

BIG DATA

Ce qu'il
faut savoir

D'un côté, il y a l'intelligence artificielle, la vision par ordinateur, les assistants vocaux, les chatbots, la reconnaissance faciale: l'innovation. De l'autre, il y a des géants technologiques qui s'échangent de l'information sur nos habitudes de consommation, nos goûts et nos comportements sans trop se soucier de respecter notre vie privée. Et entre les deux,

il y a les données. Beaucoup de données. Tellement, qu'on ne parle plus de données, mais de mégadonnées.

Ni bonnes ni mauvaises, c'est ce qu'on décidera d'en faire, et comment on le fera, qui déterminera si elles rendront notre vie plus simple, ou plus complexe...



1 Quand on dit méga, on veut dire... MÉGA !

C'est en 2006 qu'est apparue pour la première fois l'expression big data, d'abord traduite par «données massives», puis par «mégadonnées». À l'époque, Facebook était encore naissant, et Amazon n'était qu'un libraire qui vendait ses livres par Internet. En numérisant et stockant des données de toutes sortes dans des serveurs, on a commencé à réaliser l'énorme quantité de données créées, chaque jour, par l'activité humaine...

Depuis le début des années 1980, soit grosso modo à partir du moment où l'informatique est entrée dans notre quotidien, on estime que le volume de données double tous les 40 mois. Suivant cette logique, en 2020, on aurait en banque quelque chose comme 44 zettaoctets de données numériques.

Dit autrement, s'il nous prenait la folle envie de «consommer» toutes ces données sous forme de films, il faudrait que tous les habitants du Canada y passent toute leur vie, de leur naissance à leur mort. Vous avez d'autres projets. Nous aussi!

D'autant plus qu'entre temps, on aura encore multiplié par 20 la quantité de données produites! 🚀

« On estime que le volume de données double tous les 40 mois. »

JE SUIS VOLONTAIRE!



Soit $44 \times 1\ 180\ 591\ 620\ 717\ 411\ 303\ 424$ octets.
Ou pour simplifier, à 44×2 à la puissance 70 (270).
Un zettaoctet équivaut à mille milliards de téraoctets.
À lui seul, le téraoctet permet de stocker environ 500 heures de vidéo.

Qui sont les poids lourds des mégadonnées?

3

L'industrie des mégadonnées est composée d'entreprises expertes:

- en collecte de données;
- en stockage de données;
- en analyse de données et en stratégies pour s'en servir!

NETFLIX, par exemple, analyse les données recueillies auprès de ses abonnés pour personnaliser son offre client. Vous avez terminé les 236 épisodes de *Friends*? Netflix vous propose des choix en fonction de ce que vous avez déjà visionné, mais aussi selon ce que les autres utilisateurs avec des goûts similaires aux vôtres ont regardé.

NETFLIX va même plus loin. On le sait, l'entreprise investit des milliards dans la création de séries et de films originaux et exclusifs à sa plateforme. **Comment détermine-t-on si la prochaine production à succès sera un thriller politique, un drame fantastique ou une comédie d'animation? Eh oui, grâce aux mégadonnées.** Netflix analyse les thèmes et les sujets les plus populaires auprès de certains groupes d'abonnés (en fonction de l'âge, du sexe, de la localisation, etc.), et en tire des conclusions qui influenceront sa production.

Des mégadonnées à l'intelligence artificielle, il n'y a qu'un pas...

2

Les mégadonnées : c'est un tas de chiffres, d'articles et de contenus pêle-mêle, stockés un peu partout sur Internet, dans tous les formats, provenant de nombreuses sources très différentes. Bref, en plus d'être gros, c'est aussi extrêmement bordélique.

Trier, filtrer, puis analyser ces données est une tâche inhumaine. Même Marie Kondo s'y perdrait ! C'est là qu'entre en scène l'intelligence artificielle.

Avec les bons algorithmes, un système informatique parvient à repérer les données similaires, à les regrouper et à les étudier pour en extraire des informations autrement inaccessibles. L'intelligence artificielle peut, par exemple, détecter la présence de maladies et de cancers sur des radiographies, à un stade si précoce qu'ils échappent à l'œil humain.

L'IA n'a aucune connaissance du fonctionnement du corps humain. Mais en comparant les images qui lui sont soumises à une volumineuse banque d'images similaires, elle détecte des anomalies, que même le meilleur médecin, avec les meilleurs instruments ne parvient pas à déceler. ✨

C'EST DONC DIRE QUE LES MÉGADONNÉES PEUVENT SAUVER DES VIES.

