

# Into the Void: Understanding Online Health Information in Low-Web Data Languages

Hellina Hailu Nigatu<sup>1\*</sup>, Nuredin Ali Abdelkadir<sup>2,3</sup>, Fiker Tewelde<sup>4</sup>,  
Stevie Chancellor<sup>2</sup>, Daricia Wilkinson<sup>5</sup>

<sup>1</sup>UC Berkeley

<sup>2</sup>University of Minnesota

<sup>3</sup>Distributed AI Research Institute

<sup>4</sup>Addis Ababa University

<sup>5</sup>Arizona State University

## Abstract

Data voids—areas of the internet where reliable information is scarce or absent—pose significant challenges to online health information seeking, particularly for users operating in low-web data languages. These voids are increasingly encountered not on traditional search engines alone, but on social media platforms, which have gradually morphed into informal search engines for millions of people. In this paper, we introduce the phenomenon of *data horizons*: a critical boundary where algorithmic structures begin to degrade the relevance and reliability of search results. Unlike the core of a data void, which is often exploited by bad actors to spread misinformation, the data horizon marks the critical space where systemic factors, such as linguistic underrepresentation, algorithmic amplification, and socio-cultural mismatch, create conditions of informational instability. Focusing on Tigrinya and Amharic as languages of study, we evaluate (1) the common characteristics of search results for health queries, (2) the quality and credibility of health information, and (3) characteristics of search results that diverge from their queries. We find that search results for health queries in low-web data languages may not always be in the language of search and may be dominated by nutritional and religious advice. We show that search results that diverge from their queries in low-resourced languages are due to algorithmic failures, (un)intentional manipulation, or active manipulation by content creators. We use our findings to illustrate how a data horizon manifests under several interacting constraints on information availability.

**Supplementary Material** — <https://osf.io/dk7zc/>

## 1 Introduction

Social media platforms such as YouTube and TikTok are used as sources of information, including in critical settings like healthcare (Ozdemir et al. 2021). Users rely on these platforms to gain general information on health-related concerns (Chirumamilla and Gulati 2021), to seek support from others who may have similar health conditions (Naslund et al. 2014; Freeman, Caldwell, and Scott 2023), or to

learn more about planned medical procedures (Roberts et al. 2023). As such, these platforms have gradually morphed into important sources for health-related information.

Certain features of social media platforms make them appealing channels for health communication. For health information providers, social media platforms offer instant access to a wide range of audiences (Jafar et al. 2023). This quick and broad reach of social media platforms makes them especially attractive in public health crises such as pandemics (Talie Fenta, Bogale, and Anagaw 2024). For health information seekers, social media platforms provide a unique space for user-generated medical content that is “relatable” and easier to digest for the general public (Milton et al. 2023). However, while these affordances benefit the health information landscape, they could also be a source of harm. The far and immediate reach of social media platforms means that health misinformation can spread just—if not more—as quickly and as widely as credible health information (Sylvia Chou, Gaysynsky, and Cappella 2020). Further, while users value the relatability of social media health information, they at times find personal anecdotes untrustworthy (Quintero Johnson, Yilmaz, and Najarian 2017).

In the complex health information seeking landscape, language is a critical component affecting how health information is retrieved, distributed, and consumed. As Nigatu et al. argue, languages—and their speakers—may lack economic, human, data, and technological resources, creating a disparity in what language technologies are available to the speakers. For instance, most African languages are *low-web data languages*—i.e., they have limited digitally available data, which inherently limits the online health information available for their speakers (Hu et al. 2025). The lack of information in a critical setting like healthcare is arguably a form of harm by itself, as it creates a disparity in health information access (Ukonu and Ajaebili 2021). The health information gap means that speakers of these languages may be unable to access necessary—and at times life-saving information (Al Shamsi et al. 2020). This problem is especially precarious in African communities, where access to healthcare services is limited (Oleribe et al. 2019). Further, healthcare information available to these communities might be dominated by content from unverified sources. Prior work

\*corresponding author: [hellina\\_nigatu@berkeley.edu](mailto:hellina_nigatu@berkeley.edu)  
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

has shown that, despite high demand for natural cures by African users, search results provide limited high-quality resources on natural remedies (Abebe et al. 2019).

