

Data science en santé et protection des données du Métavers

Par Adel MEBARKI

Co-fondateur et directeur général de Kap Code

L'usage des réseaux sociaux en santé connaît un essor permanent dans une société de plus en plus connectée. Ces plateformes sont devenues de véritables outils dans le parcours de vie des patients. De la recherche d'informations à la constitution de communautés, les réseaux sociaux s'imposent comme un élément permanent du « parcours de soins numérique » des patients et des parties prenantes en santé. Cette production spontanée de données de la vie réelle fait l'objet d'une multitude de recherches à des fins de santé publique. Mais à l'ère annoncée du Métavers, plusieurs questionnements éthiques et sociétaux se posent autour de l'exploitation de ces données sensibles.

Introduction

Le Web communautaire connaît un essor considérable dans la vie sociale des citoyens. Avec une utilisation moyenne de l'ordre de 2 h 30⁽¹⁾ par jour, force est de constater que ces outils numériques ont drastiquement changé le paradigme social de nos sociétés (voir la Figure 1).

Cette tendance, observée depuis plus d'une décennie, a connu une accélération fulgurante avec la pandémie de Covid-19. Selon une étude d'Harris Interactive⁽²⁾, 40 % des internautes français déclarent avoir créé

un compte sur un réseau social ou téléchargé une application de messagerie instantanée pendant le confinement.

L'usage des réseaux sociaux en matière de santé ne déroge pas à la règle ; ces plateformes sont devenues de véritables outils dans la gestion du parcours de vie des patients. De la recherche d'informations à la constitution de communautés, les réseaux sociaux s'imposent comme un élément permanent du « parcours de soins numérique » des patients et des parties prenantes en santé.

Les patients s'organisent en communautés avec pour objectifs de partager leurs retours d'expérience entre pairs et de mettre en commun leurs connaissances.

⁽¹⁾ <https://wearesocial.com/fr/blog/2022/01/digital-2022-une-nouvelle-annee-de-croissance-exceptionnelle/>