DOCTEUR
VITE ON LE
PERD !

DEUX MINUTES !
J'AI FAIT MÉDECINE
PAS PROGRAMMATION !



Les fabricants de vêtements et d'accessoires connectés sont aussi de grands utilisateurs de mégadonnées. Apple, entre autres, mise sur le potentiel de ses technologies dans le secteur de la santé. Des millions d'exemplaires de sa montre Apple Watch lui ont dit la façon dont les gens marchent, courent, respirent... et parfois, trébuchent. Si bien que la société californienne sait désormais quand les gens chutent assez gravement pour se blesser. Elle envoie automatiquement une alerte, et si le propriétaire de la montre ne réagit pas, elle appelle les secours. Quelques dizaines de vies, un peu partout dans le monde, auraient ainsi été sauvées par l'Apple Watch.

Apple regroupe également toutes les données fournies par son Apple Watch, pour les confier à des centres de recherche universitaire spécialisés dans l'étude de différentes maladies, dont l'Alzheimer et le Parkinson. ✨



4 Avec de grands pouvoirs...

« Le Cambridge a éclaboussé de Mark les risques d'un

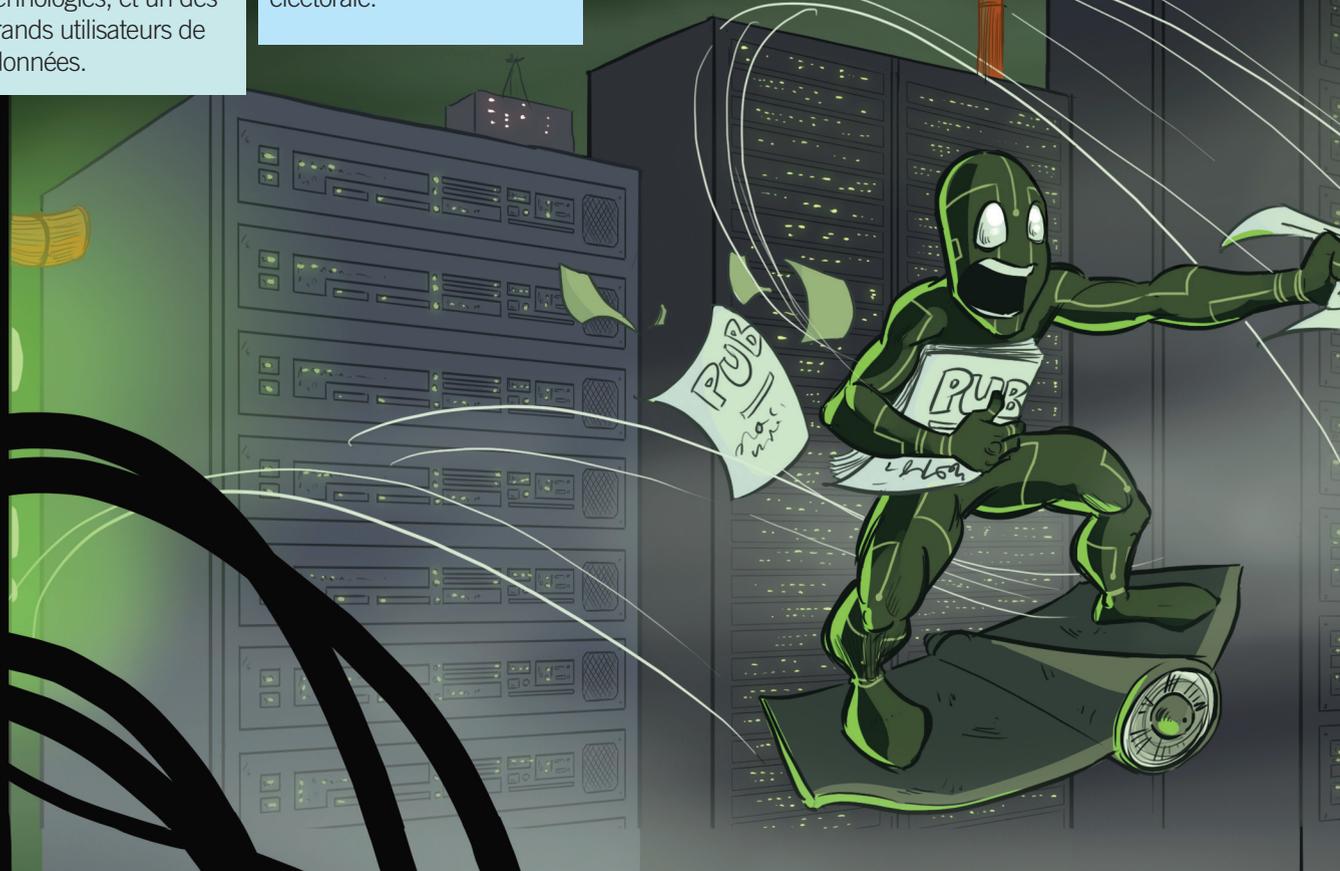
On le soulignait plus tôt: au début, Amazon n'était qu'un modeste site de vente en ligne parmi d'autres. En collectant, tous les jours, des données sur la façon dont les acheteurs et les vendeurs visitaient son site, Amazon est parvenu à améliorer leur expérience de magasinage. Puis, toujours à partir de ces données, le détaillant s'est mis à recommander d'autres produits à ses clients.

