The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns

Reveilhac, Maud; Blanchard, Arnaud

International Journal of Information Management Data Insights 2 (2022) 100068

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns

Maud Reveilhac[a,*], Arnaud Blanchard[b]

[a] Lausanne University, Faculty of Social and Political Sciences, Institute of Social Sciences, Life Course and Inequality Research Centre
[b] Lausanne University, Faculty of Law, Criminal Science, and Public Administration, Institute of Private Law

ARTICLE INFO

Keywords:
Health technologies
Social media
Topic modelling
Word embeddings
Comparative research

ABSTRACT

Drawing from recent advances in the field of health sociology, our study highlights topics and framings of health technologies (HT) diffused online by more than 4,000 identified actors actively involved in HT discussions on Twitter. Adopting an exploratory approach, we distinguish between health institutions, specialists, and advocates, and we assess key topics and framings promoted online by these actors. First, we show that the geographical distribution of important actors correlates with the citizens' reliance on social media to seek health information. Then, relying on 'state-of-the-art' methods in textual analysis, we identify prevalent online topics and show that the United States focuses more on risk management and private funding, whereas Europe focuses more on health literacy, practitioners, and start-ups. Furthermore, institutions focus more on indirect, global, and strategic problematics, whereas specialists are more concerned with direct and concrete problems. We also use creative visualisations displaying semantic relationships along important dimensions of HT, notably in terms of concerns related to technological priorities, professional skills, and privacy issues, as well as a possible shift in concerns related to privacy issues before and after the COVID pandemic. We conclude by discussing future research paths, particularly by giving insights into what are potential further survey interests.

1. Introduction: studying health care issues using social media

Platforms such as Facebook and Twitter have recently attracted the attention of enterprises and public institutions working in the field of health technology (hereafter HT) as potential communication channels for promoting their policies and products (Lupton, 2012). In our study, we aim to gain a better understanding of the major actors leading the online HT debate and, thereby, the prevalent topics and discursive frames they emphasise on social media. Providing answers to these questions is paramount as the internet in general – and social media in particular – is increasingly important for citizens who want to inform themselves about health-related issues (OECD, 2020).

We adopt a sociological approach which focuses on the role of prominent actors in the depictions of HT in publicly accessible discourses. Here, the reliance on social media by important actors in the field of HT is likely to provide a fertile source of information about the current public debate concerning HT. Indeed, social media have become important platforms through which HT companies and professionals can position themselves and get in touch with a public audience (Lupton, 2012). Meanwhile, social media are also becoming an important source of information from which citizens get health-related information (see (European Commission, Brussels 2015); Weber Shandwick, 2018).

Our study thus contributes to shedding a complementary light onto studies focusing on health-related practices on social media (Lupton, 2012) and on the potential of social media applications for health behaviour and information (Koteyko et al., 2015). It raises two main research questions: Who are the important actors in the HT field that are active on Twitter? What important topics and framings of HT are promoted by these actors online?

There are three essential motivations for undertaking this study. First, there is still little empirical knowledge about who is involved in social media conversations concerning HT and how these conversations relate to business and to raising public awareness of these technologies Kushwaha et al. (2021). study emerging management areas that are supported by big data (including social media). They show that one of the most significant areas of development relates to healthcare management and two aspects in particular are studied: research about the usage of sensor-generated data to help in addressing diseases and the use of health data to manage patients. To complement this managerial

* Corresponding author at: University of Lausanne: Universite de Lausanne, Switzerland
E-mail addresses: maud.reveilhac@unil.ch (M. Reveilhac), arnaud.blanchard@unil.ch (A. Blanchard).

https://doi.org/10.1016/j.jjimei.2022.100068
Received 17 September 2021; Received in revised form 8 March 2022; Accepted 9 March 2022

approach, we argue that it is also important to have greater awareness of which common understandings and practices are promoted on social media to raise the public medical awareness and, thereby, to generate trust in the business overall.

Second, Grover et al. (2018) show that discussions about HT on social media tend to be skewed towards computing algorithms, while they show no differences in discussions of acute and chronic diseases, nor in discussions of communicable and non-communicable diseases. In addition to this social technical perspective, we argue that there is a need to better segregate HT related tweets with respect to the professions (or profiles) of the authors. Indeed, literature indicates that health care promotions fulfil different aims according to their target audiences. For instance, healthcare firms aim to improve patient trust and satisfaction (Jiang, 2019), whereas professionals aim to provide guidance to physicians (Peluchette, Karl & Coustasse, 2016). In addition, it is conceivable that the aim of companies is to put forward their latest technology for better patient outcomes, while influencers aim to give support to particular groups and products.

Third, another major motivation underlying the proposed study is to further investigate which aspects and concerns of the public debate could lead to the development of public opinion survey items. Indeed, relying on online data from major actors has the potential to complement existing analyses. To date, most large-scale quantitative research on HT has been conducted in the form of surveys conducted by government or national institutes (WHO, 2015) or eHealth professionals from multiple European countries (HIMSS Analytics, 2019). In addition to HT experts' answers to well-defined survey concepts, it is worth considering what major actors in the field consider important to share with the wider public on social media platforms.

To conduct our empirical analyses, we rely on messages from more than 4,000 identified actors active in HT discussions on Twitter. In a first descriptive step, we identify important actor groups active on social media to promote HT and investigate whether the geographical distribution of these actors correlates with the general public reliance on social media to seek health information. Our data suggest that institutions (e.g., governmental agencies or private enterprises) and specialists (e.g., physicians or experts) are the two major groups involved in the online HT discussions. Our data further show a correlation between the retweet share of major actors' messages and the general public's reliance on social media to seek health related information.

The key findings of our research demonstrate a positive correlation between the share of retweets of major actors involved in HT and the public share relying on social media to seek health information. It also identifies prevalent topics about HT found in tweets addressing technological priorities, professional skills, and privacy issues. Word embedding enables us to demonstrate that current challenges lie in the relationship between patients and professionals, notably patients' empowerment and access to health data. It further suggests that the COVID pandemic led to a shift away from concerns related to (cyber)security towards a focus on data storage and computing.

Another contribution of our study is to promote a computational approach to disclose topics and frames in the field of HT. Therefore, in a second research step, we rely on 'state-of-the-art' computational social science methods and creative visualisations. These methods are already used widely in the fields of linguistics and digital humanities. However, they remain underused in the field of sociology. Our article thus contributes to the promotion of these methods within the field and also provides a detailed explanation of how they can be implemented in practice to address other research questions. In our study, we investigate which salient topics are discussed online and how their prevalence differs in terms of geographical coverage and actor type. Additionally, we provide a more fine-grained view of the framing of specific aspects of HT in relation to important dimensions and relationships. For instance, we look at the framing of HT in terms of challenges and opportunities, technological advances, as well as privacy concerns. We differentiate these framings by actor group and by period (e.g., pre- and post-COVID pandemic).

2. Study background: the study of HT perceptions through quantitative and qualitative methods

2.1. Public opinion about health technologies

HT can be defined as healthcare innovations relying on continuous data collection and algorithmic evaluation. In recent years, social media apps and other mobile devices have increasingly been adopted by health professionals to 'personalise' health treatment by sending people tailored messages in relation to their individual health concerns and conditions (Fagerlund et al., 2019). Qualitative studies have thus explored the experiences of organisations in the development of disruptive health services. For instance, the study of Sterling and LeRouge (2019) investigates the integration of telemedicine services. While qualitative studies have the advantage of advancing our understanding of how to encourage people to voluntarily share health information with the authorities (e.g. governmental agencies, organisations, practitioners, or experts) and of the new deployment of business models and strategies, they generally lack generalisability in terms of concerns of the general population towards health technologies.

In the meantime, privacy issues have been raised in discussions about the use of personalised computerised technology (e.g., Lyon, 2010), thus underlying the variety of concerns and expectations of different actors and stakeholders. Against this background, one strand of research focuses on the psychological mechanisms underlying the intention to use personal health devices. For instance, Tsai et al. (2019) aim to explain why people accept or reject telehealth usage. Their study suggests that technology anxiety takes on a critical role. More recently, the experimental study from Ross (2021) about COVID-19 contact-tracing apps showed that the intention in using health apps was positively related to chronic prevention focus and that this relationship was mediated by privacy and information security concerns.

2.2. The role of social media for assessing health information

To investigate the publicly accessible discourse about HT, studies have relied on social media data to study citizens' interest in, and their responses to HT. For instance, a study by Grover et al. (2018) investigated Twitter discussions on 'technology-enabled health' to identify top technologies and their relationship with specific diseases. The authors could confirm the role of technologies for treating, identifying, and healing various diseases, while being skewed towards computing algorithms. Another study by Lee et al. (2019) analysed health technology trends and sentiments related to health information technologies in tweets so as to examine the opinions of members of the public and identify their needs. Relying on an ontology and sentiment dictionary, they showed that social media constitute a useful tool for studying the public's responses to new HT. Their study makes a strong contribution to assess public concerns towards HT, notably because of the lack of survey data of the topic.

Social media platforms do not only play an important role in citizen information and expression of opinion, but they are also a means used by institutional actors and specialists to maintain public relations, promote products, and construct social events around specific interests (Lupton, 2012). In view of investigating professionals' perceptions and uses of HT, qualitative studies have focused on the perception and use of these technologies by practitioners and physicians (e.g., Brandt et al., 2018; Johansen, Holm, & Zanaboni, 2019). These studies relied on semi-structured interviews with convenient samples of general practitioners to uncover perceptions, as well as on digital health records and electronic health consultations. Other quantitative studies relied on survey data from health professionals (e.g., IT staff, administrative staff, clinicians, CIOs, CEOs, physicians, nurses, professionals from consulting companies and from eHealth related sectors). For instance, the *Annual European eHealth Survey* is conducted two to four times a year to provide insights into specialists' current and expected developments within eHealth in Europe (HIMSS Analytics, 2019).

According to the literature review of Zhang et al. (2020), social media act as a research context for public health research when it is 'mere reference', used to recruit participants and for data collection. The authors also note that, while qualitative and quantitative methods are frequently used, 'state-of-the-art' computational methods play a marginal role. Furthermore, their review shows that discourse (as well as behavioural) data on social media (e.g., Twitter) have essentially been used by professionals and organisations for public health management, such as disease surveillance, assessment, and control. Concerning HT (eHealth specifically), the authors underline that social media have substantially altered how individuals seek and share health information, discuss health issues, and engage in health behaviours. This constitutes a primary motivation for further investigating the discursive content of online messages posted by actors actively taking part in the promotion and discussion of HT. For instance, social media can be used for promoting open innovation in digital health through hashtag-based campaigning Kletecka-Pulker et al. (2021). investigated the impacts of the biomedical hashtag #DHPSP to promote visibility of patient safety and personalised medicine. The authors found that the campaign achieved high visibility with a large body of Twitter users participating in the online debate. Moreover, the campaign resulted in an increase of member enrolments and website visitors.