In this paper, we address a growing concern surrounding online information seeking: the proliferation of data voids. Golebiewski and boyd define data voids as “voids [that] occur when obscure search queries have few results associated with them” (Golebiewski and boyd 2019). In low-web data languages, data voids are not merely gaps in search results but pose significant risks in critical settings like health information, where poor search results can have severe consequences. To understand how platform design, language, and sociocultural context intersect to create or exacerbate information breakdowns, we posit the following questions:

- **RQ1:** What are the most common characteristics of search results retrieved for health-related queries in low web-data languages
- **RQ2:** What is the quality of health information retrieved for low-web data languages?
- **RQ3:** When search queries and results do not align, what are the common characteristics of those results?

Based on our findings, we introduce the concept of *data horizons*—the threshold where the quality and relevance of search sharply diminishes, marking the boundary of the start of a data void and information collapse. A data horizon is created from tensions among systemic constraints on information access, disintegrating the fabric of the information landscape into a potential data void. Contrary to a data void where external factors shape whether the void surfaces, at the data horizon, users are exposed to problematic content due to inherent limits in systemic structures. It is a tipping point in the user experience, beyond which the quality and credibility of information rapidly decline, and the space becomes more vulnerable to manipulation by nefarious actors.

We find that search results for low-web data health queries are dominated by natural remedies and religious content. We also find that most of the health information is produced by individual medical professionals whose credentials are not always easy to verify. Our findings also show that the health information gap in low-web data languages is exacerbated 1) algorithmically due to the inherent limits of information in the low-web data languages, (2) (un)intentionally through the spread of unverified remedies, even for chronic diseases with known interventions, and 3) by content creators who use non-medical terms to boost their content. Our contributions can be summarized as follows:

- We provide qualitative and quantitative insights into characteristics of search results for health-related queries in low-web data languages (Section 4.1) and the quality of health information that is available (Section 4.2).
- We identify 5 sources for search results that diverge from their queries when seeking health information in low-web data languages (Section 4.3).
- We introduce the concept of *data horizons*—boundaries right before a data void, where the information fabric disintegrates as a result of the tension caused by systemic constraints in information access (Section 5).

- Based on our findings, we provide recommendations for stakeholders in protecting individuals from crossing the data horizon into the void (Section 6).

## 2 Background and Related Work

In this section, we first provide a literature review on health communication on social media platforms. Then, we discuss the current landscape of online health information for low-web data language speakers. Finally, we present information on our two languages of study.

### 2.1 Health Communication and Social Media

Health communication refers to “the study and use of methods to inform and influence individual and community decisions that enhance health.” (Freimuth and Quinn 2004). For health communication to be effective, it has to be (1) understandable by the general public, requiring careful interpretation of dense technical and research concepts (Dahm 2012), (2) culturally relevant and contextual (Brooks, Manias, and Bloomer 2019; Tang et al. 2024), and (3) transmitted through channels of communication that are widely accessible by the community (Lee et al. 2020).

With the pervasive use of social media, platforms like YouTube and Facebook have opened avenues to reach a wider set of audiences across geographies (Jeyaraman et al. 2023). However, social media platforms have also opened the door for widespread misinformation, which in the context of healthcare information could lead to direct physical and mental harm (Ghenai and Mejova 2018). Usually, such misinformation is disseminated by non-medical professionals (Bautista, Zhang, and Gwizdka 2021). Yet, getting accreditation on social media platforms as a credible health information content creator is not an easily available process for all medical professionals. For instance, YouTube has a set of criteria for content creators to be registered as a source of health information on the platform (YouTube 2021). Currently, the criteria and processes are available for healthcare providers from 10 countries<sup>1</sup>; for healthcare providers outside those geographic locations “organizations with pre-existing, standard vetting mechanisms such as accredited hospitals, academic medical institutions, public health departments, or government organization [...] can submit a request.” (YouTube 2021) However, the request may take “months to show up in our [YouTube’s] features.” (YouTube 2021). Hence, depending on their location, some communities may have a hard time accessing contextual, verified health information.

The challenges in distributing verified health information result in a health information gap—where quality and verified health information is (1) not available or (2) not accessible to certain communities. Despite this gap, individuals increasingly rely on online platforms for their health information needs (Zhang, Milton, and Chancellor 2023). In addition to their accessibility, social media platforms create a space where users find “relatable” health information through shared personal experiences (Milton et al.

<sup>1</sup>The countries are USA, UK, Germany, Japan, India, Indonesia, Canada, South Korea, France, Brazil, and Mexico (YouTube 2021)

2023). However, not all health information needs are met through personal stories: for instance, some creators rely on (pseudo)anonymity to reduce risks associated with the self-disclosure of sensitive information while some users may avoid consuming health information based on personal experiences due to the anonymity of content creators making it difficult to verify the credibility (Milton et al. 2024). As a result, there is a tension where users require authoritative, verified health information and at the same time seek “relatable” content on social media platforms (Milton et al. 2023; Viviani and Pasi 2017). Further, online health information allows for seeking information on non-conventional medical information—for instance, users may seek and consume information on traditional, complementary, and integrative medicine (TCIM) (Trübner et al. 2025). Although such information may be useful and perceived as more reliable and accessible in some contexts, prior work has also raised concerns for how limited instructions and credible sources for information could create (un)intended risks for health information seekers (Karusala and Anderson 2022).