Résultat: **aujourd'hui, près de la moitié des achats faits sur Internet en Amérique du Nord ont lieu sur le site d'Amazon.** La société fondée par Jeff Bezos en banlieue de Seattle en 1994 est aujourd'hui une des plus grosses pointures du monde des technologies, et un des plus grands utilisateurs de mégadonnées.

Facebook est un autre grand collecteur de données personnelles. Et si la collecte de données n'est, par définition, ni bonne ni mauvaise, le scandale Cambridge Analytica qui a éclaboussé l'entreprise de Mark Zuckerberg prouve les risques d'un usage abusif.

Au début de 2018, on découvre que la firme britannique a mis la main sur les données personnelles de dizaines de millions d'utilisateurs du populaire réseau social. À l'aide de ces données acquises à leur insu, Cambridge Analytica a, pendant plusieurs mois, envoyé des messages publicitaires ciblés, pour le compte de politiciens américains en campagne électorale.

Dans certains cas, la firme connaissait si bien les habitudes de consommation, les goûts, l'allégeance politique et même les déplacements des utilisateurs, qu'elle parvenait à rédiger des messages publicitaires « sur mesure » pour toucher leur corde sensible. Une stratégie à ce point efficace que **des gens proches de l'entreprise se seraient vantés de pouvoir faire élire la personne de leur choix.** Sans commentaires. ✨



... viennent de grandes responsabilités

5

scandale
Analytica qui
l'entreprise
Zuckerberg prouve
usage abusif. >>

Les experts s'entendent pour dire que cet épisode sombre du numérique a marqué un virage important dans notre perception de la technologie. **Pour la première fois, le grand public a pris conscience du danger de partager de l'information sur sa vie privée.** Dans les universités et les centres de recherche, où on travaille sur les mégadonnées et leurs applications, on a commencé à se questionner plus sérieusement sur la responsabilité morale de la technologie. Sur son éthique.

On s'est notamment demandé si c'était une bonne idée d'exploiter les mégadonnées simplement parce que c'est possible de le faire. S'il ne serait pas plus raisonnable de s'interroger d'abord sur les effets de ce nouveau « marché » sur le public, la société, l'environnement. Cette réflexion a mené à plusieurs gestes importants, dont la création de la **Déclaration de Montréal pour un développement responsable de l'intelligence artificielle.**

Cette déclaration a été signée par des personnalités parmi les plus en vue du secteur des technologies. **L'objectif : éviter les dérapages comme celui de Cambridge Analytica.** Elle dresse une liste de principes et de valeurs à respecter par les chercheurs qui travaillent au quotidien avec les mégadonnées. De la protection de l'intimité et de la vie privée, à la promotion de l'inclusion et de la diversité, en passant par des considérations de développement durable. ✦

Cette déclaration peut être consultée ici
www.declarationmontreal-iaresponsable.com



6 Ce qui se passe dans le frigo... ne reste pas souvent dans le frigo

Ou sur l'ordinateur. Ou sur la tablette. En fait, tout ce qui est connecté à Internet transmet des données à des serveurs qui accumulent toute cette information, pour diverses raisons. Si votre frigo vient avec une connexion WiFi, dites-vous qu'il y a un serveur quelque part dans le monde qui sait combien de fois par jour vous l'ouvrez pour prendre le carton de lait.

