



Big data : comment les algorithmes influencent notre usage d'Internet ?

Repères

Le CSEM et l'éducation aux médias en bref



Le Conseil supérieur de l'éducation aux médias (CSEM) a pour missions principales de promouvoir l'éducation aux médias et de favoriser l'échange d'informations et la coopération entre tous les acteurs et organismes concernés par l'éducation aux médias en Fédération Wallonie-Bruxelles ; notamment les secteurs des différents médias, l'enseignement obligatoire et l'éducation permanente. Le CSEM assure une large diffusion de toutes ces initiatives via le site internet www.csem.be.

Le big data en bref

Littéralement le *big data* (mégadonnées en anglais) désigne des bases de données dont la taille est devenue si gigantesque qu'il est impossible de les appréhender par des moyens humains. Pour traiter ces informations, les entreprises du web recourent à des formules mathématiques appelées algorithmes. Le terme *big data* est moins utilisé pour désigner l'accumulation de données que l'impact discret, mais considérable, des algorithmes sur notre usage d'Internet.

L'éducation aux médias (EAM) permet à chaque citoyen d'améliorer sa compréhension des médias, d'en maîtriser les usages et de poser un regard critique sur ceux-ci. L'objectif de l'EAM est de transmettre à tous (enfants, ados, adultes) des savoirs, savoir-faire et des compétences indispensables à une utilisation éclairée des différents médias. Les mini-dossiers du CSEM se penchent plus spécifiquement sur le média internet.

<http://csem.be/csem/missions>

http://csem.be/csem/textes_positionnement/education

Le CSEM et le *big data*

Le CSEM propose des pistes concrètes pour comprendre et tenter de maîtriser l'influence du *big data* et des algorithmes sur notre vie en ligne. Cet outil s'adresse aussi bien aux familles qu'aux professionnels de l'éducation et de l'animation.



Le business des algorithmes

Grâce au *big data* et aux algorithmes qui permettent d'exploiter ces données, les entreprises actives sur le web influencent nos comportements pour générer des bénéfices plantureux.

Les Belges passent en moyenne plus de quatre heures par jour sur Internet et cette consommation est souvent plus importante chez les jeunes¹. Pendant tout ce temps passé en ligne, nous laissons des traces sans toujours nous en apercevoir. Les liens sur lesquels nous cliquons, les achats que nous réalisons, les photos que nous partageons, les vidéos que nous consultons, les messages que nous rédigeons, l'endroit depuis lequel nous surfons et bien d'autres informations qui nous concernent viennent nourrir d'énormes bases de données (*big data*). On estime que les internautes génèrent autant d'informations en deux jours actuellement que ce qu'a produit l'humanité entière depuis sa naissance jusqu'au début du web².

Pour traiter ces informations, les entreprises utilisent des algorithmes. Ces formules mathématiques sont omniprésentes sur le web : elles hiérarchisent l'information des moteurs de recherche comme Google, décident de ce qui apparaîtra ou non dans notre fil d'actualité sur Facebook, font apparaître des publicités ciblées et nous suggèrent de nouveaux produits à glisser dans le panier quand nous faisons nos courses sur des sites de commerce, nous recommandent des livres, de la musique ou des films à découvrir sur Spotify, Netflix, Amazon et autres plateformes qui vendent des services ou des produits en ligne.

Les algorithmes nous proposent une navigation personnalisée en comparant notre profil à ceux d'utilisateurs qui sont censés nous ressembler pour prédire nos comportements et devancer nos attentes. Sans ces services, nous passerions beaucoup plus de temps à trouver l'information dont nous avons besoin. Mais l'objectif *in fine* est avant tout d'influencer nos comportements dans un but commercial, nous pousser à consommer des produits ou passer plus de temps sur des sites. Le *big data* peut aussi être utilisé pour servir des campagnes ou de la propagande politique.

Algorithme : procédure informatique destinée à résoudre un problème. Il existe toute sorte d'algorithmes, plus ou moins complexes. La machine à laver fournit un exemple d'algorithme simple que nous utilisons tous les jours. Si je choisis le programme lavage délicat -> l'eau chauffe à 30 degrés -> si je choisis le programme coton -> l'eau chauffe à 60 degrés -> quand le lavage est fini -> mettre essorage court pour programme délicat -> etc.