The current state of the literature shows that, despite social media's important role in spreading information and opinions about health applications and technologies, the role of social media as tools to spread HT awareness by actors actively involved in online discussions about HT has been little researched. At the same time, data availability and accessibility to various platforms are changing the nature of information systems studies. Particularly, Kar and Dwivedi (2020) underscore the need to explain beyond what is observed by moving towards why the observations happen. In our study, we seek to examine who the major actors producing HT content online are and what topics and framing of HT are prevalent in their online messages. Furthermore, we study changes in content before and after the COVID-19 pandemic which enables us to overcome the limitation that cross-sectional data can only be used to observe the relationship at a certain time.

2.3. Text classification methods for retrieving textual information

The large amount of data obtained from social media platforms makes it challenging to summarize the information in an interpretable way. This issue is especially salient in explorative research when content categories or semantic groups are not defined a priori by researchers. To address this difficulty, there is a need to apply unsupervised natural language processing techniques. Our study relies on two of them, namely topic modelling (hereafter TM) and word embeddings.

TM is widely used for producing data insights (Garg et al., 2021a). In fact, topic modelling consists of grouping together a collection of words in a way where each group represents a topic in a document. TM is beneficial for analysing the content of a corpus of documents with a knowledge discovery perspective (Bundschus, Tresp & Kriegel, 2009). However, one big issue with TM is determining the adequate number of topics to consider or opt for. Recently, several studies adopted TM analyses on tweets to identify public concerns. This trend has increased significantly with the COVID-19 pandemic with the need to rapidly identify important themes of discussion and public concerns. For instance, Abd-Alrazaq et al. (2020) examined the tweets posted in English related to COVID-19 from February to March 2020 by adopting Latent Dirichlet Allocation. Furthermore, Cinelli et al. (2020) collected data related to COVID-19 on Twitter, Instagram, YouTube, Reddit, and Gab to examine public engagement on the topic of COVID-19. They extracted all of the topics related to COVID-19 by generating word embedding and then analysed the topics. Moreover, Mahdikhani (2021) introduced a novel approach to extracting the features from tweets and to predicting their retweetability using supervised machine learning algorithms. In our study, we pay particular attention to how the extracted topics are distributed among countries and actor groups on social media to enhance the validity of our findings.

Word embeddings enable us to achieve dimensionality reduction using an unsupervised learning algorithm for obtaining vector representations for words. However, it is also used to achieve accurate text classifications (Singh et al., 2022). Recently, the popularity of word embedding techniques – such as *Word2Vec* (Mikolov et al., 2013) – have been increasing in various applications because of its capturing of word semantics and syntactics. For instance, cosine similarity measure is used to compare the found lists of resources and expand the queries (Garg et al., 2021b). Since *Word2Vec* treats each word equally in a corpus (or a document), it cannot distinguish the importance of each word. Therefore, it is useful to combine it with a weighting scheme to improve a given information retrieval task. In our article, we combine it with relative frequency. Furthermore, the word embedding approach evaluates the similarity score between words, but it does not answer why as to a similarity occurs. In our article, we propose to several visualizations that enable us to support similarity justifications between words.

3. Methods and data

3.1. Identification of major actors involved in the public HT debate on Twitter

The *R* library *rtweet* was used for data crawling and for natural language processing. Using the *rtweet* library, we extracted users whose profile description on their Twitter accounts contained specific keywords. The list of keywords was built upon the selection of relevant hashtags and words using *tf-idf* as a method of keyword extraction from Twitter conversations. In information retrieval, *tf-idf* means term frequency-inverse document frequency and serves as a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. The resulting list contains the following queries: 'healthtech', 'health AND (technology OR technologies)', 'digitalhealth', 'digital AND health', 'mhealth', 'medtech', and 'ehealth'. The term 'ehealth' refers to healthcare practices supported by electronic processes and communication and includes the networking of oIT staff, administrative staff and clinicians from health facilities, professionals from health-IT related software and consulting companies. Dating back to at least 1999, the usage of the term covers not just Internet medicine but also virtually everything related to computers and medicine. The term 'mhealth' is an abbreviation for mobile health and encompasses the practice of medicine and public health supported by mobile devices (e.g., mobile phones, personal digital assistants, wearables). The user accounts were retrieved using the search queries. Then, every account retrieved was manually checked for its relevance, coded according to an actor group category by two coders. The actor categories are the following: 'institution', 'specialist', or 'advocate'. The coders further assigned the country of emission when the location field allowed them to do so. The actors identified as relevant for our study are included in our sample (N=4,120).

3.2. Selection of relevant tweets and pre-processing steps

From each of these Twitter users, we then collected up to the most recent 3,200 tweets (which corresponds to the rate limit authorised by Twitter API) which left us with more than 7.5 million tweets in total. To keep only the most relevant tweets about HT, we applied the following search query: '.*health.* | .*medicine.* | .*medical.* | .*patient.* | .*technolog.* | .*medtech.*'. We also only selected tweets from our corpus that had been posted since January 2019. We applied several preprocessing steps including the removal of stop-words (e.g., 'the', 'our', 'of', 'at'), of special characters and symbols (e.g., '#', '@', emojis, emoticons), of punctuation, and of links (e.g., 'http(s)', 'www'), as well as the splitting of concatenated expressions (e.g., 'HealthTech' becomes 'health tech') and the lowercasing of the text.

3.3. Identification of salient topics surrounding HT

We conducted TM to provide more prompt and accurate insights into trends related to HT. TM enables us to extract dominant or salient topics in the tweets collected for the study. For instance, it can automatically identify important health topics related to HT and other important themes for the actors whose tweets were retrieved. A 'topic' consists of a cluster of words that frequently occur together. The logic behind TM uses contextual clues to connect words with similar meanings and to distinguish between the uses of words with multiple meanings (Blei, 2012). TM thus aims to reduce the complexity of the tweets to 'core' meanings so that we can identify what a given tweet is about. Topic models maximise the equation p(topic|document) x p(word|topic) for all given tweets in our corpus. It thus combines document classification (p(topic|document)) and keyword generation (p(word|topic)). Documents and words are given, topics are fitted iteratively starting from a random configuration. We used the popular implementation algorithm of Latent Dirichlet Allocation as implemented in the *Mallet* software to conduct TM (McCallum, 2002). We set the number of topics to be extracted to 150, which appeared to be the most relevant number of topics after several attempts. The extracted number of topics demonstrates a good internal and external coherence, which are two criteria proposed by Grimmer and Stewart (2013) to assess the reliability of the topic extraction. Each topic is represented by a list of top related keywords, which then need to be manually labelled with a view to proposing a possible interpretation. The distribution of topics can be assessed for external parameters, such as actor group and location.

3.4. Identification of framings of HT along important dimensions

We also aim to uncover framings of HT along important dimensions of the debate. To do so, we apply word embedding (hereafter WE) analyses which enable us to better understand the relationships between words. Therefore, instead of extracting a fixed number of topics as in TM, WE lets us choose how expansive the explored space should be as it provides a low-dimensional representation of the meaning of words (Sahlgren & Lenci, 2016). The underlying logic of WE implies that the model 'learns' scores for each word in the text for some arbitrary number of characteristics (also called dimensions). The WE method represents words as vectors, where each word gets a series of scores that position it in a multi-dimensional space. WE is thus useful for retrieving important synonyms and associations surrounding important dimensions of HT. It is also well-suited to build information retrieval contexts while letting us choose how wide the discursive space should be. We relied on the *R* library *wordVectors* (Schmidt, 2017) to train WE models (the models that we employed uses the function *train_word2vec*). This library enables us to achieve matrix operations that are useful in exploring embeddings, including cosine similarity, nearest neighbour, and vector projection with some caching that makes them much faster than the simplest implementations. The input must be in a single file and pre-tokenised, and the algorithm relies on the existing word2vec code implemented by Google in the C language (Mikolov et al., 2013). The algorithm produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space.

We followed some advice on the optimal set of parameters to use for training as defined by Mikolov et al. We used skip-gram as argument type which is better for infrequent words. We used hierarchical softmax as training algorithm. We produced 100 dimensions of the word vectors and used the argument window of 10, which is appropriate for skip-gram. More vectors usually mean more precision, but also more random error, higher memory usage, and slower operations. We used 3 threads to run the training process on. Furthermore, we did not use any minimal word frequency and we made no use of the epoch (or iter) parameter which provides passes to make over the corpus in training.

We can use visualisations to obtain a concept map plotting similar words close to each other. Words that are found in most discourses appear near the centre of the map, those which are restricted to very few documents appear on the fringes of the axes. We built models for the whole dataset, but also for each actor group ('specialists', 'institutions', and 'advocates') in view of generating an additional interpretative dimension related to the actors. We also applied stemming using the *textstem* R package (Rinker, 2018).

The proposed approach for conducting our research is summarised in Figure 1. Three main stages have been followed. Stage one captures the profiles and the relevant tweets using a list of search-queries. Stage two delivers insights from the tweets through various techniques, namely TM and WE. Stage three presents the findings in form of graphical representations and innovative visualisations.

4. Results

4.1. Identifying the main actors involved in health technologies who are active on Twitter

The common population of users on social media platforms consists of non-affiliated users, users self-identifying with an organisation in their profile, official organisational accounts, influencers, fake accounts, and bots. Our sample of social media users active in the field of HT is divided between 40% institutions (public and private), 40% specialists (or practitioners), and 20% advocates. To be included in our sample, advocates must refer explicitly and primarily to HT in their profile description. For instance, journalists who cite HT as one of their minor interests are not included in our sample. Neither do we include users with either an irrelevant profile description or a very minor interest for HT.

The profile descriptions allow us to derive shared characteristics among the different groups of users. Among organisations, there are as many public as private actors (including: universities, research institutes, hospital services, health authorities, private organisations, or corporations). Institutions use Twitter to promote their services (e.g., technological advancements) or policies (or regulations). With respect to specialists, they are essentially CEOs, CIOs, practitioners, research fellows working in universities, or private entrepreneurs. Specialists, in particular, rely on Twitter to publicise their research, their research agenda (e.g., events, conferences, webinars, etc.), and new challenges associated with their practice. Our sample of Twitter users thus reflects similar specialist positions as the respondents covered by expert surveys (e.g., HIMSS Analytics, 2019).