## 2.2 Health Information and African Communities

While the health information gap is a global phenomenon, it does not affect all communities equally. Several studies have demonstrated that marginalized groups are more likely to seek health information online (Ukonu and Ajaebili 2021; Nangsangna and Vroom 2019). Yet, there is limited online health information available for such communities.

Particularly for African communities, access to quality online health information is limited due to multiple interacting factors: First, as discussed in Section 1, African languages are often categorized as low-web data languages whose lack of digital data is further exacerbated in the context of healthcare (Hu et al. 2025). Secondly, online health information that is available may not account for the cultural and social backgrounds of African communities (Usman et al. 2024). Thirdly, there is a lack of verified health information from authoritative institutions for African communities (Abebe et al. 2019).

Prior works show that health information seeking behaviors are affected by the socio-cultural backgrounds of information seekers. In a large-scale analysis of search queries from the 54 African countries, Abebe et al. (2019) found that women were more likely to search for content about pregnancy and breastfeeding, while men were more likely to search for “cure news.” Socio-cultural backgrounds may also influence the medium through which users access online health information: Ukonu and Ajaebili (2021) found that women in Southeast Nigeria (financially) depended on their spouses for avenues to access health information. Njenga, Mogambi, and Oriaso (2024) found that the elderly in Kenya relied on relatives to seek online health information on their behalf as the platforms were inaccessible to them.

## 2.3 Study Context

As we have summarized in the sections above, search queries may surface irrelevant or unverified information for speakers of low-web data languages. This increases the

probability of low-web data language health queries returning “obscure” results, thus, intensifying the health information gap into a potential data void. Golebiewski and boyd (2019) showed that in the context of health, media manipulators posting anti-vaccination conspiracy content associated their content with other videos that were health-related. In the context of high-web data languages like English and socio-cultural context of the US, Golebiewski and boyd (2019) state that algorithmically “government and health professional content is more highly weighted for search queries related to vaccination.” However, in the context of low-web data language speakers, verified health information from authoritative institutions is not always available, thereby, affecting the ranking of available content.

While prior work has explored online health information seeking behavior, or data voids, or the impact of socio-cultural aspects in more of an isolated capacity, there is an urgent need to investigate the intersection of this trio. In this study, we address this gap and carefully examine the anatomy of this space, revealing a new phenomenon that has not been explored—a data horizon. We explore this phenomenon by collecting data for health queries in two low-web data languages: Amharic and Tigrinya. Amharic is spoken in Ethiopia by an estimated 57.5 million people worldwide (Basha et al. 2023). Tigrinya is spoken in Ethiopia and Eritrea by an estimated 10 million people worldwide (Haile, Ali, and Hadgu 2023). Both languages are Afro-Semitic languages and use the Ge’ez writing script, which is an Abugida writing system—each character in the script represents a consonant and a vowel<sup>2</sup>.

We selected YouTube and TikTok as platforms of study. Prior work has demonstrated that TikTok has garnered more views as a source of information for certain health conditions (Alkhodair et al. 2024). Prior research has also explored the use of YouTube as a source of health information and found that the majority of the content is not posted by medical professionals (Clarke et al. 2023). In online health information literature, the two platforms also present a contrasting case for video information sharing, in length and quality of video, where TikTok has shorter, lower quality videos (Munoz et al. 2024).

## 3 Methods

We used a mixed-methods approach to answer our research questions: we used topic modeling to automatically extract salient topics from our collected data and used qualitative analysis to interpret the patterns in our data. In this section, we first describe how we collected our dataset and give details on each of our methods.

### 3.1 Dataset Preparation

Our study goal is to understand the characteristics of the health information gap on YouTube and TikTok in two low-web data languages. Below, we first describe how we prepared queries in the two low-web data languages. Then we describe how we collected data from the two platforms using the queries we prepared.

<sup>2</sup><https://www.omniglot.com/writing/ethiopic.htm>

Disease Category	Examples	Total Count
Communicable	Meningitis, Tuberculosis	6
Non-Communicable	Diabetes, Cancer	7
Reproductive Health	Chlamydia, Gonorrhea	5
Women and Children	Cervical cancer, Measles	11
Mental Health	Anxiety, Depression	4
Other	Vaccination, Malnutrition	5

Table 1: Health conditions used to create search queries in the two languages of study.

