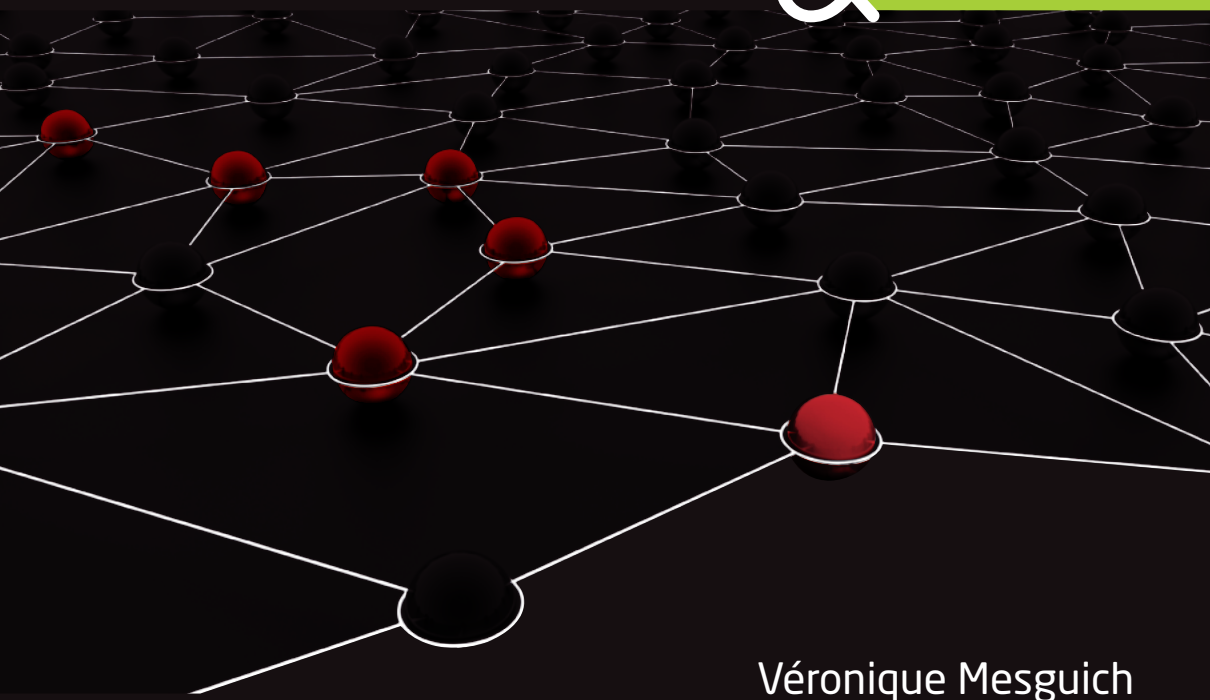
Information & stratégie



Véronique Mesguich

Rechercher l'information stratégique sur le web

Sourcing, veille et analyse à l'heure de l'IA

3e édition

Préface d'Anne-Marie Libmann

adbs

deboeck SUPÉRIEUR

Information & stratégie

Véronique Mesguich

Rechercher l'information stratégique sur le web

Sourcing, veille et analyse à l'heure de l'IA

3e édition

Préface d'Anne-Marie Libmann

adbs

deboeck SUPÉRIEUR

RESSOURCES NUMÉRIQUES

Liste thématique des sources et outils cités dans le livre (voir p. 240)

Accédez directement à votre ressource :

Scannez le code avec votre téléphone ou votre tablette

OU

Tapez l'URL dans votre navigateur

Pour toute information sur notre fonds et les nouveautés dans votre domaine de spécialisation, consultez notre site web : **www.deboecksuperieur.com**

Couverture et maquette intérieure : cerise.be
Mise en page : PCA

© De Boeck Supérieur s.a., 2024
Rue du Bosquet, 7 – B-1348 Louvain-la-Neuve

Tous droits réservés pour tous pays.
Il est interdit, sauf accord préalable et écrit de l'éditeur, de reproduire (notamment par photocopie) partiellement ou totalement le présent ouvrage, de le stocker dans une banque de données ou de le communiquer au public, sous quelque forme et de quelque manière que ce soit.

3e édition

Dépôt légal :
Bibliothèque nationale, Paris : mai 2024
Bibliothèque royale de Belgique, Bruxelles : 2024/13647/054

ISSN 2295-3825
ISBN 978-2-8073-6140-9

PRÉFACE

Le monde de l'information se transforme sous nos yeux au gré des révolutions qui, chacune à leur tour, apportent leur lot de progrès tout en fragilisant les fondations établies. Depuis l'irruption d'Internet dans notre quotidien et nos métiers, cette transformation semble s'accélérer, déconstruisant et remodelant sans relâche nos modes d'accès et de gestion de l'information, redéfinissant les critères d'évaluation de celle-ci, les modèles économiques des acteurs du secteur et questionnant la place et le rôle des professionnels de l'information.

Dans ce contexte de mutation constante, ChatGPT et autres modèles de langage avancés émergent comme des catalyseurs d'une nouvelle révolution, celle de l'intelligence artificielle au service de la connaissance. Dotée d'une capacité inédite à comprendre, analyser et synthétiser des volumes d'informations phénoménaux, l'IA offre déjà la possibilité de naviguer dans l'océan numérique avec une efficacité et une puissance nouvelles.

On pressent d'emblée que les bouleversements à venir ne vont pas se limiter à la transformation de l'expérience de recherche et de l'accès à l'information. L'intelligence artificielle nous pousse à réévaluer en profondeur toutes les facettes de notre rôle d'intermédiaires de l'information. Cela inclut notamment nos interactions avec les clients, la capacité à appréhender et analyser les champs disciplinaires spécifiques, le traitement, l'analyse et la présentation des données, et la performance globale de notre organisation documentaire.

On comprend que ce nouveau partenaire, ce « copilote » comme Microsoft l'a opportunément désigné, nous incite à repenser nos processus et stratégies de manière critique et créative. Il nous poussera aussi à valoriser bien plus encore l'apport unique de l'humain.

Cette valeur se trouvera principalement dans notre capacité à poser les bonnes questions, à comprendre les réponses fournies par des systèmes boostés à l'IA, et à appliquer ces connaissances dans des situations concrètes et variées.

Cette transition ne pourra être réalisée par les professionnels que par une préparation active et un engagement à renforcer et à renouveler leurs compétences techniques en permanence. Le nouveau livre de Véronique Mesguich apparaît comme un outil inestimable dans ce processus de préparation.

Depuis maintenant près de deux décennies, Véronique Mesguich, spécialiste du domaine de la veille et de l'intelligence économique, poursuit son exploration des profondeurs du monde numérique. Chacun de ses ouvrages revêt une importance particulière, non seulement en tant que témoin de la révolution numérique, mais

aussi pour sa capacité à démystifier et à transmettre les concepts les plus complexes de manière claire et accessible.

Ce nouvel ouvrage ne fait pas exception. L'auteur revisite, en y intégrant l'intelligence artificielle, l'ensemble des outils et méthodes de recherche et de traitement d'information et établit un nouveau standard de référence permettant à chaque professionnel d'évaluer et de mettre à niveau ses connaissances et pratiques.