Our corpus is mainly composed of social media users from the United States (50% of users are from the United States), followed by users from Europe (40%), and a residual share from other countries (10%), including Canada, New Zealand, India and African countries. According to a spring 2019 *Pew Research Center* survey (Schumacher & Kent, 2020), the social media penetration rate is more pronounced in the United States than in Europe. In European countries, the use of social media varies significantly between countries Figure 2. below illustrates the distribution of the share of retweets in our corpus in relation to the national share of respondents from representative samples of national populations seeking health information on social media (we used the survey data from

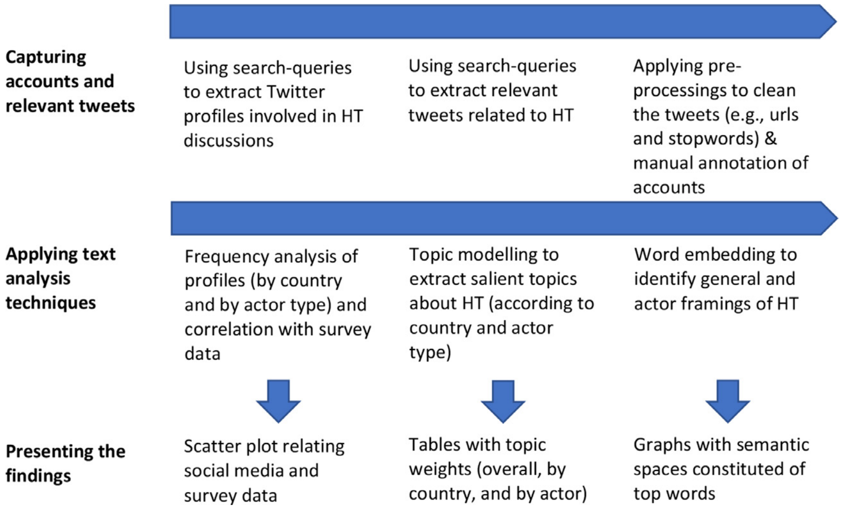

Figure 1. Proposed approach for conducting the research.

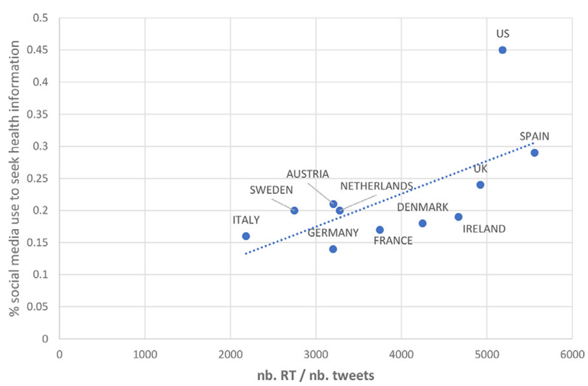

Figure 2. Relationship between the share of retweets from health technology actors (x-axis) and the reliance on social media to seek health information (y-axis).

the 2014 *Eurobarometer*[1] to plot European countries (European Commission, Brussels 2015), and survey data from the 2013 *Great American Search for Healthcare Information*[2] for the United States) Figure 2. shows a positive correlation between the share of retweets about HT and the share of national social media users relying on social media to seek health information (Pearson correlation of 0.7 only for European countries and of 0.65 with the United States included).

4.2. Extracting salient topics surrounding health technology

In this section, we assess more precisely what the important topics addressed on Twitter by the users included in our sample are. The topics extracted are multi-fold and range from ‘trendy’ topics gathering interest on Twitter to recent advances in the field of HT and policy regulations. The vast majority of topics are relevant for our analysis. More than 80% of the topics extracted have a clear interpretation and are mutually exclusive. The remaining 20% are related to news, or to summaries of events, and are difficult to differentiate (the full manual coding and the mean topic weights that refer to the text mass that this topic covers, presented in percentage terms by regions – ‘United States’ and ‘Europe’ – and actor type –‘specialist’, ‘institution’, and ‘advocate’ – can be found in Annex 1). The findings from TM shows salient themes addressing the patient–doctor relationship, patient-centred initiatives and needs, healthcare systems, innovative solutions, big data challenges, market opportunities, and customer experience (see Annex 1).

Because there are major differences in the health systems prevailing in the United States and in Europe, we also assess whether these differences are reflected in the prevalence of topics in a cross-cultural perspective. For instance, OECD data (2019) show that the amount of money Americans spend on healthcare services is higher than in any of the other developed countries in the world. At the same time, only 23% of Americans think that they get the best care possible, compared to an average 70% of EU citizens who are satisfied with the quality of healthcare. We therefore expect to find differences between cultural contexts, especially in terms of which topics are emphasised to meet the expectations of patients, citizens, and communities. To test this hypothesis, we assess the difference between the United States and Europe in the prevalence of the topics extracted (see Annex 1 for the topic weights for the two regions ‘United States’ and ‘Europe’). The topic weights show differences in topic prevalence between European countries and the United States. For instance, the latter places greater emphasis on risk management and private funding, whereas European countries focus more on health literacy, practitioners (as opposed to scholars), and start-ups.

We also expected to find different topic salience across actor types, which we test using the topic weights (see Annex 1 for the topic weights for the actor types ‘specialist’, ‘institution’, and ‘advocate’). Specialists tend to focus more on concrete and direct challenges and topics. For instance, they focus on subjects such as patient happiness and patient monitoring, as well as on the latest technological developments and the COVID pandemic response. Furthermore, specialists have a direct interest in learning/training/teamwork, which are additional direct concerns in their daily practice. In contrast, institutions focus more on indirect problematics, such as corporate policies, projects, and finances (e.g., funding, market growth, profits margins, and marketing), as well as on more strategic or global topics such as general policies and health concerns (e.g., smoking, home care, and pregnancy). Whereas institutions and specialists have a scientific and economics-oriented discourse about HT, advocates spread content mostly related to highlights, wellness, well-being, and wearables.

[1] For more information on the survey report, see: https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_404_sum_en.pdf

[2] For more information on the survey report, see: https://www.webershandwick.com/wp-content/uploads/2018/11/Healthcare-Info-Search-Report.pdf

Table 1
Cosine distances between health issues and HT related terms ('healthtech' and 'medtech').

Disease or health issue category ('term used')	cosine	cosine rank
Weight disorders ('obesity')	0.061	1
Addiction disorders ('addiction')	0.060	2
Cardiovascular diseases ('heart')	0.050	3
COVID ('covid')	0.047	4
Mental disorders ('mental')	0.041	5
Diabetes I & II ('diabetes')	0.039	6
Liver problems ('liver')	0.038	7
Hypertension ('hypertension')	0.035	8
Vascular diseases ('vascular')	0.028	9
Gerontology ('gerontology')	0.020	10
Oncology ('oncology')	0.019	11
Neurological pathologies ('brain')	0.017	12
Alzheimer's ('alzheimer')	0.017	13
Lung diseases ('lung')	0.016	14
Blood diseases ('blood')	0.010	15
Neurology ('neurology')	0.004	16
Sexually transmitted diseases ('aids', 'hiv')	0.004	17

4.3. General framing of health technology on social media

In this section, we apply WE using different strategies to extract relevant framing related to HT. Compared to TM, which provides one particular idea of a given theme, WE models enable us to search for relationships embedded in words. They can thus provide us with an overview of families of related terms, i.e. words that are found in similar contexts. In this respect, WE is a good strategy to reveal word relationships. It separates and clusters words that are semantically similar. A way to make sense of the WE is to build a text network to derive the similarities between each pair of words. Based on this network, we can build a visualisation of word relationships. This visulalisation is also referred to as a 'conceptual map'. The Figure 6 in the Annex 2 displays such a map based on top terms of our corpus of tweets. It shows pairings where words with similar meanings are nearby[3]. For instance, 'io' and 'robotics' (see upper middle pane) clearly have something in common and are plotted next to each other. Terms that appear together (e.g., 'interoperability' and 'telemedicine') cluster together on the chart (see lower middle pane).

In the following analyses, we rely on WE to obtain ways of interacting with the vector space beyond word pairings in order to build information retrieval contexts. For instance, we can thus assess the distance between two words, or between one word and several related words. In our application, we used a list of diseases and health issues for which we calculated the distances to HT related terms (notably, 'healthtech' and 'medtech'). This enables us to demonstrate that certain health issues – such as obesity, addiction, heart disease and COVID – are perceived as more 'well-suited' in terms of HT (see Table 1 containing cosine distances between our list of health issues and HT related terms). It will be important to take this result into account when interpreting the next analyses as Table 1 indicates what health 'domains' are likely to be prevalent in our corpus of HT-related tweets.

WE can also be used to highlight connections between concepts in terms of word-vector relationships. This lets us plot a number of terms in a given discursive space. However, instead of specifying vocabulary items, we can also create text visualisations corresponding to word relationships. For instance, just as 'patient' and 'professional' are individual vectors, 'patient – professional' can also be represented in a semantic space. We can simply indicate this by comparing our words to a new vector defined as the difference between the two words ('customer' and 'industry') within the same vector space. This enables us to score any words based on their relationships in order to create word representations specific to any desired word relationship.

Figure 3 displays a semantic space composed of two-word relationships: 'patient – professional' and 'challenge – opportunity'. The relationship 'challenge – opportunity' aims to illustrate an important opposition in health data usages. The increased availability of HT offers opportunities to improve important aspects relating to diseases and injuries, but HT can also be framed with respect to emerging challenges and concerns, either from the patients' or the professionals' perspective Figure 3. captures distinctions between these two continuums.

We will now explain the methodology applied to extract the words plotted in Figure 3. A similar methodology will be used for the subsequent figures (also refer to Schmidt (2017) who presented the method on which we elaborated to build our own analyses). Regarding Figure 3, we first extracted top words mostly associated to HT using the following query: 'healthtech | medtech | digitalhealth | ehealth | digihealth'. Then, we extracted the top words closed to opportunities (query: 'opportun | solute | advanc') from which we subtracted the top words closed to challenges (query: 'challeng | difficulti'). This forms the 'opportunity vector'. We also extracted the top words closed to patients (query: 'patient'), from which we subtracted the top words closed to professionals (query: 'profession'). This forms the 'patient vector'. On this basis, we calculated the cosine similarities between the 'HT vector' and the 'patient vector', as well as between the 'HT vector' and the 'opportunity vector'. Because of the big differences in the frequency of individual words, we weighted the cosine scores by the relative frequency of each word. For readability purposes, we only plotted the top 130 words.

The shape of the word distribution on Figure 3 shows that HT tend to be framed as opportunities on the patients' side and as challenges on the professionals' side. Words on the upper left pane (such as 'digitalmentalhealth', 'clearhead', 'behaviourchang') are related to the patient and the opportunity space. Words on the lower right pane (such as 'patientcentr', 'patientexperi', 'clinicaltri') are related to the professional and challenge space. These words indicate areas in which there is a need to improve the application of HT, with a focus on digital and virtual HT (e.g., 'telehealth' and 'virtualcar') Figure 3. enables us to assess further salient trends. First, on the patient side, there are words related to concerns about data safety and the guarantee of their privacy (e.g., 'dataprivaci'), as well as a call for more ethics (e.g., 'techforgood') and medical knowledge (e.g., 'digitalhealthliteraci' and 'mindblow'). Second, this trend is shared by the professionals who emphasise patient empowerment (e.g., 'patientdrivenhealthcar').

Overall, our findings tend to indicate a positive 'tonality' (or connotation) at the word level. This can result from the fact that the actors included in our sample are more likely to be favourable than critical toward HT. To assess this possible bias, we conducted a sentiment analysis of the tweets from the three groups of actors using the R package *sentimentR* (Rinker, 2019). We found that there is a general pattern toward positive language about HT (see Figure 7 in the Annex 3). However, the three groups significantly differ in their mean sentiment, with institutions and specialists relying on a more positive language than advocates (significance level of *Student-test* for *p-value* <0.05; see also Annex 4 to see the distribution of sentiment by actor group). This means that our analyses are more representative of the perspective of actors who are rather supportive of HT, thus under-representing views from other Twitter users who are critical (or sceptical) about the benefits of HT.