**Preparing Queries** We first collected diseases and health conditions, prioritizing health conditions most prevalent in the Ethiopian and Eritrean communities according to the World Health Organization (WHO), Center for Disease Control (CDC), and the Ministry of Health of Ethiopia (MOH). This resulted in 35 diseases across 5 categories. Table 1 shows the five disease categories along with the number of diseases in each category and some example health conditions. For YouTube, the first and second authors, who are native speakers of Amharic and Tigrinya, respectively, prepared queries in each language separately from the list of diseases and health conditions. After preparing the queries independently for each language, the authors discussed the queries to understand the similarities and differences in the prepared queries across the languages. We focused on four themes in the list of queries we prepared: prevention, symptom, treatment, and cure. For each health condition, we prepared 2-4 queries to account for spelling differences, paraphrases, and alternative disease names<sup>3</sup>. We refined our queries in weekly meetings where we discussed potential ambiguities and resolved confusions or disagreements. By the end of our process, we had 219 queries for Amharic and 246 queries for Tigrinya for YouTube. For TikTok, we used the names of the diseases and health conditions instead of the full question queries, as the TikTok API returns search results for individual keywords. As a result, we did not have four stages of disease progression for TikTok.<sup>4</sup>

**Collecting Data** We used the YT-DLP API<sup>5</sup> to collect data from YouTube and the TikTok Research Tools API<sup>6</sup> to collect data from TikTok. For YouTube, we first collected the top 10 search results for each of our queries. We then collected channel information for videos returned for our search results. Our data collection resulted in 1560 unique search results and 728 channels. Since the TikTok API only allowed for a maximum of 30 days when indicating time range, we focused our data collection on the most recent 3 months. In total, we collected 397 unique videos

<sup>3</sup>For instance, some diseases are referred to in their English name, transliterated to the Ge'ez script, and with their local language name. Hence, we prepare separate queries with the transliterated and the local language name.

<sup>4</sup>More details on query preparation and sample queries can be found in the Supplementary Material.

<sup>5</sup><https://github.com/yt-dlp/yt-dlp>

<sup>6</sup><https://developers.tiktok.com/doc/about-research-api>

from October 2024 to December 2024 from TikTok. Since the TikTok API required us to specify the location of the content creators, we collected data from the same 5 locations in (Nigatu and Raji 2024): United States of America, United Arab Emirates, United Kingdom, Saudi Arabia, and Ethiopia. Prior work indicates that these countries are best-suited for studying online activity of Ethiopia, Eritrea, and the diaspora (Nigatu and Raji 2024).

### 3.2 Analysis Methods

As stated above, we used a mixed methods approach relying on both automated and human analysis to answer our research questions. Below, we give details on the three methods we used to analyze our data.

**Topic Modeling** To understand what topics emerge from our collected search results, we trained Latent Dirichlet Allocation (LDA) (Blei 2003) models. We set the number of topics to correspond to the number of diseases in each of the disease categories to identify the topics that are salient in each of the categories (see Table 1). We trained models for the video titles and descriptions, and identified the salient terms in our data.

**Content Analysis** We conducted two qualitative content analyses on the data. First, to assess the credibility of the health information available in the two low-web data languages, we used the CRAP framework (Beestrum 2007). The CRAP framework is recommended by the WHO and the National Academy of Medicine to evaluate the credibility of health information online (Raynard S. Kington et al. 2021). The CRAP framework has four axis for evaluation: Currency, Reliability, Authority and Purpose. We used the guiding questions from (Raynard S. Kington et al. 2021) to perform our analysis<sup>7</sup>. For this analysis, we first subsampled our dataset for qualitative analysis. For YouTube, we selected the most viewed videos for each disease category (n=5) and each disease progression stage (n=4). With this, we had 20 videos per language for a total of 40 videos across the two target languages. For TikTok, we selected the most viewed video in each disease category (n=5), resulting in 10 videos across the two languages of study. This resulted in a total of 50 videos. We then used content analysis to understand the common characteristics of search results and to identify the cause for search results that diverge from their queries.

