement contribuer à ces objectifs, d'une façon à optimiser le rapport entre l'effort et le résultat?

Les exemples les plus réussis sont les plus pragmatiques. Il s'agit de définir

allouées de façon efficace dans la durée, et que les projets d'exploitation de la data ont une chance d'aboutir à des résultats business tangibles.

Il existe deux types principaux de cas d'usage que nous pourrions qualifier ainsi: les « data-powered solutions » et les « business improvement use cases ». Les premiers sont des produits finis résultant d'une combinaison intelligente de données. Les deuxièmes sont des algorithmes d'analyse aboutissant à des conclusions de nature à informer une décision.

Dans le but d'illustrer ces deux types de cas d'usage par des cas réels, prenons l'exemple de la société Edenred. Edenred est une entreprise française, leader mondial de l'argent fléché. Parmi ses ac-

tivités, Edenred opère une plateforme digitale d'intermédiation pour proposer des avantages aux salariés tels que les titres-restaurant, les titres-cadeaux, titres-mobilités, etc.... De ce fait, Edenred établit une connexion digitale entre trois types de parties prenantes: les entreprises clientes (qui octroient les avantages aux salariés), les salariés (qui en bénéficient comme complément de salaire et y ont accès grâce à une carte de paiement et une application mobile), et les partenaires marchands (les commerces

taurant: le trafic de visiteurs, le panier moyen, la part de marché au sein d'une zone de chalandise, des benchmarks concurrentiels, etc... Cela permet au marchand de mesurer son activité et de s'ajuster à sa concurrence et à Edenred d'apporter des services à valeur ajoutée à ses partenaires et de renforcer leurs fidélité. La réalisation d'un tel tableau de bord nécessite la modélisation et le traitement d'un grand nombre de données internes et externes, et une visualisation simplifiée.

## Comment la data peut-elle alors concrètement contribuer à ces objectifs, d'une façon à optimiser le rapport entre l'effort et le résultat?

où cet argent est dépensé). Edenred se rémunère ainsi en prenant une commission de gestion auprès de ses clients et une commission d'apport d'affaires auprès de ses partenaires marchands. Du fait des millions de transactions quotidiennes, Edenred collecte une grande quantité de données, dont l'exploitation pertinente peut être créatrice de valeur

Dans le cas d'Edenred, une « data-powered solution » est par exemple un tableau de bord dynamique, accessible à tous les partenaires marchands, donnant une lecture de leur activité liée au titre-res-

Un deuxième cas d'usage chez Edenred, porte cette fois sur le « business improvement » et traite le sujet du « churn », c'est à dire de la perte de clients. On dit que l'acquisition d'un nouveau client coûte 3 à 5 fois plus cher que sa fidélisation. Alors comment prédire qu'un client a une forte probabilité d'être perdu? Quels signaux avant-coureurs doivent être identifiés? Quelles séries de données doit-on collecter et croiser pour engendrer des alertes? Quelles actions prendre par anticipation sur la base de ces alertes afin d'éviter la perte du client. Ce sont les travaux menés par Edenred pour créer des systèmes

d'alerte pour ses commerciaux, sur la base de la modélisation algorithmique d'un grand nombre de données: la fréquence et le montant des commandes, les taux de défaillance de société par secteur, le nombre d'appels reçus, etc... Ce deuxième « cas d'usage » permet d'optimiser les coûts d'acquisition client en agissant suffisamment tôt sur les motifs

d'insatisfaction facteurs de « churn ».

Si nous devons simplifier le propos, nous pourrions le résumer ainsi: une approche pragmatique de la data, fondée sur des préoccupations business réelles, est le meilleur chemin vers une exploitation rentable des gisements de ce « nouveau pétrole ».

## 05

## La prospective et la donnée: « Foresight with Data »



### Aymar De la Mettrie

Foresight, new Markets,  
Technologies & Innovation

La prospective c'est imaginer l'avenir pour s'y préparer. Cette pratique redouble plus que nécessaire dans un monde incertain où des phénomènes s'accroissent et basculent. Ce qui caractérise ce nouveau monde c'est l'omniprésence de la donnée.

L'injonction « speak with data » nous impose de générer, stocker, échanger, mais surtout sortir des conclusions de la masse de données disponibles. Il n'est plus une semaine sans que l'intelligence artificielle nous impressionne par sa capacité à apprivoiser de grand volumes de données ou même de challenger la créativité humaine par de la génération de vidéos, d'images ou de texte.

Les nouvelles approches de prospective pourraient-elles un jour aller jusque-là? Le « foresight with data » pourrait-il un jour s'appuyer sur de la donnée pour tout anticiper y compris les surprises à venir?

Force est de constater que les réflexions

prospectives en relation internationales ou en intelligence économique, sont souvent très pauvres en simulations. Il n'est pas rare de rester au stade du texte forcément synthétique donc simplifiant, parfois quelques graphiques qui donnent l'impression que l'avenir se trouve dans la droite issue de la prolongation du passé. Ces matériaux sont souvent longs à produire et fournissent une seule analyse moyennée et le décideur ne sait souvent comment faire face aux risques car il ne connaît pas les conséquences de ses choix.

Pourtant, des approches théoriques, des données et des outils existent pour imaginer les avenirs, simuler les différentes hypothèses. La donnée peut être au service de l'analyse et la puissance de la simulation au service de la décision. L'objectif est de fournir une base analytique aux réflexions des analystes et aux décideurs.

Dans le cadre des relations internationales, pour reprendre une phrase clef

de Raphael Chauvancy<sup>1</sup> le « temps des blocs est révolu » dans le jeu des alliances entre pays ou l'on voit apparaître les « connivences fluctuantes » décrites par Bertrand Badie. C'est le temps des « conjonctures fluides » de Michel Dobry. En d'autres termes les relations entre les pays peuvent être analysées comme un ensemble fluctuant et versatile. Un terrain de jeu ou des joueurs vont se déplacer les uns par rapports aux autres, concentrer leurs efforts, s'allier ou s'opposer en fonction de la position de la balle et de l'équipe dans laquelle ils sont.

Pour simuler ces nouveaux comportements il faut trois éléments: premièrement les multiples relations ou les distances avec les autres joueurs, les règles de calcul de décision du joueur et enfin un outil de simulation. La simulation informatique permettra un « wargame » multiacteur basé sur des chiffres et des phénomènes complexes en faisant jouer les différents acteurs.

La modélisation des relations s'appuie sur les 6 essences du politique de J Freund, auxquelles se rajouteront des notions plus qualitatives issues de l'analyse; comme la direction ou l'intention stratégique, les facteurs de puissance, etc.

Pour nourrir ce cadre, on y intègre les données par pays mais surtout les relations entre pays. Ces chiffres sont accessibles: données d'échanges commerciaux, énergétiques, etc. au travers de sites internet (tradeconomics, oecd, etc...). Ces données transactionnelles peuvent être complétées par des notations plus subjectives fournies par des index internationaux (GPI, WGI, FSI, etc..).

L'ensemble de ces données est vaste, il faut sélectionner les données remises à jour de manière périodique, celles qui sont disponibles et consistantes entre elles, c'est à dire les méthodes utilisées pour obtenir les données sont les mêmes quel que soit le pays.

Enfin un outil logiciel produit des simulations et des cartographies afin de nourrir les échanges entre analystes. Il y a différentes familles de logiciels dont les logiciels de simulation multi agents (ex Net Logo<sup>2</sup>). Chaque « agent<sup>3</sup> » suit les règles qu'on lui donne, il est possible de faire jouer un ensemble très vastes d'agents. C'est ainsi que l'on simule le comportement de bancs de poissons, d'essaims d'oiseaux, le comportement de la foule qui sort d'un stade, ou encore l'embouteillage spontané des voitures sur une

autoroute.

Pour la simulation des relations internationales, chaque pays respectera les règles de distance/proximité avec les autres pays, ses facteurs d'attraction ou de répulsion avec d'autres selon le sujet et les mécaniques de décision sur l'appréciation entre pertes et gains.

Plusieurs limites ou défauts surgissent en premier lieu: bien entendu le monde n'est pas réductible à un ensemble de règles mathématiques. Les théories de « l'acteur rationnel<sup>4</sup> » ont montré leurs limites et les contraintes intérieures des pays si elles peuvent aussi être simulées augmentent la complexité du problème. Certains

états « avant » et « après ». Les chiffres permettent aussi de dépasser certains biais. Sortir du prisme égocentré pour ouvrir le domaine des réactions intersubjectives<sup>5</sup>. c'est à dire simuler ce que les autres pourraient faire sans tomber dans la subjectivité de ce que l'analyste souhaiterait qu'ils fassent. La facilité à produire des simulations permet aussi un travail de groupe collaboratif et valorise l'expertise.

Le point clef de la prospective avec ou sans « data » reste le même: la capacité d'analyse des scénarios du futur pour proposer des décisions, et cette capacité reste du ressort de l'homme. La pros-

### **Le point clef de la prospective avec ou sans « data » reste le même: la capacité d'analyse des scénarios du futur pour proposer des décisions.**

évènement ne peuvent pas être modélisables et sont donc impossibles à prédire.

Cependant les gains de cette approche compensent largement ses limites. L'usage de données objectives permet de quantifier des phénomènes et d'offrir un panel de scénarios variés, l'utilisation de représentation graphiques et de simulation permet de montrer de multiples

pective actuelle reste trop cantonnée à un monde d'experts isolés adeptes des points de vue simplifiés alors que les données et les outils informatiques disponibles permettent des simulations nombreuses, complexes et variées. Tout en restant imparfait, l'objectivation mathématique du comportement apporte une fraîcheur aux problèmes complexes et

1 - L'Occident n'est qu'une création des Etats-Unis pour justifier leur leadership en Europe

2 - Net Logo est en accès libre, <https://ccl.northwestern.edu/netlogo/>

3 - Ou « turtle » pour reprendre la terminologie Net Logo

4 - Une excellente synthèse: <https://journals.openedition.org/conflicts/579>

5 - Qui concerne les relations de personne à personne, chaque personne étant considérée du point de vue de sa subjectivité (opinions, croyances, sentiments).

permet de mieux comprendre les phénomènes. Avec l'échange collaboratif de ces variétés il est possible de préparer l'avenir et de confronter des opinions, leurs causes et conséquences probables. Le « Foresight with Data » remet l'analyste au service du décideur dans le tempo d'un monde qui accélère. En effet, rien n'est plus efficace que les multiples préparations aux avènements possibles pour acquérir de la résilience et devenir « antifragile » comme le dit N Taleb depuis 2013.

## 06

## Le Big Data au sein des Institutions financières sous le prisme de la gestion des risques



**Timothé Hervé**

Risk Manager,  
Banque de France

Le Big Data au sein des Institutions financières sous le prisme de la gestion des risques

En l'espace de quelques décennies, les infrastructures de marché et les moyens de paiement ont subi de profondes mutations, combinant exigences strictes en matière de gestion des risques et performances informatiques exponentielles. L'utilisation de plus en plus fréquente des technologies de traitement de données est marquée par le dynamisme de la vague actuelle des nouvelles technologies (Big Data, Blockchain, Cloud, IA) et par l'arrivée de nouveaux acteurs sur le marché.

Dans ce contexte, de constante évolution des marchés, les régulateurs jouent un rôle important en termes de stabilité du système financier, mais également de sécurité et d'efficacité des moyens de paiement et des infrastructures de marché.

À ce titre, nous examinerons les enjeux du big data au travers du prisme de gestion des risques.

### Innovations technologiques :

L'émergence de technologies basées sur le traitement, l'analyse et la transmission permanente et continue de grands volumes de flux, est un levier majeur d'innovation. En effet ces développements constituent un défi considérable pour le secteur financier, notamment pour les infrastructures de marché dans la création de nouveaux services et l'amélioration de l'efficacité des services existants en tirant profit des données.

Quelques exemples de technologies :

- Les fournisseurs de **Cloud** (AWS, Microsoft) propose différents types d'offres dédiées au Big Data (IaaS, PaaS, SaaS) permettant aux utilisateurs de réaliser leur travaux d'analyse et de data visualisation (ex Splunk).
- La **Blockchain** via une technologie complémentaire à celle du big data a permis d'introduire un nouveau moyen de paiement celui des cryptos actifs.

Comme dans beaucoup de secteurs

autres que la finance l'essor du big data représente un enjeu majeur sur le plan commercial via l'agrégation d'informations sur les comptes et services d'initiation de paiement, qui analysent le comportement du client en vue de suggérer des rééquilibrages des transactions et des offres bancaires (carte, découvert, crédit, placement, etc.) adaptées à son profil.

Du côté des commerçants, des solutions similaires permettent d'analyser les comportements des clients afin d'offrir une optimisation de parcours (par exemple,

### Big Data au sein de la gestion des risques:

L'application première de ces technologies dans les services de paiement consiste en l'identification des transactions à risque afin de lutter contre le blanchiment d'argent et le financement du terrorisme (renforcement du dispositifs (LCB-FT)). Le Big Data permet une démarche préventive voire prédictive des risques et non à posteriori ce qui bouleverse la dimension et l'efficacité de la gestion des risques en passant d'une analyse de l'échantillon à un examen

## En Finance, l'essor du Big Data représente un enjeu majeur sur le plan commercial via l'agrégation d'informations sur les comptes et services d'initiation de paiement.

en présélectionnant un mode de paiement en fonction des achats envisagés), ou encore d'améliorer le ciblage qualité du ciblage des campagnes promotionnelles et programmes de fidélité.

Mais ce développement technologique ne se limite pas au plan commercial ou marketing. L'industrie de la finance joue un rôle de précurseur en gestion des risques, et a d'ores et déjà introduit le Big Data dans ses processus.

complet exhaustif des données.

En plus des solutions d'authentification forte déployées par les émetteurs de paiement, la technologie de cotation du risque de transaction peut également être utilisée pour déterminer si une transaction doit être bloquée, suspendue ou exécutée.

Au travers du Big Data, le scoring qui se base sur des scénarios de fraude généralement connus représente un enjeu

considérable dans les processus de gestion des risques et souligne l'intérêt de ces technologies.

Au niveau individuel dans le cadre d'un virement, la prise en compte des données (ex: type de compte à débiter, montant, nouveau compte à créditer ou non crédité, etc), ainsi que les données sur le profil du titulaire du compte et de l'institution permettront d'identifier et de détecter via la donnée des scénarios de risque.

Dans ce sens la directive européenne DSP 2 stipulent 5 critères utilisables pour effectuer des analyses de risques :

- Identification d'un comportement anormal ou de dépenses anormales;
- Détection d'informations inhabituelles sur l'appareil ou le logiciel utilisé;
- Identification d'un virus pendant une session qui a nécessité l'authentification du client;
- Identification d'un scénario de fraude;
- Le titulaire du compte se trouve dans un endroit anormal ou un lieu à haut risque.

Dans un cadre de flux agrégés, les institutions détentrices de comptes peuvent utiliser les informations observées sur

l'ensemble de leurs clients (taux de rejet taux de rejet des prélèvements, bénéficiaires ou bénéficiaires ou destinations inhabituelles pour les virements, etc.). Ce croisement d'informations permet, le cas échéant aux institutions de notifier à certains clients l'apparition de transactions identifiées comme suspectes

### Risques et missions des régulateurs face à ces technologies:

L'innovation peut poser des problèmes en matière de fraude et de sécurité, en raison de sa nature numérique et du cyber environnement dans lequel elle évolue. L'augmentation rapide du nombre d'acteurs impliqués dans les processus financiers et de paiement, la circulation accrue des données personnelles et à la prolifération des "points de défaillance" potentiels sont des éléments justifiant les risques existants liés à ces innovations.

Ces nouveaux risques "cyber" suscitent de vives inquiétudes dans l'ensemble du secteur, y compris pour les infrastructures de marché et les technologies de paiement éprouvées, (paiements par carte en ligne, qui représentent plus des deux tiers de la fraude par carte cartes en France). De plus, les technologies les plus récentes qui n'ont pas encore été testées à grande échelle, telle que la blockchain, sont susceptibles de

créer de nouveaux risques de sécurité qui justifient une surveillance précoce et permanente.

Dans ce contexte, les autorités publiques jouent un rôle clé permettant de profiter pleinement de l'innovation tout en atténuant les menaces qu'elle génère. L'innovation technologique n'est bénéfique pour l'économie dans son ensemble si elle est réalisée dans un environnement sécurisé.

En plus de leur mission de surveillance, les banques centrales de l'Eurosystème et la BCE jouent également le rôle de catalyseur et d'opérateur d'infrastructure de marché. Dans leur rôle de catalyseur ils surveillent les efforts déployés par le secteur dans le développement de nouveaux services et de processus innovants, et apportent leur soutien aux initiatives du marché. En tant qu'opérateurs d'infrastructures de marché, l'Eurosystème et la Banque de France ont lancé un certain nombre de programmes innovants visant à améliorer l'efficacité des infrastructures de marché qu'ils exploitent. Data lake initiative de la Banque de France qui met les nouvelles technologies (Intelligence artificielle, gestion Big Data...) au service d'un système d'informations moderne pour un suivi efficace des effets de la politique monétaire et des risques prudentiels. Cela permet

des analyses plus fines et plus en rapport avec les besoins du moment.

À titre d'exemple l'AMF (Autorité des Marchés Financiers) a mis en place la plateforme ICY, mettant la data au cœur de ses préoccupations et ainsi permettant le développement des capacités d'analyses et de stockage. L'objectif pour les régulateurs est d'identifier par le biais d'algorithmes les anomalies présentes sur les marchés financiers, mais également de participer à la prévention et à la détection de toute suspicion de fraude ou de détournements.

La vague actuelle d'innovations technologiques et l'émergence de nouveaux acteurs participent à la création de nouvelles opportunités pour l'industrie financière, toutefois des risques et des défis spécifiques, au niveau opérationnels, juridiques et financiers inhérents à ces nouvelles technologies sont également présent. Les régulateurs se doivent d'être également à l'affût de ces nouvelles technologies, afin d'observer les opportunités et les risques liés ces nouvelles technologies afin de réaliser un suivi de leur développement et de leurs potentiels tout en assurant la stabilité et le respect des réglementations en vigueur.

## 07

## Données bancaires: la transformation bancaire au cœur du marché unique de la donnée



### François Marchessaux

Senior Partner, Franz Partners,  
Conseil en Stratégie & Management

L'histoire des banques universelles d'aujourd'hui est faite de rapprochements d'entités variées, de marques et de réseaux internationaux initialement développés par métiers dans une logique de proximité client, d'expertise sectorielle ou régionale toujours soumis aux exigences de régulateurs locaux.

artificielle comme facteur d'agilité et de compétitivité de la banque dans un environnement qui pèse sur le bilan des banques.

- à long terme, l'enjeu est le positionnement des banques dans l'économie de la donnée, la mise à disposition d'une donnée de confiance sur le domaine

**Dans cet environnement structurellement fragmenté, le partage de la donnée entre les différents métiers de la banque comme avec les clients et partenaires émerge comme un enjeu clé pour les banques.**

Dans cet environnement structurellement fragmenté, le partage de la donnée entre les différents métiers de la banque comme avec les clients et partenaires émerge comme un enjeu clé pour les banques.

bancaire et les possibles diversifications de la banque dans les échanges de données entre clients provenant d'industries variées.

- à court terme, l'enjeu est le développement du conseil et de l'intelligence

## La banque pilotée par la donnée : ou la généralisation d'un conseil bancaire centré sur l'intelligence artificielle comme facteur de compétitivité

A court terme, les banques universelles espèrent tirer parti de la donnée pour se différencier par le conseil spécifique à chacune de leurs activités.

La remontée des taux directeurs et la reprise de l'inflation dans un contexte de remise en question des grands équilibres géopolitiques et d'aversion au risque croissante des investisseurs face aux aléa (notamment ESG) font peser de grands risques sur le bilan des banques. Le risque

se développe sur les moyens de paiements et la banque doit ouvrir l'accès au détail des comptes de ses clients à des Fintechs spécialisées susceptibles de lui ravir une relation client désormais digitale.

Des quatre grands rôles bancaires: le bailleur, qui fournit les fonds - à crédit ou en fonds propres -; le coffre-fort - qui assure la tenue et la sécurisation des comptes; la diligence qui assure la liquidité et la disponibilité des fonds où se trouve le besoin et le conseil qui oriente vos placements ou vous permet de les liquider au meilleur prix; c'est du conseil que de la plupart des acteurs espèrent leur salut.

## La remontée des taux directeurs et la reprise de l'inflation dans un contexte de remise en question des grands équilibres géopolitiques et d'aversion au risque croissante des investisseurs face aux aléa (notamment ESG) font peser de grands risques sur le bilan des banques.

des obligations augmente alors que la performance des produits dits risqués est grevée par les nouvelles exigences réglementaires et marché.