C'est ainsi que pourra se faire l'adaptation à un environnement de plus en plus technique, marqué par une diversité croissante des sources et formats d'information hétérogènes, et la mise à disposition d'outils de plus ou en plus ouverts et puissants.

Ce socle de connaissances techniques constitue la pierre angulaire pour relever le défi que représente l'intégration de l'intelligence artificielle dans les processus de veille, de recherche et d'analyse au sein des entreprises. L'obstacle majeur à surmonter sera précisément cette intégration, exigeant un investissement considérable dans l'apprentissage des nouvelles technologies et un engagement résolu à embrasser, et parfois à être à l'avant-garde du changement.

L'ouvrage de Véronique Mesguich offre donc une ressource précieuse pour accompagner les professionnels dans cette transition complexe, en mettant en lumière l'importance d'une approche méthodique et d'une intelligence à deux dimensions : celle de l'esprit humain et celle, tout aussi innovante… de l'intelligence artificielle.

Par Anne-Marie Libmann,
Directrice des opérations FLA Consultants
et de BASES Publications

INTRODUCTION

La révolution numérique se déploie depuis déjà plusieurs décennies, et comme toute révolution, avance par bonds et par étapes. Cette révolution a bouleversé plus que tout autre secteur le vaste univers de l'information. Depuis une trentaine d'années, la nature des contenus informationnels a profondément évolué : il ne s'agit pas d'une simple mutation technique, mais d'une véritable révolution – que l'on a souvent comparée à l'invention de l'imprimerie – dans les modes de production, de consommation, de stockage, d'échange d'informations de toutes natures. Le déploiement fulgurant et récent des intelligences artificielles génératives vient ajouter de nouveaux défis et opportunités, mais aussi de nombreux risques. Ces nouveaux paradigmes concernent tant les professionnels de l'information que les utilisateurs, que ce soit dans le monde de l'entreprise, de l'enseignement ou tout simplement pour les besoins de la vie quotidienne. Comme toute révolution, la numérisation de l'information voit surgir de nouveaux modèles économiques, de nouveaux acteurs et une multitude de services.

« Organiser l'information mondiale et la rendre universellement accessible et utile » : telle est la mission que s'étaient fixée les deux cofondateurs de Google dans les premières années d'existence du célèbre moteur. Alors que Google compte aujourd'hui plus de 25 ans d'existence – un âge canonique à l'aune du web –, peut-on considérer que la mission est accomplie ? Et comment cette mission peut-elle s'accomplir de façon efficace face aux volumétries toujours plus importantes de documents et données, et aux progrès dans le domaine des intelligences artificielles ? À l'origine, la pratique de la recherche d'information est issue d'une longue tradition, à la fois universitaire et professionnelle, de requêtes dans des bases de données structurées. Aujourd'hui, selon John Battelle, cofondateur du célèbre magazine *Wired,* le « Search » ne se limite pas à quelques mots saisis dans un formulaire en ligne. Il s'agit plus d'un mode, « d'une méthode d'interaction avec les mondes physiques et virtuels »[1].

Immense et protéiforme, le web peut être perçu aujourd'hui à la fois comme un formidable moyen d'accès à la connaissance et au savoir, mais aussi (et ce n'est pas contradictoire) comme une gigantesque « base de données d'intentions » marketing. Google illustre parfaitement cette double approche et son ambiguïté.

En 2006, nous avions, avec la consultante Armelle Thomas, coécrit la première édition de l'ouvrage *Net Recherche*, guide de la recherche et de la veille sur Internet. L'objectif était de présenter aux internautes (qui n'étaient pas encore équipés de smartphones !) la variété des méthodes de recherche web ainsi que les différents moteurs (ils étaient plus nombreux qu'aujourd'hui…). Cinq éditions

1 http://www.wired.co.uk/article/the-future-of-search

se sont succédé au fil des années et ont suivi l'évolution du web et son impact sur la recherche d'information : déploiement des réseaux sociaux grand public et professionnels, du web de données, avènement du web temps réel, des applications et des contenus multimédias, de l'OSINT ou intelligence des sources ouvertes…

En ce début du XXIe siècle se déroule une nouvelle révolution numérique, résultant des avancées en matière d'intelligence artificielle et traitement des données massives, que l'on ne saurait limiter au seul ChatGPT. Sommes-nous en train d'intégrer une nouvelle ère, ou plutôt selon la belle expression de Yuval Harari[2], une « religion des données numériques » dans laquelle l'univers consiste en un flux de données ? À l'heure où les géants du numérique investissent massivement dans l'intelligence artificielle, la réalité augmentée, la reconnaissance vocale, la datavisualisation, qu'en est-il de la recherche d'information aujourd'hui ? Effectuer une recherche sur Internet n'a jamais été aussi simple pour les utilisateurs ; pourtant, derrière cette apparente simplicité se cache une grande complexité, dans la nature des contenus comme dans les dispositifs techniques. Il ne s'agit pas de verser dans un optimisme béat devant les progrès apportés par les géants du numérique, ni de les diaboliser de façon stérile. Il est important, en revanche, que l'on soit un professionnel d'entreprise, un enseignant, un étudiant ou un simple internaute, d'être conscient des formidables capacités des outils de recherche et de veille, mais aussi de leurs limites et des moyens de les pallier. La culture de l'honnête « homo numericus » passe aujourd'hui par la compréhension du fonctionnement des outils, de la maîtrise de leurs fonctionnalités avancées et de la diversité des sources d'information.

Dans l'actuelle société de l'information, dans l'écosystème complexe du web actuel, à l'heure où ChatGPT envahit les applications, comment optimiser ses recherches, collecter l'information stratégique ? Comment trouver par exemple des données financières sur ses concurrents, détecter les dernières innovations et tendances d'un secteur, suivre les nouvelles technologies, identifier des influenceurs sur les réseaux sociaux, surveiller des offres d'emploi, repérer des cours en ligne ? Au temps de la « post-vérité », comment juger de la qualité d'une information ou de la fiabilité d'une image ? Comment prévenir la redondance et la perte de temps ? Comment éviter de passer à côté d'une information potentiellement stratégique ? Comment dégager des « signaux faibles » et représenter graphiquement la complexité d'une situation ?

Cet ouvrage a pour objectif de présenter, à travers des articles clairs et synthétiques, les solutions, méthodes et sources indispensables pour répondre à ces besoins professionnels et stratégiques.

Un premier chapitre fait le point sur le paysage du web et de l'information numérique aujourd'hui. Ce chapitre permet de décrypter des termes liés au web actuel : machine learning, intelligences génératives, metavers, big data, etc., et fait le point sur l'évolution des réseaux sociaux, notamment concernant l'évolution de

2 Y. Harari, *Homo deus. Une brève histoire de l'avenir*, Albin Michel, 2017.

Twitter/X. Plusieurs notions techniques fondamentales et standards du web sont également expliqués.