¹ « Les Belges et Internet », Digimedia, 2016

² « À quoi rêvent les algorithmes. Nos vies à l'heure des big data », Dominique Cardon, La République des idées, octobre 2015



Données et vie privée

La vie privée est-elle devenue un concept presque démodé ? Dévoiler des informations personnelles sur Internet est d'autant plus facilement entré dans les mœurs que c'est souvent le seul moyen d'accéder à une série de services et d'outils en ligne. Pourtant, l'usage que la plupart des entreprises du web font de nos données devrait nous inciter à plus de prudence.

Services GPS, messageries, agendas, carnets d'adresses, réseaux sociaux, moteurs de recherches, espaces de stockage à distance (cloud)... Les entreprises actives sur le web ne mettent pas ces outils « gratuitement » à la disposition des consommateurs par philanthropie. Les données à caractère personnel que ces sites récupèrent constituent pour elles une mine d'or. **Sans discernement, les entreprises collectent toutes les données qui leur tombent sous la main bien qu'une partie seulement leur sera utile.** En 2016, le marché des données des citoyens s'élevait à 60 milliards d'euros rien que pour les 28 pays de l'Union européenne³.

En s'inscrivant sur un site, rarissimes sont les internautes qui ont la patience de sortir une loupe pour lire les conditions d'utilisation déclinées dans des pop-up kilométriques. Par ailleurs, plusieurs scandales dévoilés par la presse ont montré que l'usage fait des données peut parfois dépasser largement ce à quoi l'utilisateur pense consentir. **Facebook, par exemple, a vendu à ses partenaires commerciaux l'accès aux données les plus personnelles d'une partie de ses utilisateurs sans leur autorisation**⁴. Des dictatures utilisent les données des réseaux sociaux pour surveiller des citoyens, mais cela se pratique aussi dans des pays démocratiques, comme l'ont démontré les révélations d'Edward Snowden dénonçant comment les services secrets américains collaboraient avec les géants du web pour avoir accès aux données des internautes américains. Quant aux fuites, elles sont monnaie courante. En mars 2018, un lanceur d'alerte canadien a

dénoncé comment l'entreprise Cambridge Analytica aurait exploité une faille de sécurité de Facebook pour récolter les données de 87 millions d'internautes et influencer les élections aux États-Unis et le référendum en Angleterre.

Dévoiler un morceau de sa vie privée pour accéder à des services en ligne est rentré dans les mœurs. **Que ce soit au niveau des relations, des études, du travail ou des loisirs, ces outils font dorénavant partie de notre monde au quotidien et si nous ne cédon pas sur notre vie privée, nous risquons de nous exclure de toute une série d'opportunités.**

Les quelques exemples cités ci-dessus devraient pourtant nous encourager à faire preuve de plus de vigilance. D'autant plus que la récolte d'informations peut s'étendre à des domaines parfois intimes ou sensibles : données de santé glanées via des SMS, des mails ou des objets connectés, données financières, orientations politiques, sexuelles ou religieuses, etc.

Pour en savoir plus sur la protection de votre vie privée, le carnet de la collection Repères « Qui suis-je sur le web ? Construire son identité numérique » peut être téléchargé ou commandé gratuitement sur le site du CSEM : <http://www.csem.be>

³ Source : International Data Corporation (IDC)

⁴ « As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants » Gabriel J.X. Dance, Michael LaForgia and Nicholas Confessore, New York Times, December 2018



RGPD, QUATRE LETTRES POUR PROTÉGER LA VIE PRIVÉE

Pour mieux protéger la vie privée de ses citoyens, l'Union européenne a adopté, en mai 2018, le règlement général sur la protection des données (RGPD). Cette base légale permet, par exemple, à un internaute de demander aux entreprises de lui communiquer les données qu'elles ont conservées à son propos. Elle lui donne aussi le droit, si certaines conditions sont remplies, d'en obtenir la modification voire l'effacement. L'ensemble des droits donnés par le RGPD et les démarches à suivre pour les appliquer est décrit sur : https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens_fr

Vous trouverez également des informations à ce propos sur le site de l'Autorité de protection des données, l'organe de contrôle indépendant chargé de veiller au respect de la protection des données en Belgique : <https://www.autoriteprotectiondonnees.be/reglement-general-sur-la-protection-des-donnees-citoyen>

Quand la technologie formate la pensée

En balisant notre parcours sur le web, les algorithmes nous empêchent de profiter d'Internet dans toute sa diversité.