The first, second, and third authors of the paper conducted the qualitative coding for both analyses. The first author is a native Amharic speaker (L1) and has basic conversational skills in Tigrinya (L3). The second author is a native Tigrinya speaker (L1) and is a fluent Amharic speaker (L2). The third author is a native Amharic speaker (L1). The authors divided the videos among themselves and analyzed each video by opening and viewing the content. The authors first met to review an example video—which was randomly sampled outside of the 50 videos—and discussed the application of the CRAP framework to the sample video. This

<sup>7</sup>Guiding questions can be found in Supplementary Material.

allowed the authors to resolve ambiguities about the framework. The authors then independently applied the CRAP framework to their assigned videos and met to discuss their observations. Once the authors went through the first round of annotations, they met with the last author to iteratively discuss and refine the themes that emerged. We did not calculate inter-rater reliability as each video was annotated by a single person; instead, we relied on the iterative nature of our analysis to reduce disagreements. Finally, the first and second authors wrote up the final findings together, going over each theme identified and discussing general observations.

## 4 Results

In this section, we present the findings from our analysis. We organize this section according to our research questions: we first answer the common characteristics in content retrieved for health queries in low-web data languages. We then present our analysis of the quality of health information using the CRAP (Beestrum 2007) framework. Finally, we give insights into the characteristics of retrieved content that diverges from the search queries.

### **RQ1: What are the most common characteristics of search results retrieved for health-related queries in low web-resourced languages?**

We identified two high-level characteristics in search results for health queries in the two low-web data languages: (1) language mixing—where the returned results for some health queries are in languages other than the language of the search, and (2) nutritional advice and religious content returned for health-related queries. Across the two platforms, we found that TikTok results included more specific advice (e.g. videos referring to particular nutritional items as a solution) as compared to YouTube, where the videos were more general. Overall, we find that the search results may not address the health information needs of speakers of low-web data languages. Below, we go into the details of these findings.

**Language Mixing:** We find that searching for health-related queries in low-web data languages may surface search results in languages other than the language of the search. From our topic modeling analysis, we find that salient terms for Amharic and Tigrinya search results for YouTube include Arabic and English terms (Figure 1). Qualitatively, we find that Arabic videos in our data were about Reproductive Health and STIs: for instance, searching for symptoms of chlamydia in Amharic returned an Arabic YouTube video about the disease; searching for symptoms of measles in Tigrinya returned an Arabic video about gonorrhoea. For language mixing with English, we find that some videos that were in the language of search might have their titles in English in addition to the language of search: for instance, a video about symptoms and treatments for cancer in Tigrinya had the title in both Tigrinya and English. For Tigrinya search queries, we also observed that some of the videos returned are in Amharic: for instance, the search result for a Tigrinya query about meningitis included videos

about the disease in Amharic. As a result, a person searching in a low-web data language may not be able to access the health information they seek, as it is not available in their language.

**Salient Topics:** We find that videos in the two languages mostly included nutritional advice and natural remedies. This is evident in the salient terms from our topic modeling (Figure 1), where our analysis reveals that terms like ‘foods’, ‘ginger’, and ‘rosemary’ appear as salient terms across the two platforms. We also find that religious content appears in the list of salient terms, more dominantly for Amharic search results on TikTok (Figure 1). Qualitatively, we found that religious content is primarily returned for mental health-related queries, ranging from teachings by religious leaders to prayer videos targeted towards “relieving stress” without support for professional intervention or suggestions for available resources. For instance, the query “what are medications for anxiety?” returned a 1 hour and 34 minute prayer video on YouTube with 1.8 million views. This suggests that there might be a tendency among consumers of health information in low-web data languages to engage with religious content for mental health-related topics.

**Comparison of Search Results Across Platforms:** Search results from both platforms were similar in terms of including videos that were in languages other than the language of the search. We also observed that search results across both platforms included nutritional advice and religious content. However, we find that the salient terms in the results returned for TikTok tend to be more granular— as Figure 1, shows we observe more specific food and natural remedy terms like ‘ginger’, ‘rosemary’, and ‘lemon juice’ in both Tigrinya and Amharic when queries were conducted across the five disease categories whereas the salient terms on YouTube were more generic such as ‘food’ more broadly. We also observed spiritual and religious terms like ‘devil’, ‘evil spell’, ‘Islam’, and ‘God’ for communicable, non-communicable, and reproductive disease categories in Amharic on the TikTok search results. Further, as Figure 1<sup>8</sup> shows, TikTok results had more salient terms that were not in the language of search; this is especially more pronounced for Tigrinya, where the salient terms for reproductive health, women and children, and mental health queries included Amharic words.