4.4. Actor framing of health technology on social media

In this section, we apply WE to extract relevant actors' framing of opportunities and challenges associated with HT. To do so, we can also retrieve similarity scores while keeping the information about the actor type. To maintain discursive distinctiveness between actors, we trained word vectors separately for tweets from each actor ('specialist', 'institution', and 'advocate'). Merging the scores for each actor enabled us

[3] We relied on the python library *texplot* and on the *Gephi* software (see implementation by McClure, 2015).

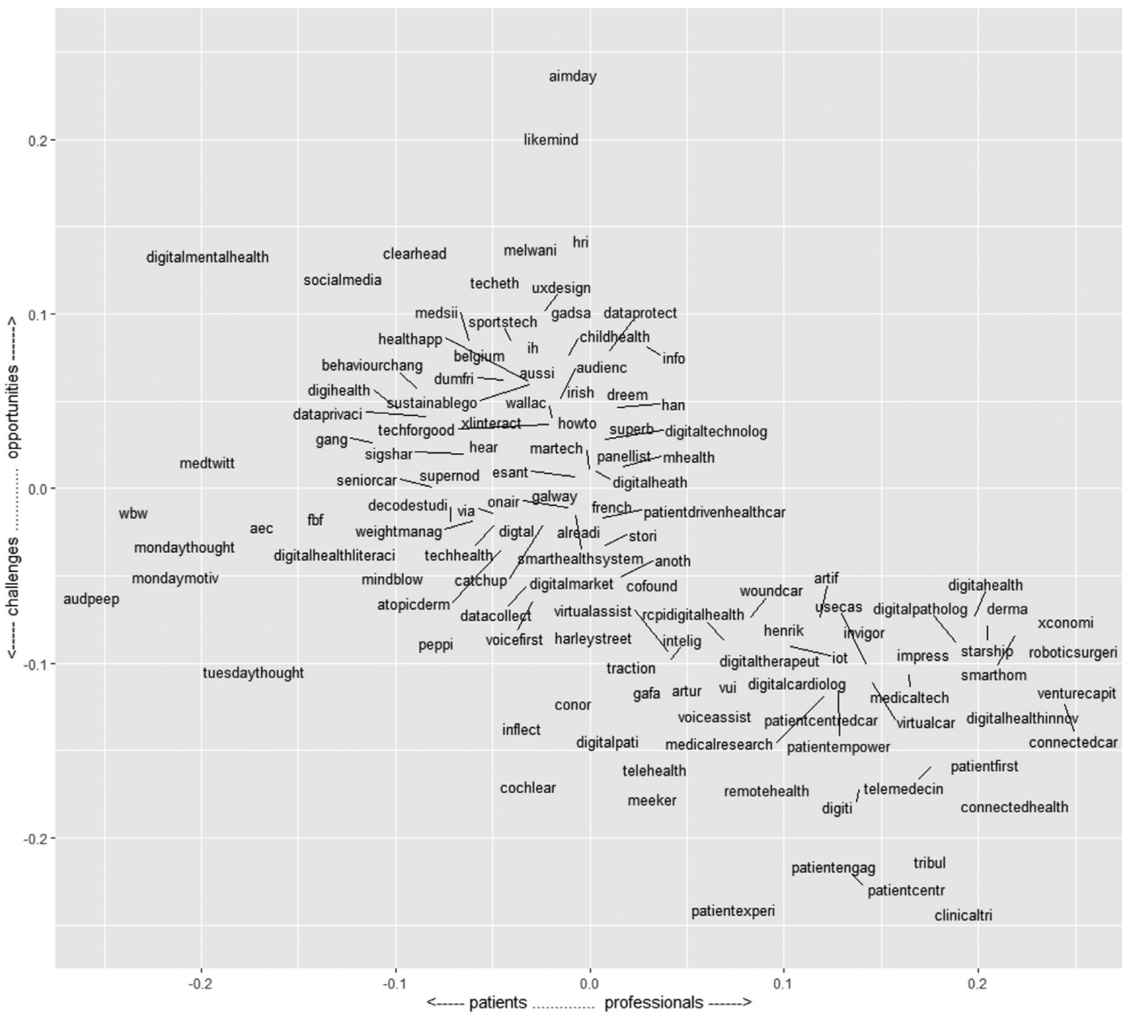

Figure 3. Semantic space composed of the two-word relationships 'patient – professional' (x-axis) and 'challenge – opportunity' (y-axis).

to identify terms that are shared among all actors ('shared' words) and terms that are more salient for a given actor compared to the other actors ('specialist', 'institution', and 'advocate') Figures 4 to 5 are based on this logic and display the discursive differences for each group of actors.

Figure 4 focuses on the similarity scores associated with new technologies and with privacy in relation to the terms 'professional' (y-axis) and 'patient' (x-axis). Top words are plotted in this discursive space and coloured according to the actor type. Specialists' framings especially emphasise concerns related to their daily practices and research (e.g., 'medtechinnov', 'showcas' and 'futurofhealth'). In contrast, institutions' framings mainly emphasise business opportunities (e.g., 'charitesummit', 'investor' and 'standout'), but they also focus on the opportunities offered by the collection and analysis of health data (e.g., 'healthanalyt' and 'healthinfo'). The advocates mainly emphasise the concrete applications (e.g., 'videoconferenc', 'healthit' and 'voitech') and domains of HT (e.g., 'biotech', 'prosthes' and 'ophtalmolog'). The shared discursive space provided by Figure 4 is in favour of more predictive medicine and new research skills, notably with the reliance on artificial intelligence and big-data analytics.

In Figure 4 (right pane), we show the top words associated with the term 'privacy'. There is a trend to associate 'privacy' concerns to the 'patient' side (x-axis) rather than the 'professional' side (y-axis). Furthermore, shared words demonstrate that data-driven technologies raise data privacy discussions associated with professionals' obligations (e.g., 'compli', 'transpar', and 'ethic'). Specialists also emphasise data access and algorithms to analyse these data. On their part, institutions are more concerned with security issues (e.g., 'hitsecur' and 'cyber'), as well as with data sharing and authenticating strategies (e.g., 'patientaccess' and 'authent'). Advocates emphasise the need for accountability (e.g., 'inform'), confidentiality, interoperability and security (e.g., 'cyberattack', 'protect') concerning HT.

HT based on continuous data collection and algorithmic evaluation have gained importance during the COVID pandemic (Scott et al., 2020). The growing interest in continuous data collection and the algorithmic evaluation of personal health data exacerbates concerns about data privacy. To highlight recent important trends, we use similarity scores associated with privacy concerns based on their distance from HT before and after the COVID pandemic.

Figure 5 shows a discursive shift between before and after the COVID pandemic, with focus moving from the professionals' to the patients' side. Furthermore, there is also an evolution from concerns related to (cyber)security to data storage and computing between the 'pre-covid' and the 'post-covid' periods. The 'pre-covid' period also rassembles more words associated with ethical considerations (e.g., 'liberti', 'imbal', and 'dilig'). In a similar vein, the 'pre-covid' period also emphasises non-discrimination issues. The legal orientation of HT discussion is present both before and after the pandemic (see shared terms in black: 'law',

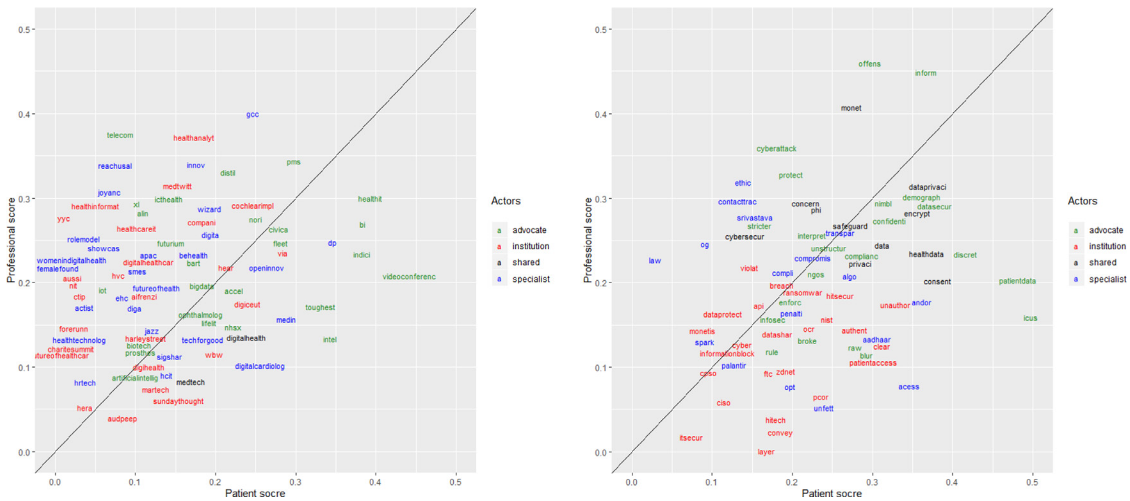

Figure 4. Top words associated with technology (left pane) and privacy (right pane) by similarity to patient (x-axis) and professional (y-axis) by actor type.

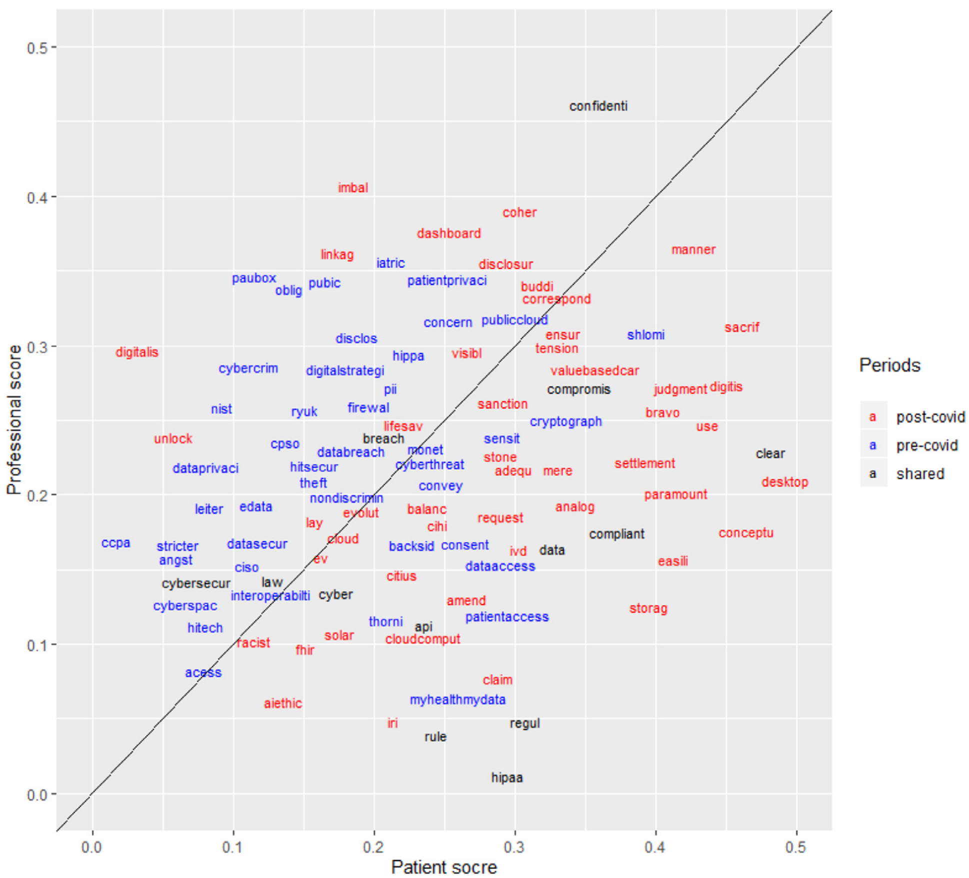

Figure 5. Top words associated with privacy close to patient (x-axis) and professional (y-axis) before and after the COVID pandemic.