⁽²⁾ <https://www2.harris-interactive.com/social-life-2020>

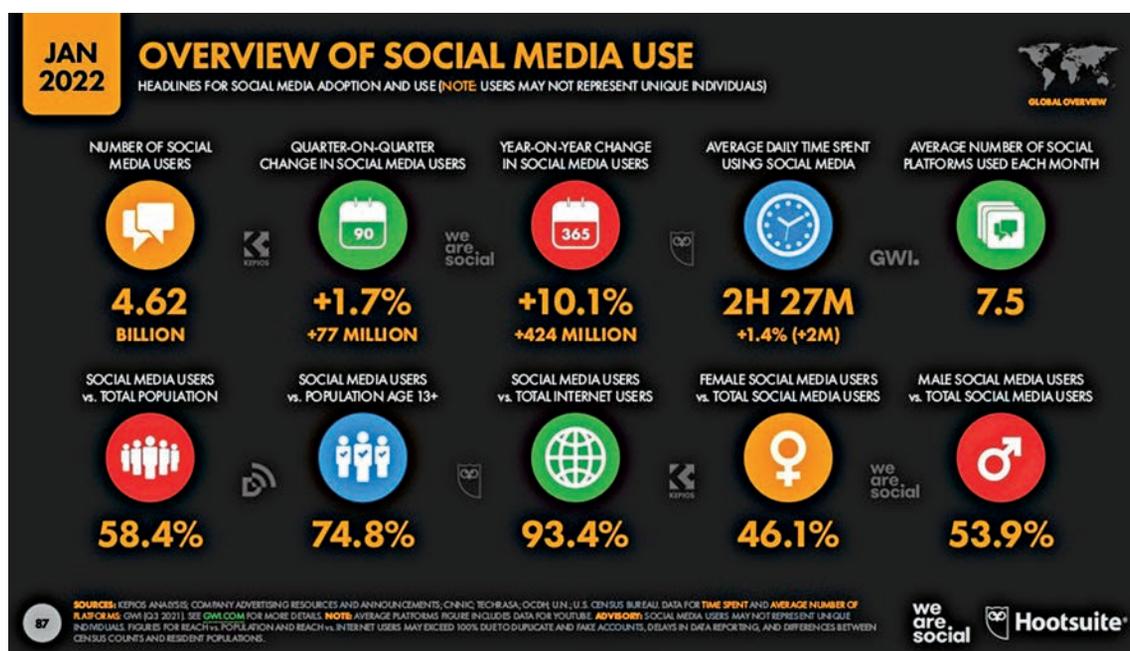


Figure 1 : Aperçu chiffré de l'utilisation des réseaux sociaux – Source : We Are Social.

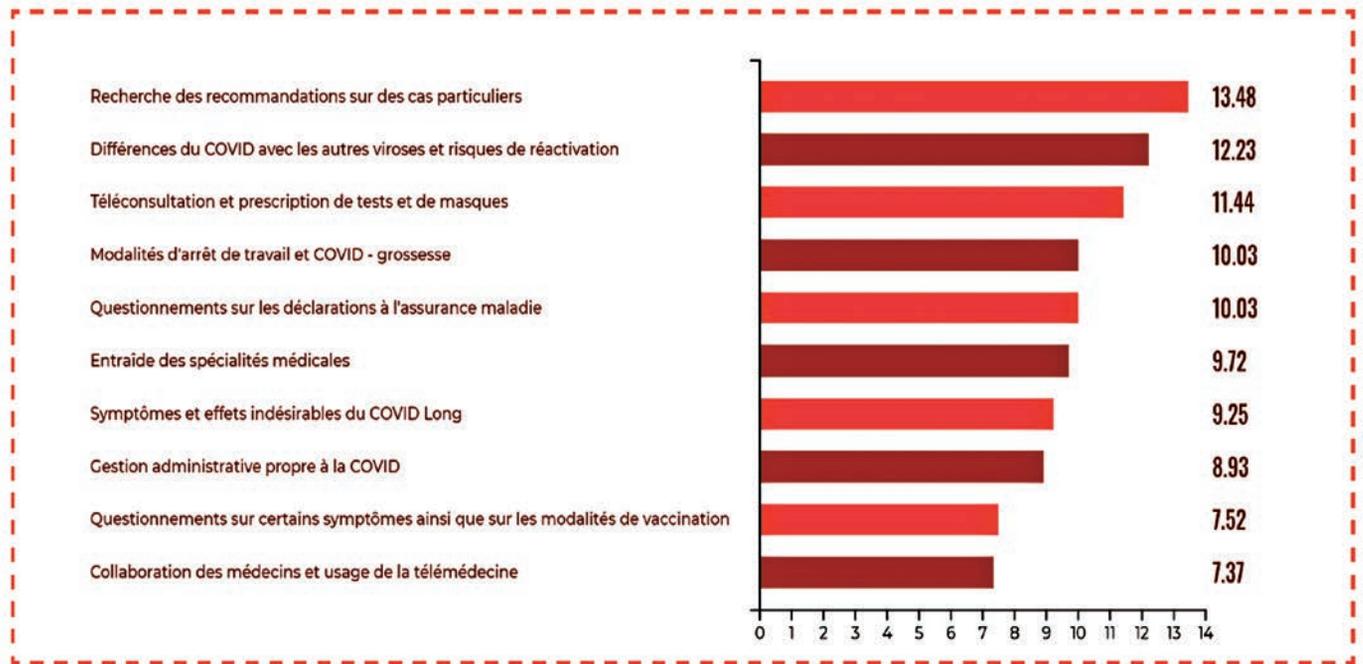


Figure 2 : Usage fait de Twitter par les professionnels de santé durant la pandémie de Covid-19 – Source : Kap Code.