Encouragée par les dernières directives européennes, une nouvelle concurrence

En effet le conseil est présent dans chaque métier bancaire, ses revenus sont basés sur des commissions qui suivent l'inflation. Le conseil bancaire influence directement la qualité des arbitrages et des portefeuilles clients; enfin, il concourt à la fidélisation client dans un

monde toujours plus digitalisé où pèse le risque de désintermédiation.

Il s'agit de comprendre les comportements, d'analyser et de comparer les performances, d'évaluer le rendement et les risques d'une situation et d'un contexte particulier voir individualisé.

Désormais capable d'identifier de nouveaux comportements clients et de nouveaux segments de clientèle, d'anticiper la valeur à long terme et de fournir des recommandations pertinentes, l'Intelligence Artificielle est le principal levier d'un conseil bancaire performant et généralisé.

Depuis quelques années les banques concentrent leurs premiers efforts sur l'expérience client. Les initiatives de marketing personnalisé et d'assistants virtuels sont largement déployés; mais le niveau de conseil délivré reste encore somme-toute assez limité.

## Au sein des banques universelles, l'évolution de l'accès aux données métiers sur le chemin critique

Le conseil bancaire est naturellement centré sur le client et consommateur de données variées. Au sein des banques universelles, l'évolution de l'accès aux données métiers et la capacité à relier des données de sources diverses sont

actuellement sur le chemin critique du développement de la compétitivité des banques.

La pertinence des recommandations de la banque dépend: de sa capacité d'analyse des préférences individuelles et de l'appétence au risque de chaque client; de la comparaison de comportements individuels et collectifs; de la prise en compte d'une situation financière globale - intégrant des revenus et positions bancaires, dettes en cours et variété des classes d'actifs détenus en portefeuille -; de l'analyse des services consommés et de la mise en relation de ces informations avec des données de marché.

De même que l'intelligence artificielle a besoin de grands volumes de données, la modélisation du risque intègre une information toujours plus variée et les moteurs de prédiction en temps réels un volume croissant d'information. Pour surpasser leurs concurrents, les IA (Intelligence Artificielle) bancaires ont besoin de données d'actualité, fiables et qualifiées, en quantité...

Or, le partage de données bancaires s'est longtemps cantonné aux opérations interbancaires et au Reporting légal ou financier. Ce partage s'est construit souvent à sens unique au fil des réglementations, métier par métier, et majo-

ritairement sous la contrainte. C'est, par exemple, le récent renforcement des exigences de contrôle client (KYC) qui a permis d'établir une première vision Client partagée entre les différents métiers.

Tirés par la recherche d'efficacité opérationnelle et les évolutions réglementaires dans le domaine des paiements (DSP2), les projets d'hébergements mutualisés et d'ouverture aux tiers ont favorisé la mise en place d'infrastructures de partage de l'information (datacenter, cloud-privé, datawarehouse et API...).

### **Toutefois la donnée bancaire reste éclatée, disséminée au sein d'architectures applicatives silotées, pensées par métier, pour une clientèle spécifique dans un contexte réglementaire particulier.**

Les dernières réglementations en matière de respect de la vie privée (GDPR), de maîtrise de la fraude (LCB/CFT) et de maîtrise du risque bancaire (BCBS239) ont nécessité le "lignage" de données sur l'ensemble des métiers bancaires. Elles ont permis aux banques de mieux connaître leurs données et d'appréhender le travail à mener pour les qualifier.

La mise en place de filières transverses spécialisées dans la Donnée et la Cyber-sécurité caractérisées par la nomination de Chief Data Officer (CDO) et de Chief Information Security Officer (CISO) au sein de chaque entité bancaire constituent aussi un premier pas vers une gouvernance du partage des données.

Toutefois la donnée bancaire reste éclatée, disséminée au sein d'architectures applicatives silotées, pensées par métier, pour une clientèle spécifique dans un contexte réglementaire particulier. Les échanges reposent encore sur des systèmes spécialisés, pensés pour un

usage ou un Reporting particulier.

Deux priorités émergent pour qu'un conseil bancaire basé sur l'intelligence artificielle puisse soutenir significativement la compétitivité d'une banque :

- La conception d'architectures d'échanges transverses et d'espaces de données hybrides et partagés pensés pour l'intelligence artificielle (confiance, actualité, qualité et volumétrie) sont

sur le chemin critique du recours à l'intelligence artificielle comme levier de compétitivité.

- Les filières « Données », aujourd'hui centrées sur la conformité, doivent évoluer vers le métier dans une recherche récurrente d'opportunité et assurer la gouvernance de données à valeur ajoutée.

Les banques n'en sont pas si loin, les efforts réalisés ces dernières années leur ont mis le pied à l'étrier.

Toutes, ou presque, s'engagent dans une approche incrémentale tirée par les premiers usages mais centrée sur la donnée et son exploitation par l'intelligence artificielle.

Le recrutement de data-scientistes et la transformation sont engagés depuis plusieurs années dans la plupart des banques. Pour exemple la Société Générale revendique déjà le déploiement de 330 cas d'usages fondés sur l'intelligence artificielle. La banque, qui se veut désormais « pilotée par la donnée », estime à 230 M€ la valeur générée par cette stratégie.

La difficulté sera de tenir le cap dans une période particulièrement mouvementée pour les marchés.

### **La création d'un marché unique de la donnée dans un espace où l'industrie bancaire manipule de grands volumes de données, souvent fortement valorisées**

A long terme, la Commission européenne entend créer un marché où les données pourront circuler au sein de l'UE et entre les secteurs en respectant les dispositions relatives à la protection de la vie privée, au droit de la concurrence avec des règles d'accès claires et équitables.

S'attaquant aux problèmes juridiques, économiques et techniques à l'origine d'une sous-utilisation des données, la création de ce marché unique de la donnée pourrait générer 270 milliards d'euros de PIB supplémentaire pour les États membres d'ici à 2028. En sécurisant les échanges d'information et l'accès à une intelligence artificielle « de confiance », l'UE deviendrait leader d'une société centrée sur les données.

La création de ce marché constitue une opportunité unique dont les banques pourraient s'emparer.

Premièrement, la donnée bancaire est fortement valorisée.

L'émergence d'offres d'open-data, de places de marché dédiées à l'échange de données, le développement du « da-

ta-brokerage » et de Fintechs dont les revenus sont principalement assurés par l'exploitation de données offrent de nouveaux espaces d'observation, plus dynamiques et plus ciblés, sur la valeur avérée de la donnée bancaire. Une donnée sécurisée et qualifiée dont la conformité réglementaire est assurée.

Pour exemple, l'émergence des places de marché dans le darknet, permet de hiérarchiser la valeur de certaines données personnelles. Sans surprise les données bancaires sont sur le podium, aux côtés des données de santé mais loin devant la donnée de consommation fournie par les plateformes internet.

**Les banques sont d'ores-et-déjà d'importants acteurs du marché de la donnée et il serait impensable de les en exclure, cela d'autant plus que, depuis quelques années, le fort développement des banques en ligne et le développement des services associés en font des plateformes numériques sensibles aux évolutions à venir du cadre européen.**

En second lieu, la banque dispose d'une forte expérience des pratiques d'achat de données et d'acquisition d'information - bancaire ou non-bancaire - qu'elle pourrait valoriser dans un marché unique de la donnée.

En effet, les banques intègrent quotidiennement d'importants volumes de données de marché issus de places de marché, d'opérateurs de paiements, d'organismes spécialisés ou de services publics... Les métiers bancaires accèdent à des sources de données de plus en plus diversifiées pour répondre aux nouveaux usages ou offrir de nouvelles expériences clients. Pour exemples: les filiales de leasing, désireuses d'évoluer vers la mobilité, acquièrent des données d'usage des véhicules; les filiales de banques privées accèdent aux bases de données juridiques et aux réseaux sociaux; les banques d'investissement exploitent la donnée de presse et les filiales d'assu-

rances exploitent les données météorologiques issue de tiers étatiques...

Toute banque produit et diffuse quotidiennement de l'information réglementée à destination des instances de supervision financière et d'institutions variées

aux plans régional et international. La banque dispose d'une forte expérience de la dissémination ciblée d'information qualifiée, y compris en temp-réel.

En synthèse, les banques sont de d'ores-et-déjà d'importants acteurs du marché de la donnée et il serait impensable de les en exclure. Cela d'autant plus que depuis quelques années, le fort développement des banques en ligne (i.e. en France +30% de croissance moyenne pondérée en nombre de comptes sur la période 2014 - 2017) et le développement des services associés en font des plateformes numériques sensibles aux évolutions à venir du cadre européen.

### De nouveaux règlements européens qui pourraient inciter les banques à agir

Enfin, si les banques ne sont pas directement visées par les projets de réglementation régissant l'usage des données en Europe, plusieurs d'entre elles pourraient avoir un impact sur les activités bancaires.

- Le Data Act (DA) vise à permettre l'accessibilité des données personnelles d'usage produites par des objets ou services connectés. Or il apparaît qu'à date, les comptes, moyens de paiement électroniques et applications de smart-

phone pourraient entrer dans la définition de 'produit' et de 'services relatifs' inscrits dans le texte. Si tel est le cas, les directives de la Banque Centrale Européenne - et notamment la directive PSD2 - devront faire l'objet de mise à jour sur lesquelles les banques auront intérêt à se positionner.

- Le règlement à l'étude pour déterminer les conditions d'usage de l'IA dans l'Union Européenne (AI Act) prévoit une gestion particulière des systèmes d'IA dits critiques. Les systèmes d'IA utilisés pour évaluer la solvabilité ou l'accès au crédit sont concernés. En sus des obligations d'éthiques et de transparence obligatoires pour tous, ce règlement pourrait imposer aux banques des mesures supplémentaires de déclaration et de contrôle (ex-ante) ainsi que des normes particulières d'entraînement des modèles IA et de documentation des résultats de traitement. La directive 'EU Liability' qui vise à faciliter l'application de la compensation de droit civil pour les dommages causés par les systèmes d'IA. Cette directive pourrait elle aussi concerner les banques, en s'appliquant aux actifs détenus par les clients en cas d'arbitrage automatique à base d'IA.
- Le Data Governance Act (DGA) doit, lui, faciliter le partage de données entre acteurs d'un même secteur éco-

nomique, entre les industries et avec les États membres. La mise en œuvre des concepts d'espaces de données partagés décrit dans ce texte pourrait nécessiter d'importants efforts de normalisation et de standardisation dans le secteur financiers. La réalisation de ces espaces de données exige des chantiers de grande ampleur dans les banques au niveau des architectures comme de la gouvernance des données. Le DGA prévoit également la création de tiers de confiance chargés de recueillir le consentement des clients sur les échanges de données entre parte-

tecture et une gouvernance orientées données.

En sus des initiatives sectorielles d'espace de données partagés impulsées par l'Europe, la déclinaison bancaire des réglementations relatives au marché unique de la donnée pourraient inciter les banques à s'engager. L'expérience qu'elles ont acquise en matière d'échange de données, de cartographies et d'harmonisation des flux comme les investissements qu'elles ont réalisés en matière d'intelligence artificielle, de cybersécurité et de conformité devraient permettre

## **L'intégration progressive de l'intelligence artificielle au conseil présent dans chaque métier bancaire constitue, dans le contexte actuel, un important levier de compétitivité pour les banques.**

naires. Les banques pourraient, là aussi, créer des structures spécialisées s'inspirant du modèle de déclaration des bénéficiaires qu'elles opèrent dans le domaine des virements.

L'intégration progressive de l'intelligence artificielle au conseil présent dans chaque métier bancaire constitue, dans le contexte actuel, un important levier de compétitivité pour les banques. C'est aussi un puissant levier de réorganisation des échanges intra-groupe sur une archi-

aux banques de prendre un certain leadership dans la mise en place du marché unique de la donnée.

Les banques, particulièrement les banques universelles, y ont intérêt : à court terme, pour satisfaire la compétitivité de leurs activités dans un environnement profondément modifié ; à plus long terme aussi parce que la donnée bancaire, par nature qualifiée et sécurisée, devrait y être fortement valorisée.

## 08

## Quelques réflexions sur l'usage de la donnée dans le domaine de l'assurance



### Maxence Jeunesse

Directeur Scientifique de l'Intelligence Artificielle, Covéa

Les assureurs sont par nature des exploitants de données. En effet, l'équilibre financier de l'assureur repose sur l'adéquation entre les primes collectées et les charges liées aux sinistres. Le cœur de la science de la tarification assurantielle repose sur la connaissance du risque. Cette connaissance est obtenue par l'exploitation des données de sinistres (montant, fréquence) mises en relation avec les caractéristiques des risques assurés (segmentation géographique, type de logement ou d'automobile). En 2020, l'indice de disruption (Accenture, 2020), place l'assurance comme un secteur vulnérable. Déjà en 2015, (Hurley, 2015) envisage que la technologie et la possibilité de collecter et traiter une quantité toujours plus grande de données vont s'accompagner d'innovations, comme le développement de l'assurance télématique. Pourtant, en septembre 2022, en un an, l'indice HSCM Public InsurTech a perdu plus de la moitié de sa valeur en un an, par rapport à une

baisse de l'ordre de 30% sur la même période pour le NASDAQ. En étudiant les trajectoires des entreprises innovantes de l'assurance Troy, Valoo ou Metro-mile, les nouveaux usages des données télématiques ne se sont pas traduits systématiquement par des succès dans le domaine de l'assurance à l'usage. On peut arguer à juste titre que la mutualisation des risques s'accommode mal de l'individualisation du tarif promise par la collection et l'utilisation des données de chaque assuré.

La lutte contre la fraude est un des cas d'usage de la donnée les plus répandus. La part des assureurs disant utiliser des modèles prédictifs pour la fraude est passé de 55% à 80% selon une étude de l'éditeur logiciel (SAS, 2016). On voit au travers de la lutte contre la fraude que les assureurs innovent en matière d'exploitation de la donnée.

Selon l'étude (CapGemini, 2021), la gouvernance centralisée de la donnée et de

sa valorisation, la collaboration avec les InsurTech, l'utilisation de l'« open data » et la maîtrise dans la construction de plateformes technologiques ouvertes d'accès de données par API sont les facteurs différenciants dans la réussite des assureurs maîtrisant l'usage de la donnée.

On peut voir d'ailleurs dans les rapports annuels d'assureurs français, que les nouveaux usages de la donnée tournent autour de l'enrichissement par des données externes, à l'instar d'AXA qui innove avec un outil de souscription

tion, comme l'évoque Sarah Bouquerel dans (CNP, 2022). Mais également des courtiers (BOULLE, 2018) qui lancent des plateformes d'évaluation des risques plus ouvertes, mais les réassureurs également comme Munich Re avec flowin.

Sur la thématique de l'open-data, le défi du changement climatique pousse les assureurs français à travailler de concert pour que soient construites des cartographies des aléas naturels, et que ces données soient accessibles comme par exemple le site Géorisques (franceassureurs, 2022)

## **Les initiatives en matière d'utilisation innovante de la donnée ne manquent pas, mais il ne s'agit pas encore de la révolution promise il y a quelques années**

digital enrichi par la data (AXA, 2021). On note aussi une exploitation de données internes non structurées, et un usage plus organisé des données clients pour faire la différence par la personnalisa-

On peut ainsi voir que les initiatives en matière d'utilisation innovante de la donnée ne manquent pas, mais il ne s'agit pas encore de la révolution promise il y a quelques années.

## 09

## Les données de santé et leurs enjeux



### Shirine Benhenda

Experte en Biologie moléculaire,  
données OMICS, PhD

A quoi peuvent servir les données dans le domaine de la santé et de la recherche médicale? Initialement, les données médicales générées permettent de poser un diagnostic, d'établir le pronostic et de définir une stratégie thérapeutique pour un patient. Il existe des recommandations thérapeutiques émanant de l'HAS (Haute Autorité de Santé) afin d'assurer une prise en charge homogène des patients sur le territoire.

Prenons l'exemple du cancer où une dizaine d'années sont nécessaires pour le développement d'un médicament thérapeutique, de la R&D à la mise sur le marché, coûtant autour de 3 milliards de dollars. Le développement de nouvelles thérapies innovantes permet ainsi d'accroître le nombre de traitements disponibles pour les patients atteints d'un cancer et de continuellement améliorer les recommandations de prise en charge thérapeutique bénéfiques pour le patient. Nous vivons un changement de paradigme en cancérologie où l'on ne

traite plus UN unique cancer mais DES cancers voire des cancers uniques. De ce fait, il est essentiel de développer des traitements adaptés à cette multitude de cancers que l'on découvre/apprend à mieux caractériser par l'amélioration des outils de diagnostic (e.g. NGS pour la génomique). Si on considère à minima 25 organes principaux dans l'organisme humain et au minimum 3 sous-types de cancer par organe principal et une dizaine d'années de la conception à la mise sur le marché, il faudrait au minimum 750 ans et au minimum 225 milliards de dollars d'investissement sans compter les échecs d'études cliniques pour produire un traitement adapté à toutes ces situations. A cela, il faut ajouter le développement les traitements associés aux rechutes. Au delà de l'investissement économique que cela représente, c'est surtout le temps nécessaire pour aboutir à des résultats qui est primordial. Et c'est bien ce même temps que les patients atteints d'un cancer n'ont pas.

Y a-t-il un moyen de faire mieux tout en gardant la même qualité et sécurité de la prise en charge des patients et tout en leur permettant de bénéficier des nouvelles innovations thérapeutiques plus rapidement? La réponse à cette question est oui et se trouve dans 5 mots clés principaux: données, structuration, diversité, volume et partage.

modèle reproductible et soutenable dans le temps?

Les données issues des études cliniques sont essentielles pour valider l'efficacité d'un traitement innovant, cependant, elles ne reflètent pas toujours la réalité de la population globale atteinte d'un cancer, celle-ci étant bien plus diversifiée. Les données de la vie réelle

## Peut-on faire mieux tout en gardant la même qualité et sécurité de la prise en charge des patients et en leur permettant de bénéficier des nouvelles innovations thérapeutiques plus rapidement?

La donnée est multi-sources (hôpital, clinique, pharmacie, mutuelle, assurance maladie, objets connectés, réseaux sociaux etc..). Plusieurs questions se posent: quelles sont les données nécessaires à capturer/collecter pour faire des progrès médico scientifiques?

Ces données sont sous différents formats (valeur numérique, valeur sémantique, imagerie, lame, fichier vcf (fichiers électroniques contenant des informations de type nom des patients, adresses), papiers, document digitalisés (e.g. PDF) etc..). Quels sont les moyens à mettre en place pour une collecte rapide, voire instantanée de ces données identifiées? Comment les structurer avec un

sont peu représentées. Les données issues des études cliniques sont pour la plupart structurées pour répondre aux hypothèses émises dans le protocole d'étude. Capturer/collecter et structurer les données des patients atteints d'un cancer en dehors des études cliniques est primordial afin de mieux étudier cette diversité de patients atteints de cancers et d'y adapter les traitements innovants.

Cette diversité de population dans la vie réelle oblige à collecter des données d'un grand nombre de patients atteints de cancers. Plus le volume est grand, plus les résultats statistiques seront précis.

Ce volume, autrement dit le Big Data en

santé soulève deux autres points: le stockage et la durée de stockage des données. Le stockage des données est coûteux et peu écologique. Il existe deux modèles. D'une part centraliser les données signifie dupliquer les données des hôpitaux et les stocker dans une structure unique (peu écologique). Fédérer les données signifie que chaque hôpital conserve les données et les met à disposition via un modèle de partage de données (plus écologique). Combien de temps stocker les données? aujourd'hui la loi prévoit 10 ans pour certaines données de santé.

Pour atteindre ce volume, il est essentiel de collecter, accéder, utiliser, partager et stocker ces données. Ces différentes actions soulèvent des questions réglementaires, législatives et éthique. Qui peut collecter les données? Qui peut accéder et utiliser les données? Ces données anonymisées ou pseudonymisées peuvent-elles être commercialisées? Combien de temps les données doivent-elles être stockées? Quel niveau de sécurité et protection pour ces données de patients? Quel droit de regard le patient a sur ses données? Depuis avril 2017, il est possible d'accéder aux données du SNDS (Système National des Données de Santé) qui regroupe les données de la SNIIRAM, PMSI, CNSA

et des données de l'assurance maladie complémentaire et ce après autorisation de la CNIL<sup>1</sup>.

L'enjeu de la donnée en santé est donc multiple:

- Améliorer le parcours de soin du patient afin d'assurer une prise en charge optimale où que soit cette prise en charge.
- Réaliser des études épidémiologiques en tant réel ce qui pourra faciliter les études en cas de pandémie (e.g. COVID 19).
- Améliorer la R&D grâce aux données en caractérisant mieux les patients qui ne répondent pas aux traitements ou qui rechutent rapidement comparativement à ceux qui répondent aux traitements.
- Améliorer les études cliniques qui aboutiront à moins d'échecs et à une mise sur le marché de nouveaux traitements pour des patients atteints de cancer mieux ciblés.
- Améliorer la prise en charge médico-économique des patients grâce à une connaissance profonde des patients atteints de cancer.
- Développer des outils diagnostiques et des outils prédictifs en intégrant une

<sup>1</sup> <https://www.snds.gouv.fr/SNDS/Documentation-technique>

analyse multi-paramétrique.