Le deuxième chapitre est entièrement consacré à la recherche : on y détaille le fonctionnement des moteurs, l'évolution de leurs algorithmes et l'intégration des intelligences génératives à travers les nouvelles solutions Google Gemini, Copilot ou Perplexity. Les concurrents classiques de Google y sont présentés, ainsi que les nombreux moyens d'optimiser les recherches sur les pages web et les différents réseaux sociaux. Un panorama des sources stratégiques et méthodes de sourcing complète ce chapitre, qui se termine par de nombreux conseils méthodologiques.

Le troisième chapitre est consacré à la veille stratégique et les différentes phases du processus : après une étude de la phase préparatoire (plan de veille) sont présentées les multiples façons de collecter l'information, des flux RSS aux alertes automatisées, en passant par la surveillance des réseaux sociaux. Un tableau comparatif des outils de collecte automatisée termine ce chapitre.

Le quatrième chapitre expose les nombreuses méthodes d'analyse de l'information pour la prise de décision. L'analyse peut se baser sur les modèles classiques, mais met en œuvre également des méthodes automatisées, parmi lesquelles on peut citer la datavisualisation, la fouille de données ou l'analyse du sentiment exprimé entre autres sur les réseaux sociaux.

Le chapitre 5 est composé de mini-études de cas, présentées sous forme de « recettes » correspondant à des besoins récurrents pour les entreprises et organisations : trouver des données financières sur les concurrents, rassembler des informations sur des personnes, repérer les innovations, identifier des experts, etc. Ce chapitre se conclut par un récapitulatif méthodologique général, et déroule pas à pas, à partir d'une étude de cas, les bonnes pratiques de recherche, veille, analyse, stockage et diffusion de l'information stratégique.

La conclusion présente les perspectives d'évolution futures et esquisse les contours d'un véritable plaidoyer pour une « littératie numérique », en évoquant les compétences indispensables aujourd'hui et demain.

Cet ouvrage s'adresse à tout internaute confronté à des problématiques de recherche, collecte ou analyse de l'information utile. Professionnels d'entreprises ou organisations dans tous secteurs, enseignants, universitaires, étudiants, porteur de projets : chacun y trouvera des méthodes, astuces, bonnes pratiques, et une aide au choix des solutions les plus appropriées.

« Google peut vous donner 100 000 réponses, un bibliothécaire peut vous donner LA bonne » : cette formule de l'auteur britannique Neil Gaiman date du début des années 2010. L'intégration de l'IA va-t-elle permettre aux moteurs de fournir cette bonne réponse en citant correctement des sources fiables ? Ou bien faut-il craindre un monde dystopique dominé par les algorithmes et des contenus

informationnels produits artificiellement en masse ? Quelles que soient les futures orientations des différents outils de recherche, notamment dans le domaine de l'IA, la société de l'information et de la connaissance requiert une bonne maîtrise de leur fonctionnement ainsi qu'une véritable littératie numérique. C'est ce que souhaite apporter cet ouvrage, dont nous vous souhaitons bonne lecture.

CHAPITRE 1

UN UNIVERS NUMÉRIQUE INFINI

Intelligence artificielle, big data, machine learning, algorithmes, « fake news »… Tous ces termes se sont invités dans nos vies numériques et sont devenus indissociables de l'Internet de ce premier quart du XXIe siècle. La recherche d'information et la veille sont depuis plusieurs années déjà largement impactées par la transformation numérique, et ce phénomène va s'amplifier avec l'évolution toujours rapide d'un web toujours plus complexe. Ce premier chapitre fait le point sur les notions fondamentales liées à cette transformation, et présente les enjeux actuels pour les entreprises et les organisations.

1. LE PAYSAGE DU WEB AUJOURD'HUI

1.1 Web 2, web 3, web 4… où va le web ?

Le grand magazine *Wired* titrait à sa une en 2010 « *The web is dead*[1] », pour mieux rajouter en sous-titre « *Long live the Internet* ». Si, dans les années 2020, le web est loin d'être mort, malgré le titre un peu provocateur du magazine, force est de constater l'évolution constante du web depuis sa création à la fin des années 1980. L'histoire d'Internet est désormais bien connue et nous nous contenterons d'en rappeler les événements marquants. Ce « réseau de réseaux » interconnectés, créé dans les années 1960 par le département des projets spéciaux de l'armée américaine, constitue l'infrastructure de nombreux services : le courrier électronique, les échanges point à point, les transferts de fichiers… et le World Wide Web, créé en 1989 par Tim Berners-Lee. Les premiers sites à proprement parler apparaissent en 1991, ce sont encore des ensembles de pages statiques, reliées entre elles via des liens hypertextes, et présentant des contenus encore très textuels. Pourtant, dès les années 1990 se développent les premiers sites de commerce électronique. La personnalisation et la mutualisation percent déjà à cette époque à travers les forums et les « sites personnels », ancêtres des blogs. Le fameux « éclatement de la bulle Internet », au début des années 2000, est suivi d'une courte période d'essoufflement du web. Ce n'est que pour mieux repartir avec une déferlante de plateformes toujours plus dynamiques, interactives et multimédias. La publication sur le web est rendue plus simple, interactive et s'effectue de plus en plus en temps réel. C'est la grande vague du « web 2.0 ». Ce phénomène, qui doit son

1 https://www.wired.com/2010/08/ff_webrip/

nom à une conférence tenue en 2004 à San Francisco, ne peut se limiter au succès planétaire des réseaux sociaux. Il s'agit bel et bien d'un recentrage du web autour de l'utilisateur lui-même, qui devient plus actif et à même de publier, commenter ou partager des contenus textuels ou multimédias. De nouveaux usages du web apparaissent, notamment via les applications iPhone ou Android, du fait de la généralisation des smartphones et tablettes à partir de la fin des années 2000.

Après le web 2, le web 3 ? Le web des années 2010 poursuit en effet les tendances de la décennie précédente en accentuant ses caractéristiques : temps réel, mobile, multimédia. La notion de « web 3 », beaucoup moins médiatisée et plus difficile à appréhender que le « web 2.0 », élargit la vague précédente du « web social » vers la notion de « web de données » (web of data) : Tim Berners-Lee le définit comme un immense réseau de données ouvertes liées entre elles par des liens d'interopérabilité[2]. Il s'agit d'une application des technologies et standards du web sémantique de façon à échanger et relier des données structurées sur le web. Ce web sémantique, basé sur des standards édictés par le World Wide Web Consortium (W3C) est surtout destiné à des machines qui vont y effectuer des traitements automatisés.

De plus en plus de données sont générées, en temps réel, non seulement par les humains (via entre autres les médias sociaux), mais aussi par des machines, capteurs ou objets connectés. C'est le phénomène des « big data », ou données massives, caractérisées par la règle des « 3 V » : volume, variété, vitesse. Des modèles de traitement sont mis en place pour analyser en temps réel ces énormes flux de données et en extraire des signaux faibles et hypothèses prédictives.

En ces années 2020 se développe le web des objets connectés, appelé également « Internet of Things », « IoT » ou parfois « web 4.0 ». La croissance du marché de l'IoT n'a pas été aussi importante que prévu, mais le nombre d'objets connectés continue de croitre, pour le grand public comme en entreprise. La 5G propose un débit plus important qui va accélérer ce déploiement. Le web des objets repose sur des connexions entre des capteurs connectés à des objets du monde physique et à leurs « jumeaux numériques », c'est-à-dire une représentation digitale d'un objet, d'une personne ou d'un lieu.