### **RQ2: What is the quality of health information retrieved for low-web data languages?**

We identified characteristics in relation to the four axis of the CRAP framework in regards to the channels and accounts that post content returned for health related queries in the two low-web data languages: (1) Currency: while content creators post frequently, they do not always post health related content; (2) Reliability: content creators rarely cite sources for their claims; (3) Authority: content creators were

<sup>8</sup>We put a screenshot of the Table due to LaTeX restrictions on using Ge’ez characters.

Disease Categories	YouTube Salient Words		TikTok Salient Terms	
	Amharic	Tigrinya	Amharic	Tigrinya
Communicable	HIV, መከላከያ(prevention), መፍትሄዎች(cure), መንገዶች(ways), ምግቦች(foods)	HIV, ቪ ኤይ ኤድስ(AIDS), ኣፈዋውሱኦ(cure), ምክልኻሉ(prevention)	ጠረ(anti), የአፍ(mouth), መጥፎ(bad), ቅዱስ(holy), ዓይን(eye), ኢጋንንት(devil), የጥርስ(teeth), መንፈስ(spirit), ኩፍር(), ኢስላም (Islam)	ጅንጅብል(ginger), ጽማቕ(juice), ሮዝሞሪ(rosemary), መዓንጣ(gut), ምሕቃቕ()
Non-Communicable	ኢንፌክሽን(infection), የስኳር(diabetes), የኩላሊት(kidney), እጥረት(lack), ቫይታሚን(vitamin)	ልቢ ድኻም(heart attack), ኩላሊት(kidney), ኢንፌክሽን(infection), ቫይታሚን(vitamin)	መናፍስት(spirits), ልጅ(child), ስጋ(meat), መንፈስ(spirit), አምላክ(God), የሎሚ ጭማቂ(lemon juice)	tumor, ወሲብ (sexual intercourse), ኃላፊ (responsible), ስንፈተ(dysfunction), የማህፀን (womb)
Reproductive Health	የሽንት ቧንቧ(urinary tract), ኢንፌክሽን (infection), አባላዘር(sexually transmitted), transmitted sexually disease, ምልክቶች (symptoms)	ሽንት ቧንቧ(urinary tract), ኢንፌክሽን(infection)	ስሜት(sensation) የማቃጠል (burning), ኢንፌክሽን(infection), የሽንት(urinary), የብልት(reproductive organ), ደግምት(evil spell), ሴት(woman), ተቀመጦ (made from)	ስሜት(sunsation), የማቃጠል (burning) ኢንፌክሽን(infection), የሽንት(urinary), የብልት(reproductive organ), ቧንቧ(tube), የቁጥኝ (syphilis)
Women and Children	ምልክቶች(symptoms), የማህፀን ካንሰር (cervical cancer), የጡት ካንሰር (breast cancer)	ጥንሲ(fetus), ማህፀን ሕማም (cervical pain), ክላሚድያ(chlamydia), ፍንጣጣ(syphilis), ምግቦች(foods)	ሽጉርቲ(Onion), በሽታ (disease) ጳዕዳ(white), ጸባ(milk), ጤናማ(healthy), ማር(honey), ጅንጅብልን (ginger)	የማህፀን(womb), ጅንጅብል (ginger) ቅንፍር, ሽጉርቲ, ጳዕዳ(white), ጸባ(milk)
Mental Health	የአእምሮ(mental), disorder ጭንቀት(stress), መፍትሄ(solution), bipolar, anxiety	መንፈስ(spirit), ውሽጢ(internal)	እንቅልፍ(sleep), እፀ(drug), ጭንቀት(stress), ህመም(disease), ፍርሃት (fear), ንግግር (discussion)	እፀ(drug), ማስጠላት(to hate), ጭንቀት(stress), የመንፈስ (of the soul)

Figure 1: Topic modeling results demonstrating salient words for both platforms (YouTube and TikTok) across different disease categories in the two studied languages (i.e. Amharic and Tigrinya). Terms that are not in the search language are highlighted in yellow .

mainly individual medical doctors or religious leaders; (4) Purpose: some channels used their content to drive traffic to their business. Below, we go into detail of our findings for each of the four axis of the CRAP framework.

**Currency:** Currency is a measure of the recency of the information. On both platforms, we find that most of the accounts and channels post frequently. However, not all the posts in a given channel are health-related. For instance, one channel, which posted on average 5 times a month in the past 9 months, posted vlogs about their personal life in addition to medical content. Across both languages, we find that most health-related videos returned (n=15) were from 2022, with one video from 2024. The oldest video returned for YouTube was from 2018, which was a video on handling stress by a religious leader. For TikTok, most of the videos (n=4) were from December 2024.