Aujourd'hui les données de vie réelle ne sont que très peu collectées car cela a un coût sur le court et moyen terme, cependant, leur exploitation a un impact majeur sur le long terme à savoir sur la qualité de vie des patients, sur les modes de prises en charge hospitalière et sur la caisse d'assurance maladie. L'exploitation de toutes les données de manière structurée et fédérée permettrait d'optimiser le parcours de soin du patient d'un point de vue bien être, diagnostique, thérapeutique, économique et sociétal.

## 10

## Data et Justice: quelles données pour quels enjeux?



### Jessica Vonderscher

Procureure de la République,  
Tribunal judiciaire de Belfort

Parlez de Justice et de Data, évoque immédiatement certaines grandes thématiques. Open Data des décisions de justice, justice prédictive et protections des données personnelles ne sont jamais très loin lorsque l'on aborde ce sujet. De nouveaux acteurs privés viennent régulièrement accélérer le besoin de transformation de la justice sur cette thématique en proposant de nouveaux services à destination tant des professionnels du droit que des citoyens : outils d'aide à la rédaction d'actes juridiques, base de données améliorée, blockchain et intelligence artificielle en tête. Mais de quelles données parle-t-on exactement en évoquant le sujet de la data dans la Justice? Quelles utilisations en faisons-nous? Et surtout, comment pouvons-nous aller vers davantage d'efficacité en alliant simplicité d'utilisation et proximité de l'utilisateur?

**Des données bien plus nombreuses à organiser que l'Open Data des décisions.**

Les données dont disposent les professionnels du droit sont tout d'abord celles issues de l'ensemble des décisions rendues par nos tribunaux et cours. Ces dernières années, plusieurs réformes ont conduit à la mise à disposition progressive de l'ensemble des décisions de justice gratuitement et sous forme électronique sur le site de la Cour de cassation. Cet Open Data n'a pas fait l'unanimité à ses débuts bien que cette ouverture s'inscrive dans un objectif de transparence et de confiance dans l'institution judiciaire. Face à l'évolution des technologies de l'information, elle était désormais indispensable pour garantir l'accès de tous au droit, celui-ci n'étant pas uniquement constitué par les textes mais aussi par la jurisprudence des tribunaux bien que nous ne soyons pas dans un système de Common-Law. La contrepartie de cette diffusion massive de décisions réside dans la nécessaire protection des données personnelles en procédant à l'anonymisation du jugement publié. Les éléments d'identité sont donc occultés

(nom, date de naissance, etc.) ainsi que, le cas échéant, certains éléments de la décision qui permettraient d'identifier les parties.

Les décisions de justice ne sont cependant pas les seules données dont nous disposons notamment en ce qui concerne la matière pénale. Chaque procédure pénale sortant d'un commissariat ou d'une gendarmerie est enregistrée a minima dans un logiciel national. Suivant la décision pénale du magistrat, certaines data pourront être reprises dans d'autre base de données. Il en ira ainsi de la peine prononcée notamment les différentes interdictions faites à la personne condamnée ou encore des modalités d'exécution de ladite peine.

## **Le choix de la multiplication des bases de données peut se comprendre par un souhait de protection accrue des données privées. Le résultat est peut-être moins évident.**

Cassiopée, Parcours, Genesis, Pharos, REDEX, FINIADA, FIJAIS, TAJ, PPN, .... Autant de noms qui laissent, à quelques exceptions près, peu d'indices sur le contenu réel des données qui y figurent.

Ces bases de données sont de deux ordres. Certaines sont des outils de gestion métier regroupant tant des éléments

sur l'identité des personnes mises en cause ou des plaignants que des éléments sur l'infraction commise ou sur la réponse pénale apportée. Ils sont complétés par des outils de suivi de l'activité et de la charge de travail qui doivent permettre d'assurer le pilotage des services.

Les données des outils de gestion métier permettent quotidiennement aux magistrats de disposer des informations utiles pour prendre leur décision au vu des antécédents pénaux des personnes mises en cause et de garantir le suivi de l'exécution des peines prononcées.

In fine, les choses pourraient être assez simples si l'ensemble des données utiles pouvaient se retrouver dans un seul et

même logiciel ou si, a minima, lesdits logiciels pouvaient communiquer entre eux. Cependant, faute d'une véritable vision et d'une stratégie efficiente sur ce sujet ces dernières décennies, l'ensemble de la data à la disposition des personnels de Justice est dispersée dans de très nombreux outils pour certains parfaitement obsolètes.

## **Une fiabilité à améliorer pour davantage d'efficacité.**

Le choix de la multiplication des bases de données peut se comprendre par un souhait de protection accrue des données privées. Le résultat est peut-être moins évident. Peu de citoyens savent en effet que leur nom figure dans l'un ou l'autre fichier géré par le ministère de la Justice ou le ministère de l'Intérieur et ne le découvre bien souvent que de manière fortuite, au cours d'une demande administrative sans lien actuel avec une quelconque infraction.

Par ailleurs, la difficile interconnexion entre les différents logiciels entraînent des reprises de données fastidieuses par l'ensemble des acteurs de la chaîne pénale avec un risque d'erreur accrue et une perte de temps colossale.

Ainsi, lorsqu'un magistrat du parquet est saisi d'un dossier, il doit tout d'abord rentrer plusieurs fois le nom de la personne mise en cause dans plusieurs applicatifs afin d'obtenir l'ensemble des informations utiles pour traiter sa procédure: les précédentes condamnations (CJN), les éventuels antécédents sur le logiciel pénal des juridictions (Cassiopée) avant de vérifier, si nécessaire, l'état de l'exécution des peines en cours (Appi) et, si la personne est incarcérée, sa fiche pénale

(Genesis).

La fiabilité des logiciels collectant nos données n'est pas toujours optimale face au manque de moyens matériels et humains. Il peut ainsi arriver que des informations soient manquantes ou que d'autres ne soient pas mises à jour régulièrement. Cela peut impacter fortement une enquête en cours mais aussi plus généralement le pilotage des services judiciaires. Concernant une enquête en cours, l'absence d'interconnexion ne permet ainsi pas toujours au service d'enquête de connaître directement la dernière adresse de la personne recherchée, dernière adresse qui aura pourtant très vraisemblablement été enregistrée par l'un des logiciels de l'administration pénitentiaire si la personne a déjà été condamnée par le passé.

Dans un rapport de 2018, la Cour des comptes avait déjà donné l'alerte en insistant sur l'insuffisance des outils d'analyse et de suivi et le manque de fiabilité des données disponibles au sein des services judiciaires.

L'efficacité du partage de l'information est pourtant indispensable pour garantir que tous les éléments d'une situation ont bien été pris en compte et améliorer la réactivité des services. Imaginons un instant un outil qui permettrait d'assurer

la véracité des éléments de personnalité présentées au tribunal correctionnel: nous aurions la certitude de l'adresse du prévenu, de sa situation familiale et professionnelle et de ses revenus au lieu de nous fonder uniquement sur ses déclarations. De telles données existent mais dans d'autres fichiers que ceux dont nous disposons et auxquels nous n'avons pas accès à ce jour. Ce partage est techniquement faisable tout en préservant la protection des données dès lors que nous pourrions créer, par exemple, une plateforme sur le modèle du Service national des enquêtes administratives de sécurité (SNEAS) regroupant des personnels des différentes administrations pour établir une liste de données vérifiées dont doit disposer un tribunal pour prendre sa décision.

### Vers une gouvernance renouvelée de la data.

Le pilotage de la data au sein de la Justice doit nécessairement être questionné et amélioré. Cela va d'ailleurs de pair avec la disparation programmée du papier: la numérisation efficiente de nos procédures ne peut se faire qu'avec une gestion de la donnée optimale pour gagner en efficacité et en sécurité juridique tout en garantissant les libertés individuelles.

Plusieurs pistes peuvent être étudiées

comme la mise en place d'un véritable service de la data au sein du ministère de la Justice. Actuellement, tant la conception des logiciels collectant nos données que le traitement et l'analyse de celle-ci sont réalisées par différents services au sein de chaque direction du ministère. Un seul et unique acteur interne au lieu de cet éparpillement aurait pour avantage premier que l'ensemble des services, d'un bout à l'autre de la chaîne pénale, disposeraient des mêmes données et de statistiques cohérentes.

Au-delà de la mise en place de ce service unique de la data, la question se pose de façon similaire au niveau déconcentré. Être magistrat ou directeur d'un établissement pénitentiaire n'est quasiment jamais synonyme de compétences en statistiques. Si nous voulons avoir une donnée fiable et un traitement utile de celle-ci, nous ne pouvons plus faire l'économie de la mise en place de statisticiens a minima au niveau régional et départemental.

L'amélioration du traitement des datas de la Justice constitue ainsi un des enjeux majeurs des prochaines décennies pour apporter une aide à la décision renforcée. Logiciels et algorithmes ne remplace-

déploiement. Une donnée de qualité c'est permettre à tous de retrouver le temps nécessaire pour être au contact des parties, la relation humaine étant le cœur même de nos métiers. La data n'est

## L'amélioration du traitement des datas de la Justice constitue un des enjeux majeurs des prochaines décennies pour apporter une aide à la décision renforcée

ront jamais magistrats et avocats mais ils peuvent indéniablement améliorer leurs conditions de travail si tant est que la technologie mise en œuvre soit de qualité ainsi que l'accompagnement lors de son

ainsi pas tant une finalité en soi qu'un levier pour améliorer la qualité du service public de la Justice et asseoir avec davantage de sérénité l'autorité judiciaire dans ses missions régaliennes.

# 11

## L'intelligence artificielle au service des Finances publiques



### Franck Duval

Administrateur des Finances publiques,  
Directeur adjoint du pôle gestion fiscale,  
DDFiP 92

Le numérique constitue une opportunité pour offrir un meilleur service aux usagers des différents services de la Direction générale des Finances publiques - DGFIP (impôt des particuliers et des professionnels, cadastre, publicité foncière et enregistrement, collectivités territoriales, aide aux entreprises en difficultés, ...).

L'utilisation de l'intelligence artificielle (IA) est un progrès majeur de ces dernières années. Elle est un facteur d'amélioration du service à l'utilisateur et de l'efficacité des politiques publiques.

Le recours à l'analyse des données et à l'IA, techniques basées sur le recouplement des données, l'analyse statistique et l'apprentissage automatique permet un traitement optimal d'un grand volume de données détenues par la DGFIP. L'objectif est d'associer des compétences métiers et informatiques sur une plateforme informatique dédiée, réunissant des données décloisonnées, issues des

applications de la DGFIP, d'autres administrations, de l'open data, de l'échange international d'informations entre administrations, des données des réseaux sociaux, ...

Quelques exemples ci-dessous illustrent cette utilisation de l'analyse des données et de l'IA dans le domaine fiscal.

Dans le cadre du projet « foncier innovant », il est possible de détecter les constructions ou aménagements non déclarés à partir des vues aériennes de l'Institut national de l'information géographique et forestière (IGN). Les algorithmes permettent d'extraire des images aériennes publiques de l'IGN, les contours des immeubles bâtis ainsi que des piscines. Un traitement informatique croisé vérifie, ensuite, à partir des déclarations effectuées par les propriétaires, si les éléments détectés sur les images sont correctement imposés.

L'IA est également mise en œuvre pour détecter d'éventuels comportements

frauduleux concernant les déclarations d'impôts déposées par les entreprises et les particuliers, pour aider à la programmation des contrôles fiscaux et mener des actions de contrôle plus ciblées. Ce dispositif permet de détecter à grande échelle des indices de fraudes en croisant un grand nombre de bases de données (déclarations de revenus et de résultats, fichiers des comptes bancaires, échanges automatiques d'informations entre Etats, réseaux sociaux, caractéristiques des entreprises contrôlées et des redressements effectués, reporting pays par pays pour les entreprises de plus de 750 M€

Galaxie qui permet de visualiser au niveau national, d'une part, des liens existants entre des entités professionnelles (liens de participation) et entre des entités professionnelles et des personnes physiques (dirigeants, associés ou actionnaires), et d'autre part, des éléments de contexte sur la situation patrimoniale et fiscale de ces personnes (revenu fiscal, identité du conjoint, numéro SIREN, catégorie de chiffre d'affaires, impôts et taxes auxquels sont assujettis les entreprises,...). Il permet de passer au crible de nombreuses informations pour lutter contre la fraude fiscale.

### **L'IA et la data science permettent par exemple à la DGFIP de dicter d'éventuels comportements fiscaux frauduleux, de mieux contrôler le paiement des dépenses de l'Etat et de prévenir les difficultés financières des entreprises.**

de CA de leur activité, effectifs, bénéfice et impôt acquitté, ...). Elle permet aussi de modéliser des comportement frauduleux et d'affiner des méthodes de profilage. Le dispositif couvre 200 thématiques pour les risques fiscaux des professionnels et 50 pour ceux relatifs aux particuliers. La part des contrôles ciblés par l'IA est ainsi passée de 13% en 2018 à 50% en 2022.

La DGFIP s'est aussi dotée d'un outil

Les outils d'analyse de données et l'IA ont été également très utiles pendant la crise sanitaire dans le cadre de l'octroi du fonds de solidarité (FDS) aux entreprises afin d'opérer un premier tri dans les demandes reçues, à partir de critères simples, afin de traiter les demandes en très grand nombre dans des délais très courts et de détecter des fraudes manifestes. Les critères de filtrage ont notamment permis de repérer des demandes

multiples ou la discordance des numéros SIREN, de distinguer ce qui relève de l'erreur humaine ou d'une intention frauduleuse.

L'IA et la data science permettent par exemple à la DGFIP de dicter d'éventuels comportements fiscaux frauduleux, de mieux contrôler le paiement des dépenses de l'Etat et de prévenir les difficultés financières des entreprises. Grâce au recours à l'IA, en capitalisant sur l'historique des contrôles passés et des irrégularités réellement détectées par les comptables, la DGFIP automatise la sélection des demandes de paiement les plus à risques et réduit ainsi le coût budgétaire des contrôles et le risque financier pour l'Etat.

La DGFIP a également mis en œuvre le projet « Signaux Faibles » qui repose sur un algorithme capable de cibler les fragilités des entreprises afin de mettre en place des actions d'accompagnement le

plus en amont possible des difficultés financières. Ce dispositif a permis de détecter des entreprises en difficulté et d'en sélectionner certaines d'entre elles pour les contacter et leur proposer un soutien.

En conclusion, l'utilisation de IA a permis des progrès majeurs dans l'exploitation efficace des données. Il ne faut néanmoins pas perdre de vue que les productions issues de l'IA ne sont que des outils d'aides à la décision. Cela suppose donc toujours in fine une intervention humaine pour valider la conclusion qui peut être tirée d'une anomalie détectée. Elle nécessite également des mesures d'accompagnement au changement.

Enfin, il est nécessaire pour une institution publique ou privée de garder la maîtrise du fonctionnement de l'utilisation de l'IA même en cas de recours à des prestataires privés pour certains développements informatiques.

## 12

## Propriété intellectuelle et Data: réflexions d'ordre général



### Henri Michel Reynaud

Avocat et expert Propriété  
Intellectuelle auprès de l'OCDE

La Propriété Intellectuelle est un système permettant de protéger et de valoriser les créations de l'esprit humain par différents droits.

Ces droits sont privatifs et confèrent à leurs titulaires des monopoles sur leur utilisation: marques, brevets, dessins, modèles et récemment noms de domaine. Les titulaires peuvent ainsi générer des revenus en donnant leurs droits en licence, c'est-à-dire en permettant à des tiers partenaires de les utiliser moyennant le paiement d'une redevance.

Le système est donc très protecteur des auteurs, créateurs, inventeurs et responsables qui y investissent et permet surtout d'organiser le processus de création, de rémunération et de valorisation des œuvres de l'esprit au sens large.

Sans ce système, l'utilisation des créations de l'esprit serait simplement et uniquement soumise à la loi de l'offre et de la demande, avec tous les risques et tous les abus inhérents à ce type de situation. Or,

depuis une vingtaine d'années, nos sociétés modernes ont intégré des outils informatiques de plus en plus indispensables à leur fonctionnement, impliquant de partager sur tous types de réseaux tout un tas de données que chaque personne doit divulguer, le plus souvent gratuitement et de façon publique (non anonyme), afin de pouvoir utiliser pleinement toutes les fonctionnalités de ces outils. Ces données appelées le plus souvent de leur nom anglais « data » sont ainsi récoltées par les propriétaires des outils précités, lesquels les revendent à d'autres ou à des institutions publiques, ou encore les donnent en licence, à l'instar des droits de propriété intellectuelle.

Aux côtés de ces données dites personnelles existent les programmes informatiques qui font fonctionner les outils éponymes (applications, progiciels, logiciels SaaS, etc..).

Il existe donc deux types bien distincts de data, à savoir, d'une part, les données

personnelles et, d'autre part, les codes et programmes informatiques.

Toutes ces données sont devenues en l'espace de quelques décennies une sorte de matière première immatérielle (*immaterial commodity*) que de nombreux acteurs de la vie économique monnayent librement, aucun cours ni aucun prix n'étant indiqué à l'avance, aucun marché ni aucune bourse n'existe à ce jour pour fixer le prix des données elles-mêmes, contrairement aux « autres » matières premières.

Enfin, l'utilisation généralisée des smartphones et des ordinateurs a induit le fait de visualiser les objets et/ou des œuvres protégées par des droits de propriété intellectuelle sur de nombreux écrans, ces œuvres et ces objets devenant de plus en plus intangibles à nos sens, alors que leur existence matérielle et physique n'a cependant jamais été altérée.

Les écrans, en particulier ceux de nos smartphones et ordinateurs portables ou fixes, sont devenus la source principale, voire unique, de représentation du réel.

Dans ce contexte, la data joue le rôle de matière première ou de brique constituant le réel, une sorte de code constitutif des objets de la vie courante : meubles,

vêtements, véhicules, objets de décoration ou outils en tout genre. Ce biais cognitif veut que l'on confonde aujourd'hui un objet avec la data qui compose sa représentation à l'écran.

Pour cette raison, une confusion s'est opérée entre « IP » (intellectual property) et « IT » (internet technology), les départements juridiques des cabinets, des entreprises et des institutions ayant parfaitement intégré les deux notions au sein de leurs départements « IP-IT ».

Cependant, il n'y pas de véritables points communs entre l'IP et l'IT, contrairement à ce que la majorité des non initiés suppose, hormis le cas très particulier des noms de domaine, j'y reviendrai.

Contrairement au domaine du commerce et de la libre concurrence, dont la Propriété Intellectuelle constitue une exception, inscrite comme telle dans le Traité de Maastricht<sup>1</sup> par exemple, il n'existe aucun texte à ma connaissance identifiant la propriété intellectuelle ou n'importe quel droit de propriété intellectuelle parmi ceux relatifs à la data.

Au sein du RGPD par exemple, il n'est fait qu'une seule fois référence à la propriété intellectuelle, précisément au considérant 63 du préambule pour bien

indiquer que le droit garanti par le RGPD de pouvoir avoir accès à ses données personnelles ne doit pas porter atteinte aux libertés d'autrui et notamment le secret des affaires ou la propriété intellectuelle « notamment au droit d'auteur protégeant le logiciel ».

Ce texte, loin de construire des passerelles entre data et PI, confirme au contraire l'opposition qui existe en réalité entre les deux notions, en tous cas pour tout ce qui touche aux données personnelles. Il convient donc à ce stade de rappeler de façon claire que la Propriété Intellectuelle ne protège pas les données personnelles.

Toutefois, ainsi qu'il a été dit plus haut, les data regroupent également les codes et les programmes informatiques utilisés par nos appareils électroniques, lesquels sont bien protégés par la propriété intellectuelle.

La protection de ces data particulières, regroupées sous le nom de software a subi une évolution que l'on pourrait qualifier de ratée.

Je ne vais pas ici retracer en détails la genèse de la protection du software mais, en résumé, il convient de comprendre que le logiciel a été protégé par défaut par un droit qui ne correspond ni à sa nature technique ni à la fonction d'ingé-

nier de celui qui le crée.

En effet, en France et dans le reste du monde, le « logiciel » fut d'abord exclu de la protection par le brevet dans les textes gouvernant ce droit de propriété intellectuelle (article 52 (2) et (3) de la Convention sur le Brevet Européen du 05.10.1973 qui précise que « ne sont pas considérées comme des inventions au sens du §1 (qui définit quelles sont les inventions brevetables)

En 1973, ces programmes d'ordinateurs étaient considérés comme des présentations d'informations lesquelles sont également exclues de la brevetabilité dans le même § c) de l'article 52 (2) CBE.