Fin 2021, la notion de « web 3 » fait sa réapparition, mais dans un contexte différent : il s'agit plutôt en l'occurrence d'un web décentralisé, basé sur les technologies de blockchain. Les technologies mises en œuvre dans cette nouvelle conception du web 3 (qui semble rester là encore une niche) concernent en grande partie des actifs financiers numériques, sous forme de cryptomonnaies ou jetons non fongibles (NFT).

Le web 3 pourrait présenter l'avantage de juguler la concentration des services du web, actuellement aux mains de quelques

2 https://www.w3.org/DesignIssues/LinkedData.html

géants du numérique via leurs plateformes, mais pose des problèmes en termes de régulation et de prévention entre autres de la cybercriminalité.

De façon un peu schématique, on pourrait retracer l'évolution du web comme une superposition de multiples interconnexions : des documents liés par des liens hypertextes, des personnes liées entre elles par des réseaux sociaux, des données liées entre elles par des liens d'interopérabilité, et enfin des objets interconnectés entre eux. Le Gartner Group désigne sous l'appellation « Intelligent digital mesh[3] » (maillage digital intelligent), la complexité des interactions entre les personnes, contenus, services et dispositifs du web.

Fin 2022, un nouvel entrant suscitait un engouement sans précédent autour des intelligences artificielles génératives. Il s'agit bien sûr de ChatGPT dont nous détaillerons le principe ci-après. Rappelons préalablement quelques éléments fondamentaux concernant l'intelligence artificielle.

1.2 IA : quelques fondamentaux

L'intelligence artificielle est présente depuis de nombreuses années (et bien avant l'irruption des IA génératives) dans l'univers du numérique et du web ! Le domaine de l'intelligence artificielle est extrêmement vaste et très évolutif : selon la définition qu'en donnait l'un de ses pères fondateurs, John McCarthy[4], on peut définir l'IA comme la science et l'ingénierie de machines capables de réaliser et d'automatiser des tâches nécessitant normalement l'intelligence humaine. Dès la fin des années 1950 apparaît le terme de machine learning (ou apprentissage automatique), qui donne aux ordinateurs la capacité d'apprendre grâce à des algorithmes. Rappelons qu'un algorithme n'est rien d'autre « qu'une séquence d'instructions utilisée pour résoudre un problème. La transmission de cet algorithme va permettre de ne pas avoir à inventer une solution à chaque fois[5]. » Le principe n'est pas nouveau et est utilisé par les mathématiciens depuis des milliers d'années. L'arrivée de la micro-informatique, puis la démocratisation d'Internet, ainsi que la prolifération de contenus numériques ont contribué à un essor sans précédent des algorithmes.

Il existe de nombreuses méthodes d'apprentissage en machine learning. On distingue l'apprentissage supervisé, à partir d'exemples, dans lequel les algorithmes exploitent des données étiquetées existantes, de l'apprentissage non supervisé, lorsque les algorithmes doivent découvrir par eux-mêmes des règles et points communs entre des éléments ou données non étiquetés ou annotés. Quant à l'apprentissage par renforcement, il s'agit d'un mode d'apprentissage statistique inspiré de l'apprentissage humain. Comme dans l'apprentissage non supervisé, la machine apprend de façon autonome, mais doit expérimenter des décisions pour atteindre

3 https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/

4 https://www-formal.stanford.edu/jmc/whatisai.pdf

5 S. Abiteboul et G. Dowek, *Le temps des algorithmes*, Éditions du Pommier, 2017.

un objectif et maximiser des expériences positives sous forme de « récompenses ». Cette technique est utilisée surtout dans le domaine des jeux vidéo et de la recommandation de contenus. Les IA génératives telles que ChatGPT mettent en œuvre plusieurs types d'apprentissage : supervisé, non supervisé et par renforcement.

Le « deep learning », ou apprentissage profond, est un sous-ensemble du machine learning basé sur des « réseaux de neurones profonds », inspirés par le fonctionnement du cerveau humain. Un algorithme de deep learning peut identifier des caractéristiques communes à de nombreuses données et concevoir des modèles prédictifs. Le deep learning est surtout utilisé pour la reconnaissance visuelle, les systèmes de recommandation et le traitement du langage naturel (Natural Language Processing, ou NLP). L'objectif du NLP est de permettre aux ordinateurs de comprendre, interpréter des textes, et de répondre à des questions ou d'effectuer des traductions.

Les techniques d'IA ont progressé rapidement au cours de ces dernières années, en suscitant à la fois espoirs et inquiétudes. Tout en reconnaissant les progrès des technologies, le regretté grand spécialiste du cosmos Stephen Hawking avait émis dès 2014 des craintes vis-à-vis d'une intelligence artificielle qui serait capable de s'améliorer et de se reproduire sans les humains, et à terme de surpasser, voire remplacer les humains[6]. Plus récemment, en début 2023, une lettre ouverte[7] signée par des milliers de scientifiques et experts en intelligence artificielle a été publiée. La lettre demandait une pause dans le développement de l'intelligence artificielle afin de permettre une réflexion approfondie sur les risques potentiels de cette technologie. Les signataires de la lettre ont exprimé leur inquiétude face à la possibilité que l'IA soit utilisée pour créer des armes autonomes, des systèmes de surveillance de masse ou des technologies de contrôle social. Ils ont également souligné que l'IA pourrait avoir des conséquences négatives sur l'emploi, l'économie et la société. D'autres experts estiment au contraire que cette demande de pause pourrait retarder le développement de technologies prometteuses.

Les géants du numérique misent en effet énormément sur l'intelligence artificielle depuis plusieurs années. Leurs services recherche et développement investissent massivement sur des algorithmes d'apprentissage automatique destinés à améliorer leurs produits et services, tels que les moteurs de recherche, les recommandations de contenus, la reconnaissance vocale, la traduction automatique… Ces avancées techniques concernent également la vision par ordinateur pour des applications telles que la reconnaissance faciale, la détection d'objets, la réalité augmentée et la réalité virtuelle. Plus généralement, les retombées de l'IA concernent de très nombreux secteurs : transports, médecine, pharmacie, finance, commerce électronique, marketing, éducation…

6 http://www.bbc.com/news/av/technology-30299992/stephen-hawking-full-interview-with-rory-cellan-jones

7 https://futureoflife.org/open-letter/pause-giant-ai-experiments/

1.3 IA génératives et modèles de langage

Si le robot conversationnel ChatGPT connaît un succès fulgurant depuis fin 2022, on sait moins que le modèle de langage GPT a été élaboré dès les années 2015 par OpenAI.

C'est en effet à cette époque qu'OpenAI a été créé en tant qu'association à but non lucratif par plusieurs entrepreneurs, dont Elon Musk, qui en est parti en 2018. Devenue une entreprise en 2019, la start-up a bénéficié d'investissements importants de la part de Microsoft, en échange d'un partenariat privilégié.

GPT-2 a été mis au point dès 2019, suivi par des versions de plus en plus puissantes ; GPT-3 en 2020 et en 2023 GPT-4, capable de traiter à la fois du texte et des images.