Les algorithmes se révèlent indispensables pour trouver son chemin dans la masse d'informations présentes sur Internet si bien que nous sommes devenus dépendants de ces formules discrètes qui anticipent nos choix et prennent des décisions à notre place. À partir des traces que nous laissons sur Internet et du « portrait » qui en émerge, les algorithmes filtrent les contenus et les classent pour nous proposer un parcours personnalisé au sein des possibilités infinies offertes par le web, avec pour conséquence d'amenuiser cette diversité qui fait justement tout l'intérêt du réseau. Imaginez que vous vous promeniez dans une librairie qui contient tous les ouvrages du monde, mais où un vendeur s'obstinerait à vous proposer uniquement des livres qui vous cantonnent dans un seul registre !

Les algorithmes contribuent de plus en plus à peser sur nos choix culturels, nos opinions politiques, voire nos vies sentimentales pour ceux et celles qui fréquentent les plateformes de rencontre. Ils nous enferment dans la case créée par notre « profil » et, sans nous en rendre compte, nous pensons de moins en moins à ce que nous souhaitons réellement ni à ce que nous pourrions découvrir d'autre.

Les enjeux soulevés par l'omniprésence des algorithmes sont aussi économiques. Des sites web, notamment des médias d'information, ont parfois vu leur trafic chuter dramatiquement du jour au lendemain parce que Google ou Facebook avaient changé l'algorithme dont ils dépendent pour leur visibilité.

Des formules obscures

Les internautes sont en droit d'exiger plus de transparence concernant le fonctionnement de ces formules qui ont une influence non négligeable sur nos comportements.

Malheureusement, il est presque impossible de savoir comment fonctionnent ces algorithmes. Sur quelles données basent-ils leur décision ? Quel objectif leur a-t-il été assigné ? Qui est responsable en cas de fraude ou de *bugs* ? Sans parler du fait que certains algorithmes ont acquis de tels degrés de complexité et sont modifiés si régulièrement que même les spécialistes ne peuvent pas toujours expliquer comment le programme est arrivé à telle décision. Nous avons confié une partie importante de notre vie numérique à des boîtes noires. Mais les algorithmes ne sont pas des formules neutres. Aux États-Unis, l'algorithme d'Amazon ne proposait pas les mêmes services de livraison aux clients habitant dans des quartiers populaires. Prenant en compte pour prendre ses décisions des critères comme la distance jusqu'au centre de traitement, la demande locale, le nombre de clients « Prime », l'algorithme d'Amazon accentuait des discriminations existantes⁵. Plusieurs autres exemples montrent que les algorithmes peuvent contribuer à renforcer les préjugés sociaux, racistes ou sexistes de la société, que ce soit à cause des humains qui les ont conçus ou par des biais existants dans les bases de données qui les nourrissent.

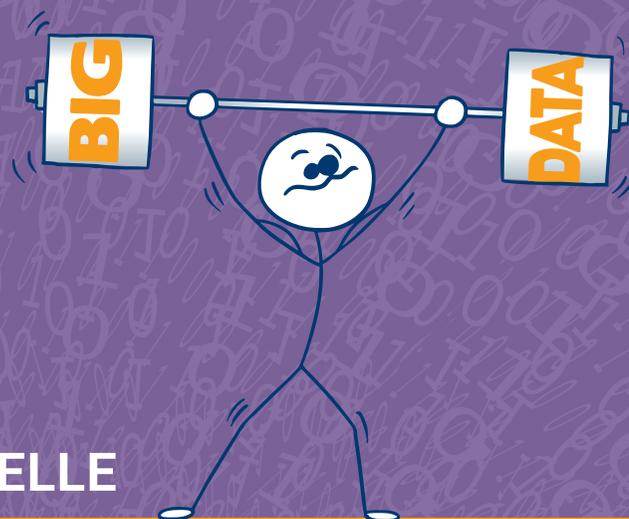
La révolution engendrée par le *big data* et ses algorithmes fait l'objet de nombreux questionnements. Davantage de transparence doit être exigée de la part des sociétés qui recourent à ces procédés⁶. Le modèle économique des grandes entreprises actives sur le web repose sur le secret et la propriété intellectuelle, mais il existe aussi des algorithmes dont le code est libre d'accès. Pour contrecarrer certains des effets pervers cités ci-dessus, des expériences sont aussi menées pour créer des algorithmes qui évitent de reproduire les stéréotypes ou qui favorisent la découverte culturelle. D'autres projets encore ont pour objectif d'encourager la diversité dans l'univers très masculin des programmeurs.