Ce texte constitue la première fracture juridique entre la présentation d'informations (donc de la data) et la Propriété Intellectuelle, qui exclut ainsi de son système la présentation d'informations.

Ce texte perdu aujourd'hui dans la dernière révision de la CBE de novembre 2020 (17ème édition).

Les progrès de la science, notamment informatiques, furent tels dans la seconde moitié du XXème siècle que le « programme d'ordinateur » évolua vers un produit technique complexe, rebaptisé « logiciel ».

Le législateur, conscient que ce logiciel

1 - Articles 34, 35 et 36 du Traité sur le fonctionnement de l'Union Européenne, dans sa dernière version : Traité 2016/C202-01 du 07.06.2016

est avant tout une création de l'esprit, dut donc faire entrer cette création nouvelle dans le système de la Propriété Intellectuelle, sans toutefois pouvoir utiliser le droit de brevet qui lui était pourtant tout naturellement dédié.

C'est ainsi qu'en France par exemple (mais dans la plupart des autres pays membres de l'OMPI également), une loi dérogoire aux lois existantes sur le droit d'auteur incorpora le logiciel dans la Propriété Intellectuelle par le biais du droit d'auteur (Loi du 03.07.1985, aujourd'hui article L112-1 du Code de la Propriété Intellectuelle).

Cette loi définit donc le logiciel comme une œuvre de l'esprit et son créateur comme un auteur, mais la protection de l'œuvre et le régime du créateur de logiciels sont différents de ceux accordés aux autres œuvres artistiques.

Le droit d'auteur ayant vocation à protéger les œuvres artistiques, créées avec un souci esthétique, on s'aperçoit bien que le logiciel et a fortiori le programme d'ordinateur ne rentrent pas naturellement dans cette catégorie.

Avec le développement de l'informatique, les data constituées de logiciels, d'algorithmes et de programmes informatiques voient donc leur protection assurées par un droit d'auteur qui ne leur

correspond pas vraiment.

Cette situation prévaut encore aujourd'hui pour tous types de créations informatiques, quels que soient les métiers et les acteurs des sociétés du secteur, des plus modestes aux plus grandes multinationales (GAFAM).

Devant cet anachronisme juridique, les acteurs de la vie économique comme les pouvoirs publics tentèrent d'apporter des solutions visant à mieux protéger le logiciel et/ou le programme informatique (algorithme, mais aujourd'hui par exemple Blockchain et ses utilisations: cryptoactifs, NFT, etc...).

La première solution fut de contourner la loi sur le brevet et le fameux article 52 CBE. Une jurisprudence très abondante naquit pour tenter de contourner, en Europe, ce texte ô combien embarrassant.

L'initiative vint des Etats-Unis, où la loi sur le brevet (en particulier 35 US Code § 101) est rédigée en termes plus vagues que l'article 52 CBE, ce qui permit notamment à Microsoft de faire breveter son logiciel DOS, qui évoluera ensuite en Windows, dès 1985.

La protection par brevet est directement à l'origine du succès commercial de Microsoft, comme de celui d'IBM et

des autres sociétés du secteur de cette époque, qui, grâce au système des licences évoqué au début de cet article, put contrôler l'utilisation de son logiciel par les tiers et engranger de substantielles redevances.

Tout d'abord bloquées en Europe par la CBE (Convention sur le Brevet Européen), les entreprises américaines cherchèrent donc à trouver une faille ju-

brevets au monde.

« Microsoft »: une revendication relative à un programme (en l'occurrence à des "instructions exécutables par ordinateur") sur un support déchiffrable par ordinateur échappe elle aussi nécessairement à l'exclusion de la brevetabilité visée à l'art. 52(2) CBE.

Une seconde approche de pouvoir protéger des data non personnelles consista à

## Les GAFAM, grand pourvoyeurs et grands utilisateurs de data, figurent parmi les premiers déposants de brevets au monde.

ridique dans l'application de l'article 52 pour pouvoir bénéficier dans nos pays de la protection par brevet de leurs logiciels, bien mieux adaptée à ces produits que le droit d'auteur.

La Chambre de Recours de l'OEB (Office Européen des Brevets) rendit donc de très nombreuses décisions qui permirent, au fil du temps, de vider de sa substance l'article 52 CBE et permettre ainsi, aujourd'hui, de protéger par brevet la quasi-totalité des logiciels<sup>2</sup>.

De nos jours les GAFAM, grand pourvoyeurs et grands utilisateurs de data figurent parmi les premiers déposants de

créer un droit sui generis pour toute nouvelle catégorie identifiable de produits « software ». Cette seconde approche correspondit à l'essor de l'Internet civil à la fin des années 1990. Les batailles juridiques faisant rage autour de la brevetabilité des logiciels, les pouvoirs publics décidèrent de créer des lois nouvelles mieux adaptées aux nouveaux « produits software » plutôt que de faire évoluer les lois existantes jusqu'à les priver de leur substance, comme nous l'avons vu avec la brevetabilité des logiciels.

Ce fut aussi le cas, évoquons-le subrepticement, des bases de données, elles aussi protégées par le droit d'auteur alors que

2 - Encore récemment : Grande Chambre de Recours de l'OEB, décision T424/03 23.02.2006.....

ce droit ne leur correspond pas vraiment.

Les premiers outils software nouveaux à bénéficier de cette approche furent les noms de domaine. Aux Etats-Unis,

tifs, qui lui-même vise en premier lieu les marques, mais également les dénominations sociales ou les noms commerciaux.

Même si l'intégration des noms de

## La deuxième grande catégorie d'outils software à bénéficiaire de cette approche (créer de nouvelles règles spécifiques adaptées aux nouveaux outils) fut l'algorithme.

l'ICANN (autorité propriétaire et gestionnaire de l'Internet civil, tel que nous le connaissons et utilisons) édicta donc un Règlement à portée internationale visant à protéger les noms de domaine de premier niveau (appelés gTLD) gérés par cette institution ainsi qu'à régler les conflits entre les titulaires de ces noms de domaine et les propriétaires de marques antérieures.

Ce Règlement baptisé UDRP (Uniform Dispute Resolution Policy) allait permettre d'enclencher un mouvement mondial visant à harmoniser toutes les législations nationales relatives aux noms de domaine.

En France comme dans la plupart des pays de l'OCDE, les noms de domaine sont aujourd'hui considérés comme des signes distinctifs des entreprises et, ainsi, incorporés au Droit des Signes Distinc-

domaine au sein de la Propriété Intellectuelle fut l'objet d'une construction jurisprudentielle dérivée de l'application de l'UDRP (Uniform Domain-Name Dispute-Resolution Policy), le nom de domaine possède actuellement une définition et une place cohérentes au sein du système de la PI, contrairement au logiciel ou à la base de données pendant des décennies.

La deuxième grande catégorie d'outils software à bénéficiaire de cette approche (créer de nouvelles règles spécifiques adaptées aux nouveaux outils) fut l'algorithme.

Toujours définitivement exclu de la brevetabilité, l'algorithme est aujourd'hui protégé en France par la loi 2018-670 du 30.07.2018 sur le secret des affaires, laquelle transpose en droit interne la Directive Européenne 2016/943 du

07.06.2016. Cette loi rattache cependant la protection de l'algorithme au Droit commercial au sens large et fait ainsi sortir l'algorithme du champ de la Propriété Intellectuelle.

En résumé, l'algorithme est protégé contre la copie en application des règles de la concurrence déloyale et cette copie est constitutive d'une faute, mais elle n'est pas une contrefaçon.

La réflexion logique qui découle de ces approches visant à protéger les nouveaux outils software (le sort de la protection des données personnelles étant fixé, nous

software nouveau n'est pas protégé par la PI, mais pourrait l'être par le Droit commercial et en particulier la loi sur le secret d'affaires.

Cependant, contrairement aux algorithmes dont le caractère non visible et non transparent correspond bien à cet outil juridique, ce n'est pas le cas de la Blockchain dont les caractéristiques existentielles sont justement de n'être ni secrète ni invisible.

Une réflexion devra donc être menée sur la protection juridique de l'outil « Blockchain », le Droit actuel paraissant

## Une réflexion devra donc être menée sur la protection juridique de l'outil « Blockchain », le Droit actuel paraissant mal adapté à cet objectif.

l'avons vu) est donc d'intégrer ou pas ces nouveaux outils dans un corpus législatif existant en Propriété Intellectuelle (principalement droit de brevet) ou de créer de nouvelles règles pouvant permettre de protéger cet outil nouveau par la Propriété Intellectuelle (nom de domaine) ou par d'autres règles (algorithme). Le cas qui vient immédiatement à l'esprit est la Blockchain.

Pour l'instant, il est certain que cet outil

mal adapté à cet objectif.

Enfin, il ne faut pas tomber dans le biais cognitif évoqué plus haut et confondre la data constituée des briques qui composent l'outil software et la data constituée des briques représentant à l'écran une œuvre, un objet ou un produit protégé par un droit de Propriété Intellectuelle.

A cet égard, SARRUT Avocats a récemment fait interdire l'usage d'un site « pi-

cassol.io » permettant à des utilisateurs de miner des NFT contrefaisant certaines œuvres de Pablo Picasso ainsi que la marque « PICASSO » en utilisant les règles de la Propriété Intellectuelle : cela ne signifie pas que la data utilisée sur et par ce site constitue elle-même une propriété intellectuelle.

## 13

## Le Big Data et les mutations du monde du travail



**Gilbert Cette**

Économiste et Professeur à  
NEOMA Business School

Le Big Data a favorisé l'émergence de plateformes numériques qui, dans certaines activités comme la livraison ou le transport urbain, ont pris une place considérable et ont suscité beaucoup de débats et d'inquiétudes concernant l'avenir du salariat. C'est l'un des aboutissements parmi les plus marquants et débattus des évolutions technologiques récentes. Ces évolutions ont facilité la réalisation d'algorithmes sophistiqués, permettant la mise en contact d'un donneur d'ordre et d'un prestataire, organisée par une plateforme numérique qui peut suivre la prestation et les conditions de sa réalisation et pratiquer la facturation du client et la rémunération du prestataire. Uber, dans le transport urbain, est la première plateforme de ce type. Elle n'a été créée qu'en 2010 et fut imitée ensuite par de nombreuses autres plateformes, aussi bien dans le transport urbain que dans la livraison à domicile où Deliveroo est l'un des acteurs parmi les plus connus. D'autres activités de services pourraient être concernées par l'émergence de pres-

tations organisées par des plateformes numériques du même type. On peut par exemple penser à la coiffure mais aussi à de nombreuses prestations de services à domicile, comme certains services de santé, de garde d'enfants, de ménage, ...

Les travailleurs des plateformes de VTC, type UBER, ou de livraison, type Deliveroo, relèvent d'un statut d'indépendants et non d'un statut de salariés. Ils sont fortement subordonnés et pâtissent, en comparaison avec les salariés, d'un important déficit de protections. Ce déficit ne se limite pas au domaine social (chômage, retraite, ...). Il concerne également d'autres domaines comme la rémunération minimale (les salariés sont protégés par un salaire minimum, le SMIC en France, sans équivalent pour les indépendants), le droit au repos et les conditions de travail (par exemple, le droit national et supranational définit des amplitudes maximales au temps de travail des salariés, sans équivalent pour les indépendants), la contestation d'une séparation

qui prend pour eux la forme d'une déconnexion (la séparation à l'initiative de l'employeur prend pour les salariés la forme d'un licenciement dont les motifs peuvent être contestés en Prud'hommes).

Ce déficit de protections des travailleurs indépendants des plateformes comparés aux salariés a nourri deux types d'angoisse.

## Le déficit de protections des travailleurs indépendants des plateformes comparés aux salariés a nourri deux types d'angoisse.

Une première angoisse est quantitative : elle repose sur la crainte que le travail indépendant médiatisé par les plateformes se substitue progressivement au travail salarié. La part, dans l'emploi total, des travailleurs indépendants fortement subordonnés et pâtissant d'un déficit de protections augmenterait au détriment du travail salarié. En d'autres termes, cette angoisse est celle d'une ubérisation croissante du monde du travail, permise par les développements technologiques et le Big Data. Mais l'analyse des données statistiques sur les trois dernières décennies amène, au moins pour l'instant, à rejeter cette crainte. Dans les 39 pays

actuellement membres de l'OCDE, à l'exception de trois d'entre eux, la part de l'emploi indépendant dans l'emploi total est demeurée stable ou a diminué sur les trois dernières décennies. Les trois pays qui font exception sont les Pays-Bas, le Royaume-Uni et la France. Cette et Lopez (2022)<sup>1</sup> ont montré que dans ces trois pays l'augmentation de la part de l'emploi indépendant dans l'emploi total s'explique par la mise en œuvre de

politiques très volontaristes simplifiant la réglementation et accordant des avantages fiscaux à l'emploi indépendant, dès le début des années 2000 (soit bien avant l'émergence des plateformes) aux Pays-Bas et au Royaume-Uni, et à partir de 2008 en France (avec la création du statut d'auto-entrepreneur puis de micro-entrepreneur. Il apparaît donc que l'angoisse quantitative trouve peu de confirmation dans les chiffres.

La seconde angoisse est qualitative : il est difficilement acceptable que des travailleurs, dont la part demeure certes faible, pâtissent structurellement d'un déficit

de protections, comparés aux salariés, alors même qu'ils sont très fortement subordonnés, par l'usage d'algorithmes informatiques facilités par les innovations technologiques et le Big Data, à des plateformes numériques qui ne sont pas juridiquement leurs employeurs.

Les réponses des différents pays à ce déficit de protection sont contrastées en Europe. Un grand nombre de pays, dont ceux de l'Est de l'Europe, demeurent dans une logique de laisser-faire qui maintient le déséquilibre de protections évoqué plus haut. D'autres pays, comme le Royaume-Uni et l'Italie, ont créé une nouvelle catégorie de travailleurs, intermédiaire entre le salarié et l'indépendant. Ce choix complexifie davantage la situation puisqu'une frontière parfois floue entre salarié et indépendant fortement subordonné est remplacée par deux frontières tout aussi floues entre cette nouvelle catégorie et chacune des deux autres. Et les travailleurs relevant de cette nouvelle catégorie subissent toujours un appauvrissement de leur protection, comparativement aux salariés. Un dernier groupe de pays, comme l'Espagne, a décidé de basculer ces travailleurs dans le salariat. Mais un tel choix, préconisé par certains dans le débat public en France, appelle deux remarques. Tout d'abord, dans leur très grande majorité, ces travailleurs

préfèrent le statut d'indépendant à celui de salarié. Le travail indépendant est souvent pour eux l'occasion de mettre un pied sur le marché du travail formel dont ils sont éloignés. C'est un premier effet très positif de l'émergence de ce type de travail. Ensuite, ce basculement fait perdre des formes de flexibilité utiles à ces activités, ce qui peut les brider fortement au détriment du consommateur. Deliveroo a d'ailleurs décidé de cesser ses activités en Espagne. Rappelons que, dans de nombreux pays, l'émergence des VTC a permis de contourner le monopole des taxis qui s'opposaient, parfois avec violence, à toute réforme de leur profession. L'émergence des VTC a ainsi permis de faire cesser, dans le transport urbain privé, le rationnement de la demande par l'offre, ce qui est favorable à l'activité et à l'emploi. C'est un second effet très positif de l'émergence de ce type de travail.

La France a choisi une autre voie, préconisée par Jacques Barthelemy et moi-même (voir par exemple op. cit. 2017), consistant à amener les partenaires sociaux à construire eux-mêmes les normes protectrices adaptées à ces activités. L'autorité des relations sociales des plateformes d'emploi (ARPE), présidée par Bruno Mettling, a ainsi organisé, du 9 au 16 mai 2022, les premières élections professionnelles dans ces activités. Le

1 - Gilbert Cette et Jimmy Lopez (2022): « Promoting self-employment: Does it create more employment and business activity? », Labour, vol 36, pp. 94-114.

taux de participation y a été plus faible qu'aux élections TPE de 2021 (5,4 %) : il est de 1,8 % pour les 84 000 livreurs inscrits et de 3,9 % pour les 39 000 conducteurs de VTC. Mais il s'agit d'une première et la sensibilisation à l'importance de ce scrutin n'était pas encore élevée parmi une population peu familiarisée au vote et, surtout pour les livreurs, très instable. Une nouvelle consultation aura lieu dans deux ans avant la mise en place

La France a donc choisi une approche privilégiant la négociation collective pour construire des normes protectrices, adaptées aux spécificités des plateformes et de leurs travailleurs. Il faut souhaiter un plein succès à cette approche dès les négociations de 2022. Et qu'elle pourra constituer une voie reconnue par la directive sur les plateformes d'emploi dont un projet est en discussion en Europe. Les normes issues de la négo-

### **La France a donc choisi une approche privilégiant la négociation collective pour construire des normes protectrices, adaptées aux spécificités des plateformes et de leurs travailleurs.**

d'un cycle de quatre ans. Dans les deux activités, livraison et VTC, une grande majorité des suffrages s'est portée sur des organisations favorables au statut d'indépendant. Dès la fin de l'année 2022 et après la désignation des représentants des plateformes, les négociations s'amorceront sur les questions difficiles de la rémunération minimale, des conditions et horaires de travail, de la transparence des algorithmes ...

ciation collective sont les plus à même de concilier au mieux protection des travailleurs et efficacité économique. Le Big data a favorisé l'émergence et le développement de certains types d'activités favorables à la croissance et à l'emploi. Né dans la civilisation de l'usine, le droit du travail doit cependant encore s'adapter pour que ses protections s'étendent à ces nouvelles formes d'emploi, sans en brider le dynamisme.

# 14

## Le chef d'entreprise au centre du jeu de la donnée



### Benjamin Meisse

Chargé d'études senior,  
CEGEDIM,  
Lauréat du Trophée K2 "Communication  
et marketing", édition 2017

Le 27 janvier 2022, Geoffroy Roux de Bézieux, Président du MEDEF, résumait sur Thinkerview en quelques mots: "il y a un sujet qui nous pose un problème, c'est la domination des GAFAM. Car toute la croissance de l'économie vient du numérique et nous sommes entre les mains de quelques monopoles mondiaux"<sup>1</sup>.

Ces empires de l'information ont en effet un avantage comparatif, la donnée, pierre angulaire de leur modèle économique, permettant un ciblage publicitaire très précis. Pour Facebook, cela représente 98 % de ses revenus, et c'est dans l'ensemble de ces données qu'ils collectent et possèdent, qu'ils puisent pour maintenir leurs monopoles et s'étendre à de nouveaux marchés. De fait, que les hommes en soient conscients ou non, qu'ils le veuillent ou non, la donnée façonne aujourd'hui le monde économique matériel comme immatériel<sup>2</sup>.

Comment façonne-t-elle l'univers des dirigeants d'entreprise?

En elle-même, une donnée brute n'a aucune valeur. Elle ne l'acquiert que par sa numéricité, sa volumétrie, son interopérabilité et sa fiabilité. C'est précisément ce dernier point qui font des données financières les données les plus prisées par les dirigeants.

Fiables, les données financières le sont notamment grâce à la nomination obligatoire d'un commissaire aux comptes qui certifie que les comptes annuels sont réguliers et sincères et donnent une image fidèle du résultat des opérations de l'exercice écoulé ainsi que de la situation financière et du patrimoine de la personne ou de l'entité à la fin de cet exercice<sup>3</sup>.

En outre, ces données normées sont d'autant plus facilement assimilables par les dirigeants qu'elles sont leur langage courant, qu'elles permettent de quan-

1 - <https://www.youtube.com/watch?v=x6qsHdJNn1E>

2 - Voir à ce propos les ouvrages de Guillaume Pitron: La guerre des métaux rares et L'enfer numérique.

3 - C. com., art. L. 823-9.

tifier le réel et l'immatériel, qu'elles écrivent le passé et permettent parfois d'anticiper l'avenir.

Dès lors, les dirigeants leur accordent leur confiance et peuvent (faire) traiter les données financières pour permettre deux choses : accroître leur faculté d'arbitrage et réduire les risques quant aux stratégies économiques et commerciales à adopter. Si l'arbitrage s'exerce toujours

plus faible que les mentions légales de la plupart des médias d'informations (les omissions, les inexactitudes, les carences, pour ne pas dire les intox) impliquent qu'elles ne peuvent conduire les possesseurs de ces médias à en être responsables. La démarche est ici l'inverse de celle d'un commissaire aux comptes ; l'information n'y est pas certifiée juste mais susceptible d'être erronée, voire fausse.

### **L'enjeu pour le dirigeant est d'optimiser par la donnée et par l'information sa capacité à déterminer le couple risque / rendement de sa stratégie pour l'entreprise.**

sur le temps, en fonction des budgets et des ressources humaines, il se décline sur une grande variété de thèmes : acquisition des ressources, vente de la production, acquisition de capital, croissance externe, etc.

