Recommendation Algorithms, a Neglected Opportunity for Public Health

REVUE MÉDECINE ET PHILOSOPHIE

Lê Nguyên Hoang[1], Louis Faucon[1] and El-Mahdi El-Mhamdi[2] *
[1] Ecole Polytechnique Fédérale de Lausanne, Switzerland, Emails: firstname.lastname@epfl.ch,
[2]École Polytechnique, France Email: elmahdi@elmhamdi.com

ABSTRACT

The public discussion on artificial intelligence for public health often revolves around future applications like drug discovery or personalized medicine. But already deployed artificial intelligence for content recommendation, especially on social networks, arguably plays a far greater role. After all, such algorithms are used on a daily basis by billions of users worldwide. In this paper, we argue that, left unchecked, this enormous influence of recommendation algorithms poses serious risks for public health, e.g., in terms of misinformation and mental health. But more importantly, we argue that this enormous influence also yields a fabulous opportunity to provide quality information and to encourage healthier habits at scale. We also discuss the philosophical, technical and socio-economical challenges to seize this immense opportunity, and sketch the outlines of potential solutions. In particular, we argue that it would be extremely helpful if public and private institutions could publicly take a stand, as this may then generate the necessary social, economical and political pressure to massively invest in the research, development and deployment of the potential solutions.

KEYWORDS : recommandation algorithms, public health, artificial intelligence.
DOI : 10.51328/103

Introduction

Artificial Intelligence (AI) promises major advances in medicine and public health, from advancing our knowledge of molecular biology (Senior et al., 2020; Gupta et al., 2020) to monitoring the progress of large-scale pandemics (Cavlo et al., 2020); from treatment development (Ong et al., 2020) to scalable diagnosis instruments (Rajpurkar et al., 2017). However, such developments also raise ethical concerns, especially in terms of software security, privacy and misuse (Fernández-Alemán et al., 2013). Moreover, one may argue that this line of work has been somewhat under-delivering, at least in contrast to the massive investments and hype that accompany the "AI and health" slogans (Shortliffe, 2019).

On the other hand, AI algorithms have been widely deployed on highly influential large scale platforms such as Facebook, YouTube or Twitter. The website Statista (Clement, 2020) reports that, in 2019, "the average daily social media usage of internet users worldwide amounted to 144 minutes per day". Moreover, what social media users are exposed to seems to be extremely dependent on the AI algorithms that the big tech companies use for content recommendation. YouTube Chief Product Officer[1] Neal Mohan reported that 70% of YouTube views result from algorithmic recommendation, as opposed to user's search, user's subscription feeds or external links (Solsman, 2018).

The algorithms designed to provide such recommendations are called *recommendation algorithms.* Recommendation algorithms typically survey the content published on their platforms and the activity of the platforms' users to organize users' news feeds, and to suggest new content, accounts and groups to consume, follow and join. Critically, such algorithms are *customized*. They provide tailored recommendations to different users, which

[1] As of January 2018, when the interview in the references was conducted.

makes them challenging to study, especially for external researchers (Aral, 2020).

In a widely debated experiment involving 689,003 Facebook users, Kramer et al. (2014) showed that a tiny modification of the Facebook newsfeed algorithm sufficed to slightly change users' behaviors within a single week. Namely, by simply removing 10% of negative posts on the Facebook newsfeed, their analysis revealed that users started posting more positive contents. Yet, Hohnhold et al. (2015) also showed that it usually takes weeks, if not months, to observe important user behavior changes on Google search after a modification of advertisement placements. These two studies, among others, suggest that large-scale human behaviors can be significantly modified by what social media algorithms choose to expose billions of their users to. Given the scale of the problem, Milano et al. (2020) argue that society at large should now be regarded as an important stakeholder of what social media algorithms recommend.

This evidently raises important ethical concerns, especially in terms of political manipulation and misinformation, as will be discussed in the next section of this paper. However, in most of this paper, we will mostly stress the fact that social media should also be regarded as an urgent opportunity to be seized by global health actors. Indeed, for many diseases, including obesity, COVID-19 and mental health, the information that patients are exposed to and the habits that they adopt are arguably some of the best available treatments. In fact, many health agencies have already massively invested, or are asked to massively invest, in public service announcements to mitigate these diseases (Nestle and Jacobson, 2000; COCONEL Group, 2020). Yet, such announcements have arguably failed to fully take advantage of the opportunities offered by social media.

In this paper, we argue that it is urgent that a lot more attention be paid to such opportunities, both by computer scientists and big tech companies, but also and equally importantly by philosophers, doctors and public health agencies. We believe that, to seize such opportunities, it is critical for all of these entities to recognize the importance of recommendation algorithms for global health, so that added social and legal pressures are put on social media companies. In fact, we argue that it would be extremely helpful if, for instance, such entities could publicly declare that making recommendation algorithms beneficial for public health has become a top healthcare priority.