'regulation', 'compliant' and 'rule'), although it seems to have taken a more punitive orientation in the 'post-covid' period (e.g., 'judgment', 'sanction'). This can be explained by the fact that the 'post-covid' period seems to be characterised by terms related to emergency (e.g., 'lifesav').

5. Discussion of the main findings

In the first research step, we identified the major actors leading the public debate on HT on Twitter. We showed that the most represented actors are institutions and specialists (80% of our corpus) who are mainly located in the United States. We also found a positive correlation between the share of retweets from major actors' tweets and the share of the public relying on social media to seek health information. The lesser representation of HT advocates provides a partial explanation of the low proportion of topics related to news or lighter topics

In a second step, we relied on TM to extract topics and concerns underlined by major actors involved in the field of HT. We assessed important differences across cultural contexts (i.e. Unites States versus

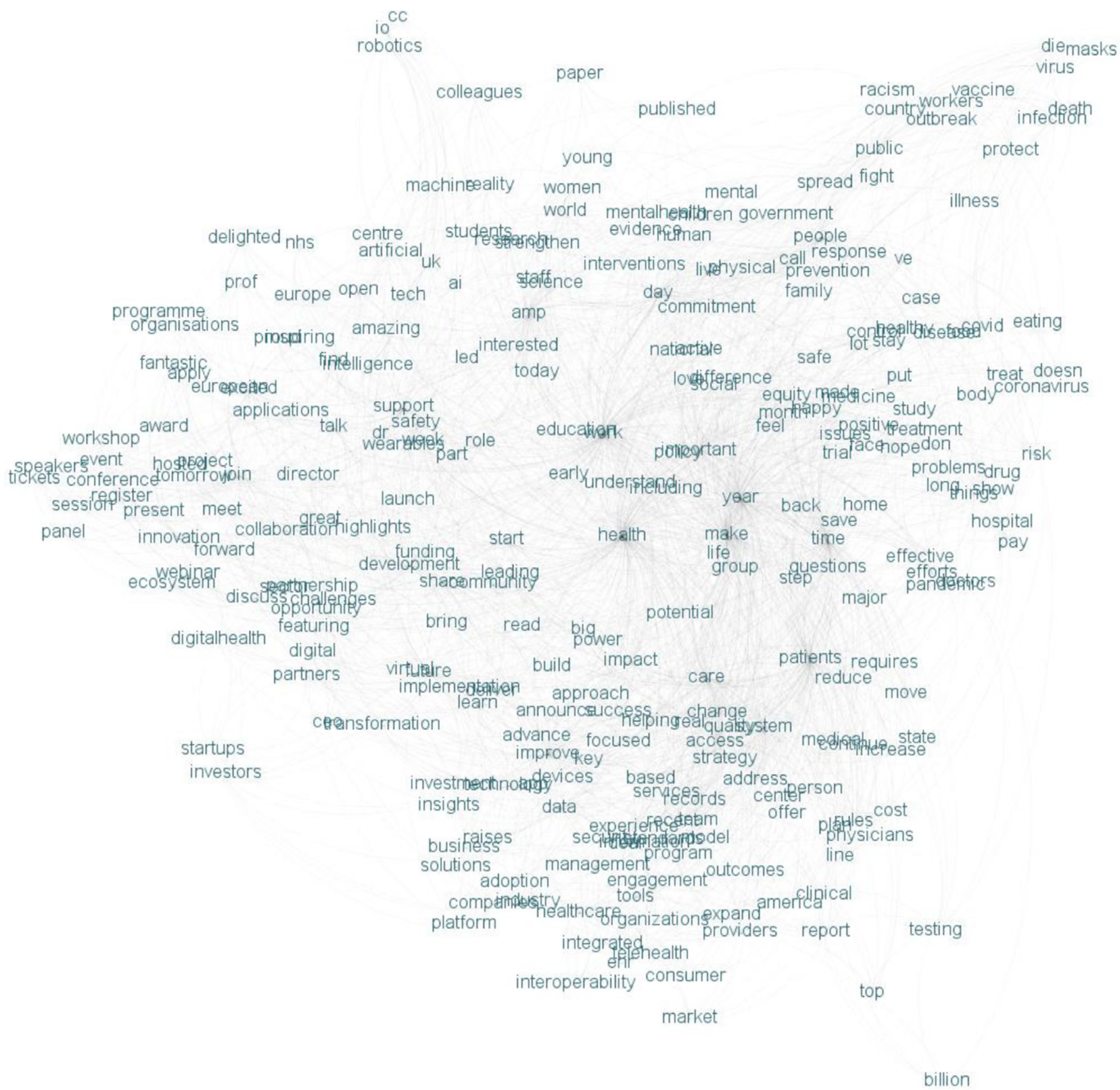

Figure 6. Semantic relations based on the top terms in our corpus of tweets.

European countries) and actor types (namely institutions, specialists and advocates). Using TM, we showed that there are important differences between the United States and Europe in the prevalence of topics related to HT. For instance, the United States focuses more on risk management and private funding, whereas Europe focuses more on health literacy, practitioners, and start-ups. The topics extracted also showed different focuses among the actors. Institutions focus more on indirect, global, and strategic problematics, whereas specialists are more concerned with direct and concrete problems. Our dataset shows no particular pattern for advocates. Advocates are also active actors in the HT field, but they focus on less substantive themes, such as wearables, well-being, and healthy lifestyles.

In a further step, we relied on WE to gather general and actor-specific understandings of HT along important dimensions (see Figure 3). The semantic space crossing two relationships, namely 'patient–professional' and 'opportunity–challenge', shows that current challenges lay particularly in the relationship between patients and professionals, both in terms of patients' empowerment and in access to health data and information. There is also an emphasis on new development opportunities (e.g., equipment and wearables). Furthermore, professionals focus on what could be well-suited domains (e.g., imaging and videos for diagnostics), whereas patients are concerned with data protection issues (e.g., in terms of artificial intelligence, demystification, and customer experience).

The discursive spaces along the 'patients' and 'professionals' dimensions display important (and perhaps opposing) challenges between these two actors in terms of technological innovation and privacy concerns. For instance, specialists and institutions focus on adapting to HT by learning and developing new applications, whereas advocates are concerned with data privacy and also insist on the importance of data protection (see Figure 4, left pane).

The new challenges regarding privacy imply that practitioners will tend to focus on their responsibilities and obligations (or liabilities) by focusing on legal, ethical, and IT security concerns (see Figure 4, right pane). There is also a clear patient demand for more control of health data (e.g., in terms of transparency, access, and interoperability). There is, thus, patient demand for a more horizontal relationship with practitioners.

We concluded by analysing a possible shift in concerns related to privacy issues before and after the COVID pandemic. We note that word scores linked to privacy have generally become more prevalent in relation to the patients' side since the beginning of the pandemic. Furthermore, we discern two broader categories of terms related to either 'legality' or 'ethics'.

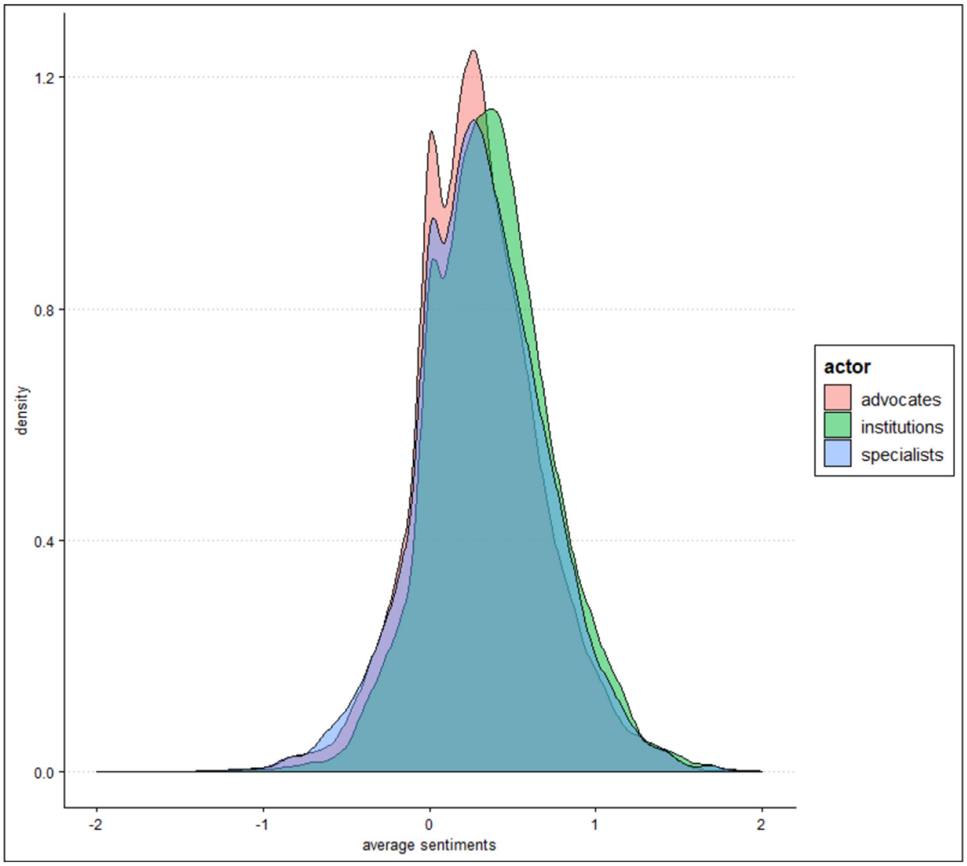

Figure 7. Density plot of sentiment score by actor type.

5.1. Theoretical contributions

In line with the literature suggesting that social media serve efficiently for health care discussions (Jiang, 2019), our findings demonstrate the usefulness of investigating HT-related discourses online. In particular, the proposed study discusses some of the opportunities and concerns expressed by users posting about HT on social media, while also discriminating the groups of users. The different groups of users that we investigated strategically use social media according to their characteristics (e.g., public or private entities, practitioners or business managers, and influencers) and according to the purpose of the information delivered (e.g., raising public awareness, selling products and services, and raising concerns). In this study, we have conceptually identified the most salient topics and framing of HT based on the words that are relevant for these groups of users.