À partir du moment où l'on sort de l'univers chiffré, c'est l'information qui prend le relais dans les grilles de lecture du dirigeant. En elle-même, une information n'a aucune valeur, tout comme la donnée. Sa valeur s'apprécie au regard d'un contexte et de qualités subjectives : sa confidentialité, son utilisabilité, son actualité, etc. Sa fiabilité est sans doute la qualité la plus mise à mal. Celle-ci est d'autant

C'est probablement ce qui fait que l'information s'avère beaucoup plus difficilement fiable que la donnée et donc plus difficilement interprétable. Elle demeure pourtant essentielle aux dirigeants d'entreprise qui consacrent énormément de temps à la recherche d'informations. Ils s'orientent vers ceux qui partagent les mêmes codes qu'eux, la complicité d'un même monde : leur réseau et les collaborateurs proches. Un échange verbal où la qualité de l'information et donc sa crédibilité est apportée par l'interlocuteur, si tant est que le dirigeant de l'entreprise puisse avoir le temps à consacrer à son réseau, chose parfois délicate pour les

patrons de TPE et PME.

L'information contrebalance ainsi l'univers du reporting loin d'être taillé sur mesure mais ayant malgré tout une utilisabilité intrinsèque. L'information est avant tout un domaine plus humain et plus collaboratif que la donnée, mais partage avec cette dernière les mêmes buts : accroître la faculté d'arbitrage et réduire les risques stratégiques. Car l'enjeu pour le dirigeant est d'optimiser par la donnée et par l'information sa capacité à déterminer le couple risque / rendement de sa stratégie pour l'entreprise.

Malgré cela, "les entreprises se rendent vite compte que non seulement leurs perceptions des marchés se révèlent rarement exactes, mais qu'en plus, elles n'ont qu'un pouvoir explicatif assez faible sur l'aboutissement d'une manœuvre concurrentielle"<sup>4</sup>. Témoignage que sans doute un surplus d'informations serait le bienvenu pour les aider à mieux apprécier leurs marchés et leurs concurrents.

## 15

## Les nouveaux métiers du digital: focus sur celui de Social Media Trader



**Yara Furlan**

Trader Social Media,  
Publicis Media

Depuis quelques mois maintenant, j'exerce dans le domaine du marketing digital, un environnement qui ne cesse d'évoluer depuis plusieurs années. Pour être plus précise, je suis Social Media Trader pour une agence média. Connaissez-vous ce métier ?

Avant d'aborder le sujet, revenons aux origines de la publicité digitale. C'est en 1994 qu'apparaît la toute première bannière publicitaire en ligne, un format inédit à cette période. Fort de son succès, les moteurs de recherches décident également de se développer en intégrant des espaces publicitaires à leur plateforme. Par la suite, de nombreux formats publicitaires ont été développés sur Internet, notamment avec l'apparition de Google. Aujourd'hui on considère comme publicité digitale toute publicité apparaissant sur Internet quel que soit le support (ordinateur, mobile, tablette).

Abordons maintenant le sujet des différents canaux publicitaires et plus précisé-

ment celui des réseaux sociaux. Historiquement, les canaux les plus connus sont le « Search » (publicité payante d'un site internet que l'on retrouve dans un moteur de recherche) et le « Display » (espace publicitaire vendu sur internet), mais avec l'évolution des différentes plateformes, d'autres sont apparus dont celui des réseaux sociaux. Avec une utilisation de plus en plus importante, la publicité sur les réseaux sociaux a énormément évolué. En 2022, 58,7% de la population mondiale et 56,4% de la population française utilisent les réseaux sociaux. En termes de dépense publicitaire, la publicité sur les réseaux sociaux est aujourd'hui le 2ème plus grand marché dans la publicité digitale. Selon une estimation, 33% des dépenses publicitaires digitales seront investis dans les médias sociaux d'ici la fin de l'année 2022.

Mais pourquoi autant de budget est-il dépensé sur cette plateforme ? Depuis de nombreuses années, les réseaux sociaux collectent les données des utilisateurs

et analysent leurs comportements en fonction de leurs actions effectuées. Toute personne qui utilise les réseaux sociaux laisse des traces numériques qui sont stockées dans une base de données. Ainsi, que ce soient les centres d'intérêts, les habitudes, les engagements ou encore les données démographiques (localisation, âge, sexe...), tout est enregistré, stocké et analysé. Autant d'informations importantes pour les marques, qui vont leur permettre de mieux cibler leurs

placements publicitaires pour le compte de notre client. Cette mission passe par un accompagnement général du client notamment au niveau des recommandations stratégiques des campagnes, comme la répartition du budget, l'optimisation, le planning, le ciblage ou encore le visuel des créations. Toutes ces recommandations ont pour objectif d'aider le client à avoir une meilleure approche publicitaire afin d'améliorer les résultats des campagnes.

portings afin de voir si des optimisations sont à faire. A la fin de chaque campagne, un bilan est envoyé au client pour qu'il prenne connaissance des performances sur les réseaux sociaux.

Le métier de Social Media Trader évolue continuellement et c'est ce qui le rend intéressant et passionnant. Avec l'engouement qu'il y a autour des réseaux sociaux aujourd'hui il est devenu indispensable pour une marque d'y être présente afin de travailler sur les différents outils que

proposent les plateformes. Ce métier est devenu un incontournable qui ne cesse d'évoluer depuis quelques années et qui tend à devenir de plus en plus stratégique.

Le métier de Social Media Trader, fait partie des nouveaux métiers du digital, et est donc très prisé par les nouvelles générations, toujours plus connectées les unes que les autres, et très apprécié des marques, désireuses de toujours renforcer leur visibilité sur Internet.

## 33% des dépenses publicitaires digitales seront investies dans les médias sociaux d'ici la fin de l'année 2022.

prospects et leurs clients afin de leur proposer des publicités en accord avec leurs besoins. Les marques ont donc un avantage considérable à diffuser leur publicité sur les réseaux sociaux.

Maintenant que le contexte est posé, et que l'univers des médias sociaux est plus clair, il est temps de parler du métier de Social Media Trader. Généralement exercer en agence média, notre principale mission est d'élaborer et de créer des campagnes médias sur les réseaux sociaux en fonction des informations communiquées par le client. En d'autres termes, notre rôle est d'acheter des em-

En tant que Social Media Trader nous intervenons sur les publicités payantes et non pas sur les publicités organiques (non payantes). Cette stratégie nous permet de cibler un marché spécifique, afin de répondre à l'un des trois principaux objectifs d'une campagne média, à savoir, la sensibilisation, la considération ou encore la conversion.

Nous avons également pour rôle d'assurer le bon déroulement des campagnes et de leurs performances sur les différentes plateformes (Meta, Snapchat, Twitter, Tiktok, Pinterest ou encore LinkedIn). Durant et après les campagnes, des analyses sont réalisées via des re-

## 16

## Data et Key Opinion Leader: un pivot de l'influence du nouveau monde?



### Jun Zhou

Entrepreneur, Lecturer & Consultant  
in Chinese Social Media,  
MSIE, HEC Paris

De l'économie du commerce électronique à la politique, le Key Opinion Leader (KOL) a joué un rôle essentiel dans la transmission d'influence numérique, mais un tel outil peut également être une épée à double tranchant selon qui le contrôle.

« flux de communication en deux étapes », introduite pour la première fois par le sociologue Paul Lazarsfeld en 1944. Le sociologue et scientifique de la communication Elihu Katz qui a élaboré la théorie du « flux de communication en deux étapes » a illustré plus tard que le fait que le leader

**Le Key Opinion Leader (KOL) a joué un rôle essentiel dans la transmission d'influence numérique.**

#### Qu'est-ce que KOL ?

KOL est l'abréviation de Key Opinion Leader, quelqu'un qui est démontré comme un influenceur numérique digne de confiance et de respect. Ce mot est couramment utilisé dans le marketing des médias sociaux, décrivant un type de marketing de commerce électronique intégré aux médias sociaux. Le concept de leader d'opinion vient de la théorie du

d'opinion soit influent est lié »

- à la personnification de certaines valeurs (qui on est);
- à la compétence (ce que l'on sait);
- à la localisation sociale stratégique (que l'on connaît) ». <sup>1</sup>

Les KOL travaillent généralement dans leur domaine spécialisé et sur les médias sociaux, et ils partagent des connais-

1 - E. Katz, "The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis," Political Opinion Quarterly, vol. 21, no. 1, pp. 61-78, 1957.

sances ou des conseils basés sur leurs spécialités. Les consommateurs écoutent le KOL parce qu'ils représentent une autorité et une crédibilité forte. Certains pourraient discuter de la différence entre KOL et influenceur. Cependant, le concept clé est le même: l'influence. Dans le dictionnaire Merriam-Webster, l'influence est définie comme « <sup>2</sup> le pouvoir ou la capacité de provoquer un effet de manière indirecte ou intangible; l'acte ou le pouvoir de produire un effet sans exercice apparent de la force ou exercice direct du commandement; l'ingérence corrom-

pue dans l'autorité à des fins personnelles ». Un indicateur commun (pas toujours fiable) pour mesurer l'influence d'un KOL ou d'un influenceur est le nombre de followers. Un abonné sur les médias sociaux est quelqu'un qui s'abonne au flux d'autres comptes pour suivre les mises à jour. En fonction du nombre d'adeptes, ils sont classés en cinq types: Nano-KOL (1K-5K), Micro-KOL (5K-20K), Mid-KOL (20K-100K), Macro-KOL (100K-1M) et Mega-KOL (plus de 1M). Avec ces catégories, le rapport coût/prix moyen par publication varie ex-

ponentiellement. Prenez Instagram par exemple: en 2020, dans le monde entier, le prix minimum par publication pour les Nano est d'environ 10 dollars, pour les Mid environ 30 dollars, mais pour les Macro et Mega KOL, les prix par publication sont respectivement de 160 dollars et 1k dollars. Gardez à l'esprit que sur différentes plateformes de médias sociaux, en raison de la dif-

férence de complexité de production, le prix par publication avec le même type de KOL avec un nombre similaire d'abonnés varie. Sur Youtube, en 2020 dans le monde entier le prix global par post (vidéo) est plus cher que sur Instagram, passant de 60 dollars pour Nano-KOL à 7k dollars pour Mega-KOL.

Average price per post of Instagram influencers worldwide in 2020, by number of followers (in U.S. dollars)

Average price per post of Instagram influencers worldwide 2020, by followers

Number of followers	Average minimum price	Average maximum price
Nano-influencers (1K-5K)	10	41
Micro-influencers (5K-20K)	15	85
Mid-influencers (20K-100K)	33	296
Macro-influencers (100K-1M)	165	1804
Mega-influencers (over 1M)	1118	

Average price per sponsored video on YouTube worldwide in 2020, by number of followers (in U.S. dollars)

Average price per sponsored video YouTube 2020 worldwide, by profile size

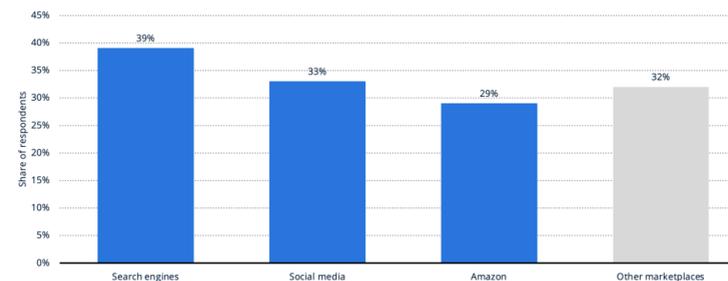
Number of followers	Average minimum price	Average maximum price
Nano-influencers (1K-5K)	60	793
Micro-influencers (5K-20K)	168	1240
Mid-influencers (20K-100k)	389	1859
Macro-influencers (100k-1M)	1306	4061
Mega-influencers (over 1M)	7332	

(Source: Statista\_Influencer Marketing)

2 - "merriam-webster," [Online]. Available: <https://www.merriam-webster.com/dictionary/influence>.

Leading sources of inspiration for online shoppers worldwide as of April 2021

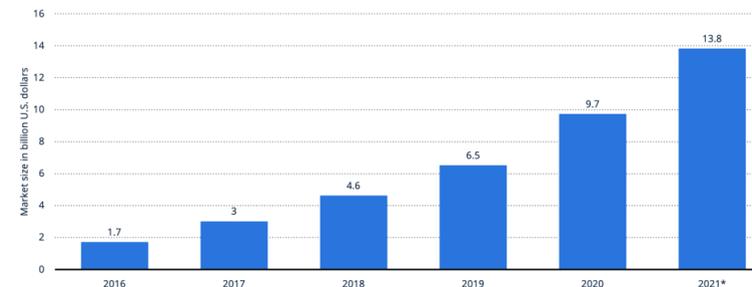
Leading sources of shopping inspiration for digital buyers worldwide 2021



(Source: Statista E-commerce dans le monde)

Influencer marketing market size worldwide from 2016 to 2021 (in billion U.S. dollars)

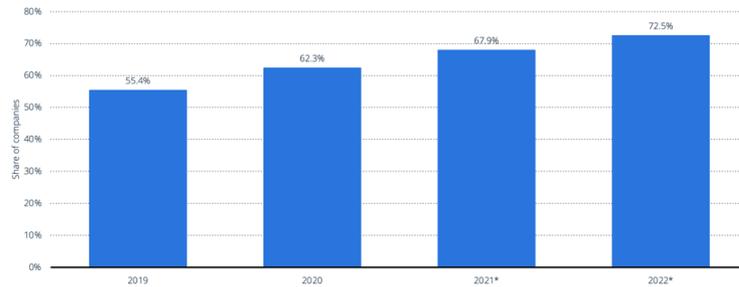
Global influencer marketing value 2016-2021



(Source: Statista Influencer Marketing)

### Influencer marketing usage rate in the United States from 2019 to 2022

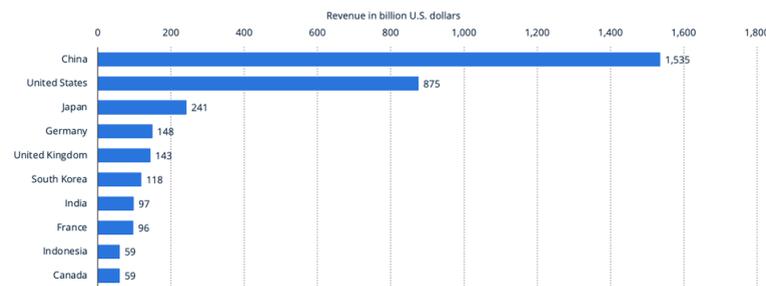
Influencer marketing penetration rate in the U.S. 2019-2022



(Source: taux de pénétration du marketing Statista Influencer aux États-Unis)

### Revenue of e-commerce worldwide in 2022, by country (in billion U.S. dollars)

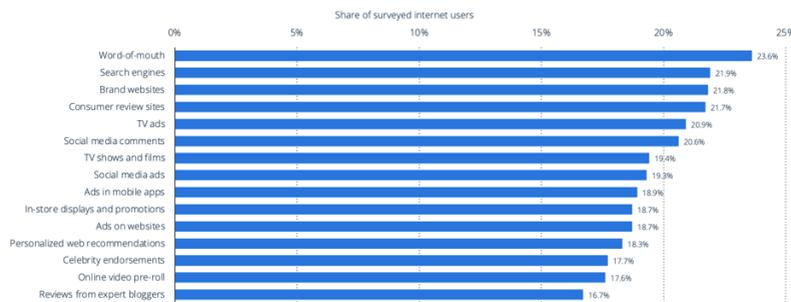
Global revenue of the e-commerce market 2022, by country



(Source: Statista E-commerce en Chine)

### Most-used information channels for finding new brands and products in China as of 3rd quarter 2021

Major brand discovery sources in China Q3 2021



(Source: Statista Sources de découverte de marque en Chine Q3 2021)

### Quel rôle important le KOL a-t-il joué ?

Le KOL s'est développé grâce aux réseaux sociaux et joue aujourd'hui un rôle essentiel dans la prise de décision des consommateurs dans le commerce électronique. Selon Statista, les médias sociaux représentent 33% en tant que deuxième source mondiale d'inspiration pour les achats en ligne en 2021. En 2021, la taille du marché mondial du KOL a triplé sa valeur à 13,8 milliards de dollars par rapport à 2018. Cette année, le taux de pénétration du marketing KOL aux

ment entre les lecteurs/utilisateurs et les créateurs de contenu, rapproche les gens et brise la barrière géographique. Les médias sociaux permettent aux individus de partager leurs opinions dans un environnement plus libre, moins contrôlé et supervisé que les médias traditionnels. L'influence du KOL sur les médias sociaux ne suscitent pas seulement son rôle clé dans l'économie du commerce électronique, mais aussi dans d'autres aspects tels que la politique.

## Le KOL s'est développé grâce aux réseaux sociaux et joue aujourd'hui un rôle essentiel dans la prise de décision des consommateurs dans le commerce électronique.

États-Unis devrait être de 73%. La Chine est de loin le plus grand marché du commerce électronique au monde avec un chiffre d'affaires de 1,5 billion de dollars (4 billions de dollars dans le monde) et le marketing KOL a été classé comme premier choix des internautes chinois pour découvrir des marques, battant les moteurs de recherche, les sites Web de marques et les sites d'avis des consommateurs.

À partir de l'ère du Web 2.0, la nouvelle façon de communiquer avec les blogs et les réseaux sociaux façonne l'engage-

Suivant le concept d'influence, le KOL politique impose un impact fondamental sur le monde. L'ancien président américain Barack Obama est probablement l'un des premiers à adopter les KOL politiques. Démarrant son compte Twitter en 2007, en 2008 il était le plus suivi au monde.<sup>3</sup> Pendant la campagne, Obama a utilisé plus de 15 sites de réseaux sociaux. Il a été le tout premier candidat à la présidence à tirer parti d'une méthode aussi puissante pour publier fréquemment ses activités quotidiennes et communiquer avec ses partisans pour des collectes

3 - J. Han, Networked Information Technologies, Elections, and Politics: Korea and the United States, Lexington Books, 2011.

de fond. Cette identité du KOL sur les réseaux sociaux lui permit de rejoindre les jeunes générations pour voter. Après Barack Obama, Donald Trump est un autre KOL politique célèbre qui fit passer son influence sur les médias sociaux à un autre niveau. Avant l'élection, beaucoup de gens connaissent Donald Trump pour sa célèbre série télévisée « *The Apprentice* » qui avait débuté en 2004 et s'était terminée en 2017. Il fut l'hôte de ce show pendant quatorze saisons avant de se présenter à la présidence. Trump a été le premier président américain à ne pas avoir d'expérience militaire ou politique. Sans aucun doute, il savait très bien comment fonctionnaient les médias et l'influence. Il rejoint Twitter en 2009 pour atteindre 88,7 millions d'abonnés, publiant environ 57 000 tweets au total.<sup>4</sup> Suite à une journée choquante dans l'histoire américaine – la prise d'assaut du Capitole des États-Unis, Trump fut banni de Twitter, Facebook et d'autres plateformes de médias sociaux. Une telle interdiction montre au monde à quel point l'influence en ligne peut être destructrice.

### Avantages et inconvénients

Les avantages et les inconvénients de la communication KOL concernent la façon de tirer parti des avantages et des inconvénients des médias sociaux en tant qu'outil de transmission d'influence.

#### Avantages – une nouvelle façon de transmettre l'influence, de façon radicale et sans frontières.

Les médias traditionnels offrent une communication unidirectionnelle tandis que les médias sociaux offrent une communication bidirectionnelle et elle est plus rapide et plus radicale grâce aux fonctions de simple clic de partage et de republication. Bien qu'il soit difficile de trouver un moyen unanime de mesurer l'influence accordée aux différences de mécanisme pour les partages et les points de vue sur les différentes plateformes de médias sociaux, il est toujours plus rentable de transmettre une influence sur les médias sociaux que sur les médias traditionnels. Dans les réseaux sociaux, la facilité de diffuser l'information d'un utilisateur à un autre permet un effet de chaîne d'influence plus rapide. Ainsi, le KOL, avec n'importe quel contexte connecte facilement les gens très rapidement avec un large public. Cette synergie

crée un mouvement massif de personnes partageant les mêmes idées. L'impact est radical et sans frontières. Imaginez qu'il y a des milliers d'années, si Platon voulait transmettre ses pensées d'Ouest en Est, il aurait fallu des décennies pour que des milliers de messagers traversent les pays et les régions. Maintenant, une lettre prend environ une semaine à deux semaines pour voyager de la France à la Chine. Cependant, un influenceur qui est bien connecté au sein des réseaux sociaux (on parle de « communautés ») peut envoyer un message n'importe où et n'importe quand et en un clic. Ce réseau de médias sociaux rend le monde plus plat et plus transparent, permettant au KOL d'étendre son influence non seulement localement mais aussi globalement.