We also present partial solutions to improve global health through recommendation algorithms, and call for further academic efforts to research, test, audit, analyze, question, correct, develop, secure, legislate, debate and deploy such solutions. Clearly, this is no easy task, but this is why massive efforts should be invested in researching solutions as soon as possible; and why advocating for the importance of recommendation algorithms seems extremely helpful.

The paper is organized as follows: In Section 2, we discuss the scale of the infodemic and why it is still arguably a neglected aspect of public health despite recent efforts. In Section 3, we develop an argument on how quality information can and should be used as a medical intervention. In Section 4, we review the impact large scale information systems could have on mental health. In Section 5, we present a series of easily implementable solutions. We conclude in Section 6.

Infodemic

A blend of *information* and epidemic, the term infodemic gained popularity during the ongoing COVID-19 pandemic[2] (Galloti et al., 2020) as misinformation campaigns around the disease gained momentum. The WHO, the UN, the UNICEF, the UNDP, the UNESCO, the UNAIDS, the ITU, the UN Global Pulse and the IFRC[3] (Joint Statement, 2020) started dedicating special efforts to battle the infodemic. In particular, WHO hosts a page dedicated to COVID-19 misinformation[4].

But the infodemic did not begin during the COVID-19 crisis. For instance, vaccine hesitancy gained enough ground that the WHO listed it (WHO, 2019) among its top ten public health crises as of January 2019 (one year before the COVID-19 outbreak). Johnson et al. (2020), after analysing discussion groups involving 100 million Facebook users, warns that their "theoretical framework reproduces the recent explosive growth in anti-vaccination views, and predicts that these views will dominate [the public opinion landscape] in a decade".

Misinformation also affects other areas of medicine, such as cancer (Loeb et al., 2019), alternative medicine (Collier, 2018) or nutrition (Myrick and Erlichman, 2020). Disturbingly, some influencers with misleading and dangerous health information are widely recommended by recommendation algorithms. For instance, despite having been reported hundreds of times back in 2016 for dangerous misinformation (Schepman, 2016; Olivier, 2020), and despite YouTube's claimed will to fight dangerous health misinformation, some YouTubers with misleading health-related content have gained over 500,000 subscribers, and accumulated millions of views in 2020 alone.

The COVID-19 pandemic has arguably made health misinformation even more problematic, especially as public health became all the more intertwined with political agendas (Biancovilli and Jurberg, 2020). Indeed, Bradshaw and Howard (2018) report that there are already important politically-motivated investments to bias public opinion. Concerningly, Vosoughi et al. (2018) provide evidence that some misinformation spreads much faster than some reliable information on social media. Unfortunately, such a phenomenon does not seem restricted to social networks with recommendation algorithms. More recently, the French COVID-19 conspiracy documentary *Hold Up* went viral on Vimeo, and through sharings of extracts from the documentary on all sorts of social medias. Its success allowed it to raise over 180,000 euros on the crowdfunding platform Ulule, and over 150,000 euros on the participative financing platform Tipeee. Machado et

[2] An interesting but not so surprising fact is that the term itself did not have its own wikipedia page before the Covid19 pandemic https://en.wikipedia.org/w/index.php?title=Infodemicaction=history, a page which is mostly a redirection to the one on the misinformation arroud the COVID-19 pandemic https://en.wikipedia.org/wiki/Misinformation-related-to-the-COVID-19-pandemic.

[3] "WHO" stands for "World Health Organization", "UN" for "United Nations", "UNICEF" for United Nations Children's Fund, "UNDP" for "United Nations Development Programme", "UNESCO" for "United Nations Educational, Scientific and Cultural Organization", "UNAIDS" for "United Nations Programme on HIV and AIDS", "ITU" for "International Telecommunication Union" and "IFRC" for "International Federation of Red Cross and Red Crescent Societies".

[4] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters.

al. (2019) report a large amount of junk news on public WhatsApp groups.

We stress the fact that misinformation need not be false information, or information from "fake news" sources (Grinberg et al., 2019). Factual evidence can deeply mislead, e.g., by telling the story of an individual who survived from a disease after they adopted some alternative medicine treatment, or of a child who sadly died a few months after they received some vaccine (Nisbett and Borgida, 1975). In fact, even statistical factual data can "lie", e.g., because of cherry-picking (Morse, 2010), misinterpretation (Kerr, 1998) or confounding variables (Simpson, 1951; Wagner, 1982). And while double-blind randomized controlled trials provide more robust and reliable signals, they too are malleable and can be hacked to provide misleading conclusions, as evidenced by the reproducibility crisis (Baker, 2016) and as argued in the case of drug testing by Stegenga (2018). This has led to harsh criticisms of today's dominant *null hypothesis statistical test* method (Amrhein et al., 2019), and a call for the research, development and use of more reliable statistical approaches and ways of phrasing research conclusions (Wasserstein et al., 2019).