**Reliability** tells us to what degree the information is based on fact versus opinion. Overall, while the videos in our analysis were fact-based—most were educational videos that list signs and symptoms or give recommendations for natural remedies—a majority of the videos did not cite their sources. We find that the videos varied widely in terms of the topics they covered, including educational content about medi-

cal conditions, personal stories about experience with health conditions, news reportings, and religious content. However, most of the videos do not cite their sources—only two YouTube videos cite their sources while none of the TikTok videos cite their sources. Some videos use images from other sources or add visual demonstrations. Most videos also gave nutritional advice—i.e. what food to eat for a particular disease. For some videos, it was hard to verify the reliability as they were (1) in another language, (2) did not contain medical information; for instance, a query about pregnancy returned a comedy skit about pregnancy cravings, or (3) was religious content; we find that all but three queries related to mental health returned religious content as the top-viewed video<sup>9</sup>.

**Authority** gives information on who the creators are and what credentials they have. Most of the channels and accounts in our analysis are run by individuals whose credentials were hard to verify. Although many of the content creators in our analysis were individual medical professionals, we find that none of the content creators were verified by the platforms as medical content creators. Additionally,

<sup>9</sup>See Supplementary Material for channel and account examples.

channels that disclose their locations all indicate being located outside of Ethiopia or Eritrea. We find that the channels that disclose their location are located in South Africa (n=1), the US (n=3) and Germany (n=1) for Tigrinya, while for Amharic, we find all channels that disclosed their location set their location to the US (n=9). We find that some creators share their office location (Google Maps link) and phone number. Interestingly, even when there is evidence that the content creator is located in Ethiopia, the channel is registered in the US. Creators also link their profiles to other platforms like Twitter/X, TikTok, and Facebook. We were able to verify 5 medical doctors by following the links they provided to other platforms; for instance, one creator had a “ECFMG Certified Medical Grad” tag on their Twitter account. Some creators post their email and Facebook accounts but do not provide any verification. For TikTok, it was difficult to verify the creators as there is no information in their descriptions, and most of the videos were voice-overs. Two channels had their first name, and one channel had a phone number to contact for asking for services. One creator had a Dr. title before their name, although there were no avenues to verify their credentials.

**Purpose** provides insights into whether the information available is pushing a particular agenda. While most channels were not biased towards a particular agenda, we noted a few videos where the content creators used the video to drive traffic to their business. In general, most of the videos were about nutrition and foods to eat for a particular disease. However, in some cases, the videos included ads or contact information for viewers who were interested in contacting the content creators. One YouTube channel and one TikTok account had an ad for a dietary supplement from a company owned by the content creators. In other videos, we observed that doctors leave their contact information for people who want personal consultations. We also found one account that used its TikTok content to derive business for their traditional medical clinic. Some content creators also add disclaimers in their videos or descriptions that people should seek medical advice when necessary. The 5 medical doctors we could verify and one of the religious leaders all had a disclaimer that the content was not a substitute for medical diagnosis and that people should talk to their doctors for such services. However, fake accounts created using the identities of medical professionals sometimes post scam content. In one video, one of the content creators we verified as a doctor shared that other people have been using their image and channel name with fake accounts to exploit people, falsely claiming they will send them medication in return for monetary payment.

**RQ3: When search queries and results do not align, what are the common characteristics of those results?**

So far, we found that search results retrieved do not always match their queries, which creates a space for harm to health information seekers—users cannot get the information they are seeking about certain medical conditions. Instead, the content retrieved may be (1) in another language, (2) about

Source	Description	Example
Borrowed Medical Terms	Some diseases that do not have terms in the local language, and instead use transliterated English words, result in unrelated content being returned.	Searching for ‘HPV’ and ‘Meningitis’ returns results that are about other diseases.
Similar Medical Terms	Some medical terms have similar writings in the local languages	Anemia translates to “Blood Decline” in Amharic which is similar to “Blood Pressure.” As a result, the videos returned are about “Blood Pressure.”
Sensitive Medical Conditions	Diseases that are culturally considered to be taboo topics, especially related to reproductive health.	Searching for information on STDs results in videos in Arabic or English
Critical Medical Conditions	Searching for information about chronic medical conditions returns videos on nutritional cures.	Searching for information on “cancer.”
Non-Medical Terms	Popular search terms that have nothing to do with the medical content and instead are popular at the time.	Political terms used in hashtags

Table 2: Five categories of terminology that lead to search results that diverge from the queries.

an unrelated medical condition, or (3) non-medical content. Further, we find that there are different degrees of human manipulation—i.e, to what extent there is an agent manipulating the health information gap for disingenuous purposes—in causing the search results to diverge from their queries. In this section, we explore in-depth the divergence of search results from their search terms.