#### Inconvénients – une méthode facile de manipulation

La transmission d'influence sur les médias sociaux est très (trop?) puissante. Il est également destructeur pour le KOL de tirer parti de son influence pour diffuser de fausses informations, induire en erreur les adeptes et manipuler le public.

Prenons l'exemple d'une pomme. Si seulement quelques personnes disent qu'elle est noire, vous penserez probablement

qu'ils sont fous. Mais si des milliers, des centaines de milliers de personnes disent qu'elle est blanche? Quelle sera votre réponse? Beaucoup de gens penseraient qu'il est évident qu'elle est rouge et ils ne seront pas d'accord sur la réponse, que ce soit blanc ou noir. Cependant, une expérience réelle réalisée dans les années 1950 nous a montré le contraire. C'est l'expérience bien connue de conformité d'Asch.<sup>5</sup> Cette expérience a démontré que les gens pensent qu'ils peuvent résister à la pression du groupe, surtout quand ils savent que le groupe a tort alors qu'en fait ils sont étonnamment sensibles à la conformité et choisissent les mauvaises réponses afin de s'intégrer au groupe. Maintenant, nous devons nous poser la même question pour les informations qui ont été partagées des milliers de fois en ligne. La croyons-nous parce que beaucoup de gens adhèrent à cette opinion? Ou la croyons-nous parce qu'il vient de quelqu'un qui est dépeint comme puissant, influent et légitime – quelqu'un qui est un KOL?

### Future KOL

La principale caractéristique de l'influence peut être résumé comme la volonté de changer la façon dont les gens pensent et agissent. Les finalités peuvent

4 - A. MADHANI and J. COLVIN, "A farewell to @realDonaldTrump, gone after 57,000 tweets," 9 January 2021. [Online]. Available: <https://apnews.com/article/twitter-donald-trump-ban-cea450b1f12f4ceb8984972a120018d5>.

5 - K. Cherry, "The Most Famous Social Psychology Experiments Ever Performed," 14 December 2020. [Online]. Available: <https://www.verywellmind.com/famous-social-psychology-experiments-2795667>.

être économiques, liées au partage de connaissances ou à des fins de contrôle de l'information. De plus en plus de personnes d'horizons divers essaieront d'être KOL et utiliseront leur influence pour servir ces objectifs. Au cours des dix dernières années, nous pouvons clairement voir le nombre croissant de célébrités, de politiciens, d'hommes d'affaires et de nombreuses autres personnes de différentes professions devenir des KOL. La compétition pour attirer l'attention du public et maintenir sa domination deviendra plus féroce que jamais. Dès lors, une question se pose: vivons-nous dans un monde « post-vérité »<sup>6</sup>? Le dictionnaire Oxford a défini la « post-vérité »<sup>7</sup> comme « se rapportant des circonstances dans lesquelles des faits objectifs ont moins d'influence sur la formation de l'opinion publique que les appels à l'émotion et aux croyances personnelles ». L'une des tactiques clés pour la création de contenu sur les réseaux sociaux consiste à créer une résonance émotionnelle avec le public et les abonnés. Il n'y a qu'une mince ligne entre les deux intentions qui sont sous-tendues: une connexion authentique et une manipulation.

À l'ère actuelle du Web 4.0, les données sont fournies pour signifier que l'intelli-

gence artificielle et l'apprentissage automatique ont fusionné dans notre vie quotidienne pour l'innovation et l'intelligence collectives. Pourtant, ces technologies ont également été utilisées dans les plateformes de médias sociaux pour construire des opinions dans une direction souhaitée. Il est tout à fait possible de créer un faux compte (un robot) KOL et des millions de faux abonnés. Prenez TikTok par exemple, au premier trimestre 2022, les faux comptes représentent plus de la moitié du total des comptes supprimés. Alors, comment savez-vous que le compte que vous suivez est la vraie personne et non un personnage ou un robot? Le KOL virtuel qui est un personnage numérique créé à l'aide d'un logiciel et géré par une entreprise n'est pas nouveau dans le monde des médias sociaux. Le concept KOL virtuel peut être facilement étendu à une personne dont les gens ne peuvent pas identifier s'il est réel ou quelle est sa véritable intention.

Thomas L. Friedman, trois fois lauréat du prix Pulitzer, dans son livre "The World Is Flat: A Brief History of the Twenty-First Century" a introduit l'idée d'"horizontalisation" - "Nous partons d'un monde où la valeur a été créée presque exclusivement dans des silos verticaux et contrôlé,

Number of TikTok accounts removed from the platform from 3rd quarter 2020 to 1st quarter 2022, by reason (in millions)

TikTok: account removed 2020-2022, by reason



(Source: Statista\_Social Networks)

du haut vers le bas, vers un monde où la valeur sera de plus en plus créée par la connexion et la collaboration, horizontalement. Nous passons d'un modèle vertical de création de valeur à un modèle. La société organisée verticalement peut également être vue comme une pyramide de « relations de pouvoir d'oppression et de résistance »<sup>8</sup>. La société organisée horizontalement est davantage vue comme un ensemble de groupes divers se rassemblant autour d'une culture commune, d'une religion, ou de diverses idéologies (ethniques, politiques, etc.). Dans ce contexte et, outre le nombre croissant

de KOL classiques, on peut se demander s'il y aura plus de KOL représentant ces groupes?

On ne peut aborder le futur des KOL sans se poser les questions suivantes: en tant que lecteurs et utilisateurs des réseaux sociaux, comment pouvons-nous porter nos propres jugements sur les informations massives alimentées par les KOL? Comment pouvons-nous encourager des KOL plus « éthiques » et tirer parti de leur influence? Enfin, qui jouera le rôle de régulateur sur les réseaux sociaux et comment?

6 - H. McQuilkin and M. Chakrabarti, "Part IV: Are We Living In A Post-Truth World?," 27 February 2020. [Online]. Available: <https://www.wbur.org/onpoint/2020/02/27/part-iv-post-truth>.

7 - "Word of the Year 2016," Oxford University Press, [Online]. Available: <https://languages.oup.com/word-of-the-year/2016/>.

8 - M. A. Martínez, "Vertical and horizontal relations: What is 'society' about?," 28 May 2018. [Online]. Available: <https://www.miguelangelmartinez.net/?Vertical-and-horizontal-relations-What-is-society-about>.

## 17

## Data visualisation & Data Storytelling: quand les compétences techniques ne suffisent plus pour exploiter ses données



**Yann Levy**

Data Analyst,  
Expert BI

**Croissance des volumes de données: pourquoi parle-t-on désormais « d'or numérique » ?**

Chaque jour, 2 500 000 000 000 000 (17 zéros!) bytes de données sous toutes les formes (interactions, déplacements, transactions...) sont générés. Soit l'équivalent d'une feuille Excel complète pour chaque être humain, chaque jour. On estime que 90% de l'intégralité des données produites l'ont été au cours des deux dernières années.

- **Compréhension de son activité:** suivi d'indicateurs financiers et non financiers, des tendances par produit ou zone géographique, de la saisonnalité...

- **Compréhension de ses clients:** exploitation de données qualitatives, algorithmes de classification pour segmenter ses clients, analyse des comportements d'achat...

- **Prévisions:** projections de séries temporelles (chiffre d'affaires, commandes...), régressions...

**On estime que 90% de l'intégralité des données produites l'ont été au cours des deux dernières années.**

Pour les organisations, l'exploitation de ces données devient un avantage concurrentiel. En effet, l'organisation qui sait collecter, mettre en qualité, analyser et exploiter ces données bénéficie d'atouts considérables:

Or, l'Humain ne peut tirer de conclusions à partir de données brutes. C'est pour cela que des outils de Data Visualisation sont utilisés pour analyser les données et présenter ces analyses sous forme de visuels qui serviront de supports de com-

munication.

### L'avènement des solutions de data visualisation : comment visualiser et comprendre ses données ?

Les recherches poussées en matière de data visualisation sont très récentes. Cependant, le besoin d'utiliser des visuels pour comprendre des données et en tirer des conclusions remonte lui à plusieurs siècles. L'exemple du Dr John Snow est souvent étudié. Dès 1854, il est convaincu que le choléra se transmet par l'eau.

Il va donc faire apparaître les cas sur une carte de Londres et démontrer une corrélation entre le nombre de cas et la proximité géographique avec une pompe à eau.

L'objectif n'est alors pas uniquement de présenter les données de façon intelligible, mais bien de présenter les données afin qu'elles permettent une prise de décision éclairée et efficace. Dante Vitagliano (fondateur de Trailmapper) définit la data visualisation comme :



Source : [www.theguardian.com](http://www.theguardian.com) – chaque trait noir représente un malade et la concentration autour de la pompe saute aux yeux immédiatement

**“ (...) le langage de la prise de décisions. Les bons graphiques transmettent efficacement les informations. Les excellents graphiques communiquent les informations et améliorent la prise de décision.”**

Les usages sont presque infinis aujourd'hui : du monde du sport, à la médecine, en passant par l'art, le journalisme et l'éducation. Toutefois, les entreprises demeurent les plus grandes utilisatrices de rapports de data visualisation. A leur échelle, la data visualisation est considérée comme la dernière étape d'une stratégie data bien plus globale (collecte de données, ERP, data lakes & data warehouses, CRM, solutions de Business Intelligence...).

Malgré toutes les avancées techniques dans le domaine, Scott Berinato (journaliste à Harvard Business Review) constate en 2020 que les entreprises ne parviennent pas à tirer des conclusions à partir des données collectées.

**Et si visualiser les données n'était pas suffisant pour prendre de meilleures décisions ?**

**L'émergence du Data Storytelling : et s'il était possible d'aller encore plus loin dans la communication avec les données ?**

La démocratisation des outils de data visualisation est très récente. Les grands outils du marché ont commencé à faire leur apparition au milieu des années 2000 (2003 pour Tableau, 2015 pour Power BI et 2016 pour Google Data Studio). Avant cette période, la réalisation de graphiques était réservée à des scientifiques ou à des profils très techniques.

Alors que les avancées technologiques ont redéfini les normes, pour beaucoup l'exercice de data visualisation se concentre sur la création de graphiques dans un outil dédié. Or, cette approche néglige l'histoire sous-jacente. C'est pour cela que le **Data Storytelling est apparu** : présenter des données au sein d'une structure narrative.

L'ouvrage fondateur sur le sujet « Storytelling with Data » a été publié en 2014 par Cole Nussbaumer Knaflic. La thèse générale invite le lecteur à **replacer les données dans un contexte en plus d'utiliser des visuels clairs et à adapter le message à son audience** (durée, niveau de vue, format, moyen de communication par exemple). On en revient à l'objectif de base de la data visualisation : communi-

quer efficacement et aider à prendre la bonne décision au bon moment.

Pourquoi c'est efficace? Depuis la nuit des temps, les Hommes ont toujours communiqué par des histoires. C'est pourquoi des histoires bien construites seront bien plus marquantes qu'une accumulation de faits et de chiffres. Aussi, les histoires sont capables de faire appel aux émotions de l'audience quand la neuroscience a prouvé que la plupart des décisions étaient guidées par les émotions plutôt que par la logique.

Alors, quelles sont les solutions à notre disposition pour s'assurer de l'impact de notre communication ? Comment être sûr que le message a été transmis et compris ? Si l'on en revient à l'article de Scott Berinato, le journaliste préconise six compétences clés à retrouver dans les équipes projet dont l'objectif final est de

communiquer le résultats d'analyses de données :

1. Préparation des données
2. Analyse des données
3. Expertise métier
4. Gestion de projet
5. Graphisme
6. Storytelling

La solution vient d'une diversité des compétences et donc des profils. L'analyse des données, aussi complexe soit-elle ne présente aucune valeur si elle ne répond pas à une problématique métier bien identifiée et si elle n'est pas présentée de manière à convaincre et/ou faire agir son audience.

**« La présentation efficace des faits est aux données ce que la cathédrale est à ses fondations » (Willard Brinton, 1914)**

## 18

## Big Data et cybersécurité



## Thibaut Heckmann

Officier de Gendarmerie, Senior Associated Researcher at Centre de Recherche de la Gendarmerie and ENS Paris-Ulm, PhD (ENS-Ulm, RHUL, Cambridge'18)  
Lauréat du Trophée K2 Cyber en 2020

Dans la plupart des cas, les cyberattaques ciblent les systèmes d'information et, plus précisément, le sous-système technique qui comprend à la fois des technologies matérielles (hardware), logicielles (software) ainsi que des équipements de télécommunication. Ces attaques ont pour objectif de voler des données, de les détruire ou de les modifier, et d'altérer le fonctionnement normal des systèmes. La donnée numérique étant devenue « l'or noir » de notre société, son architecture doit être à même de garantir sa confidentialité, son intégrité et sa disponibilité en temps réel.

A l'instar des objets volés, les données piratées peuvent se monnayer au prix fort sur le darknet, dans la mesure où elles sont susceptibles de donner lieu à diverses escroqueries, telles que de l'espionnage, du chantage, du vol d'identité ou des atteintes à la réputation. Cependant, le stade de la seule atteinte à la réputation a été dépassé et nous assistons actuellement à une prise de conscience

collective concernant le caractère vital et le risque « physique » dont est porteuse la donnée numérique.

Par son caractère dématérialisé, universel et sans frontières, les cyberattaques visant la donnée touchent un large spectre de cibles dont tout utilisateur de systèmes numériques peut en être la victime, qu'il s'agisse de particuliers, d'organismes privés ou publics. S'agissant des particuliers, les attaques sont rarement personnalisées et consistent le plus souvent à envoyer un grand nombre de requêtes dans le but de maximiser la probabilité de succès de l'attaque. Néanmoins, les citoyens demeurent assez peu sensibilisés à ce genre de problématiques, ce qui les rend d'autant plus vulnérables. Leur principal souci est la non-divulgence de leurs données personnelles. Les rançons sont souvent peu onéreuses pour inciter les victimes à payer afin de récupérer les informations qui les concernent. Concernant les organismes privés et publics, les attaques sont

beaucoup plus ciblées et le montant des rançons plus important. Le plus souvent, la nature disruptive de l'attaque pousse les organismes privés à régler la rançon afin de limiter les pertes financières consécutives à un arrêt partiel ou total de l'activité. Pour ce qui est des structures publiques, la rupture de l'activité peut entraîner des conséquences beaucoup plus dramatiques car dépassant simplement les simples aspects monétaires et touchant directement à la vie humaine ou à la santé, lorsqu'il s'agit d'établissements médicalisés. La règle officiellement affichée étant de ne jamais payer la rançon, deux risques majeurs sont alors identi-

tragiques illustrations de ces problématiques, notamment avec le cas des attaques par rançongiciel visant des hôpitaux. En général, ces attaques procèdent par une intrusion dans un système vulnérable (peu ou pas sécurisé) donnant un accès privilégié à des données médicales sensibles, ceci dans le but de les chiffrer et d'en réaliser ensuite une copie totale. Cette intrusion entraîne inévitablement une perte de confiance dans le système et dans l'intégrité de la donnée (risque de falsification des résultats d'examens, de modification d'ordonnances ou de surdosage médicamenteux). Le risque vital qui s'ensuit est non négligeable.

de porter atteinte aux données les plus sensibles, de façon à obtenir un paiement le plus rapide possible. Les hôpitaux en constituent le meilleur exemple. Demain, les cybercriminels n'hésiteront pas à s'attaquer aux systèmes de traitement

de l'eau potable, aux infrastructures d'énergie (telles les centrales nucléaires), aux sites militaires, aéroportuaires, portuaires ou ferroviaires, avec toutes les conséquences que l'on peut imaginer.

## **La sécurisation des systèmes numériques n'est plus seulement un enjeu de protection des données personnelles, mais devient un sujet majeur en ce qu'il concerne désormais la préservation de la vie humaine**

fiés : le premier est de ne jamais pouvoir récupérer les données (devenues impossibles à déchiffrer), le second porte sur la perte de confiance dans l'ensemble des données stockées (ignorance de l'ampleur de l'attaque, maintien frauduleux et continu dans le système, modification, copie ou altération des données).

L'actualité a récemment fourni de

Il s'avère donc nécessaire de sensibiliser largement à ce genre de risques à travers une formation et une information à tous les niveaux. La sécurisation des systèmes numériques n'est plus seulement un enjeu de protection des données personnelles, mais devient un sujet majeur en ce qu'il concerne désormais directement la préservation de la vie humaine.

Les attaquants n'auront en effet de cesse

## 19

**Biais, confidentialité,  
RGPD****Jill-Jênn Vie**

Chercheur à l'INRIA  
Member of the European Commission Expert  
Group on AI & Data in Education & Training

Les données d'humains, couplées avec de l'apprentissage statistique, ont la possibilité de révolutionner des domaines tels que l'éducation ou la santé: comprendre et améliorer comment les humains apprennent, ou comment ils survivent à une maladie. Cependant, c'est compromis par le fait que les données sont par leur nature même très sensibles, et que des fuites de données ont dissuadé l'opinion publique.

Mais c'est toujours dans l'intérêt des institutions d'ouvrir un maximum de données, afin de les partager avec un plus grand nombre d'experts. La recherche en systèmes de recommandation a beaucoup bénéficié de l'ouverture des jeux de données Movielens et du prix Netflix. Grâce à la mission Etalab du gouvernement, la France est le deuxième pays au monde en matière d'open data, entre la Corée et le Japon. Même le code de l'impôt ou bien la plateforme Parcoursup ont ouvert leurs algorithmes et des statistiques agrégées au niveau des ins-

titutions sont même disponibles pour Parcoursup. Afin de créer des services pour recommander un parcours universitaire à des utilisateurs, il faudrait avoir des données plus riches; il y a toujours un compromis à trouver entre utilité des données et confidentialité des participants.

C'est également vrai pour des institutions décentralisées: pour mener des études médicales qui ont du sens, il faut un échantillon représentatif de la population. Mais certains hôpitaux peuvent avoir un échantillon biaisé de la société, ainsi une méthode statistique fonctionnant dans un hôpital peut ne pas convenir à un autre hôpital. C'est pourquoi des hôpitaux souhaiteraient partager leurs données, tout en respectant la confidentialité de leurs utilisateurs, afin d'avoir un meilleur échantillon de la population et d'ajuster leurs découvertes.

Lorsqu'on ouvre un jeu de données, la méthode la plus courante pour enlever

les données sensibles et suivre le RGPD est la pseudonymisation. Mais ce n'est pas suffisant pour garantir l'anonymat des utilisateurs<sup>1</sup>. Même lorsqu'on n'ouvre qu'une partie du jeu de données, certains participants peuvent être ré-identifiés avec probabilité proche de 1. Le RGPD nous encourage à ne collecter

que des données minimales afin qu'un service fonctionne. Mais pour garantir que les décisions de nos algorithmes ne discriminent pas un échantillon de la population par rapport à un autre, il faut connaître notre population autant que possible, il y a donc là un paradoxe.

**Lorsqu'on ouvre un jeu de données, la méthode la plus courante pour enlever les données sensibles et suivre le RGPD est la pseudonymisation. Mais ce n'est pas suffisant pour garantir l'anonymat des utilisateurs.**

<sup>1</sup> [https://edps.europa.eu/system/files/2021-04/21-04-27\\_aepd-edps\\_anonymisation\\_en\\_5.pdf](https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf)

## 20

## Comment l'utilisation d'un matériel dédié au rendu graphique a permis la troisième vague de l'intelligence artificielle et ses applications concrètes?



### Jean-Baptiste Harry

HPC & AI Solution Architect & pre sales EMEA, NEC Europe

#### Un peu d'histoire de l'intelligence artificielle

L'essor de l'intelligence artificielle (IA) comme nous la connaissons aujourd'hui a été permise par l'arrivée à maturité technique de plusieurs disciplines. Celles-ci vont des mathématiques, du hardware et de ses configurations, et du software appliqué aux domaines métiers. L'objet est ici de se concentrer sur le domaine du hardware. En effet celui-ci est une des clefs qui a permis de faire passer cette discipline à des domaines d'application concrets; tant chez les utilisateurs de tous les jours que chez les entreprises. Cette avancée a permis le démarrage de la troisième vague de l'intelligence artificielle.

D'un point de vue historique le développement de l'intelligence artificielle s'est fait en trois vagues majeures. La première vague de l'intelligence artificielle démarre dans les années 50. C'est le début des premiers ordinateurs dans

les centres de recherche et la conceptualisation informatique du neurone par McCulloch et Pitts. Par manque de puissance informatique, de mémoire, de capacité de stockage sur ces ordinateurs de plusieurs tonnes, les choix de recherche se portent vers l'intelligence artificielle symbolique. Ce paradigme d'IA, basé sur des règles compréhensibles par l'homme, fut ainsi l'axe de recherche prédominant jusqu'au début des années 1990.