Given the political motivations and financial incentives to spread some information rather than others (Kahan et al., 2013), and the cost of thorough fact checking, of sound reasoning, of exhaustive literature surveying and of querying multiple experts, in the absence of quality human or algorithmic content moderation, it seems that low quality information should be expected to dominate (Aral, 2020). This raises serious concerns for global health. It seems urgent to promote a lot more quality health information. Interestingly, recommendation algorithms could be a formidable asset to do so.

But reliable information about disease prevention and treatment may not be what is most urgent to recommend. Interestingly, the *Healthy People 2030* project by the *US Office of Disease Prevention and Health Promotion* added "attaining health literacy" as one of its foundational principles and overarching goals (ODPHP, 2017). They defined health literacy as "people's capacities to find, understand, and use health information and services for informed decisions and actions". But attaining such health literacy requires repeated exposures to quality pedagogical explanations of what a reliable information search entails. Unfortunately, such explanations are currently mostly drowned within a flood of junk news. The help of recommendation algorithms to dig out and promote such pedagogical contents seems essential.

Overall, instead of merely a threat, the predominance of recommendation algorithms could also be regarded as a great opportunity to drastically improve global health information; which could then drastically improve global health. Unfortunately, thus far, the enormous potential of recommendation algorithms to do good seems very neglected (Hoang, 2020a). Typically, the Netflix documentary *The Social Dilemma* depicts a very negative view on recommendation algorithms, and barely suggests that they could be a powerful asset to do a vast amount of good. It seems urgent to also underline the great public health opportunity offered by recommendation algorithms, if designed at least partly to improve global health.

Quality Information Saves Lives

Perhaps no story highlights the importance of quality science information better than Ignaz Semmelweis' failure to convince his colleagues of the importance of hygiene. In the 1840s, Semmelweis imposed a hand-washing policy in his clinic, before delivering babies. He then observed a drastic reduction of the childbed fever death rate of the new mothers. Semmelweis had discovered the staggering effectiveness of hygiene. Unfortunately, Semmelweis failed to communicate his findings effectively. Instead, he presented flawed explanations[5] (Tulodziecki, 2013). After years of rejections, Semmelweis became increasingly angry and even accused some of his colleagues of murder (Dykes, 2016). This led to a failure to standardize hygiene.

But producing quality information is merely the first necessary step. To exploit this information, it then needs to be effectively communicated. The COVID-19 pandemic arguably illustrates some failures to communicate quality information effectively. Indeed, before sufficiently compelling data allowed to conclude which COVID-19 vaccine should be widely approved and recommended (Zimmer, 2020), the best treatment available was arguably prevention through adequate behaviors. This includes hygiene, physical distancing and wearing masks, as well as the acceptance of more drastic measures such as tracing, isolating and lockdown. Arguably, there is a lot of room for improvement on this front, especially in Western countries.

The importance of quality communication has been long recognized for other health concerns, such as addictions, nutrition or lack of physical exercise, among others. Regulations forced tobacco and alcohol industries to include a warning against risks in their advertisements, while massive investments have been made to promote healthier diets[6]. In France, the slogans "eat five fruits and vegetables per day" or "antibiotics should not be [automatically self-prescribed]" have been memorized by millions of individuals.

Unfortunately, the effectiveness of such communications is unclear. In fact, Werle and Cuny (2012) designed a randomized controlled trial, whose results revealed a negative effect of the spots "eat five fruits and vegetables per day" on teenagers' food consumption. The authors suggest that broadcasting such spots after an advertisement for unhealthy hedonic food seems to have increased the acceptability of the hedonic food. More generally, we should not exclude the possibility that clear and factual messages backfire, especially if they aim to affect the audience's beliefs or behavior. Vogelsanger (2018) shares similar concerns in the context of "climate preaching", which may appear accusatory to climate denialists and reinforce their denial.

More generally, determining what makes a message effective is arguably a very challenging research endeavor. Garcia-Retamero and Cokely (2017) found notable differences between the effectiveness of infographics, and

[5] In particular, Tulodziecki (2013) argues that Semmelweise put too much emphasis on cadaveric material being the only cause of childbed fever. Yet this claim was inconsistent with childbed fever in hospitals and with the seasonality of childbed fever, which is why Semmelweis' views got rejected by his colleagues.

[6] While anecdotal, the video below shows Amazon's Alexa algorithm recommending fast food twice to a hungry user. Billions of users of Alexa, OK Google or Siri may be nudged towards unhealthy or healthy food, because of the way such algorithms are designed. https://twitter.com/so_sroy/status/1325392314739662850

showed notable differences. Crucially, when such messages are spread at scale, even a 1% difference in, say, user engagement, ends up having a huge impact.