⁵ « Amazon same-day delivery less likely in black areas, report says », Elizabeth Weis, USA Today, 2016.

⁶ « A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation », Commission européenne, mars 2018



BIG DATA ET INTELLIGENCE ARTIFICIELLE



Avec l'avènement du *big data*, les travaux sur l'intelligence artificielle ont pris un nouveau tournant. Les premières recherches en matière d'intelligence artificielle, un concept développé dans les années 50, avaient pour objectif d'apprendre à la machine à reproduire un raisonnement. Aujourd'hui, les machines sont nourries avec des masses gigantesques de données qu'elles analysent pour prédire, décider, apprendre (*machine learning*). Plus un algorithme traite d'informations, plus il devient performant. Leur intelligence est statistique. Ils comparent plus qu'ils ne comprennent. Les premiers programmes de traduction automatique se basaient sur les règles de grammaire. Aujourd'hui, Google Translate compare des bases de données existantes de textes traduits dans différentes langues pour proposer la traduction la plus probable.

Les perspectives ouvertes par le *big data* en termes d'intelligence artificielle dépasse largement le seul domaine de l'Internet. Aujourd'hui, des programmes permettent de détecter des cancers, de recruter des candidats pour un emploi, de rendre des décisions de justice, de gérer les besoins d'électricité d'une ville pour faire des économies d'énergie. Imaginer jusqu'où ces technologies pourront mener revient à lire l'avenir dans une boule de cristal. Mais il importe de pouvoir dès aujourd'hui réfléchir aux règles qui accompagneront ces évolutions, notamment en termes de transparence et de respect de la vie privée⁷.

⁷ Déclaration de Montréal pour un développement responsable de l'intelligence artificielle, 2017, www.declarationmontreal-iaresponsable.com

Conseils : à la maison et à l'école

À moins de crypter vos mails, de fermer tous vos comptes sur les réseaux sociaux et de désactiver la géolocalisation sur votre smartphone, il semble difficile aujourd'hui de ne pas se faire déposséder de ses données privées en surfant sur le web. La révolution **big data** est difficilement contournable. Mais prendre conscience des traces qu'on laisse un peu partout sur Internet et de la façon dont celles-ci sont utilisées pour orienter notre navigation est un premier pas qui peut permettre de limiter l'influence que les algorithmes exercent sur nos vies. Par ailleurs, on voit se développer sur Internet une série d'outils plus respectueux de la vie privée.

- La plupart des moteurs de recherche hiérarchisent les résultats en fonction de la popularité des sites et des données personnelles récoltées à propos des préférences des internautes. Mais il existe des alternatives plus respectueuses de la vie privée que vous pouvez expérimenter avec votre enfant ou à l'école (voir ressources en ligne). Testez aussi un navigateur moins indiscret. À l'instar des moteurs de recherche, le logiciel utilisé pour surfer sur Internet recueille des informations à votre propos. Une entreprise comme Google, qui propose à la fois un moteur de recherche, un navigateur (chrome) et une série de services (messagerie, cartes, etc.) a accès à de très nombreuses données sur ses utilisateurs.
- À la maison, discutez avec votre enfant pour régler votre navigateur afin d'effacer les cookies à la fin d'une session si cela vous semble pertinent. Les cookies sont des petits fichiers déposés sur votre ordinateur par les sites que vous visitez pour obtenir des informations sur votre profil de navigation. Ils peuvent s'avérer pratiques, par exemple, pour retenir vos préférences et vos mots de passe, mais peuvent aussi se révéler fort indiscrets. Certains sites offrent aussi la possibilité de refuser d'emblée les cookies quand vous vous connectez. Malheureusement, votre facilité de navigation et de lecture risque d'en être altérée.
- Apprenez à votre enfant à paramétrer ses appareils et ses comptes. Lorsque vous téléchargez une application ou ouvrez un compte pour utiliser un service en ligne, prenez le temps d'aller avec lui dans les paramètres pour régler les options relatives à la vie privée. N'hésitez pas à renouveler cette opération de temps en temps, car les paramètres de partage des données peuvent changer suite à des mises à jour. Questionnez ensemble la pertinence du partage des données avec l'usage qu'il aura de cette application. Par exemple, accepter la géolocalisation pour installer une lampe de poche sur son smartphone est-il vraiment nécessaire ?
- En groupe ou en famille, faites l'exercice de comparer vos résultats pour une même recherche et d'en discuter.
- Parlez avec votre enfant, ou avec les jeunes dont vous avez la responsabilité, des intentions des entreprises actives sur Internet et de leur modèle économique.
- Faites l'exercice, au moins une fois dans votre vie, de lire jusqu'au bout les conditions d'utilisation et la politique de vie privée d'un site que vous utilisez.