La création de GPT et de modèles de langages similaires n'avait pas pour objectif principal la recherche d'informations, mais plutôt l'amélioration de la compréhension, ainsi que la génération de texte en langage naturel par des machines, et ce en toutes langues. Pour cela, le modèle de langage se base sur une architecture de neurones à même de traiter efficacement des séquences de données, comme des phrases ou des paragraphes, en conservant les relations contextuelles importantes.

C'est tout le sens de l'acronyme GPT :

G signifie « Generative » : les modèles de langage sont capables de générer de façon autonome des contenus sous forme de textes longs, ou d'autres types de données.

P signifie « Pre Trained » : les modèles sont préalablement entraînés sur un vaste corpus de données textuelles avant d'être utilisés pour des tâches spécifiques. Cette étape de pré-entraînement leur permet d'acquérir une compréhension générale du langage naturel et des connaissances implicites contenues dans les textes.

T signifie « Transformer » : il s'agit d'une architecture de réseau de neurones profonds introduite en 2017, qui permet de « comprendre » un contexte et de prédire la continuation d'un texte, c'est-à-dire quels sont les mots les plus probables à partir d'un début de phrase. Pour cela, la technologie des transformers convertit les mots (ou parties de mots) en unités de base appelées « tokens ». Les tokens sont convertis sous forme de « vecteurs », c'est-à-dire des représentations numériques qui codent les informations sémantiques et contextuelles des tokens. Les transfomers utilisent également des paramètres, que l'on pourrait définir comme des espèces de « règles » qui vont permettre au modèle de comprendre et générer du texte. Les paramètres constituent les « connexions » entre les différentes parties du modèle, et sont représentés également sous forme de vecteurs numériques.

À titre d'exemple, GPT-3;5 compte 175 milliards de paramètres. GPT-4 en compterait plus de 1 000 milliards, même si ce chiffre n'est pas confirmé.

OpenAI communique assez peu sur les sources qui ont servi à l'entraînement de ChatGPT. Selon un article scientifique[8] paru en 2020, on sait que plusieurs grands réservoirs de données ouvertes ont été utilisés : CommonCrawl (corpus de plus de 250 milliards de pages web), OpenWebtext2 (publications dans les forums Reddit), Bookcorpus (ouvrages du domaine public numérisés), et Wikipédia. ChatGPT apprend également, par apprentissage renforcé, de ses échanges avec les utilisateurs.

Les modèles de langage sont utilisés pour de nombreuses applications : générer des textes dans un style donné, traduire automatiquement d'une langue à l'autre, résumer ou synthétiser des documents, répondre à des questions…

GPT est loin d'être le seul grand modèle de langage : tous les géants du numérique s'intéressent actuellement au sujet. Il existe également des modèles de langage open source, citons notamment le modèle développé par la start-up française Mistral.

Le tableau ci-dessous récapitule les principaux modèles de langage.

Modèle de langage	Concepteur	Caractéristiques
GPT-3;5/ GPT-4	OpenAI	La version 3.5 a été mise à disposition du public fin 2022 via l'interface ChatGPT. La version 4 (2023), plus puissante, est payante
LaMDA/ PaLM/Gemini	Google	LaMDA (2021) est un modèle de langage factuel plutôt destiné à fournir des informations factuelles PaLM (2022) est plus généraliste. Gemini (2023) est multimodal (texte et multimédia)
LlaMA	Meta	Partiellement open source, polyvalent, entraîné sur un corpus de 20 langues. Code LlaMA génère du code.
Claude2/ Claude3	Anthropic	Partiellement ouvert. IA « constitutionnelle » basée sur de principes éthiques fondamentaux inspirés des droits humains
Bloom	Hugging Face, soutenu par le CNRS	Open source, entraîné sur un corpus de 46 langues, diffusé avec une licence responsable
Mistral 8x7B	Mistral (start-up française)	Modèle open source, compréhension de plusieurs langues, génération de code
YaLM	Yandex	Modèle open source anglo-russe
Ernie Titan	Baidu	Modèle de langage en langue chinoise

8 https://arxiv.org/pdf/2005.14165.pdf

La mise à disposition de ces modèles de langage auprès du grand public via des robots conversationnels suscite des débats intenses. S'ils sont capables de générer des textes, comprennent-ils vraiment le sens des questions posées (et des réponses qu'eux-mêmes fournissent) ? Ou bien peut-on les comparer à des « perroquets » qui se contentent de reproduire des contenus sur lesquels ils ont été entraînés ? Il a été constaté par ailleurs une tendance parfois à l'« hallucination », c'est-à-dire des affirmations fausses ou inventées de toutes pièces par l'IA. Les IA génératives ne sont pas exemptes non plus de biais, dans la mesure où elles peuvent reproduire des biais présents dans leurs données d'entraînement.

À partir du printemps 2023, Microsoft a intégré ChatGPT à Bing, donnant ainsi des possibilités conversationnelles au moteur, sous le nom de Bing Chat, rebaptisé depuis Copilot. Google a riposté avec son service Google Bard, devenu ensuite Gemini : nous détaillerons au chapitre 2 ces nouvelles fonctionnalités. Le nouvel assistant Windows Copilot propose en plus de générer du texte dans un style donné, à partir d'une thématique.

1.4 Les data sont-elles forcément big ?

Nous avons évoqué plus haut l'importance du « big data » dans l'univers numérique. Les chiffres atteignent en effet des sommets vertigineux dès qu'il s'agit des données : selon le cabinet IDC, la taille de l'univers numérique devrait atteindre d'ici 2035 2000 zettaoctets de données[9]. Il convient de différencier la provenance de ces données massives : elles sont générées en temps réel, non seulement par les humains (via entre autres les médias sociaux), mais aussi par des capteurs, ou encore des transactions via des systèmes d'informations ou objets connectés. Les big data ne sont pas toutes en ligne, elles peuvent être stockées dans des « data centers » ou autres « data lakes » au sein d'une entreprise.

En quoi ces données massives peuvent-elles être utiles pour la recherche d'information et la veille ? Tout d'abord, n'oublions pas que toutes les data ne sont pas forcément « big » et il existe une grande variété de qualificatifs dans le vocabulaire de la donnée :

- On parle de « small data » pour désigner des données dont la lecture et la compréhension ne nécessitent pas de recourir à des outils d'analyse complexes. Les « small data » ne s'opposent pas aux « big data », elles en constituent plutôt un sous-ensemble, plus facile à manipuler ou à appréhender.
- Le phénomène de l'« open data » concerne la libération et la réutilisation de données généralement publiques, illustrées en France par la plateforme data.gouv.fr, qui agrège des jeux de données en formats ouverts et réutilisables, provenant d'institutions et collectivités. De nombreux gouvernements, institutions ou collectivités proposent désormais leurs données dans des formats ouverts. C'est le cas également des données produites par les chercheurs. Ce phénomène est particulièrement intéressant pour la recherche d'information et la veille.