What makes this line of research all the more challenging is that the reception of a message may strongly depend on the recipient's world view. Concerningly, Kahan et al. (2013) showed that politically motivated reasoning could make individuals diverge in their analysis of univocal but tricky numerical data. This means that a purely factual piece of information can be misleading for a subpopulation. Curiously, even the individuals who are more scientifically educated, often failed to correctly analyze the data if the data contradicted their intuitions; in fact, in such a case, educated individuals then performed just as poorly as uneducated individuals. Intriguingly, however, Kahan et al. (2017) later found out that, as opposed to intelligence and data, scientific curiosity seems to successfully make individuals with diverging political identities converge on factual considerations. More empirical data on the effectiveness of different messaging to different audiences seems critical.

Unfortunately, collecting data on the effectiveness of healthcare messages, especially in their actual context of diffusion, is extremely hard. But interestingly, social media platforms seem to be in a particularly fitting position to do so. Indeed, such platforms constantly collect massive amounts of data about what contents users are exposed to, how much time they spend watching such contents, and what the users do after being exposed to the contents. To research the effectiveness of different health messages, it seems critical that health agencies work with such platform providers. If done correctly, major progress may be achieved in domains like obesity, pandemic prevention or vaccination, especially if techniques like multi-armed bandit optimization are used to optimize the search for the most effective communication contents (Berry et al., 2010).

Results from social psychology seem important to integrate too. Typically, self-affirmation theory (Badea and Sherman, 2019) has consistently shown that individuals were more likely to accept contradictory views if they first affirm their values or successes that are not questioned by the contradictory views. Interestingly, for instance, Shermann et al. (2000) showed that subjects were significantly more receptive to articles on AIDS risks, and more willing to buy condoms, if they first underwent such a self-affirmation exercise. The self-affirmation exercise consisted of writing an essay describing why the subject's most important value is so important to them, and a time when it played a particularly important role. Meanwhile, subjects in the no-affirmation group had to do so for a value of little importance to them. Remarkably, the fraction of subjects who purchased condoms went from around 25% for the no-affirmation group to around 50% for the self-affirmation group.

Interestingly, experiments run by Werle et al. (2011) also give evidence of the importance of customizing the information to be communicated. In particular, the experiments showed that, when targeting teenagers, highlighting the social risks of obesity seems more effective. To make the experiment realistic, the authors tested the effect of a repeated exposure to prevention messages in a brochure containing diverse unrelated topics. They then asked subjects to fill forms, and to choose a thank-you snack. They found out that 65% of the subjects exposed to the social argument chose the healthy snack, as opposed to 55% for subjects exposed to the health argument. This strongly suggests that an effective health campaign must deliver several arguments and must customize the argument to deliver to the target audience. As it turns out, recommendation algorithms are precisely designed and optimized to perform such a customization.

Perhaps most importantly, recommendation algorithms can promote important health messages at scale, by reaching billions of individuals. Moreover, previous randomized controlled experiments on voting turnouts (Bond et al., 2012), positive messaging (Kramer et al., 2014) and ad blindness (Hohnhold et al., 2015) have already highlighted the effectiveness of algorithms at affecting users' feelings, beliefs and behaviors. Right now, this immense power is not used for good, and arguably has very undesirable consequences. If recommendation algorithms were designed for good, their enormous impact could improve the health of millions of individuals. Realizing this may change our discourse on recommendation algorithms, and what we demand from them, formally or informally. This seems critical to seize the opportunities presented by recommendation algorithms.

Mental Health

In this section, we propose to focus on the particular challenge of mental health, partly because it has been recognized as a growing concern (American Heart Association, 2019), and partly because, as we will see, the impact of social networks on mental health has gained a lot of attention lately.

In fact, on May 14, 2020, the World Health Organization argued that "substantial investment [is] needed to avert mental health crisis". Depression and anxiety were increasing, while many mental health services were interrupted because of the COVID-19 pandemic. What is more, the sudden isolation imposed by the lockdowns and physical distancing measures was increasing pre-existing concerns about the negative impacts of technology abuse on mental health — a trend that was underlined by the increasing popularity of the term *doomscrolling* (Watercutter, 2020).

In fact, according to Heron (2017), suicide is a leading cause of death among young people of age 15 to 34 in the United States, second only to unintended injuries. Worldwide, suicides add up to nearly one million deaths per year. It is noteworthy that the exposure to suicidal stories seem to increase suicide risks (Yıldız et al., 2018; Chan et al., 2018; Swedo et al., 2020), albeit stories about suicidal ideation without suicidal behavior may actually decrease suicide risks (Niederkrotenthaler, 2010). The former case is called the Werther effect, while the latter is known as the Papageno effect (Scherr and Steinleitner, 2015).