C'est fou, non? Mais pas de panique! Ce genre d'information est transmis de façon anonyme au fabricant, qui va s'en servir pour améliorer ses produits, sans savoir qui fait quoi exactement. En sachant que les gens ouvrent plus souvent la porte du frigo que celle du congélateur, il pourra concevoir une nouvelle version de frigo avec une porte plus durable, plus légère ou plus facile à ouvrir. **Et ce qui vaut pour un frigo vaut généralement aussi pour un gadget informatique, ou même pour la plupart des applications mobiles que nous utilisons. Oui, même TikTok.**

Les créateurs de l'application de partage vidéos *TikTok* l'améliorent sans arrêt. Ils la font évoluer en fonction de ce que ses dizaines de millions d'utilisateurs aiment et aiment moins à son sujet. C'est ce qui rend le partage de vidéos sur son réseau plus divertissant ou attrayant. Résultat: on y passe plus de temps, on visionne plus de vidéos, accompagnées de plus de publicités, qui permettent aux gens de *TikTok* d'empocher plus d'argent. 🌟



7 Est-ce qu'on peut voler mes données?

En un mot: oui. Généralement, quand vous vous y attendez le moins. Par exemple, probablement qu'en ce moment même, une application sur votre cell est en train d'envoyer des données à un serveur sur ce que vous êtes en train de faire.

Quoique, pour le moment, comme vous lisez Curium, vous êtes plutôt à l'abri. 😊

Éventuellement, un cyberpirate pourrait mettre la main sur ces données (vos photos, vos contacts, etc.), et les utiliser pour créer de faux profils sur divers sites, comme Facebook. Non seulement est-ce une pratique courante, mais le vol d'identité est une des méthodes préférées des pirates pour soutirer de l'argent à leurs victimes. ⚡



8 On pourrait même les retrouver en Chine !

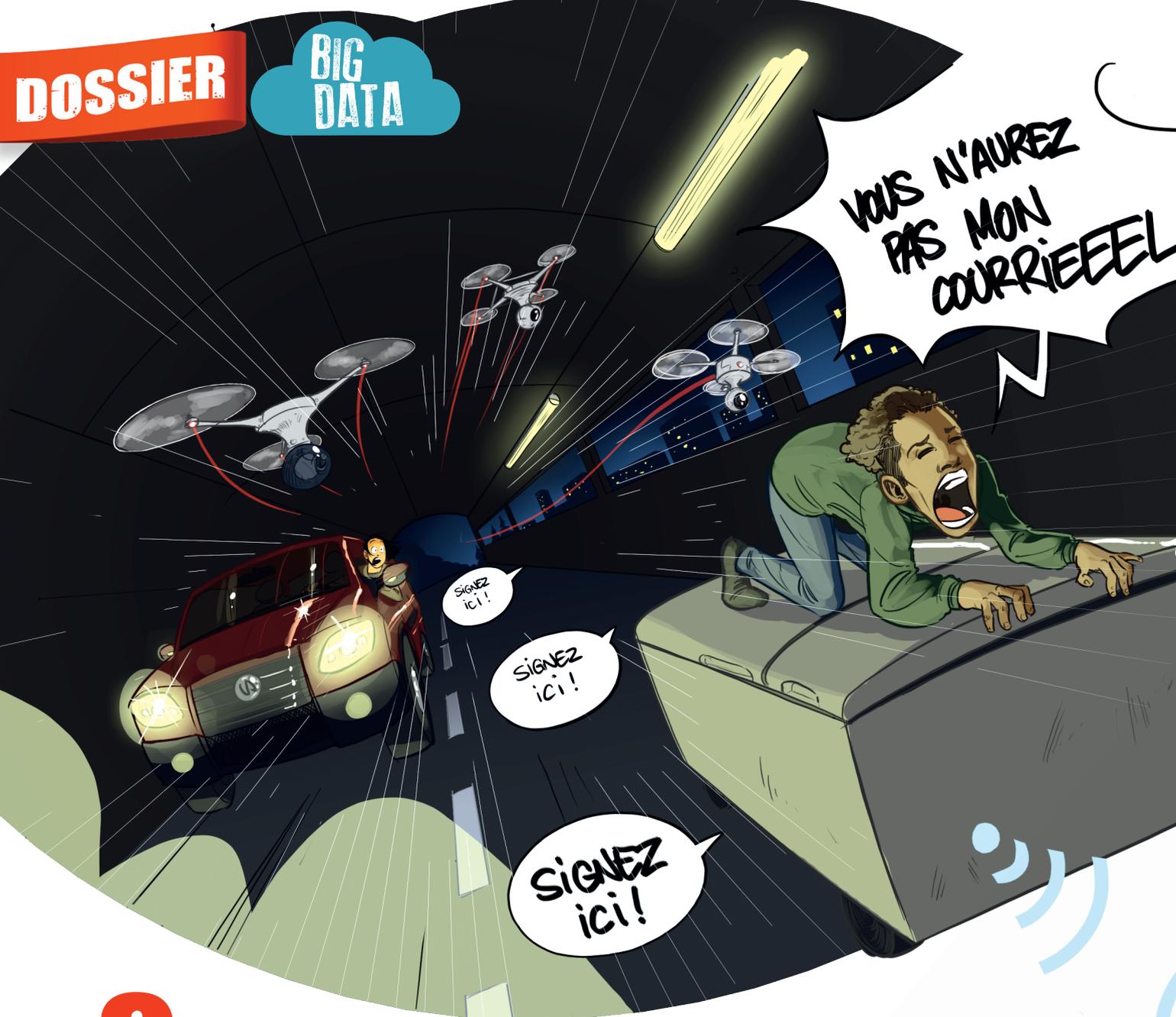
TikTok est plutôt inoffensive et, disons-le, pas mal divertissante. Il y aurait toutefois un côté sombre à cette application. Un enjeu géopolitique grave qui oppose deux superpuissances mondiales: la Chine et les États-Unis.