De leur côté, les professionnels de santé se questionnent et s'entraident également, par exemple, *via* le hashtag #DocToCoc sur la plateforme Twitter⁽³⁾. Ces interactions virtuelles multiplient les potentiels usages de ces réseaux par les patients, les aidants et les professionnels de santé (voir la Figure 2 ci-dessus).

L'usage fait des réseaux sociaux en matière de santé

En santé, les réseaux sociaux ont deux principaux usages du point de vue du système de soins : la communication ciblée et la recherche observationnelle.

Ces plateformes réunissent un nombre important de patients et des outils performants, ce qui permet un ciblage plus fin des populations par les professionnels de santé. Ainsi, les actions d'information et de prévention font l'objet de campagnes de mieux en mieux ciblées⁽⁴⁾. Ce ciblage permet notamment d'améliorer les recrutements de volontaires dans le cadre d'études observationnelles ou d'essais cliniques. À titre d'exemple, la constitution de la cohorte d'étudiants de l'étude i-Share pour suivre la santé mentale de plus de 30 000 étudiants en France s'est appuyée sur les réseaux sociaux pour leur recrutement et le maintien de l'intérêt des participants tout au long de l'étude⁽⁵⁾.

⁽³⁾ <https://www.kapcode.fr/articles/doctococ-covid19-twitter/>

⁽⁴⁾ <https://www.santepubliquefrance.fr/docs/reseaux-sociaux-et-community-management-la-prevention-a-la-rencontre-des-publics>

⁽⁵⁾ <https://presse.inserm.fr/la-sante-de-30-000-etudiants-suivis-pendant-10-ans/7640/>

L'avènement de l'infodémiologie

L'usage quotidien de ces plateformes participe activement à la mise en place de ce que le Dr Guy Faggherazzi présente comme le « Digitosome »⁽⁶⁾. Ce concept désigne l'ensemble des données générées en ligne par un individu tout au long de sa vie, que ce soit *via* son *smartphone*, ses objets et dispositifs médicaux connectés ou encore son activité sur les réseaux sociaux. Ainsi, le partage spontané par les patients de leurs données et la disponibilité de celles-ci permettent l'émergence d'une nouvelle discipline de recherche : l'infodémiologie.

Au début des années 2010, le Dr Gunther Eysenbach a défini cette discipline de recherche comme celle construite autour de l'usage des données issues d'Internet dans un objectif de santé publique⁽⁷⁾. Ce terme d'infodémiologie, né de la contraction entre les informations issues de la navigation Web et épidémiologie, désigne la capacité à appréhender les déterminants de santé des populations, grâce aux données qu'elles génèrent en ligne, dans le but de participer à la mise en place d'actions de santé publique.

Cette nouvelle discipline vise notamment à utiliser le « digitosome » des parties prenantes de la santé afin de modéliser les parcours de soins, d'évaluer la qualité de

⁽⁶⁾ FAGHERAZZI G. *et al.* (2018), « Étude mondiale de la détresse liée au diabète : le potentiel du réseau social Twitter pour la recherche médicale », *Revue d'Épidémiologie et de Santé publique*, 66 :S197-S198.

⁽⁷⁾ EYSENBACH G. *et al.* (2009), "Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search", *Communication and Publication Behavior on the Internet*, *Review J. Med. Internet Res.* 11(1):e11.

vie des patients ou de surveiller les signaux faibles de pharmacovigilance. L'infodémiologie, qui était encore méconnue il y a quelques années, connaît aujourd'hui un boom scientifique. Une équipe de recherche de l'Université du Maryland a dénombré plus de dix usages différents des réseaux sociaux et des données qui en sont issues dans le but de participer à l'amélioration de la santé publique⁽⁸⁾.

La disponibilité des données ainsi que l'amélioration des outils de *machine learning* et de traitement automatique du langage permettent le développement de ces nouveaux cas d'usage (voir la Figure 3 ci-après).