In any case, Carlyle et al. (2017) point out that content with the hashtags suicide and suicidal trigger more engagements than others. This suggests that (1) the Instagram recommendation algorithm might favor such contents and (2) users may be incentivized, consciously or not, to post such contents. This led the authors to conclude that "public health and mental health professionals should consider increased involvement on this platform". In particular, more research seems needed to better distin-

guish the contents that will likely increase suicide risks from those that will not, and to design algorithms that will promote the latter over the former.

But such dramatic mental health issues are not the only concerns to be had. Ironically, social media seem to also create loneliness and depression (Hunt et al., 2018), which increase risks for other diseases such as cancer or Alzheimer (Cacioppo and Cacioppo, 2014). Anxiety can also accompany the recurrent exposure to bad news, while anger may result from repeated exposure to aggressive opinions. In fact, according to Hubert (2014), in France, 18% of middle school students declare themselves victims of some cyber aggression. Jackson (2019) even argues that the negative bias of classical and social media may be causing *learned helplessness*, that is, users may feel that many challenges are beyond hope, which may hinder their willingness to do good. Finally, our digital experience may be causing a drastic reduction of our attention span, as suggested by measures made by González and Mark (2004) and Mark et al. (2016) on workers' average duration of online screen focus.

Interestingly, social media could be an important part of the solution to reduce the negative effects of technology abuse. As a starter, Eichstaedt et al. (2018) showed promising results for depression diagnosis, which is well known to be hard without social media data if patients do not themselves decide to consult doctors. In particular, while the use of social media seems overall beneficial to users' mental health, Holmgren and Coyne (2018) link abusive scrolling to relational aggression and depression.

But social media could be doing a lot more. They could arguably help users' mental health by promoting therapeutic contents, or by suggesting users to consult a psychiatrist. They could help users nurture their curiosity and their happiness, by favoring enthusiastic contents every now and then.

Evidently, any attempt to do so will be filled with potential pitfalls. It thus seems critical that such attempts result from a close collaboration between social media companies, medical experts and health organizations (Ginsberg and Burke, 2017), as was done for instance in Litt et al. (2020) and Ernala et al. (2020). In fact, this is only one of many ethical, technological and socio-economical challenges (Hoang and El Mhamdi, 2019) that need to be faced to seize the fabulous sanitary opportunities provided by social media. But such opportunities seem large enough to justify massive investments to research, develop and deploy potential solutions to meet these challenges.

Challenges and Potential Solutions

Given the limits of today's algorithms, to combat misinformation and promote quality information, it seems critical to rely on the judgment of experts. Several systems have been proposed, notably for fact-checking (Shamlo, 2018) or ethical decision making (Lee et al., 2019). These proposed solutions reveal several challenges.

First, there needs to be a mechanism for assessing the quality of experts to know how much their judgments can be trusted. Second, interfaces should be designed to collect quality expert judgments effortlessly. Third, potentially conflicting judgments from multiple experts should be aggregated into a unique recommendation decision. And finally, such solutions need to be actually implemented by the large social media companies. We discuss below these challenges in further details.

Identifying experts

In practice, expertise is most often certified by the degrees the experts obtained. Frustratingly, few universities enable third party websites to automatically certify the fact that a given expert obtained a given degree from them (Federal Trade Commission, 2005). Another common proxy to assess an expert's expertise is to check their publication list. It is noteworthy, however, that some platforms like Google Scholar do not allow for easy scraping of their data[7] (Else, 2018), which arguably hinders the ease to automatically verify experts' expertise. We acknowledge, however, that such proxies can be misleading (Waltman and Van Eck, 2012).

Besides, no single expert should be considered perfectly reliable, especially when they are discussing topics outside their domain of expertise. As an example, physics Nobel laureate Ivar Giaever opposes the scientific consensus on anthropogenic climate change. More generally, it seems desirable to aggregate the views of a large number of experts, rather than to take the view of a single expert for granted. One simple solution to do so would be to accept any individual with an email address from a trusted institution to register as an expert. However, to which extent should a given individual be regarded as an expert, especially on transdisciplinary questions with a moral dimension, is a question on which an agreement seems challenging to reach.

Designing an adequate interface

The role of the interface through which expert judgments are queried is arguably a very neglected research direction. After all, experts are typically busy people; it is often difficult to obtain enough of their attention to collect inputs from them. An appealing interface, which is effortless to use and which asks informative and yet easy-to-answer questions, seems critical to get the most out of experts.

One interesting proposal by Noothigattu et al. (2017) and by Lee et al. (2019) is to rely on comparison-based judgments. In this framework, the expert is repeatedly asked to choose one of two options. Social comparison theory (Festinger, 1954) argues that this better fits our natural judgment process. This can also avoid boundary effects, e.g. when users tend to give a maximal rating to too many items. Interestingly, the Bradley and Terry (1952) model allows to infer scores from such comparison-based judgments. In this model similar to the ELO system used to rank chess players, if an option is systematically preferred to another option, then the reconstructed scores of the former will be significantly larger than the score of the other option.