Another theoretical contribution relates to the methods used for investigating HT discourse in terms of topicality and framings. First, we adopted fully automated methods to collect and analyze the collected tweets, which enabled us to have significantly more variety in the topics and frames analysis than would be feasible with manual annotation. Furthermore, the use of word embedding combined with innovative visualisations along important dimensions can be applied in other fields of technologies and information management to complement managerial and social technical perspectives. This methodology will hopefully guide future researchers to perform in-depth analysis in individual HT subdomains (e.g., privacy concerns, business opportunities, crisis management).

5.2. Implications for practice

Our findings reveal that social media are not only a useful source of information about the current state of HT (e.g., business opportunities), but also about which concerns surround HT policy and the role of HT in crisis management. The practical implications of our study can thus be segregated into several audiences, namely the users, the (public or private) compagnies, and the (private) practitioners or business managers. On the basis of these findings, the different user groups can decide what aspects should be prioritised and how to frame them so as to address salient concerns. Concerning the users, our analyses reveal that they are mostly concerned about privacy and security when discussing HT. Therefore, it is important that social media platforms provide users with authentic and balanced information about HT so that users can make informed choices and find answers to their concerns. Concerning companies, we show that social media can be used as useful channels to raise public awareness by promoting specific campaigns and to monitor trends in disease conditions (e.g., COVID-19 crisis management). Concerning the practitioners, they can usefully rely on social media to provide innovative solutions to diseases while putting forward their own business.

From a methodological perspective, our findings also have practical implications for the research community. Studying topics and frames stemming from social media accounts of specific users enables us to derive the most salient dimensions of the debate about HT. However, we still know little about whether the concerns and opportunities expressed are representative of those of the general population. It would therefore be useful to complement the proposed methodology by using additional methods, such as opinion surveys. In this case, the findings of our study could serve as a basis for identifying the HT areas and aspects worth surveying at national levels while considering possible country-effects on health care systems and health promotion.

6. Concluding remarks and outlook

Our study makes two important contributions to the research on HT. First, it provides an exhaustive picture of the major actors in the HT field actively posting on social media and of what topics and framings they share with the wider public on Twitter. In this view, our study represents an important step towards a better understanding of how and why social media can impact citizens' health attitudes and behaviours. The second contribution of our study is to provide an innovative methodology for investigating important HT framings using creative visualisations.

Our study nonetheless entails several limitations that would be worth addressing in future research. First, Twitter is only one possible social media platform, with specific rules and conventions. It is less used than Facebook and allows less extended user contributions. However, Twitter data are submitted to fewer access restrictions and also cover an international population. These characteristics make Twitter data suitable for the purpose of our analysis. Nevertheless, other professional platforms, such as LinkedIn, could offer an alternative source of data for studying in greater depth how institutions and specialists in the field of HT portray themselves and recruit specific profiles.

Second, future studies could also examine the evolution of HT discussions online by accessing historical data. Our study is limited to the most recent tweets and, thus, does not allow for the study of the evolution of HT themes or concerns over time. Our corpus of actors testifies that a historical study is feasible, as the majority of Twitter accounts were created several years ago (the majority were created from 2011 onward). HT are characterised by rapid changes in the health and social care sector, and the development and impact of these changes are hard to predict. Our data already account for the current shifts in information technology and big data, automation, and artificial intelligence. This shift was brought to light in a recent study by the OECD, which identified a new demand for skills and specialisations among health and social care workers, while reducing the importance of other professional roles (OECD, 2019).

Third, we restricted our analysis to major actors, which, possibly, does not give voice to more negative or concerned opinions about the use of HT (see end of section 4.2). Therefore, future studies might include the network of Twitter followers to seek a more global view of HT as perceived by the public. In a similar vein, we encourage the development of surveys covering public reliance and concerns about HT that can complement existing surveys conducted with official health actors.

Fourth, we focused on discourses surrounding HT from the perspective of topicality and framing. However, another important discursive component relates to tonality and emotion, also referred to as opinion mining. For instance, Ridhwan and Hargreaves (2021) relied on opinion mining to investigate public sentiment about the COVID-19 outbreak in Singapore. They showed how policy measures triggered different emotions, drawing from previous studies using social media to monitor public health-related issues expressed online (García-Díaz et al., 2018). We should nonetheless note that opinion mining does not always reflect stance (e.g., favouring or rejecting a policy issue). Other meta-information, such as retweets or likes, could also be useful in measuring support for – or the contestation of – given HT aspects. This would be useful for understanding how the broader public reacts to the tweets posted by each user group.

To date, most surveys about HT have been conducted with specific groups (such as health professionals and institutions), but there are few indicators of the perception and usage of HT by the general public (or representative samples of national populations). A space has thus been incentivised for research that identifies people's experiences when taking up or resisting new digital HT. Our study provides insights about what could also be potential survey interests. Developing survey items about HT would allow for a direct comparison between spontaneous online discussions and structured survey opinions.

Despite these limitations, we are confident that the findings from our study can help major health actors (such as HT companies and practitioners) to better target their campaigns while considering the concerns expressed by the different online audiences. This is in line with findings from Obembe et al. (2021) who studied tourist public responses on social media to crisis communications during the early stages of COVID-19. Indeed, the authors have shown that online publics played a key role in shaping the narratives of the crisis, thereby facilitating public engagement. However, a combination of analytical strategies and data sources is needed to take the next steps beyond the 'what has happened' to the 'why it happens' (Kar & Dwivedi, 2020).

Data availability statement

Original Twitter data were collected for the analysis and are not available for public access.

Supplemental online material

Please refer to the Annex 1 ("Topic modelling results (manual labelling, topic weight, and top terms").

Declaration of Competing Interest

We have no conflicts of interest to disclose.

Funding

None.

Acknowledgements

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Annex 1: Topic modelling results (manual labelling, topic weight, and top terms)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
1	patient needs	patients	patients, can, need, know, don, like, think, just, people, healthcare	0,04533	0,02712	0,02367	0,02962	0,01770	0,03451
2	/	/	health, forward, looking, great, digital, amp, today, event, day, innovation	0,03731	0,01924	0,02988	0,02225	0,02373	0,02409
3	/	/	time, patients, get, will, can, now, long, need, just, right	0,03511	0,01946	0,01824	0,02241	0,01382	0,02491
4	digital transformation	innovations	digital, healthcare, health, new, technology, will, transformation, future, care, innovation	0,03449	0,02283	0,02094	0,02454	0,02099	0,02080
5	public system	health system	health, public, amp, people, medicine, science, need, will, just, don	0,03322	0,01903	0,01892	0,02383	0,01199	0,02978
6	patient technol-ogy/support	patients	care, patients, health, can, help, home, providers, technology, learn, patient	0,03059	0,02080	0,01402	0,01531	0,02069	0,01455
7	specialists (CEO/professor/etc)	actors	health, amp, ceo, director, prof, president, healthcare, professor, john, founder	0,02784	0,01502	0,01352	0,01365	0,01640	0,01168
8	crisis response	security	health, covid, pandemic, crisis, response, coronavirus, public, healthcare, care, can	0,02745	0,01729	0,01387	0,01614	0,01587	0,01518
9	services/programs	innovations	health, care, nhs, digital, amp, across, innovation, support, new, social	0,0274	0,00660	0,03375	0,01683	0,01926	0,01425
10	patient teams	patients	great, work, patient, amp, team, patients, safety, thanks, thank, see	0,02728	0,01248	0,02057	0,01362	0,01437	0,01846
11	care system	health system	care, health, healthcare, value, based, patient, system, amp, approach, systems	0,02711	0,01772	0,01326	0,01424	0,01576	0,01562
12	report/litterature	information	health, new, read, report, research, article, amp, paper, published, medicine	0,02711	0,01390	0,01688	0,01438	0,01368	0,01804
13	webinars	information	join, webinar, register, amp, health, free, will, learn, now, next	0,02657	0,01453	0,01933	0,01248	0,02085	0,01241
14	patient safety/experience	patients	patient, care, improve, can, healthcare, outcomes, safety, amp, experience, quality	0,02643	0,01989	0,01279	0,01437	0,01871	0,01316
15	partient access to records	patients	patients, can, app, health, patient, access, online, video, nhs, help	0,0262	0,01354	0,01888	0,01666	0,01720	0,01534
16	patient care	patients	patients, can, data, make, amp, making, patient, need, care, decision	0,02617	0,01544	0,01450	0,01462	0,01459	0,01509
17	quality systems (social needs)	health system	health, care, amp, access, social, need, services, quality, systems, communities	0,02596	0,01563	0,01268	0,01314	0,01556	0,01672
18	impact of technology	innovations	healthcare, technology, make, will, can, future, look, change, like, impact	0,02554	0,01447	0,01434	0,01575	0,01317	0,01487
19	telemedicine	innovations	telehealth, telemedicine, care, virtual, patients, remote, visits, covid, pandemic, patient	0,0249	0,02012	0,01027	0,01788	0,01459	0,01553
20	health solutions	innovations	healthcare, amp, challenges, together, solutions, innovation, health, technology, industry, can	0,02394	0,01265	0,01518	0,01232	0,01515	0,01132

(*continued on next page*)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
21	mental health	health domains	health, mental, can, help, day, awareness, mentalhealth, week, world, support	0,02246	0,01308	0,01549	0,01222	0,01664	0,01344
22	health panels/discussions	information	join, health, will, amp, register, healthcare, live, panel, today, don	0,02229	0,01704	0,01301	0,01635	0,01675	0,01192
23	european public system	health system	health, digital, global, amp, european, public, systems, europe, national, policy	0,02198	0,00887	0,01793	0,01323	0,01408	0,01328
24	research papers	information	health, digital, new, evidence, study, review, interventions, based, use, research	0,02177	0,01058	0,01556	0,00995	0,01158	0,01708
25	artificial intelligence	innovations	tech, health, intelligence, artificial, healthcare, learning, digital, technology, machine, via	0,02039	0,01281	0,01786	0,02307	0,01202	0,01561
26	big data	innovations	data, health, analytics, use, research, patient, can, big, real, using	0,01953	0,01197	0,01139	0,01099	0,01158	0,01107
27	partnerships	industry	health, excited, proud, team, announce, new, healthcare, work, see, part	0,0195	0,01160	0,01056	0,01118	0,01157	0,01054
28	blockchain industry	industry	healthcare, technology, technologies, blockchain, industry, market, via, trends, will, tech	0,01896	0,01362	0,01181	0,01342	0,01297	0,01108
29	hiring opportunities	industry	health, team, apply, looking, join, amp, research, opportunity, interested, work	0,01863	0,00921	0,01749	0,00992	0,01342	0,01290
30	/	/	health, get, day, one, time, just, week, can, amp, today	0,01848	0,01039	0,00981	0,01159	0,00918	0,01090
31	insurance	health system	health, healthcare, care, insurance, survey, patients, costs, cost, new, study	0,01839	0,01636	0,00645	0,01440	0,01033	0,01176
32	patient health record	patients	health, data, patient, records, ehr, electronic, record, systems, information, platform	0,01822	0,01264	0,01016	0,01211	0,01132	0,01126
33	world health (future of health)	innovations	health, role, amp, play, people, healthy, can, future, work, world	0,01818	0,00906	0,01155	0,00878	0,01014	0,01049
34	health workers (e.g., nurses)	actors	health, day, thank, nurses, care, patients, healthcare, workers, amp, world	0,01813	0,01355	0,00922	0,00960	0,01230	0,01176
35	family doctor (physician)	actors	patient, patients, experience, family, care, doctor, voice, physician, engagement, can	0,01811	0,01279	0,00827	0,01203	0,00918	0,01182
36	virtual events	information	health, conference, will, event, join, healthcare, register, annual, visit, week	0,01805	0,01036	0,01358	0,01008	0,01547	0,00801
37	risks (pandemic, mental, etc)	security	health, mental, social, people, amp, issues, risk, impact, can, covid	0,01798	0,01138	0,01005	0,01022	0,00975	0,01387
38	tracing (for covid)	covid	health, public, covid, coronavirus, cases, testing, contact, new, will, tracing	0,01784	0,01111	0,00860	0,01256	0,00898	0,01362
39	learning ressources	innovations	health, care, learning, resources, amp, new, free, available, healthcare, professionals	0,01775	0,00715	0,01598	0,00856	0,01257	0,00886
40	project development	industry	research, health, new, amp, funding, innovation, will, support, projects, project	0,01771	0,00775	0,01430	0,00798	0,01295	0,00854