La seconde vague a commencé fin des années 1980. Celle-ci fut permise par l'amélioration des puissances de calcul disponible permettant d'utiliser les neurones en réseau afin d'aborder la résolution de problèmes plus complexes. Mais l'avancée majeure de cette vague fut l'introduction de l'algorithme de rétro-propagation par Yann Le Cun. Cette technique mathématique de correction des pondérations du réseau a permis la mise en place de l'automatisation de la phase d'apprentissage telle qu'on la connaît aujourd'hui.

Enfin la troisième vague de l'IA, telle qu'on la connaît aujourd'hui, date du début des années 2010. Le redémarrage des recherches est dû principalement à l'apport de puissance de calcul amené par les Graphics Processing Units (GPU). En effet, ces puces initialement dédiées au graphisme permettent de faire du calcul tensoriel efficacement et de lancer des tâches en parallèle. L'exploitation de ces caractéristiques tant dans le domaine de l'IA que dans d'autres domaines scientifiques a généré un gain très important de puissance de calcul. L'accélération considérable des temps d'apprentissage a permis d'affiner les paramétrages des réseaux et d'en tester facilement de nouveaux plus demandeur en puissance de calcul. Les débuts se firent sur des

sance de calcul d'un point de vue applicatif les réseaux de neurones se sont diversifiés et améliorés. Alors qu'au début, principalement axés sur l'analyse d'image, ils permettent aujourd'hui de faire des chabots, de la recommandation de produits, de l'analyse de tests et de l'analyse de sentiments. Actuellement les applications basées sur de l'IA sont des « workflow » enchainant plusieurs réseaux de neurones spécifiques complexes. On arrive aussi à de très grands réseaux de neurones entraînés pour la résolution de tâches multiples. Mais ceux-ci nécessitent des temps longs d'entraînement (apprentissage), toujours plus de ressources coûteuses à l'achat (matériel informatique) et énormément d'électricité pour cette mise au point.

## Le modèle de référence GTP (Generative Pre-trained Transformer) d'open AI cofondé par Elon Musk correspond à des milliards de paramètres à calibrer.

cartes GPU dédiées au graphisme mais s'en est suivi le développement de puces GPU Nvidia principalement destinées au calcul à haute performance (HPC) et à l'IA pour arriver au matériel que l'on connaît actuellement, principalement dominé par Nvidia.

En parallèle de l'augmentation de puis-

### L'évolution actuelle de l'intelligence artificielle

Un des axes d'amélioration de ces réseaux de neurones a été la complexification de ces réseaux pour répondre à l'augmentation des jeux de données de validation. La conséquence fut des graphes plus importants, des tenseurs plus importants et

plus nombreux donc un besoin de puissance de calcul plus important.

A titre d'exemple le modèle de référence GTP (Generative Pre-trained Transformer) d'open AI cofondé par Elon Musk correspond à des milliards de paramètres à calibrer.

Certes ces applications offrent des possibilités formidables mais le besoin de ressources de calcul est énorme. L'utilisation de nouvelles puces efficaces pour le calcul tensoriel fut donc un des éléments clef pour le développement de l'IA. Aujourd'hui Nvidia principal fournisseur de carte GPU est le leader incontesté de ce marché et a permis en rendant le calcul tensoriel accessible l'essor des applications d'intelligence artificielle. En dix ans l'action NVIDIA a eut une croissance de 7000%.

### Quelles puces pour faire de l'IA ?

L'évolution des modèles vers les réseaux de neurone de plus en plus important en taille oriente le marché vers des puces de calcul de plus en plus puissantes. Sur ces modèles les calculs sont principalement une suite de calculs tensoriels et l'efficacité sur ces opérations est un critère essentiel pour la performance des applications d'IA comme pour les codes de calcul. Depuis des nombreuses années

les constructeurs développent des bibliothèques optimisées pour le traitement de ces opérations sur leurs puces. Les constructeurs équipent maintenant leurs puces de cœurs de calcul dédiés à ces opérations comme les GPU Nvidia (A30, A100, H100) avec cœurs de calculs optimisés et dédiés seulement à ce type d'opération (tensor core).

Ainsi les GPU de Nvidia sont les puces qui dominent ce marché du matériel pour l'intelligence artificielle.

### Des puces mais pas que ...

Quand on parle de hardware pour l'intelligence artificielle on pense souvent aux puces mais de nombreux autres éléments ont révolutionné le domaine de l'IA. Dans ce cadre de calcul haute performance et hautement parallèle, il est essentiel de comprendre que les accélérations sont le plus souvent limitées par les éléments les plus lents. Ainsi pour totalement bénéficier des accélérations des puces de calculs, il faut pouvoir leur emmener les données à calculer d'une manière efficace, de pouvoir écrire les résultats en mémoire etc. C'est le même principe qu'une chaîne de production.

Les points suivant introduisent les principaux éléments qui ont permis et/ou ont évolués grâce et avec l'essor de l'IA:

- **Les serveurs**: développement et optimisation de nouveaux serveurs pour l'utilisation des accélérateurs par les constructeurs comme NEC, Atos, DELL, HPE, IBM ... Ces acteurs conçoivent et pilotent la construction de centres de calculs.
- **Le stockage**: l'intelligence artificielle a besoin de données pour être entraîné donc stockage, de stocker les paramètres des modèles (700 GB pour GTP3). Ces besoins de stockage sont de deux types. Le stockage local près des puces de calcul sur les serveurs (training ou inférence) avec les disques à mémoire flash ou la RAM et le stockage centralisé sur des baies de calcul hautes performances.
- **Les réseaux rapides faibles latences et hautes performances**: ces réseaux déjà présents dans le HPC permettent sur les supercalculateurs d'échanger de gros ou petits volumes de données avec des latences de l'ordre de la micro seconde avec des débits de 100 à 400 GB/s. L'importance de ces réseaux est telle que Nvidia a acquis Mellanox, leader et spécialiste mondial de ce secteur. C'est actuellement un composant en pénurie retardant de nombreux projets.
- Indirectement les télécommunications via la 5G et les besoins important

de transferts rapides de données

- **Le edge computing**: calculer hors du data center à l'emplacement physique de l'utilisateur ou de la donnée ou sur un centre intermédiaire proche. Cela peut se faire sur le terminal final comme un téléphone, des machines-outils, des voitures ... Cette branche du calcul a aussi amené le développement de puces dédiés à l'inférence.
- **La partie exploitation**: conteneurisation et exploitation as a service avec des solutions comme docker et kubernetes qui ont eu un impact dans de nombreux domaines d'application de l'informatique du web en passant par l'IA jusqu'au domaine scientifique des codes de calculs. Le déploiement par conteneur permet de mettre à disposition du métier des suites logicielles construites dans des environnements performants et validés avec un simple clic.
- **La partie logicielle**: pour être performante la couche logicielle doit être optimisée pour le matériel sur lequel elle sera exécutée. Ces étapes nécessitent des optimisations de haut niveau, spécifiques aux langages, aux matériels, aux librairies constructeurs et sont coûteuses en temps et ressources. Pour pallier ces contraintes, Google et Facebook ont respectivement dévelop-

pé deux des Frameworks open source majeurs du domaine TensorFlow et Pytorch. Ces couches logicielles permettent ainsi de concevoir facilement des applications avec des connaissances métier et de réseaux de neurones.

- **La collecte des données** dans le web, dans les objets connectés...

sont prometteurs mais ne sont encore murs. Meta (Facebook) propose maintenant des serveurs complets conçus en interne pour entrer dans leur Metavers (vente clef en main).

La partie « matériel » est un des facteurs importants de l'IA car sans matériel pas de calculs. L'embargo pour le marché

## La course à la puissance et son coût ont amené les GAFAM et le chinois BADU à concevoir leurs propres puces dédiées à l'intelligence artificielle.

### La place importante des GAFAM

Il existe ainsi une amélioration constante du matériel pour l'intelligence artificielle. Cette course à la puissance et son coût ont amené les GAFAM (Apple, Google, Facebook, Tesla, AWS) et le chinois BADU à concevoir leurs propres puces dédiées à l'intelligence artificielle. Il est à noter que ces puces sont produites pour un usage interne (sauf dans le cas d'Amazon Web Services) afin de répondre aux besoins de ces entreprises et de garder leur avantage concurrentiel. Les nouveaux axes de recherche d'intelligence artificielle et d'informatique quantique

Chinois montre l'importance stratégique du matériel dans ce domaine. Concernant l'évolution de l'IA plusieurs questions se posent :

- Le futur rôle des GAFAM / BADU / TESLA qui développent leur propres puces et les principales couches logicielles du domaine (TF, Pytorch...).
- L'impact énergétique et financier, la hausse du cout de l'énergie vont-ils réorienter les choix technologiques?
- Quel sera l'impact écologique des data centers?

# 21

## La révolution de la data pourrait être à plusieurs vitesses : les dirigeants, salariés et citoyens doivent en être conscients



### Anu Kuhanathan

Professeur associé, NEOMA Business School, co-directeur du MSc Finance & Big Data

La data est l'un des grands enjeux des années à venir. Une course technologique s'est engagée depuis bientôt une décennie. Les entreprises et les institutions publiques se sont lancées dans des projets ambitieux pour tirer parti de l'immensité des données qui sont produites, disponibles et peuvent désormais être exploitées afin d'optimiser l'expérience des clients, la productivité des travailleurs, la rentabilité du capital, le bien-être des usagers et des citoyens etc. Cette course a parfois tendance à ressembler à une ruée vers l'or - anarchique, conflictuelle, pleine de vendeurs d'huile de serpents. Toutefois, une chose est certaine : cette course à la data porte en elle le germe de nombreuses sources d'inégalités. Ces inégalités peuvent être "classées" dans trois grandes catégories.

Tout d'abord, le déploiement de la data peut exacerber l'inégalité entre Nations. En effet, le grand public aura certainement noté le débat récent sur la question de la "souveraineté numérique". Ce débat

n'est que le reflet de la prise de conscience par certains Etats, notamment en Europe des inégalités naissantes. Les géants du numérique américains contrôlent et amassent des quantités de données pharaoniques sur les citoyens/consommateurs du monde entier, le déploiement des satellites basse altitude est dominé par les acteurs américains et les Chinois vont très certainement contrôler les infrastructures des futurs réseaux 5G. En outre, chez ces deux superpuissances, la réglementation sur l'usage et la protection des données est relativement faible, ce qui confère aux entreprises et institutions de ces pays, un avantage important dans l'exploitation des données. Alors qu'en Europe, nous sommes avant tout des clients et pourvoyeurs de données, avec une réglementation nettement plus restrictive. Il ne s'agit pas ici de faire le procès de la régulation européenne, qui a ses mérites, mais ce déséquilibre entraînera des avantages économiques, industriels voire diplomatiques à certains pays sur d'autres.

La seconde source d'inégalité venant de la data concerne les entreprises. En effet, les grandes entreprises ont investi ces dernières années dans des *data labs*, acheté des données externes, mis en place des processus pour mieux collecter et stocker leurs données etc. Or comme nous l'a révélé la crise du Covid-19, les

grassement et se permettent de recruter des *data scientist* en leur offrant des rémunérations que beaucoup de PME ne pourraient se permettre de supporter. En somme, il s'agit d'un microcosme bien particulier dont les pratiques ne peuvent être généralisées à l'ensemble des entreprises.

## Une chose est certaine : cette course à la data porte en elle le germe de nombreuses sources d'inégalités

petites entreprises sont à la traîne sur ces sujets digitaux. Les TPE-PME ont pour beaucoup manqué le virage digital, alors la data... Si nos plus petites entreprises, qui, rappelons-le, représentent l'essentiel de l'emploi, se retrouvent désavantagées face aux grandes entreprises dont elles sont souvent les fournisseurs, leurs modèles économiques, leurs rentabilités et leurs activités pourraient être mis sous pression. Un contre-argument au déséquilibre que je pointe ici serait de mentionner les *start-ups*, qui sont des petites structures innovantes et dont beaucoup cherchent à vivre d'un modèle économique fondé sur la data. Néanmoins, il est utile de rappeler que la majorité de ces entreprises ne soufflent pas leur 4e bougie et que celles qui prospèrent en apparence sont souvent financées

Enfin, la data pourrait exacerber les inégalités entre individus (consommateurs, citoyens, usagers). En effet, les modèles de Machine-Learning utilisent les données passées pour "apprendre" et faire des prévisions ou prendre des décisions pour le futur. Sauf que voilà, le passé reflète aussi nos biais du passé. Si une banque, n'octroyait que peu de crédit aux résidents d'un certain code postal (au hasard, de Seine Saint-Denis, par exemple) et qu'elle décide de développer un modèle de décision automatique d'octroi de crédit, il est possible que le modèle discrimine contre les résidents de cette zone géographique, perpétuant une inégalité ancrée dans le passé (encore une fois, il n'est pas question ici de débattre du bien-fondé des décisions passées). Autre exemple, lors de mes fonctions passées, j'ai travaillé avec

une équipe de *data scientist* sur un modèle de pricing automatique pour une compagnie d'assurance. L'objectif de l'entreprise était d'avoir une prime personnalisée en fonction du client - véritable Graal pour tout assureur. L'entreprise avait acquis des données bancaires, des données sur les achats en ligne etc. pour mener à bien le projet. Après les premiers tests, on s'est rendu compte que les modèles discriminaient systématiquement les femmes. Nous avons donc retiré la variable « Gender » des modèles. Pas de chance, les modèles discriminaient de nouveau les femmes, cette fois, ils avaient détecté la caractéristique du genre à travers les

comportements d'achats (transaction chez une enseigne de lingerie, un parfumeur, un esthéticien etc.). Face à ces problèmes multiples, d'ordre réglementaire mais aussi éthique le projet a finalement été mis en "sommeil". Voilà, le type de dilemme auquel les entreprises auront à faire face. Les données et les modèles ne feront que perpétuer nos biais, ce qui pourrait renforcer les inégalités économiques et sociales voire même être une entrave à l'égalité des chances - imaginez si les sélections scolaires ou d'embauche embarquaient automatiquement ce type de biais.

## 22

## La souveraineté numérique au fondement de notre performance nationale



### Franck DeCloquement

Expert en intelligence stratégique. Membre d'honneur du CEPS et de la CyberTaskForce. Enseignant. Spécialiste Cyber



### Aurélie Luttrin

Présidente,  
Eokosmo

La souveraineté numérique, souvent invoquée mais finalement très mal comprise et encore moins appliquée, est pourtant l'épicentre de notre sécurité nationale, mais aussi l'un des piliers essentiels de notre performance collective en tant que Nation.

**Pourquoi la question de notre souveraineté numérique est-elle si essentielle au fond, au point d'être actuellement très âprement débattue ?**

Parce que l'enjeu sous-jacent demeure la capacité pour une Nation démocratique

**La souveraineté numérique [...] est l'épicentre de notre sécurité nationale mais aussi l'un des piliers essentiels de notre performance collective en tant que Nation.**

Pour une entreprise, un citoyen ou un État stratégique et souverain, elle se définit comme le fait d'être en capacité de disposer de la pleine maîtrise de ses données. Et ceci suppose aussi de savoir se défendre juridiquement, économiquement et technologiquement d'autres États fournisseurs – ou entreprises étrangères – pour le stockage, la captation et l'exploitation des dites data.

comme la nôtre de rester libre et autonome face aux visées inquisitrices et prédatrices d'autres acteurs mondiaux.

Depuis que nous sommes rentrés – peu ou prou – dans une 4ème révolution industrielle (celle du numérique), le monde qui nous entoure et ses modalités de fonctionnements ont changé. « *Il faut rendre l'humanité attentive aux grands bou-*

leversements que la quatrième révolution industrielle va provoquer. La quatrième révolution industrielle bouleverse notre société dans ses fondements » (Klaus Schwab, Fondateur du World Economic Forum, 8 Janvier 2016, Le Temps), « Avec [cette] révolution industrielle, le pouvoir réside désormais dans la détention de la donnée et non plus dans la détention du capital » (La transformation numérique, s'adapter ou disparaître, Thomas M. Siebel)

**Cette donnée, qui est souvent présentée comme le nouvel or noir du 21ème siècle, quelle est-elle au juste ? Pourquoi est-elle si convoitée, tel l'anneau du pouvoir ?**

Chaque jour, avec nos smartphones, nos ordinateurs, nos objets connectés, l'ensemble des capteurs déployés dans nos entreprises, dans nos réseaux, dans nos lieux d'habitation, nous générons à l'échelle mondiale plus de 2,5 trillions d'octets de données. Quand nous découvrons que nous disposons de trois fois plus d'objets connectés (IOT) que d'êtres humains sur la surface de notre planète<sup>1</sup>, nous réalisons très vite que notre monde physique voit peu à peu advenir l'émergence de son « jumeau », tel un véritable double numérique. Ainsi, tout comme

nos bâtiments connectés produisent des données retraçant leur mode de fonctionnement, chaque être être humain connecté produit un avatar numérique qui duplique au sens propre sa vie réelle, avec à la clef l'accès à ses plus intimes secrets et signatures comportementales. Elon Musk, fondateur de TESLA, prédisait lui-même que, « dans 25 ans, il y aura plus de nous dans le Cloud que dans notre corps ».

**Mais une donnée en soi n'a pas de réelle valeur, seule la manière de l'exploiter va être source de richesse et de pouvoir.**

Les premiers à l'avoir compris seraient les États-Unis et la Chine, et dans leurs sillages, leurs très nombreux chevaux de Troie et proxy : les GAMMA (Google, Apple, Meta, Microsoft et Amazon), et les BHATX (Baidu, Huawei, Alibaba, Tencent, Xiaomi). En effet, tous ces mastodontes géopolitiques ont très vite saisi que les pouvoirs politique et économique résidaient désormais dans la maîtrise des jumeaux numériques : maîtrisez les données des hôpitaux, des entreprises, des citoyens, des administrations, des services publics, c'est maîtriser l'adversaire géopolitique et le concurrent économique. L'annexion subreptice ne

sera pas territoriale cette fois, mais l'effet produit sera tout aussi catastrophique.

**C'est ainsi que les entreprises technologiques sont très vite devenues les bras armés des États dont elles dépendent, et se sont lancées à cet effet dans la construction rapide d'immenses monopoles plurisectoriels pour annexer le terrain à bas bruit, de manière déterminante et particulièrement discrète.**

nouvelle vision politique et d'imposer des monopoles sectoriels avec, en arrière-plan, l'objectif de devenir le numéro un de l'intelligence artificielle.

« *Celui qui deviendra leader en ce domaine sera le maître du monde* » a affirmé le président Russe Vladimir Poutine, le 1er septembre 2017 devant des étudiants de la ville de Yaroslavl en Russie.

## Les entreprises technologiques sont très vite devenues les bras armés des États dont elles dépendent, et se sont lancées à cet effet dans la construction rapide d'immenses monopoles plurisectoriels.

Ce monopole de la donnée est parfois renforcé par la réglementation, comme c'est actuellement le cas aux États-Unis, avec l'adoption du CLOUD Act (Clarifying Lawful Overseas Use of Data Act). Quand 80 % des sociétés du CAC 40 hébergent leur Cloud chez Amazon, et que bon nombre de collectivités territoriales ou de ministères publics stockent leurs données auprès de sociétés américaines, l'inquiétude est bien là. Pour comprendre l'essence de leur stratégie commerciale, il faut comprendre leurs aspirations profondes mêlées d'une volonté de dépasser les limites de l'humain, de proposer une

Pour ce faire, la mainmise sur les données (des citoyens, des entreprises et des États) doit être une priorité nationale. Une course à l'appropriation du nouveau carburant est lancée pour une domination politique, économique et sociale. Une domination qui répond à un mode opératoire précis : la stratégie du criquet.

Le mode opératoire est toujours le même : sa première phase consiste en la conclusion d'un contrat avec un acteur du secteur cible (partenariat, contrat de service), la deuxième phase en l'étude du mode de fonctionnement du secteur considéré (chaîne de valeurs, technique,

1 - Selon le rapport de David Mercer, Strategy Analytics, en date du 14 mai 2019, « *Global Connected and IoT Device Forecast Update* », il y aurait 22 milliards d'objets connectés fin 2018 avec une prévision de 40 milliards en 2025.

process...), la troisième phase consiste à phagocytter le partenaire, la quatrième phase opère la création progressive d'un lien de forte dépendance, voire d'un monopole dans le secteur cible. Ayons toujours à l'esprit cette citation du très offensif et libertarien Peter Thiel, cofondateur de PayPal et Palantir Technologie: « la concurrence est une idéologie qui déforme notre pensée ».