Quid de l'éthique et de la protection des données personnelles ?

Cependant, l'évolution sociologique que nous connaissons à travers l'usage quotidien des outils numériques engendre une dimension préoccupante. Déjà présent depuis plus d'une décennie, le phénomène d'infodémie vient rebattre les cartes de l'information de santé sur les réseaux sociaux. Véritable raz de marée qui s'est développé tout au long de la pandémie de Covid-19, les contenus de désinformation ont eu une influence négative non négligeable sur l'acceptabilité des politiques sanitaires partout dans le monde et ont amplifié le phénomène de défiance institutionnelle. Cela pose de véritables questions quant à la pertinence des contenus d'information en santé disponibles sur Internet et quant aux moyens dont disposent les pouvoirs publics pour informer convenablement leurs populations.

L'OMS a notamment initié des travaux portant sur la lutte contre la désinformation en collaborant avec les différentes plateformes afin de mettre en place des fils

d'actualités « validés »⁽⁹⁾. Qualifié de menace majeure de la prochaine décennie, le phénomène de la désinformation fait écho à de nombreux questionnements éthiques et réglementaires autour des réseaux sociaux et de leur usage au quotidien.

De même, la disponibilité et les conditions d'accès aux données issues des réseaux sociaux posent question. Le récit du scandale Cambridge Analytica dénoncé en 2018⁽¹⁰⁾ a mis la lumière sur les manquements des politiques des plateformes en matière de mise à disposition des données. Même si, depuis, d'importants changements ont été apportés pour protéger au mieux les citoyens, un certain nombre de ces questions restent en suspens.

Comment peut-on à la fois accepter de partager autant de données avec des opérateurs tels que Facebook ou Twitter, et permettre, en même temps, aux citoyens de protéger leurs données personnelles ? Comment peut-on réguler les finalités et le cadre méthodologique du traitement de ces données ? Comment peut-on œuvrer pour faire en sorte qu'un consentement individuel direct soit possible ?

À l'ère du Métavers

Encore au stade conceptuel, le Métavers est la future évolution du numérique. Il désigne la prochaine évolution d'Internet, qui reposera sur des espaces virtuels, persistants et partagés accessibles *via* une interaction en 3D. Cette vision, si elle se réalisait, changerait la réalité sociale et médicale de demain.

⁽⁸⁾ CHEN J. & WANG Y. (2021), "Social Media Use for Health Purposes: Systematic", *Review J. Med. Internet Res.* 23(5):e17917.

⁽⁹⁾ <https://www.who.int/fr/news-room/spotlight/let-s-flatten-the-infodemic-curve>

⁽¹⁰⁾ <https://siecledigital.fr/2020/10/08/cambridge-analytica-une-enquete-prouve-que-les-donnees-tirees-de-facebook-ont-permis-de-cibler-les-electeurs/>