Ressources en ligne :

Le RGPD expliqué aux jeunes dans un kit pédagogique

Le site www.jedecide.be, l'Autorité de protection des données et Mediawijis ont ainsi développé un kit pédagogique permettant aux enfants de se familiariser de manière ludique avec les principes de base du nouveau règlement relatif à la protection des données.

Ce kit comprend :

- un clip animé www.jedecide.be/les-jeunes/RGPD ;
- un poster éducatif www.jedecide.be/les-parents-et-lenseignement/support-pedagogique
- des fiches pédagogiques www.jedecide.be/les-parents-et-lenseignement/support-pedagogique
- un carnet des copains de classe www.jedecide.be/les-parents-et-lenseignement/support-pedagogique.

Les moteurs de recherche suivants déclarent ne pas enregistrer les données de navigation des internautes :

Qwant est un moteur de recherche français qui fonde sa communication sur le respect de la vie privée. Petite particularité, l'option « social » permet de faire une recherche sur les tweets.

Searx est un méta-moteur de recherche qui utilise Google et Bing, mais ne stocke pas les données de navigation des internautes.

DuckDuckGo est un moteur de recherche qui propose aussi une notation des sites en fonction de leur degré de respect de la vie privée.

Les navigateurs suivants mettent en avant le respect de la vie privée dans leur communication :

Firefox est un logiciel de Mozilla, une organisation sans but lucratif qui lutte « pour contrecarrer les plans des grandes entreprises et rendre Internet accessible à tous et partout ». Le code source est ouvert : n'importe qui (de compétent) peut aller le lire et voir comment le navigateur fonctionne réellement. Les dernières versions proposent un très bon paramétrage par défaut, qu'il est encore possible d'affiner au besoin. Les modules Ublock origine (bloqueur, entre autres, de publicités) et Privacy badger (contre les traqueurs) peuvent être ajoutés au navigateur.

Epic Privacy Browser est configuré par défaut pour bloquer tous les cookies, ne pas enregistrer l'historique des sites et des recherches effectuées

UR Browser s'engage à ne pas collecter d'autres données que celles qui sont techniquement nécessaires et à travailler avec des partenaires qui respectent la législation européenne sur la protection de la vie privée.

Big data, big consommation d'énergie

Les données que nous générons en surfant sur Internet sont stockées sur des serveurs dans de gigantesques hangars (data centers). On estime qu'Internet engloutit au total 10 % de la consommation mondiale d'énergie, dont 30 % sont brûlés rien que pour faire tourner ces centres de données⁸.

⁸ « Numérique : le grand gâchis énergétique », par Laure Gailloce, 16 mai 2018, journal du CNRS

LA COLLECTION REPÈRES DU CSEM

Notre collection a pour objectifs d'informer et d'outiller, de façon succincte et concrète, les parents et les professionnels de l'éducation.

Parmi les thèmes traités :

- le cyberharcèlement
- l'identité numérique
- la liberté d'expression
- la désinformation
- s'informer en ligne
- Big data

Retrouvez tous nos dossiers sur

<http://www.csem.be/collectionreperes>



Une initiative du Conseil supérieur de l'éducation aux médias

CSEM
Boulevard Léopold II, 44-6E635
1080 Bruxelles
www.csem.be - contact@csem.be



2019 – Les ressources proposées dans cet ouvrage sont correctes à la date de parution.

Editeur responsable : Patrick Verniers - CSEM - Boulevard Léopold II, 44-6E635 - 1080 Bruxelles - Design : www.racg-bxl.be