**Nous sommes en guerre économique avec des alliances qui se créent au fil du temps, et pouvant, à terme, déstabiliser la France, l'Europe et nos populations dans leur ensemble, politiquement, économiquement et socialement.**

Rappelons l'affaire Naval Group, la cession massive de données de santé des citoyens français au Cloud Microsoft Azur via le Health Data Hub, la gestion massive de données sensibles de nos entreprises par la BPI au géant Amazon, le passage sous pavillon américain de pépites technologiques et stratégiques incontournables comme par exemple Exxelia. La liste exhaustive de toutes les atteintes aux intérêts nationaux est longue et suffisamment préoccupante pour nous conduire désormais à agir sans plus attendre pour stopper l'hémorragie de nos forces vitales.

Des blocs se constituent : les États-Unis,

le Royaume-Uni, l'Australie, le Canada et la Nouvelle-Zélande depuis le 5 mars 1945 avec le Traité UK USA (United Kingdom – United States Communications Intelligence Agreement), la Chine qui étend son influence en Europe, en Afrique avec sa nouvelle route de la Soie (Belt and Road Initiative), la Russie qui renforce ses liens avec l'Inde (pas moins de 28 accords commerciaux signés depuis le 6 décembre 2021) et la Chine (une alliance a été annoncée, le 15 septembre 2022 en marge du Sommet de l'OCS – Organisation de Coopération de Shanghai).

**Et la France et l'Europe dans tout cela ?**

**Les vers sont dans le fruit.** L'État et les collectivités territoriales françaises sont plus perméables que jamais aux stratégies d'influences étrangères.

Combien de ports, d'aéroports passés sous pavillon chinois dans le cadre de la nouvelle routes de la soie. Combien de collectivités ont subventionné des entreprises chinoises comme Huawei sous motif qu'ils créaient de l'emploi en occultant complètement le droit chinois qui impose désormais un transfert automatique des données et des clés de chiffrement stockées par les sociétés chinoises du numérique (Région Grand Est – Médiapart 25 février 2021).

Combien de collectivités territoriales ont cédé à des partenariats gratuits avec Google ou Microsoft ? (Rennes France 3 régions 22 janvier 2018 – Université d'Angers Infos jeunes 8 décembre 2021).

portant indubitablement atteinte à nos intérêts en matière de sécurité nationale.

À travers les tensions étatiques, la souveraineté numérique ne doit plus être le leitmotiv de quelques représentants du secteur technologique français, **mais bien la clé de voûte d'un nouveau**

## **La souveraineté numérique ne doit plus être le leitmotiv de quelques représentants du secteur technologique français, mais bien la clé de voûte d'un nouveau système politico-économique à repenser, à réinventer.**

**L'Europe, elle-même, ne développe pas une stratégie claire** avec, d'un côté, la promotion haut et fort de la souveraineté numérique et, de l'autre, un accord surprenant sur le transfert des données personnelles aux États-Unis conclu le 25 mars 2022 alors même que les Américains n'ont pas changé leur réglementation en matière de captation pour « suspicion de violation du droit américain » (CLOUD Act), et qu'au contraire, ils l'ont même renforcée en sourdine à travers leur législation secrète sur la sécurité nationale et le Freedom Act... au mépris total du droit des Nations européennes.

**Un accord sur fond de crise énergétique qui se résume à un troc énergie contre data** aussi dangereux que déséquilibré, et

**système politico-économique à repenser, à réinventer.**

À la fois question de sécurité nationale, condition de survie de notre démocratie et fondement d'une croissance économique pérenne, la souveraineté numérique n'apparaît plus comme une option et un gadget de plus, mais comme un impératif et un devoir absolu sans lequel l'État français entrera dans un état de déliquescence programmé.

**Nos concitoyens souhaitent-ils un État en crise perpétuelle ? Rien n'est moins sûr.**

Alors, quel plan d'action avons-nous l'obligation de mener pour défendre les intérêts de la Nation ?

### 1. La souveraineté numérique, une question de sécurité nationale et de survie de la Nation

**Qu'on se le dise une bonne fois pour toute, la souveraineté numérique est une composante de la souveraineté nationale.**

La souveraineté est définie en droit comme « la qualité de l'État de n'être obligé ou déterminé que par sa propre volonté, dans les limites du principe supérieur du droit, et conformément au but collectif qu'il est appelé à réaliser ». En France, cette autorité appartient à la Nation, au peuple qui l'exerce par ses représentants.

Si la Nation n'a pas la maîtrise de ses données alors que les données sont le nouveau pouvoir, la Nation n'a plus de pouvoir.

Une réforme constitutionnelle s'avère donc nécessaire pour prendre en compte les évolutions technologiques et les enjeux de la 4ème révolution industrielle. Il convient d'instaurer une hiérarchie des normes efficaces qui permettra plus facilement de sanctionner et de contrôler

tous les actes qui iraient à l'encontre de ce principe fondamental.

En attendant une refonte de notre Constitution, un arsenal juridique est déjà à notre portée pour défendre les intérêts de la Nation. Le Code pénal porte en son sein tout un bloc juridique qui pourrait être des plus efficaces s'il était invoqué et appliqué.

Pour rappel, par intérêts fondamentaux de la Nation, il faut entendre, selon l'article 410-1 du même code, « *son indépendance, (...) l'intégrité de son territoire, (...) sa sécurité, (...) la forme républicaine de ses institutions, [les] moyens de sa défense et (...) sa diplomatie, (...) la sauvegarde de sa population en France et à l'étranger, (...) l'équilibre de son milieu naturel et de son environnement et [les] éléments essentiels de son potentiel scientifique et économique et de son patrimoine culturel* ».

L'article 411-6 du Code pénal, quant à lui, dispose que « *le fait de livrer ou de rendre accessibles à une puissance étrangère, à une entreprise ou organisation étrangère ou sous contrôle étranger ou à leurs agents des renseignements, procédés, objets, documents, données informatisées ou fichiers dont l'exploitation, la divulgation ou la réunion est de nature à porter atteinte aux intérêts fondamentaux de la nation est puni de quinze ans de détention criminelle et de 225 000 euros*

*d'amende* ».

À la lecture de ces textes, beaucoup doivent s'interroger.

Si nous voulons renouer avec la performance, deux phases s'imposent : une phase pédagogique et une phase plus coercitive, l'une n'allant pas sans l'autre.

### De l'urgence d'une grande vague de formation et de sensibilisation

Un plan d'urgence de formation et de sensibilisation va devoir être conçu et déployé. La Nation doit elle-même se mettre à niveau pour comprendre le mode de fonctionnement de la nouvelle ère dans laquelle nous sommes entrés.

Magistrats actuels et futurs magistrats à l'ENM, Universités, Grandes Écoles, chefs et cheffes d'entreprise, élus et élues, fonctionnaires d'État, territoriaux, hospitaliers, toutes et tous doivent désormais maîtriser la géopolitique des données, les fondements de l'intelligence économique, l'écosystème des données et de leur exploitation.

Nous ne devons plus avoir des directeurs généraux d'établissements publics ni des élus affirmant qu'ils ont choisi Amazon ou Google ou Huawei parce que « c'était gratuit », ni de directeurs d'incubateurs proposant fièrement aux startups fran-

çaises des entreprises américaines spécialisées en cybersécurité comme accompagnatrices incontournables de nos pépites en gestation.

### Il faut sonner le glas de l'hémorragie de données stratégiques, en partance vers la Chine et les États-Unis.

Notre Nation est prise en étau et victime d'un jeu de go de la part des puissances étrangères du numérique.

Nos collectivités territoriales, nos entreprises sont devenues des cibles de choix dans la guerre mondiale des données et la course à l'IA.

D'où l'urgence de sensibiliser la Nation car nous sommes face à des forces étrangères dont le mode opératoire est toujours le même :

- Analyser les cibles (faiblesses psychologiques, faiblesses économiques et sociales, modes de fonctionnement, réseau, entourage familial et professionnel, identification des besoins du territoire visé).
- Utiliser les failles psychologiques et répondre à un besoin précis du territoire (stratégie court terme).
- Pénétrer les territoires, les appauvrir (à moyen et long terme) pour mieux phagocyter l'État.

Une grande réforme des enseignements au collège et lycée sera également nécessaire pour former les plus jeunes et les futures générations d'acteurs de la Cité.

**L'éducation nationale est le berceau de la démocratie, le numérique souverain son bras armé pour défendre élèves et corps enseignants, et construire une éducation performante répondant aux besoins des élèves en les accompagnant dans leur devenir de citoyens, libres et éclairés.**

**L'éducation nationale est le berceau de la démocratie, le numérique souverain son bras armé pour défendre élèves et corps enseignants, et construire une éducation performante répondant aux besoins des élèves en les accompagnant dans leur devenir de citoyens, libres et éclairés.**

Ces politiques pénale et éducationnelle devront s'accompagner d'une véritable stratégie du renseignement.

Arrêtons avec la naïveté et l'application de certains principes des Lumières obsolètes en ces temps troubles, ce qui implique de revoir la stratégie des échanges universitaires, la participation à certains Think Tank à l'aune de cette guerre mondiale du sens que nous subissons.

Il est urgent de concevoir une véritable stratégie coordonnée du renseignement, et **d'allouer les fonds nécessaires à la protection de la Nation pour éviter les compromissions.**

Cela ne signifie naturellement pas le retour à un repli sur soi et cesser toute alliance. Au contraire, c'est enfin redonner ses lettres de noblesse aux alliances véritables et aux partenariats qui impliquent un rapport d'égalité et non de vassalité ! Respectons-nous et les autres nous respecteront enfin !

**2. La souveraineté numérique, cette base indispensable pour garantir la protection de notre démocratie.**

**L'enjeu de la souveraineté numérique est bien la Liberté et le pouvoir d'émanciper nos contemporains.**

Si des puissances étrangères détiennent les données de nos populations, comment celles-ci pourraient-elles être libres, et

maîtriser leur destin commun ?

Quelle légitimité ont certains élus, certaines institutions pour déléguer le pouvoir de la Nation à des puissances étrangères en velléité de maîtriser ou de manager nos décisions ?

Outre le sujet de la constitutionnalité de telles actions, se pose la question de l'avenir de notre démocratie et la survie de notre modèle unique de société.

**Entre ingérences massives, attaques cyber et fake news, l'avenir nous paraît bien sombre.**

Le résultat : une captation massive et à bas bruit de nos données et des perméabilités toujours plus grandes aux ingérences étatiques adverses.

**Ce « Soft Power », qui n'a de doux que le nom, fonctionne et opère confortablement à mesure que la confiance entre la Nation et ses représentants s'étiole.** La nature ayant horreur du vide, certaines puissances étrangères ont fait preuve d'un remarquable opportunisme en prenant la place laissée vacante par nos institutions défaillantes et nos représentants, sempiternels perdants du grand jeu.

**Ce « Soft Power », qui n'a de doux que le nom, fonctionne et opère confortablement à mesure que la confiance entre la Nation et ses représentants s'étiole.**

**Instagram, Facebook, Twitter, TikTok, ces réseaux sociaux sont en réalité de véritables armes cybernétiques et cognitives de soft power** (pouvoir doux – capacité de faire faire à autrui ce qu'il n'aurait pas fait dans un premier temps), détruisant à **petit feu notre liberté de choix, l'intérêt général**, nos dispositions à l'indépendance politique au profit d'une atomisation de la Nation grignotée par des valeurs qui nous viennent d'ailleurs.

Aujourd'hui, nous sommes face à une crise politique unique dans la Vème République à travers la remise en cause de la représentation nationale et de ses représentants politiques bien peu soucieux de la défense de l'intérêt général, lui préférant la valse des ego...

**Or, « le gouvernement sans chef n'est pas nécessairement une absence de gouvernement ; en fait il peut devenir dans certaines circonstances le plus tyrannique et cruel**

**entre tous** » Hannah Arendt (Condition de l'homme moderne, p.79).

Construire notre souveraineté numérique c'est aussi en tant que Nation ne pas laisser autrui nous dicter notre présent et notre futur, c'est créer un projet et des visées communes dans lequel les français.es pourraient enfin se retrouver et s'unir.

### 3. La souveraineté numérique, la nouvelle condition pour l'accomplissement d'une croissance économiquement pérenne.

Enfin, la souveraineté numérique est le prérequis indispensable à toute stratégie de croissance économique durable.

**Sans souveraineté numérique, pas de souveraineté alimentaire, ni industrielle.**

Martelons qu'une entreprise française qui transfère toutes ses données vers des puissances étrangères donne le pouvoir aux dites puissances et creuse indubitablement sa propre tombe.

Nous commençons à en voir les prémices dans la grande distribution ou le secteur automobile...

Avant de nous lancer dans les stratégies économiques de courte vue, brouillonnes

et confuses, un corpus d'actions devra être au préalable déployé dont notamment :

- Élargir la liste des Opérateurs d'importance vitale.
- Limiter les investissements toxiques et contrôler toutes les prises de participation dans les secteurs sensibles.
- Fixer des objectifs précis et opérables en termes de recherche et de développement.
- Soutenir l'innovation à impact avec de la commande, au lieu de « subventionner » à tout va des mastodontes français qui ont jusqu'à présent été très gourmands en financements publics, mais guère performants en productions efficaces et viables.
- Concevoir un plan stratégique coordonnant l'action des entreprises/centres de recherche et investisseurs, afin de servir un seul objectif: la construction de la souveraineté numérique.

**Aujourd'hui, nous ne pouvons plus nous permettre d'avoir des investisseurs qui s'enorgueillissent d'avoir au capital de licornes des BHATX.**

Tout comme il faudra se pencher très sérieusement sur les conditions d'octroi de subventions européennes, étatiques ou

régionales, afin d'éviter tout subventionnement d'entreprises qui ne respecteraient pas les principes de la souveraineté numérique.

atteinte à l'intérêt national.

Le plan d'action proposé n'est évidemment pas exhaustif et mériterait naturellement d'être bien plus étoffé.

## Nous n'avons plus véritablement le choix. Construire notre souveraineté numérique est une urgence absolue qu'il va nous falloir très vite monter à l'échelle.

Il en est de même de tous les incubateurs plus ou moins sponsorisés, ou proxy de puissances étrangères en deuxième rideau, qui vont devoir incarner cette souveraineté sous peine de sanctions pénales dont nous préconisons l'application.

**Les entrepreneurs du digital, les startupers doivent aussi œuvrer dans leur recherche d'excellence et d'innovation dans l'intérêt de la Nation,** et ne plus uniquement rêver d'être rachetés par des Américains ou des Chinois, si ces rachats peuvent porter

**En résumé,** nous n'avons plus véritablement le choix. Construire notre souveraineté numérique est une urgence absolue qu'il va nous falloir très vite monter à l'échelle, corrélativement en cela aux enjeux véritables qui nous font face. Et nous devons entamer à cet effet un virage serré, en durcissant nos positions si nous voulons toujours exister sur le plan international à échéance brève, et tout simplement survivre aux crises que nous traverserons encore telles des brindilles embarquées sur un bateau ivre...

# Bibliographie

- Accenture. (2020). Why does insurance rank so high on the Disruptability Index? <https://insuranceblog.accenture.com/>.
- AXA. (2021). Rapport intégré. [https://www-axa-com.cdn.axa-contento-118412.eu/www-axa-com/4076309b-00e2-4b92-8012-f0ff8c25d1f0\\_axa\\_ri2021\\_vf\\_accessible.pdf](https://www-axa-com.cdn.axa-contento-118412.eu/www-axa-com/4076309b-00e2-4b92-8012-f0ff8c25d1f0_axa_ri2021_vf_accessible.pdf).
- Begenau, J. F. (2018). *Big data in finance and the growth of large firms*. Cambridge, MA: NBER.
- Bessen, J. E. (2017). *Information technology and industry concentration*. Boston: Boston University School of Law.
- BOULLE, R. (2018). Gras Savoye Willis Towers Watson lance risk & analytics. [www.argusdelassurance.com/](http://www.argusdelassurance.com/).
- Bresciani, S. F. (2021). Bresciani, Stefano, et al. "Using big data for co-innovation processes: Mapping the field of data-driven innovation, proposing theoretical developments and providing a research agenda. *International Journal of Information Management*.
- CapGemini. (2021). The data-powered insurer: Unlocking the data premium at speed and scale. *CapGemini Resarch Institute*.
- Cavallo, A. (2017). Are online and offline prices similar? Evidence from large multi-channel retailers. *American Economic Review*.
- Cavanillas, J. M. (2016). *New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe*. Springer Nature.
- Chakravorti Bhaskar, B. A. (2019, January 24). Which Countries Are Leading the Data Economy? *Harvard Business Review*.
- Cherry, K. (2020, December 14). *The Most Famous Social Psychology Experiments Ever Performed*. Retrieved from <https://www.verywellmind.com/famous-social-psychology-experiments-2795667>
- CNP. (2022). Rapport annuel 2021/2022. *CNP Assurances*.
- franceassureurs. (2022). 20 projets pour accompagner les Français dans l'évolution de leurs modes de vie. <https://www.franceassureurs.fr/>.

- Goodridge, P. a. (2015). How does big data affect GDP? Theory and evidence for the UK.
- Han, J. (2011). *Networked Information Technologies, Elections, and Politics: Korea and the United States*. Lexington Books.
- Hurley, R. P. (2015). Insurance disrupted, General Insurance in a connected world, Deloitte. London: *The Creative Studio, Deloitte*.
- IDC. (2021). *Global datasphere*. [https://www.idc.com/getdoc.jsp?containerId=IDC\\_P38353](https://www.idc.com/getdoc.jsp?containerId=IDC_P38353).
- Katz, E. (1957). The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. *Political Opinion Quarterly*, 21(1), 61-78.
- MADHANI, A., & COLVIN, J. (2021, January 9). A farewell to @realDonaldTrump, gone after 57,000 tweets. Retrieved from <https://apnews.com/article/twitter-donald-trump-ban-cea450b1f12f4ceb8984972a120018d5>
- Martínez, M. A. (2018, May 28). *Vertical and horizontal relations: What is 'society' about?* Retrieved from <https://www.miguelangelmartinez.net/?Vertical-and-horizontal-relations-What-is-society-about>
- McAfee, A. E. (2012). *Big data: the management revolution*. Harvard Business Review.
- McQuilkin, H., & Chakrabarti, M. (2020, February 27). *Part IV: Are We Living In A Post-Truth World?* Retrieved from <https://www.wbur.org/onpoint/2020/02/27/part-iv-post-truth>
- merriam-webster. (n.d.). Retrieved from <https://www.merriam-webster.com/dictionary/influence>
- PAPPAS, N. (2016, April 4). *Socrates, Cynics and Flat-Nailed, Featherless Bipeds*. Retrieved from <https://archive.nytimes.com/opinionator.blogs.nytimes.com/2016/04/04/of-socrates-cynics-and-flat-nailed-featherless-bipeds/>
- Philippon, R. M. (2019). *The economics of Big Data and artificial intelligence*. Emerald Publishing Limited.
- SAS. (2016). *Fraud, Coalition Against Insurance. The state of insurance fraud tech-*

nology. SAS.

- Trabucchi, D. a. (2019). Data-driven innovation: Switching the perspective on Big Data. *European Journal of Innovation Management*.
- United States Holocaust Memorial Museum, W. D. (2020, December 8). "FINAL SOLUTION": OVERVIEW. Retrieved from <https://encyclopedia.ushmm.org/content/en/article/final-solution-overview>
- United States Holocaust Memorial Museum, W. D. (2021, July 16). EICHMANN TRIAL. Retrieved from <https://encyclopedia.ushmm.org/content/en/article/eichmann-trial>
- *Word of the Year 2016*. (n.d.). (Oxford University Press) Retrieved from <https://languages.oup.com/word-of-the-year/2016>

# Les enjeux du Big Data

Nos groupes de travail essaient de rendre compte d'une idée force du K2: la connaissance appartient à tout le monde.

Il nous est apparu que le point de vue de l'expert, auquel nous sommes habitué ne devait plus être exclusif. Un besoin croissant du citoyen de connaître par lui-même se fait sentir en tout domaine. C'est la raison pour laquelle nos groupes de travail sont composés de spécialistes mais aussi de nombreuses personnalités qui ne disposent pas, a priori, de compétences particulières sur le sujet traité mais qui souhaitent y réfléchir. Nous pensons ainsi que les efforts de réflexion et de partage de professionnels issus de tous horizons dans une démarche pluridisciplinaire permettront de retrouver de l'unité dans la connaissance et donc une meilleure compréhension globale des choses. Telle est la démarche qu'ont entreprise les membres du Groupe K2 Les enjeux du Big Data.

L'impact du Big Data sur tous les secteurs du monde s'apparente à une véritable révolution industrielle et est en passe de changer les rapports de force économique au sein des sociétés mais aussi entre pays. L'ensemble de ses ramifications est tellement vaste qu'il faudrait plusieurs ouvrages pour tous les traiter. Le présent rapport se concentre sur les aspects du Big Data qui nous semblent essentiels pour bien appréhender ses conséquences à court et moyen terme.



ISBN : 9782493056498  
Ce livre ne peut être vendu

Cercle K2, 10, rue des pyramides  
75001 Paris, France  
cercle-k2.fr