Aggregating potentially conflicting judgments

Unfortunately, we should expect experts to disagree on many topics, as evidenced by the lack of consensus in, say, moral philosophy. One solution to nevertheless reach a collective decision is to aggregate individual judgments

[7] It is unclear to which extent scraping is allowed by Google's Term of Service: "Google reserves the right to suspend or terminate your access to the services or delete your Google Account if [...] we reasonably believe that your conduct causes harm or liability to a user, third party, or Google — for example, by [...] scraping content that doesn't belong to you".

from multiple experts through some voting mechanism. Such voting mechanisms are the object of study of computational social choice theory (Brandt et al., 2016).

In particular, this field has highlighted the importance of properties such as strategy-proofness, which demands that honesty be an optimal strategy. This seems necessary to incentivize experts to provide high quality judgments. Interestingly, such aggregation of different experts' opinions also allows the system to be robust, because trust wrongly placed in a particular expert would be compensated by judgments from other experts. Some solutions such as *majority judgment* (Balinski and Laraki, 2011) have, to some extent, such properties.

Socio-economical challenges

Recently, the Tournesol framework has been proposed by Hoang (2020b) and aims to combine all the partial solutions discussed above. But even if a platform like Tournesol successfully identifies quality contents to recommend, this identification will have little impact if it is not used by actual large-scale recommendation algorithms.

It is noteworthy, as well, that any intervention must anticipate the risks of an exodus of social platform users to other less moderated platforms, such as *4chan*, *Parler* (Culliford and Paul, 2019) or *Bitchute* (Trujillo et al., 2020). To achieve this, the reliability of the content should not be the only feature that should matter. It seems critical as well that the content be engaging. More generally, it seems important that social media platforms strike a happy balance between entertaining users and delivering reliable information.

While implementing such ideas may conflict with their short-term priorities, interestingly, the social media leaders have publicly claimed their increased desire to make their platforms more beneficial to mankind (Wojcicki, 2020; Zuckerberg, 2018), which led to measurable improvements on their platforms, e.g., in terms of conspiracy theory recommendations (Faddoul et al., 2020) or added snippet links to Wikipedia or WHO. This is probably helped by increased social pressure and regulation threats.

However, to achieve more, additional support from health organizations seems greatly desirable. If they publicly declare that making recommendation algorithms robustly beneficial is a top global health priority, then we may expect an important increase of public and private investments in this research direction. Assuming that a compelling technical solution is then proposed, that social and legal pressures are large enough, there might then be a reasonable hope that such a solution will indeed be implemented by these social media companies.

Conclusion

In this paper, we argued for the importance of recommendation algorithms in improving public health. We discussed how both misinformation and social network addiction are public health challenges raised by these large scale systems, which may be aggravated by recommendation algorithms. But it is noteworthy that the absence of a recommendation algorithm, such as on direct messaging applications, still seems to expose us to such risks. Instead, we argued that recommendation algorithms should be regarded as an opportunity to have a large positive impact on public health. In particular, by better identifying which contents should be promoted at scale, provided that we could also convince or force large social media companies to promote such contents, we can vastly increase the reach of quality information. We identified several aspects of potential solutions, as well as numerous philosophical, technical and social challenges.

Our hope is that more computer scientists, technologists, companies, but also philosophers, medical doctors and health organizations, will research and promote such solutions as well, and that progress will be made to solve such challenges. In particular, we hope to have convinced our readers that it would be very valuable for the future of public health, if public and private institutions could publicly declare that recommendation algorithms have become a major risk and a massive opportunity for global health and treat these algorithms as such.

Acknowledgement

Authors would like to thank Konrad Seifert, Siméon Campos, Christophe Michel, Felix Grimberg, Mohamed Oubenal, Najib El Mokhtari, Stéphane Debove, Naomi Nederlof, Adam Shamir and Juliette Ferry-Danini for their useful feedbacks, which greatly improved the paper.

RÉFÉRENCES

Aral, S. (2020). The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt. Currency.

American Heart Association. (2019). Mental Health: A Workforce Crisis.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance.

Badea, C., & Sherman, D. K. (2019). Self-affirmation and prejudice reduction: When and why?. Current Directions in Psychological Science, 28(1), 40-46.

Baker, M. (2016). Reproducibility crisis. Nature, 533(26), 353-66.

Balinski, M., & Laraki, R. (2011). Majority judgment: measuring, ranking, and electing. MIT press.

Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). Bayesian adaptive methods for clinical trials. CRC press.

Biancovilli, P. & Jurberg, C. (2020). When governments spread lies, the fight is against two viruses: A study on the novel coronavirus pandemic in Brazil. Medrxiv.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 295-298.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4), 324-345. Bradsh

aw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. The Computational Propaganda Project, 1.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.). (2016). Handbook of computational social choice. Cambridge University Press.