9 https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/

- Le concept de « smart data » désigne un dispositif d'analyse de données en temps réel et directement depuis leur source, et pertinentes par rapport à une problématique (par exemple, analyse marketing du sentiment exprimé par les utilisateurs d'un produit, diagnostic des équipements et maintenance préventive dans l'industrie, etc.). Le « smart data » apporte ainsi une nouvelle dimension issue de l'intelligence artificielle à la « business intelligence ». Nous développerons ces points au chapitre 4.
- La notion de « dark data » (à ne pas confondre avec le « dark web ») correspond à d'innombrables données non utilisées et non analysées, souvent par méconnaissance de leur existence au sein d'une entreprise, ou de leur intérêt potentiel.
- Quant au concept de « linked data », il a été défini dès le milieu des années 2000 par Tim Berners-Lee et utilisé dans le cadre du web de données (voir plus haut) pour désigner ce vaste réseau d'informations liées entre elles. De grandes bases de données ouvertes comme DBpedia ou Geonames servent de référentiels destinés à lier les données entre elles et constituer ce « grand graphe géant » qui rendra la recherche web plus efficace et dynamique. Les liens entre ces « linked data » sont effectués par et pour des machines, de façon à rendre ces données interopérables entre elles, et faciliter la compréhension de ces données par des outils automatisés. Le principe des linked data s'appuie entre autres sur le RDF, langage de base du web sémantique.

Une étude menée par le cabinet Qlik portant sur « Les salariés européens et leurs connaissances sur la data », pointe les lacunes des Français dans le domaine de la data et les présente comme des « datanalphabètes »[10]. La « data literacy » consiste en la capacité à rechercher, lire, manipuler, mais surtout analyser des jeux de données et les utiliser dans le processus de décision.

Dans le domaine de la recherche d'information et de la veille, on aura principalement affaire à de l'« open data » en tant que données brutes à traiter, ou bien à des « smart » ou « small » data, jeux de données pertinents par rapport à une problématique. Un parallèle peut être établi avec le datajournalisme, ou journalisme de données, qui consiste à exploiter des données pour mettre en évidence des faits, des relations, des tendances…

Le « data scientist » ou le « chief data officer » sont davantage concernés par la gestion et l'analyse des big data ou l'identification des « dark data ». Quant aux « linked data », elles sont surtout exploitées par des architectes de l'information ou promoteurs du web de données.

Le chapitre 4 présentera un panorama des principales méthodes de datavisualisation, utiles aux professionnels de la veille ou du datajournalisme.

10 http://www.influencia.net/fr/actualites/tendance,etudes,francais-sont-analphabetes-data,7819.html

1.5 Réseaux sociaux et messageries instantanées : de nouveaux usages

Apparus il y a déjà une vingtaine d'années, les réseaux sociaux ont bouleversé le paysage de la publication et du partage d'information, et se sont peu à peu invités dans la panoplie des solutions de recherche d'information et de veille.

Le monde des réseaux sociaux a connu au cours des dernières années plusieurs bouleversements : disparitions, rachats, nouveaux entrants. Si le marché semble se stabiliser autour de plusieurs grands acteurs, de nouveaux venus apportent des usages innovants.

Actuellement, Facebook domine toujours le secteur (sans doute en partie grâce à la pandémie) avec en 2023 3 milliards d'utilisateurs actifs mensuels dans le monde (et 33 millions en France)[11]. Le groupe Meta détient par ailleurs non seulement Facebook et Instagram, mais aussi les applications de messagerie Messenger et WhatsApp. Nous détaillerons au chapitre 2 la stratégie actuelle du groupe, qui souhaite élargir son périmètre bien au-delà de ses objectifs originaux que sont le divertissement et la mise en relation des internautes. Confronté à un vieillissement de ses utilisateurs (ce phénomène étant relatif d'un pays à l'autre !), le réseau Facebook a été doté au fil des années de nouvelles fonctionnalités (stories, reels, marketplace…). En 2023, Meta a lancé un nouveau service de microblogage appelé Threads, dont l'objectif est de concurrencer Twitter/X.

Ce dernier est en effet en plein bouleversement depuis son rachat par Elon Musk fin 2022. Twitter, qui avait réussi à rester indépendant depuis sa création en 2006, présentait des caractéristiques intéressantes pour la recherche et la veille (ouverture, souplesse, instantanéité, variété des sources d'informations), mais aussi des aspects négatifs (influence, manipulation, propos haineux et polémiques). Rebaptisé X par son nouveau propriétaire, le réseau souffre actuellement d'une gestion pour le moins chaotique et semble s'orienter vers une sorte d'application totale centralisant plusieurs services, notamment bancaires et marchands.

Plusieurs membres de Twitter/X ont récemment quitté le réseau pour protester contre la gestion d'Elon Musk, ainsi que ses prises de position politiques. Mais les alternatives à Twitter ne sont pas si nombreuses : le réseau open source et décentralisé Mastodon peine à attirer un nombre conséquent d'utilisateurs étant donné la complexité de son interface.

Le nouveau service Threads de Meta a du mal à fidéliser son public, une fois l'effet de curiosité estompé. L'alternative la plus solide pourrait être BlueSky, dont l'interface est très proche de celle de Twitter. À l'origine, Bluesky est une initiative créée en 2019 par Twitter autour d'une architecture décentralisée. Devenue indépendante, la plateforme est ouverte à tous depuis février 2024.

11 https://www.blogdumoderateur.com/chiffres-facebook/

En raison de la viralité forte des images et vidéos, les réseaux multimédias se taillent la part du lion : YouTube dépasse les 2,4 milliards d'utilisateurs actifs mensuels dans le monde ; Instagram en compte près de 2 milliards. Nous étudierons au chapitre 2 les modalités de recherche via ces réseaux, et au chapitre 3 leur intérêt pour les différents types de veilles. Les réseaux Snapchat et TikTok ont conquis la génération « selfie » grâce entre autres au caractère momentané et très visuel de leurs contenus. La progression fulgurante de TikTok a incité ses concurrents à proposer des formats similaires de courtes vidéos (les reels sur Instagram ou les « shorts » sur YouTube). Les podcasts et formats audio deviennent également viraux grâce aux réseaux sociaux.

TikTok serait-il en passe de devenir « le Google de la génération Z » ? C'est du moins ce que suggèrent plusieurs études, dont une enquête menée en 2024 par Adobe[12]. Une majorité de jeunes de moins de 25 ans utiliseraient en effet ce réseau comme moteur de recherche, et le préfèrent à des moteurs établis comme Google. Les tutoriels vidéo, ainsi que les formats courts et personnalisés de TikTok ont un fort pouvoir d'attractivité, notamment auprès des plus jeunes générations.

La plateforme Twitch, leader dans le streaming des jeux vidéo et rachetée en 2014 par Amazon, devient le vecteur de nouveaux modes de production de contenus et d'interactions, et suscite un engouement de la part de politiques désireux de toucher un public jeune.

Autre plateforme issue du monde des jeux vidéo, Discord a également considérablement élargi son public depuis la pandémie, et a pu servir à la diffusion de cours en ligne, de messagerie professionnelle…

Du côté des réseaux professionnels, LinkedIn, racheté en 2016 par Microsoft, domine le secteur avec plus de 985 millions d'utilisateurs dans le monde, dont plus de 28 millions en France, et a pu profiter du phénomène de désaffection qui touche Twitter/X. Nous étudierons également en détail les méthodes de recherche et veille associées à ce réseau.