(*continued on next page*)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
41	school/university	education	medicine, health, students, medical, school, amp, program, university, faculty, research	0,01771	0,01441	0,00568	0,00616	0,01132	0,01008
42	equity (gender/race/etc)	education	health, women, amp, black, gender, racism, equity, disparities, sexual, diversity	0,01767	0,01223	0,00833	0,00886	0,00978	0,01203
43	hospital/emergency	actors	patients, hospital, care, covid, hospitals, patient, home, emergency, icu, new	0,01729	0,01052	0,00996	0,00991	0,00929	0,01170
44	surveys	information	please, survey, health, help, share, want, know, can, take, get	0,01728	0,00762	0,01275	0,00900	0,00957	0,01146
45	patient stories	patients	patient, patients, one, just, life, people, like, story, medicine, amp	0,01717	0,01089	0,00830	0,01233	0,00664	0,01388
46	problem solving	innovations	healthcare, system, health, technology, problem, need, one, can, care, change	0,01646	0,01004	0,00860	0,01130	0,00734	0,01161
47	precision medicine (cell/genomics/etc)	health domains	medicine, precision, new, technology, research, amp, technologies, cell, disease, cancer	0,01618	0,00973	0,01036	0,01008	0,00957	0,00999
48	challenge	innovations	health, apply, challenge, now, digital, tech, innovation, deadline, open, healthcare	0,01599	0,00972	0,01667	0,01056	0,01481	0,00846
49	cardiovascular diseases	health domains	patients, heart, disease, risk, study, can, stroke, failure, chronic, diseases	0,01584	0,01093	0,00919	0,00928	0,00961	0,01101
50	medicare/medicaid	health system	health, telehealth, medicare, state, care, new, healthcare, medicaid, services, will	0,01526	0,01606	0,00433	0,01286	0,00973	0,00918
51	gratulations/awards	information	health, award, awards, congratulations, year, innovation, best, tech, healthcare, winners	0,01432	0,00964	0,01101	0,00888	0,01182	0,00787
52	health costs/fundings	health system	health, year, healthcare, million, billion, report, per, funding, growth, digital	0,01431	0,00966	0,00746	0,00994	0,00770	0,00958
53	women in tech	education	health, tech, healthcare, companies, digital, big, via, digitalhealth, women, startups	0,01431	0,01057	0,00824	0,01164	0,00743	0,01033
54	/	/	years, last, year, week, one, months, ago, two, past, next	0,01424	0,00772	0,00666	0,00744	0,00695	0,00707
55	startups	industry	health, tech, amp, startups, innovation, companies, medtech, digitalhealth, healthcare, healthtech	0,01421	0,00684	0,01189	0,01031	0,01077	0,00680
56	medical market (de-vices/regulations/etc)	industry	medtech, medical, amp, tech, device, med, industry, devices, companies, innovation	0,01403	0,00577	0,01107	0,01151	0,00915	0,00630
57	funding platforms	education	health, digital, tech, startup, via, million, raises, funding, healthcare, platform	0,01398	0,01080	0,00957	0,01212	0,00862	0,00999
58	donations/conferences	information	health, conference, now, register, don, event, miss, join, get, digital	0,01384	0,00781	0,01216	0,00918	0,01240	0,00714
59	(cyber)security	security	data, privacy, security, health, healthcare, patient, cybersecurity, cyber, information, amp	0,01376	0,01032	0,00784	0,01112	0,00853	0,00879

(*continued on next page*)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
60	aging	health domains	healthy, hearing, health, aging, can, ageing, older, sleep, amp, exercise	0,01335	0,00891	0,00754	0,00615	0,00907	0,00779
61	/	/	podcast, health, listen, episode, healthcare, digital, ceo, new, amp, latest	0,0132	0,00975	0,00713	0,01352	0,00859	0,00670
62	covid testing	covid	patients, covid, test, study, testing, positive, symptoms, coronavirus, tests, new	0,01316	0,00998	0,00815	0,00918	0,00718	0,01246
63	healthcare company (e.g., Amazon cited)	industry	health, healthcare, company, amazon, care, digital, new, telehealth, via, united	0,0131	0,01404	0,00554	0,01918	0,00659	0,01023
64	head of medicine (chief/officer/director/etc)	actors	chief, officer, health, healthcare, director, medical, ceo, amp, technology, digital	0,01231	0,00819	0,00609	0,00739	0,00817	0,00540
65	youth wellbeing	health domains	health, people, mental, young, support, amp, help, can, social, services	0,01195	0,00390	0,01238	0,00651	0,00833	0,00700
66	safety (covid distancing)	covid	stay, healthy, home, keep, safe, can, health, amp, help, people	0,01186	0,00759	0,00633	0,00697	0,00701	0,00796
67	children health	health domains	health, children, mental, amp, child, school, kids, youth, schools, young	0,01186	0,00735	0,00611	0,00455	0,00786	0,00634
68	depression/anxiety	health domains	mental, long, health, term, patients, depression, study, anxiety, therapy, can	0,0116	0,00724	0,00707	0,00633	0,00666	0,00797
69	latest news	information	latest, newsletter, healthcare, blog, health, read, technology, news, post, featuring	0,01122	0,00791	0,00659	0,00611	0,00786	0,00679
70	food/diet/nutrition	health domains	healthy, food, health, diet, can, nutrition, eating, amp, eat, foods	0,01115	0,00879	0,00580	0,00634	0,00816	0,00826
71	surgery	health domains	patients, patient, cancer, surgery, treatment, new, first, therapy, brain, technology	0,01112	0,00836	0,00784	0,00719	0,00780	0,00801
72	clinical trial	education	patient, clinical, research, trials, patients, trial, amp, engagement, involvement, public	0,01069	0,00575	0,00704	0,00668	0,00576	0,00648
73	disrupting issues/innovations	innovations	latest, thanks, health, daily, healthcare, technology, innovation, news, disrupting, global	0,01068	0,01236	0,00530	0,00657	0,01144	0,01554
74	health and environment (climate/air/etc)	health domains	health, amp, global, climate, change, public, diseases, world, disease, air	0,01023	0,00592	0,00621	0,00530	0,00717	0,00608
75	covid staff	covid	workers, healthcare, care, ppe, covid, health, amp, help, patients, support	0,00964	0,00604	0,00505	0,00568	0,00569	0,00653
76	wearables	innovations	health, monitoring, wearable, wearables, patient, new, data, test, apple, devices	0,00949	0,00769	0,00578	0,00734	0,00649	0,00593
77	headlines	information	health, today, thanks, hit, watch, connect, headlines, edition, stay, amp	0,00946	0,00750	0,00491	0,01012	0,00484	0,00674
78	health technology	innovations	technology, health, design, digital, following, thanks, user, technologies, solutions, game	0,00943	0,00473	0,00788	0,00546	0,00670	0,00460
79	NHS	health system	nhs, health, read, digital, trust, digitalhealth, news, full, story, healthtech	0,00899	0,00328	0,01024	0,01196	0,00619	0,00365
80	mobile apps	innovations	health, mental, apps, app, help, digital, support, can, mobile, tools	0,0089	0,00553	0,00590	0,00479	0,00603	0,00501

(*continued on next page*)

n°	manual label	categorisation	top words	weight	United States	Europe	Advocates	Institutions	Specialists
81	chronic conditions/diseases	health domains	diabetes, patients, chronic, management, health, disease, conditions, type, people, care	0,00854	0,00491	0,00508	0,00422	0,00530	0,00458
82	international relations (Sinai/Australia/Germany/Canada/etc)	industry	health, digital, minister, sinai, national, new, australia, system, today, first	0,00844	0,00294	0,00467	0,00629	0,00438	0,00504
83	saving lifes (with technology)	innovations	lives, life, help, save, people, technology, healthier, can, saving, live	0,00797	0,00365	0,00538	0,00417	0,00480	0,00366
84	cancer	health domains	cancer, patients, treatment, breast, screening, amp, diagnosis, oncology, skin, awareness	0,00773	0,00562	0,00346	0,00318	0,00518	0,00425
85	interoperability	innovations	health, interoperability, data, patient, information, access, healthcare, rule, onc, amp	0,00758	0,00803	0,00230	0,00462	0,00586	0,00382
86	medications/substances (for pain and addictions)	health domains	patients, use, medication, opioid, treatment, adherence, drug, addiction, pain, substance	0,00755	0,00664	0,00288	0,00349	0,00533	0,00422
87	engineering	innovations	health, research, science, amp, technology, university, students, engineering, based, course	0,00752	0,00293	0,00588	0,00295	0,00474	0,00462
88	developing countries (india/africa/etc)	health system	health, india, amp, africa, healthcare, bharat, ayushman, pmjay, corona, south	0,00723	0,00196	0,00220	0,00252	0,00487	0,00500
89	/	/	happy, year, health, healthy, new, family, day, holiday, christmas, good	0,00694	0,00493	0,00347	0,00389	0,00484	0,00431
90	vaccines (covid/flu/etc)	covid	vaccine, vaccines, health, covid, flu, vaccination, get, amp, public, workers	0,00692	0,00472	0,00335	0,00459	0,00386	0,00477
91	generic drugs	health domains	medicines, good, latest, generic, drug, use, fda, drugs, safety, thanks	0,00653	0,00272	0,00470	0,00412	0,00499	0,00337
92	smoking	health domains	health, media, social, digital, publications, smoking, today, quit, web, twitter	0,00594	0,00545	0,00234	0,00410	0,00540	0,00251
93	training and education (simulation/imaging/radiology/etc)	education	technology, simulation, healthcare, training, radiology, imaging, aid, medical, education, amp	0,00576	0,00435	0,00278	0,00294	0,00436	0,00222
94	maternity/motherhood	health domains	healthcare, icymi, health, harlow, women, maternal, pregnant, pregnancy, mothers, babies	0,00519	0,00534	0,00179	0,00165	0,00525	0,00193
95	covid in UK (stocks management)	covid	via, coronavirus, healthinnovations, health, pharma, stocks, healthcare, brexit, stories, news	0,0051	0,00246	0,00756	0,00302	0,00296	0,00678
96	eyes (ophthalmology/optometry/etc)	health domains	eye, vision, video, patient, watch, peek, amp, ophthalmology, patients, videos	0,0046	0,00270	0,00281	0,00202	0,00289	0,00270
97	(international) congress	information	digital, now, world, book, health, congress, online, london, healthcare, conference	0,00317	0,00075	0,00654	0,00117	0,00499	0,00100
98	private clinics (arrayit)	health system	healthcare, arrayit, san, sales, team, life, usa, sciences, top, markets	0,00289	0,00392	0,00086	0,00139	0,00341	0,00090
99	chronic pain (hand/back/shoulders/etc)	health domains	pain, health, free, call, randolph, screening, hand, foot, back, register	0,00266	0,00288	0,00120	0,00099	0,00283	0,00110
100	services (premium)	patients	health, clinical, healthcare, services, testing, sales, team, providing, premium, reports	0,00148	0,00193	0,00123	0,00059	0,00233	0,00038