Depuis des années, les États-Unis accusent la Chine d'imiter ses inventions et de chiper ses idées. Officiellement, on parle de «vol de propriété intellectuelle».

Concernant TikTok (parce que, oui, TikTok est une entreprise chinoise), c'est surtout l'utilisation des données qui pose problème. Aux yeux du gouvernement américain, TikTok récolte des tonnes de données sur ses utilisateurs américains, pour ensuite les renvoyer en territoire chinois. Ce qui se passe avec les données ensuite? On ne le sait pas trop. Les plus pessimistes craignent que le gouvernement chinois les utilise pour nuire aux Américains. ⚡

« Les créateurs de l'application TikTok l'améliorent sans arrêt. Ils la font évoluer en fonction de ce que ses dizaines de millions d'utilisateurs aiment et aiment moins à son sujet. »





9 Vos données personnelles valent de l'or !

Dès qu'elles atterrissent sur le serveur d'une application mobile ou d'un site web, nos données ne nous appartiennent plus. Elles deviennent la propriété d'une entreprise qui espère probablement faire des sous à partir de cette information.

En d'autres mots, nos données personnelles valent de l'or ! **Peut-être pas tant que ça, à elles seules, mais en s'ajoutant à celles**

de dizaines de millions d'autres personnes, elles se transforment soudainement en mégadonnées qui, elles, ont beaucoup de valeur.

Certaines informations, comme une adresse courriel, une photo ou même l'adresse Internet d'un appareil connecté suffisent pour alimenter les mégadonnées. Bref, on ne s'en sort pas. Raison de plus pour jouer dans

les réglages de son mobile ou de son ordinateur et de s'assurer qu'on sait quelle application collecte quel type d'information. Et là encore, ce n'est pas simple. Qui ici lit les conditions d'utilisation des applis ? Allez. Soyez honnêtes. Ouin. C'est ça.

Nos cells deviennent des buffets à volonté de données personnelles. ★

10 Un futur autonome... et inquiétant ?

Pour être entièrement autonomes, les voitures recourent à des algorithmes qui analysent des tonnes et des tonnes de situations de tous les jours dans lesquelles les automobilistes se retrouvent. À grande vitesse sur l'autoroute. Coincé sur le pont à l'heure de pointe. Et tout le reste entre les deux.

Pour bien fonctionner, les voitures autonomes seront aussi connectées. On pourra les héler à partir d'une application mobile, puis prendre place à bord jusqu'à notre destination finale. **Dans un monde idéal**, ce système réduirait le nombre de véhicules sur la route, ce qui aiderait à régler une foule de problèmes liés au transport.

Ce scénario catastrophe n'est pas si farfelu : les premiers véhicules autonomes seront sur la route au plus tard dans cinq ans. En fait, de nombreuses applications issues de l'utilisation des mégadonnées sont attendues au fil des prochaines années. Avec un impact immédiat et majeur sur notre quotidien.



Dans un monde moins idéal, ces systèmes présentent des failles de sécurité majeures : on peut facilement les pirater. Un pirate malveillant pourrait provoquer une panne généralisée sur la route, et ainsi causer des bouchons monstres, et même des accidents mortels.



J'AI PIRATÉ SON FRIGO.

ALORS ?

IL A DU CRAN MAIS ON RÉUSSIRA À LUI VENDRE CETTE FOUTUE GARANTIE !

« Dès qu'elles atterrissent sur le serveur d'une application mobile ou d'un site web, nos données ne nous appartiennent plus. »