La sphère universitaire est également impactée par le phénomène des réseaux sociaux académiques, permettant aux chercheurs d'échanger et d'interagir entre eux (on parle alors de « science 2.0 ») et de promouvoir leurs publications de recherche. Nous verrons au chapitre 3 l'intérêt de réseaux comme ResearchGate (plutôt orienté sciences exactes) ou Academia (plutôt orienté sciences sociales) pour la veille stratégique. Espaces de convivialité, les réseaux sociaux académiques se distinguent des plateformes de dépôt d'archives ouvertes (comme HAL mis en place par le CNRS) dans la mesure où ils n'ont pas vocation à l'archivage pérenne. Certains éditeurs de publications scientifiques considèrent les réseaux sociaux académiques comme une concurrence déloyale. En 2017, ResearchGate était visé par une plainte de plusieurs éditeurs scientifiques, dénonçant une violation du droit d'auteur. Des

12 https://www.adobe.com/express/learn/blog/using-tiktok-as-a-search-engine

milliers d'articles ont dû être retirés. Le conflit juridique s'est éteint en 2023 à la suite d'un accord permettant aux auteurs ayant publié des articles de recherche dans les journaux de l'ACS et d'Elsevier de partager leurs travaux sur la plateforme ResearchGate de manière conforme au droit d'auteur.

La pandémie a favorisé par ailleurs l'usage des messageries privées : chaque jour, des centaines de milliards de messages s'échangent sur WhatsApp et autres applications de messagerie, y compris dans la sphère professionnelle.

Parmi les principaux concurrents de WhatsApp, citons essentiellement Telegram et Signal.

Créé en 2013 par deux frères russes opposants au pouvoir, Telegram est une application de messagerie cryptée réunissant plus de 800 millions d'utilisateurs dans le monde. Les chaînes sont un outil de diffusion de messages à un grand nombre de personnes simultanément. Il est possible d'y effectuer des recherches par mots-clés. Telegram est beaucoup utilisé dans un contexte de communication politique ou de militantisme public, ce qui en fait un outil intéressant pour la veille, tout en étant bien entendu vigilant sur la nature des publications.

Se pose bien sûr la question de la confidentialité des échanges reposant sur un cryptage des contenus, qui n'empêche malheureusement pas toujours les cyberattaques.

De façon plus générale, cette évolution des réseaux sociaux vers des contenus de plus en plus multimédias, privés et éphémères est-elle compatible avec la pratique de la recherche et de la veille ? Nous verrons dans les chapitres suivants comment contourner ces difficultés et rechercher ou collecter des contenus de qualité via les différents réseaux sociaux.

Car on constate depuis déjà plusieurs années la convergence – et la concurrence accrue – entre moteurs de recherche et réseaux sociaux. Une grande partie des internautes – et notamment pour des recherches très « quotidiennes » – sont en attente en effet de résultats concis et précis… que ne fournissent pas toujours les moteurs classiques, mais que les réseaux sociaux sont plus à même de fournir.

D'un autre côté, le monde des réseaux sociaux est marqué par des phénomènes de manipulation d'information ou d'influence (les fameuses « fake news »), et de redondance due aux nombreux relais de messages. Le danger des « bulles de filtres », ou enfermement algorithmique, a été pointé dès 2011 par le militant Eli Pariser : l'utilisateur de réseaux sociaux aurait tendance à s'isoler dans une sorte de bulle de « pensée unique », et n'échangerait qu'avec des personnes partageant les mêmes opinions ; les algorithmes auraient tendance à « pousser » vers l'utilisateur, afin de le fidéliser, des contenus correspondant à ses opinions. Outre le phénomène des risques de bulles de filtres, les réseaux sociaux peuvent donner lieu à des risques de manipulations politiques et d'atteintes à la vie privée. Le scandale « Cambridge Analytica » a

éclaté en 2018 lorsqu'il a été révélé que cette officine britannique avait collecté les données personnelles de millions d'utilisateurs de Facebook sans leur consentement. Ces données auraient été utilisées pour influencer l'opinion publique sur diverses campagnes politiques, notamment l'élection présidentielle américaine de 2016 et le référendum sur le Brexit au Royaume-Uni.

1.6 Où en sont les blogs, wikis et forums ?

Blogs et wikis sont deux formes de publication ne nécessitant pas de compétences techniques pour leur utilisation. Ces deux formes illustrent parfaitement les deux tendances complémentaires du web 2.0 : personnalisation et mutualisation. Un blog est bien sûr un espace personnel, dans lequel des billets sont publiés et organisés par ordre antéchronologique. L'auteur du blog est généralement le seul à publier, les billets pouvant être commentés par les internautes. Un wiki est davantage destiné à la publication collaborative : plusieurs personnes peuvent participer à l'élaboration et à la modification des contenus. L'apparence est plus neutre, et les articles sont organisés généralement sous forme d'une arborescence de thèmes.

Où en sont les blogs et wikis aujourd'hui ? Après avoir rencontré un grand succès au milieu des années 2000, la blogosphère a beaucoup évolué et s'est en quelque sorte « fondue » dans la masse du web. Peu de choses distinguent désormais un blog d'un site créé par exemple avec la plateforme Wordpress. Ceci étant, si les blogs à usages personnels ont quasiment disparu, il existe encore de nombreux blogs d'experts, de journalistes ou de chercheurs (par exemple les carnets de recherche sur le site Hypothèses.org), dont les contenus peuvent apporter des éléments éclairants, notamment pour la veille stratégique. Les blogs ont également évolué vers des formats multimédias : le « vlog » désigne un blog diffusant essentiellement des vidéos, en lien avec la plateforme YouTube. De façon générale, les blogs sont souvent liés à une présence dans le web social, les réseaux faisant fonction de « caisse de résonance » aux publications.

On rencontre sur le web relativement peu de wikis spécialisés : la célèbre encyclopédie Wikipédia (et ses produits « frères » Wiktionnaire, Wikisource, etc.) laisserait-elle trop peu de place aux autres wikis ? Nous verrons au chapitre 2 les « bonnes pratiques » concernant l'utilisation de Wikipédia pour la recherche et la veille.

Les forums ont fait leur apparition dès les premiers temps de l'Internet : le système de forums Usenet remonte en effet à la fin des années 1970 ! Les forums web ont démarré dans les années 1990, et certains restent très fréquentés, notamment sur des questions grand public comme la vulgarisation médicale, ou encore sur des thèmes ayant trait aux nouvelles technologies (assistance informatique, programmation, jeux vidéo…).

Reddit est un cas un peu particulier : à la fois forum de discussion et réseau social, cette plateforme très populaire (surtout aux États-Unis) lancée en 2005 permet à ses utilisateurs de discuter de sujets en commun, mais aussi de se connecter entre eux et

d'interagir. Reddit est organisé en communautés thématiques appelées subreddits.

À l'heure des réseaux sociaux temps réel et des messageries instantanées, des espaces de publications asynchrones comme les forums continuent ainsi à se maintenir et peuvent présenter un intérêt pour la veille, notamment dans le cadre de la collecte et analyse de l'opinion.