Cacioppo, J. T., & Cacioppo, S. (2014). Social relationships and health: The toxic effects of perceived social isolation. Social and personality psychology compass, 8(2), 58-72.

Calvo, R. A., Deterding, S., & Ryan, R. M. (2020). Health surveillance during covid-19 pandemic.

Carlyle, K. E., Guidry, J. P., Williams, K., Tabaac, A., & Perrin, P. B. (2018). Suicide conversations on Instagram™: contagion or caring?. Journal of Communication in Healthcare, 11(1), 12-18.

Chan, S., Denny, S., Fleming, T., Fortune, S., Peiris-John, R., & Dyson, B. (2018). Exposure to suicide behaviour and individual risk of self-harm: Findings from a nationally representative New Zealand high school survey. Australian & New Zealand Journal of Psychiatry, 52(4), 349-356.

Clement J. (2020). Daily time spent on social networking by internet users worldwide from 2012 to 2019. https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide

COCONEL Group. (2020). A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. The Lancet. Infectious diseases, 20(7), 769.

Collier, R. (2018). Containing health myths in the age of viral misinformation.

Culliford E. and Paul, K. (2019). Unhappy with Twitter, thousands of Saudis join pro-Trump social network Parler. Reuter.

Dykes, B. (2016). A History Lesson On The Dangers Of Letting Data Speak For Itself. Forbes.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., ... & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences, 115(44), 11203-11208.

Else, H. (2018). How I scraped data from Google Scholar. Nature News QA.

Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020, April). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

Faddoul, M., Chaslot, G., & Farid, H. (2020). A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. arXiv preprint arXiv:2003.03318.

Federal Trade Commission (2005). Avoid Fake-Degree Burns By Researching Academic Credentials.

Festinger, L. (1954). A theory of social comparison processes. Human relations, 7(2), 117-140.

Fernández-Alemán, J. L., Señor, I. C., Lozoya, P. Á. O., & Toval, A. (2013). Security and privacy in electronic health records: A systematic literature review. Journal of biomedical informatics, 46(3), 541-562.

Galloti, R., Valle, F. Castaldo, N., Sacco, P., & De Domenico, M. (2020). Assessing the risks of "infodemics" in response to COVID-19 epidemics. Nature Human Behaviour.

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. Human factors, 59(4), 582-627.

Ginsberg, D. and Burke, M. (2017). Hard Questions: Is Spending Time on Social Media Bad for Us? Facebook Hard Questions.

González, V. M., & Mark, G. (2004, April). "Constant, constant, multi-tasking craziness" managing multiple working spheres. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 113-120).

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. Science, 363(6425), 374-378.

Gupta, H., McCann, M. T., Donati, L., & Unser, M. (2020). CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning. BioRxiv.

Heron, M. P. (2017). Deaths: leading causes for 2015.

Hoang, L. N. (2020a). Science communication desperately needs more aligned recommendation algorithms. Frontiers Communication, Science and Environmental Communication.

Hoang, L. N. (2020b). Tournesol: Collaborative content recommendations. https://bit.ly/tournesol-app

Hoang, L. N. & El Mhamdi, E. M. (2019). Le fabuleux chantier : Rendre l'intelligence artificielle robustement bénéfique. EDP Sciences.

Hohnhold, H., O'Brien, D., & Tang, D. (2015, August). Focusing on the Long-term: It's Good for Users and Business. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1849-1858).

Holmgren, H. G., & Coyne, S. M. (2017). Can't stop scrolling!: pathological use of social networking sites in emerging adulthood. Addiction Research Theory, 25(5), 375-382.

Hubert, T. (2014). Un collégien sur cinq concerné par la cyber-violence.

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. Journal of Social and Clinical Psychology, 37(10), 751-768.

Jackson, J. (2019). You Are What You Read: Why changing your media diet can change the world. Unbound Publishing.

Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., ... & Lupu, Y. (2020). The online competition between pro-and anti-vaccination views. Nature, 1-4. (https://www.nature.com/articles/s41586-020-2281-1.pdf)

Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC, 2020, Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. url: https://www.who.int/news-room/detail/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation

Kahan, D. M., Peters, E., Dawson, E., & Slovic, P. (2013). Motivated numeracy and enlightened self-government. Behavioural Public Policy, 1, 54-86.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. Political Psychology, 38, 179-199.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. Personality and social psychology review, 2(3), 196-217.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., ... & Procaccia, A. D. (2019). WeBuildAI: Participatory framework for algorithmic governance. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-35.

Litt, E., Zhao, S., Kraut, R., & Burke, M. (2020). What Are Meaningful Social Interactions in Today's Media Landscape? A Cross-Cultural Survey. Social Media+ Society, 6(3), 2056305120942888.