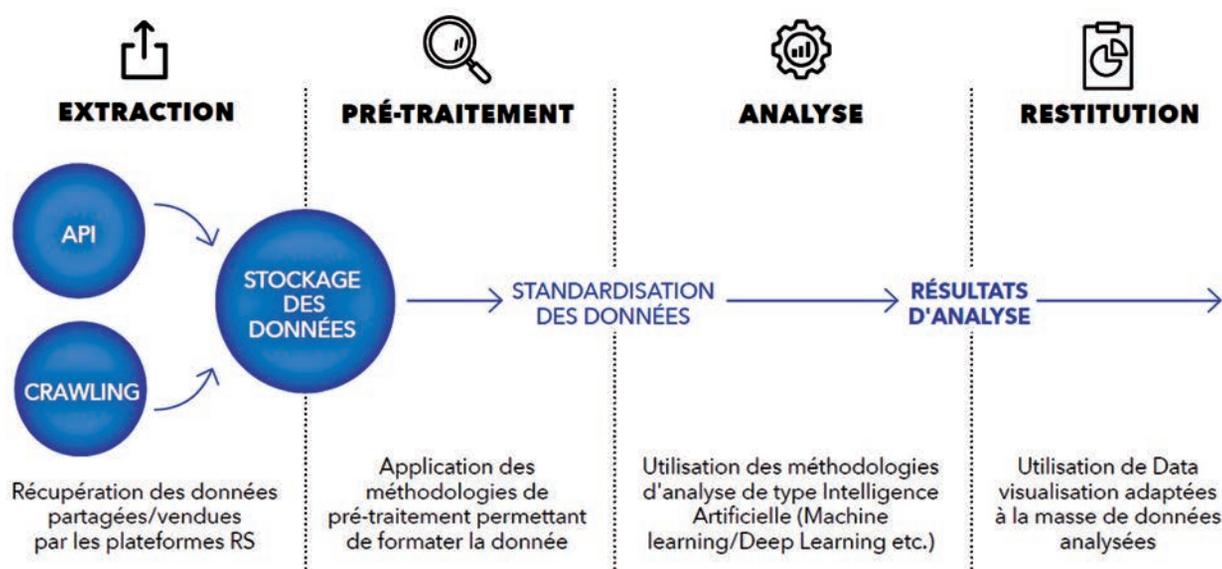


Figure 3 : Schéma d'analyse des traitements des données issues des réseaux sociaux – Source : Healthcare Data Institute.

Il sera notamment possible, dans un monde virtuel permanent, d'avoir accès aux soins et aux expertises médicales en faisant abstraction des contraintes physiques.

Encore au stade embryonnaire, cette vision viendrait décupler considérablement le digitosome des patients et ouvrirait ainsi la voie à une nouvelle forme de médecine personnalisée. La création (ou plutôt la multiplication) de biomarqueurs digitaux viendra alimenter des programmes de pré-diagnostic, de suivi médical, de dépistage ou encore de prédiction de la survenue d'une épidémie. Autant de données qui pourront alimenter les enjeux de santé publique de demain.

Toutefois, cela amènerait aussi potentiellement une centralisation plus importante de la donnée dans les mains d'acteurs déjà largement établis. La dynamique de concentration des données sensibles au sein des GAFAM que l'on connaît aujourd'hui, si elle se poursuit, les mènera à une hégémonie de plus en plus grande. De même, la multiplication de ces données accroîtra considérablement les risques de (ré)identification des personnes, tant elles vont permettre d'isoler des individus uniques au sein de la masse des internautes.

Conclusion

Les évolutions technologiques et la multiplication des données issues des réseaux sociaux offrent de véritables opportunités en matière d'amélioration des politiques de santé publique. Elles s'accompagnent aussi de multiples questionnements quant à la viabilité du modèle numérique de demain.

Le succès du Métavers et son impact positif en matière de santé ne se concrétiseront qu'avec la confiance que lui accorderont les citoyens et le respect par ses acteurs (GAFAM et acteurs publics) des libertés des premiers. Pour cela, une réflexion ambitieuse doit être menée autour de trois axes prioritaires :

- l'éthique : comment réguler et identifier les usages éthiques des réseaux sociaux ? À l'instar des études observationnelles, les comités de protection des personnes peuvent avoir un rôle important à jouer au regard de la validation du caractère éthique des projets de recherche ;
- la méthodologie : comment adopter des méthodologies de référence permettant une validation des résultats de ces recherches ? Au même titre que les essais cliniques, l'infodémiologie a un fort besoin d'un cadre méthodologique pour améliorer son acceptabilité ;
- le réglementaire : comment conjuguer recherche populationnelle et protection des données personnelles ? Au même titre que le Règlement général de la protection des données personnelles a été une démarche ambitieuse de régulation au niveau européen, nous aurons à entamer une réflexion similaire vis-à-vis de la mutation d'Internet et des enjeux de santé de demain.