1.7 Un univers numérique toujours plus multimédia

Les premiers sites apparus dans les années 1990 étaient encore très textuels, et, pour des raisons liées aux techniques et capacités de l'époque, laissaient peu de place au multimédia.

Le web regorge aujourd'hui de contenus multimédias dont voici une brève typologie :

- Images : on trouve à profusion sur le web photos, images ou dessins numérisés. Les principaux formats associés sont les suivants : JPG, GIF, TIFF, PNG… et correspondent à différents types d'encodage ou de compression. Attention, toutes les images ne sont pas libres de droits.
- Vidéos : les supports vidéo sur le web ne se consultent pas que via les réseaux sociaux ! Ils servent au divertissement (films, clips, courts métrages), mais aussi à l'enseignement avec une déferlante de cours en ligne sous la forme MOOC (massive open online course) ou SPOC (Small private online course), plus personnalisés et interactifs. Plusieurs plateformes de streaming légales sont proposées, de façon généralement payante, aux internautes.
- Sons : là encore, le streaming payant s'est généralisé avec des plateformes comme Spotify, Deezer ou Qobuz. Mais il existe également des solutions open source de publication et partage de fichiers musicaux comme Soundcloud. Quant aux podcasts, si le principe n'est pas nouveau, cette forme de diffusion de contenus audio a connu un très fort engouement depuis 2019, et de nombreuses plateformes de diffusions sont apparues, dans plusieurs domaines, en complément des médias audio traditionnels.

Face à cette déferlante de nouveaux contenus multimédias, la recherche textuelle classique via des mots-clés demeure assez peu efficace. Selon une expression reprise maintes fois, « la caméra est le nouveau clavier[13] » ! De nouvelles technologies, que nous détaillerons dans les prochains chapitres, se sont développées, parmi lesquelles :

- La recherche inversée d'images : on recherche des images à partir non pas d'un mot-clé, mais d'une autre image, ou d'une forme.
- L'analyse automatisée d'images : les géants du numérique élaborent actuellement des systèmes sophistiqués d'analyse d'images à travers des techniques d'intelligence artificielle incluant des systèmes de reconnaissance faciale, ou de recommandation automatisée.

13 https://www.wired.com/2017/05/camera-wants-kill-keyboard/

- Le « speech to text » : retranscription du contenu d'un fichier audio sous forme de texte. Cette technique est utilisée entre autres par YouTube pour générer « à la volée » des sous-titres à une vidéo, à partir d'un fichier son retranscrit sous forme de texte et traduit automatiquement (parfois de façon encore approximative !).
- Le « text to speech » : à l'inverse, ce système transforme un texte écrit en langage parlé. On trouvera des applications dans le monde des assistants virtuels et des objets connectés.
- La recommandation musicale : des moteurs de découverte et recommandation musicale sont basés sur des algorithmes spécifiques et analysent automatiquement les métadonnées associées aux fichiers musicaux.

1.8 Expériences immersives : le bilan mitigé des métavers

En 2021, la société Facebook prenait le nom de Meta : ce changement reflétait (entre autres motivations) l'ambition de l'entreprise de créer des métavers, c'est-à-dire des mondes virtuels immersifs destinés à de nombreuses applications dans les domaines du travail collaboratif, de l'éducation, du commerce en ligne, du jeu…

Rappelons à ce propos le principe de la réalité virtuelle, dans laquelle l'utilisateur se trouve plongé dans la simulation d'une situation, grâce à une immersion totale au milieu d'images 3D réagissant en temps réel aux actions de l'utilisateur, et ce via un casque de réalité virtuelle. Ces technologies sont à distinguer de la réalité augmentée, plus proche de la « vraie » réalité, puisqu'elle consiste en une superposition au monde réel de données ou images virtuelles diffusées via un casque, mais aussi un smartphone ou des lunettes spéciales.

Le metavers est encore en cours de développement, mais n'a pas à ce jour atteint le succès escompté par Meta et d'autres acteurs du secteur. Ce phénomène peut être dû à de nombreux freins techniques ou économiques, tels que le coût des équipements nécessaires, la complexité de l'utilisation, une expérience utilisateur encore peu maîtrisée, sans oublier les problèmes de sécurité des données ou d'impact écologique[14]. À ce jour, les cas d'usage des metavers concernent surtout certains jeux vidéo, ainsi que la formation professionnelle ou la maintenance industrielle. Pour autant, nous sommes encore loin d'un déploiement massif.

1.9 L'OSINT : un immense champ d'investigation

L'OSINT (ou intelligence des sources ouvertes) consiste à identifier et exploiter des sources d'information publiques. Cette notion parfois un peu floue englobe un large éventail de sources : médias, réseaux sociaux, documents institutionnels, sites classiques, forums, images satellites, etc. L'OSINT est un atout pour la recherche avancée et la veille, dans la mesure où sa pratique offre

14 https://blog.octo.com/metavers-humain-environnement-limites

Information & stratégie

L'essor des intelligences artificielles génératives révolutionne les accès à l'information, dans le monde professionnel comme dans la sphère universitaire. Plus que jamais, la maitrise des nombreuses méthodes et approches de la recherche d'information, de la veille et de l'analyse stratégique est indispensable.

Rédigé par une experte en veille stratégique, ce livre actualise les connaissances dans les domaines de la recherche d'information et de la veille sur Internet. Il permet de comprendre le fonctionnement des moteurs de recherche à l'ère de l'intelligence artificielle, d'optimiser les recherches web, d'identifier des sources d'information méconnues, de mettre en place différents types de veille stratégique et de découvrir un éventail de méthodes d'analyse automatisée de l'information.

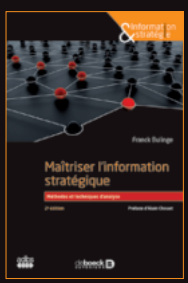
Pratique et opérationnel grâce à ses exemples concrets, ce manuel permet à tout internaute de trouver, collecter, qualifier et analyser l'information réellement utile : un véritable guide de survie dans la jungle de l'information numérique.

–

VÉRONIQUE MESGUICH est consultante-formatrice dans le domaine de la méthodologie de veille et du management stratégique de l'information. Elle intervient dans de nombreux établissements d'enseignement supérieur, dont l'École européenne d'intelligence économique (EEIE), l'École des bibliothécaires et documentalistes (EBD) et le CNAM, ainsi qu'en entreprise. Elle a dirigé l'Infothèque du Pôle universitaire Léonard de Vinci et a été coprésidente de l'ADBS.

ANNE-MARIE LIBMANN (préface) est directrice des opérations de FLA Consultants et de BASES Publications.

DÉCOUVREZ AUSSI :



www.**deboecksuperieur**.com

27,90 €
ISBN 978-2-8073-6140-9

Créée en 1963, l'ADBS (Association des Professionnels de l'information) fédère une grande variété de professionnels de l'information numérique (veilleurs, knowledge managers, gestionnaires de contenus numériques, documentalistes, records managers, etc.). Elle compte près de 3000 membres en France. En 2013, l'ADBS s'associe aux éditions De Boeck pour créer la collection « Information & Stratégie » qui allie les savoir-faire des deux partenaires.

adbs

9 782807 361409

Design **cerise.be**