Loeb, S., Sengupta, S., Butaney, M., Macaluso Jr, J. N., Czarniecki, S. W., Robbins, R., ... & Langford, A. (2019). Dissemination of misinformative and biased information about prostate cancer on YouTube. European urology, 75(4), 564-567.

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., & Howard, P. (2019, May). A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In Companion proceedings of the 2019 World Wide Web conference (pp. 1013-1019).

Mark, G., Iqbal, S. T., Czerwinski, M., Johns, P., & Sano, A. (2016, May). Neurotics can't focus: An in situ study of online multitasking in the workplace. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 1739-1744).

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. AI & SOCIETY, 1-11.

Morse, J. M. (2010). "Cherry picking": Writing from thin data, 20(1), 3.

Myrick, J. G., & Erlichman, S. (2020). How audience involvement and social norms foster vulnerability to celebrity-based dietary misinformation. Psychology of Popular Media, 9(3), 367.

Nestle, M., & Jacobson, M. F. (2000). Halting the obesity epidemic: a public health policy approach. Public health reports, 115(1), 12.

Niederkrotenthaler, T., Voracek, M., Herberth, A., Till, B., Strauss, M., Etzersdorfer, E., ... & Sonneck, G. (2010). Role of media reports in completed and prevented suicide: Werther v. Papageno effects. The British Journal of Psychiatry, 197(3), 234-243.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 32(5), 932.

Noothigattu, R., Gaikwad, S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., & Procaccia, A. (2018). A voting-based system for ethical decision making. In Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

ODPHP (2017). Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives for 2030: Recommendations for an Approach to Healthy.

Olivier, J. (2020). Le youtubeur Thierry Casasnovas (Regenere) dans le viseur de la justice pour "mise en danger de la vie d'autrui". Télé-loisir.

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. BioRxiv.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Lungren, M. P. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225.

Schepman, T. (2016). Thierry Casasnovas, le gourou du « tout cru », vous attend tranquille sur YouTube. L'Obs.

Scherr, S., & Steinleitner, A. (2015). Between Werther and Papageno effects. Der Nervenarzt, 86(5), 557-565.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Penedones, H. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

Shamlo, N. B., & Alavi S. (2018) AVOW: Expert-Sourced Fact Checking and Statement Evaluation.

Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. Personality and Social Psychology Bulletin, 26(9), 1046-1058.

Shortliffe, E. H. (2019). Artificial intelligence in medicine: weighing the accomplishments, hype, and promise. Yearbook of medical informatics, 28(1), 257.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society: Series B (Methodological), 13(2), 238-241.

Solsman, J.E. (2018). YouTube's AI is the puppet master over most of what you watch. CNET.

Stegenga, J. (2018). Medical nihilism. Oxford University Press.

Swedo, E. A., Beauregard, J. L., de Fijter, S., Werhan, L., Norris, K., Montgomery, M. P., ... & Sumner, S. A. (2020). Associations between social media and suicidal behaviors during a youth suicide cluster in Ohio. Journal of Adolescent Health.

Trujillo, M., Gruppi, M., Buntain, C., & Horne, B. D. (2020). What is BitChute? In Proceedings of the 31st ACM Conference on Hypertext and Social Media (pp. 139-140).

Tulodziecki, D. (2013). Shattering the myth of Semmelweis. Philosophy of Science, 80(5), 1065-1075.

Vogelsanger, P. (2018). Preaching, accusation, guilt and denial: Learning from Ignaz Semmelweis for climate communication. Klimaatelier.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

Wagner, C. H. (1982). Simpson's paradox in real life. The American Statistician, 36(1), 46-48.

Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, 63(2), 406-415.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05".

Watercutter, A. (2020). Doomscrolling Is Slowly Eroding Your Mental Health. Wired.

Werle, C., Boesen-Mariani, S., Gavard-Perret, M. L., & Berthaud, S. (2011). Social risk efficacy in preventing youth obesity. ACR North American Advances.

Werle, C. O., & Cuny, C. (2012). The boomerang effect of mandatory sanitary messages to prevent obesity. Marketing Letters, 23(3), 883-891.

Wojcicki, S. (2020). My mid-year update to the youtube community. YouTube Official.

World Health Organization (WHO). (2019). Ten threats to global health in 2019. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019

World Health Organization (WHO). (2020). Substantial investment needed to avert mental health crisis. https://www.who.int/news/item/14-05-2020-substantial-investment-needed-to-avert-mental-health-crisis

Yıldız, M., Orak, U., Walker, M. H., & Solakoglu, O. (2018). Suicide contagion, gender, and suicide attempts among adolescents. Death studies.

Zimmer, C. (2020). First, a Vaccine Approval. Then 'Chaos and Confusion'. New York Times.

Zuckerberg, M. (2018). Facebook Post on interference and misinformation in elections. https://www.facebook.com/zuck/posts/10104797374385071