Annex 2: Semantic relations based on the top 200 terms of our corpus of tweets

Annex 3: Distribution of sentiment by group of actor

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of medical Internet research, 22*(4), e19016. 10.2196/19016.

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM, 55*(4), 77–84. 10.1145/2133806.2133826.

Brandt, C. J., Søgaard, G. I., Clemensen, J., Sndergaard, J., & Nielsen, J. B. (2018). General Practitioners' Perspective on eHealth and Lifestyle Change: Qualitative Interview Study. *JMIR mHealth and uHealth, 6*(4), e88. 10.2196/mhealth.8988.

Bundschus, M., Tresp, V., & Kriegel, H. P. (2009). Topic models for semantically annotated document collections. NIPS workshop: Applications for Topic Models: Text and Beyond. pp. 1-4.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific reports, 10*(1), 1–10. 10.1038/s41598-020-73510-5.

European Commission, Brussels (2015). "Flash Eurobarometer 404 (European Citizens´ Digital Health Literacy)". GESIS Data Archive, Cologne. ZA5948 Data file Version 1.0.0, https://doi.org/ 10.4232/1.12194

Fagerlund, AJ, Holm, IM, & Zanaboni, P. (2019). General practitioners' perceptions towards the use of digital health services for citizens in primary care: a qualitative interview study. *BMJ Open, 9*(e028251), 1–17. 10.1136/bmjopen-2018-028251.

Garg, R., Kiwelekar, A. W., Netak, L. D., & Bhate, S. S. (2021a). Personalization of news for a logistics organisation by finding relevancy using NLP. In V. K. Gunjan, & J. M. Zurada (Eds.), *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI* (pp. 215–226). Cham: Springer. Volume 2https://doi.org/. 10.1007/978-3-030-68291-0_16.

Garg, R., Kiwelekar, A. W., Netak, L. D., & Ghodake, A. (2021b). i-Pulse: A NLP based novel approach for employee engagement in logistics organization. *International Journal of Information Management Data Insights, 1*(1), Article 100011. 10.1016/j.jjimei.2021.100011.

García-Díaz, J. A., Apolinario-Arzube, Ó., Medina-Moreira, J., Luna-Aveiga, H., Lagos-Ortiz, K., & Valencia-García, R. (2018). Sentiment Analysis on Tweets related to infectious diseases in South America. In Proceedings of the Euro American Conference on Telematics and Information Systems, *Brazil* (pp. 1–5). 10.1145/3293614.3293647.

Grimmer, J., & Stewart, B. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297. 10.1093/pan/mps028.

Grover, P., Kar, A. K., & Davies, G. (2018). Technology enabled Health – Insights from twitter analytics with a socio-technical perspective. *International Journal of Information Management, 43*, 85–97. 10.1016/j.ijinfomgt.2018.07.003.

HIMSS Analytics. (2019). "Annual European eHealth Survey 2019". eHealth Trendbarometer. https://europe.himssanalytics.org/sites/himssanalytics_europe/files/eHealth%20TRENDBAROMETER%20-%20HIMSS%20Analytics%20Annual%20European%20eHealth%20Survey%202019.pdf

Jiang, S. (2019). Functional interactivity in social media: an examination of Chinese health care organizations' microblog profiles. *Health Promotion International, 34*(1), 38–46. 10.1093/heapro/dax056.

Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research–Moving away from the "What" towards the "Why". *International Journal of Information Management, 54*, Article 102205. 10.1016/j.ijinfomgt.2020.102205.

Kletecka-Pulker, M., Mondal, H., Wang, D., Parra, R. G., Maigoro, A. Y., Lee, S., … Atanasov, A. G. (2021). Impacts of biomedical hashtag-based Twitter campaign:# DHPSP utilization for promotion of open innovation in digital health, patient safety, and personalized medicine. *Current Research in Biotechnology, 3*, 146–153. 10.1016/j.crbiot.2021.04.004.

Koteyko, N, Hunt, D, & Gunter, B. (2015). Expectations in the field of the internet and health: an analysis of claims about social networking sites in clinical literature. Sociol Health Illn. Mar*, 37*(3), 468–484. 10.1111/1467-9566.12203.

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), Article 100017. 10.1016/j.jjimei.2021.100017.

Lee, J., Kim, J., Hong, Y. J., Piao, M., Byun, A., Song, H., & Lee, H. S. (2019). Health Information Technology Trends in Social Media: Using Twitter Data. *Healthcare informatics research, 25*(2), 99–105. 10.4258/hir.2019.25.2.99.

Lupton, D. (2012). M-health and health promotion: the digital cyborg and surveillance society. *Social Theory and Health, 10*(3), 229–244. https://link.springer.com/content/pdf/10.1057/sth.2012.6.pdf.

Lyon, D. (2010). Surveillance, power and everyday life. In P. Kalantzis-Cope, & K. Gherab-Martin (Eds.), *Emerging Digital Spaces in Contemporary Society* (pp. 107–120). Houndsmills, UK: Palgrave Macmillan.

Mahdikhani, M. (2021). Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of COVID-19 pandemic. *International Journal of Information Management Data Insights, 100053*. 10.1016/j.jjimei.2021.100053.

McCallum, A. K. (2002). "MALLET: A Machine Learning for Language Toolkit". http://mallet.cs.umass.edu

McClure, D. (2015). "Textplot refresh – Python 3, PyPI, CLI app". http://dclure.org/tutorials/textplot-refresh/

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Obembe, D., Kolade, O., Obembe, F., Owoseni, A., & Mafimisebi, O. (2021). COVID-19 and the tourism industry: an early stage sentiment analysis of the impact of social media and stakeholder communication. *International Journal of Information Management Data Insights, 1*(2), Article 100040. 10.1016/j.jjimei.2021.100040.

OECD (Organisation for Economic Co-operation and Development). (2019). *The Heavy Burden of Obesity: The Economics of Prevention*. Paris: OECD Publishing. 10.1787/67450d67-en.

OECD (Organisation for Economic Co-operation and Development). (2020). "ICT Access and Usage by Households and Individuals". OECD Telecommunications and Internet Statistics. Database accessed in January 2021. https://stats.oecd.org/Index.aspx?DataSetCode=ICT_HH2

Peluchette, J. V., Karl, K. A., & Coustasse, A. (2016). Physicians, patients, and Facebook: Could you? Would you? Should you? *Health marketing quarterly, 33*(2), 112–126. 10.1080/07359683.2016.1166811.

Rinker, T. W. (2018). "Textstem: Tools for Stemming and Lemmatizing Text". http://github.com/trinker/textstem

Rinker, T. W. (2019). "sentimentr: Calculate Text Polarity Sentiment". http://github.com/trinker/sentimentr

Ross, G. M. (2021). I use a COVID-19 contact-tracing app. Do you? Regulatory focus and the intention to engage with contact-tracing technology. *International Journal of Information Management Data Insights, 1*(2), Article 100045. 10.1016/j.jjimei.2021.100045.

Sahlgren, M., & Lenci, A. (2016). The Effects of Data Size and Frequency Range on Distributional Semantic Models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 975–980). Austin/Texas: Association for Computational Linguistics. https://aclanthology.org/D16-1099/.

Schmidt, B. (2017). "Word Vectors. An R package for building and exploring word embedding models". https://github.com/bmschmidt/wordVectors

Schumacher, S., & Kent, N. (2020). *"8 charts on internet use around the world as countries grapple with COVID-19"*. Pew Research Center https://www.pewresearch.org/fact-tank/2020/04/02/8-charts-on-internet-use-around-the-world-as-countries-grapple-with-COVID-19/.

Singh, K. N., Devi, S. D., Devi, H. M., & Mahanta, A. K. (2022). A novel approach for dimension reduction using word embedding: An enhanced text classification approach. *International Journal of Information Management Data Insights, 2*(1), Article 100061. 10.1016/j.jjimei.2022.100061.

Scott, B. K., Miller, G. T, Fonda, S. T., Yeaw, R. E., Gaudaen, J. C., Pavliscsak, H. H., Quinn, M. T., & Pamplin, J. C. (2020). Advanced Digital Health Technologies for COVID-19 and Future Emergencies. *Telemedicine and e-Health, 26*(10), 1226–1233. http://doi.org/10.1089/tmj.2020.0140.

Sterling, R., & LeRouge, C. (2019). On-demand telemedicine as a disruptive health technology: qualitative study exploring emerging business models and strategies among early adopter organizations in the United States. *Journal of medical Internet research, 21*(11), e14304. 10.2196/14304.

Tsai, J. M., Cheng, M. J., Tsai, H. H., Hung, S. W., & Chen, Y. L. (2019). Acceptance and resistance of telehealth: The perspective of dual-factor concepts in technology adoption. *International Journal of Information Management, 49*, 34–44. 10.1016/j.ijinfomgt.2019.03.003.

Weber, S. (2018). "The Great American Search for Healthcare Information. Accessed online on May 2021". https://www.webershandwick.com/wp-content/uploads/2018/11/Healthcare-Info-Search-Report.pdf

WHO (World Health Organization). (2015). *"Global survey on health technology assessment by national authorities. Main findings"*. Geneva: WHO http://www.who.int/health-technology-assessment/MD_HTA_oct2015_final_web2.pdf?ua=1.

Zhang, Y., Cao, B., Wang, Y., Peng, T. Q., & Wang, X. (2020). When Public Health Research Meets Social Media: Knowledge Mapping From 2000 to 2018. *Journal of medical Internet research, 22*(8), e17582. 10.2196/17582.