We identified five cases where the search results diverge from their queries. Below, we describe these five cases listed in order of degree of human manipulation for dubious purposes. Starting with the cases where there is no human manipulation, we first present borrowed medical terms and similar medical terms, where the divergence of retrieved results from the queries can be attributed to the *linguistic* limits of the search engine in handling health queries that are in low-web data languages. We then present the cases where the manipulation is still largely in the failure of the retrieval algorithm, but may be shaped by socio-cultural facets. Here, we have sensitive medical terms where the divergence may be due to the lack of content about tabooed topics in the *cultural* context of low-web data languages. We then present the case of critical medical conditions where there may be (un)intentional interference with the search experience for critical medical cases that may require authoritative infor-

mation. Lastly, we present the case where there is active human manipulation—using non-medical unrelated terms as hashtags, content creators manipulate the retrieval algorithm to boost content that is not related to the query.

**Borrowed Medical Terms:** Some diseases and health conditions are referred to with their English name or with their scientific name. In some cases, this is because the medical condition does not have a name in the target language—for instance, AIDS is referred to as AIDS in both languages (although, it is written with the Ge’ez characters). In other cases, the medical condition might have a name in the target language but the name might not be widely used or might be used interchangeably with the English or scientific name—e.g. “cancer” has a local name in both languages but may also be referred to with its English name. In our analysis, we found that some of the search results returned for health conditions that are referred to by borrowed medical terms were unrelated to the queries. For instance, search queries about “HPV” returned videos (1) about computer problems—some related to computer viruses, (2) about different medical problems—like acid reflux, (3) sexual advice that violates the YouTube platform policy.

**Similar Medical Terms:** We also found that distinct medical conditions that have similar words are confused in the search results: for instance, for Amharic, the medical term for Anemia is “dem manes” which literally translates to “Blood Decline.” However, the search results for queries in Amharic included videos about “Blood Pressure” which in Amharic is “dem bezat.” Thus, while the videos were medical in nature, they were about an unrelated medical condition than what was searched.

**Sensitive Medical Conditions:** Search results for medical conditions that were sensitive or were about culturally tabooed topics also diverged from their queries. Particularly, searching for STDs resulted in videos in languages other than the language of search. For instance, searching for “What is the cure for syphilis?” in Amharic repeatedly returned videos about Syphilis in Arabic and English. On the other hand, searching for “How can I prevent HIV?” in Tigrinya consistently returned videos in Amharic. Language mixing in search results for sensitive medical conditions is also evident in Figure 1 where the reproductive health category has more salient terms that were different from the language of search.

**Critical Medical Conditions:** In cases where the health condition is critical—e.g. cancer or diabetes, we find that the search results were more about natural cures. Usually, the video titles promise a quick and lasting fix: for instance, a query about breast cancer cure returned Amharic videos about “fruit that kills cancer after just a few minutes” and videos about “anti-cancer foods” that will “eliminate cancer if you eat these 8 foods frequently.”

**Non Medical Terms:** We also found that content creators use non-medical, popular terms as hashtags to boost their content. For instance, a video about symptoms and cures for evil eye was returned for a query about meningitis; the video

used hashtags of popular TV shows and names of political figures in its video description, which may have boosted its ranking in the search engine.

## 5 Proposing a Data Horizon

Recall that a data void is a gap in digital information where search queries do not have high-quality nor high-quantity information, resulting in a *void* that can be manipulated by bad actors. In this paper, we investigate a data horizon—the search experience right before one goes into a data void. As briefly described in Section 1, a data horizon is the boundary where the information landscape begins to fall apart due to systemic constraints on information access. The data horizon in a health information gap for low-web data languages has five axes of resource constraints. Below, we describe each of the five axes and connect them to our findings from the previous section to demonstrate how they shape the experiences of low-web data speakers seeking health information online.

**Lack of Digital Resources in Target Language:** Due to the low-quality web data, the search results for health queries in the two low-web data languages include information in other languages or unrelated information in the target language (Section 4.1). Limits in digital resources available in the language are further reflected in search results for queries with borrowed and similar medical terms, where results returned diverge from the queries (Table 2).

**Critical Domain Setting:** In a critical domain such as healthcare, lack of verified health information may lead to life-threatening consequences (Section 2). As presented in Section 4, we find that the information returned for health queries may not always be medical content or may be about a health condition unrelated to what was in the query. Further, the information returned may not be in the language of search, creating a language disparity in access to health information.

**Lack of Verification Mechanisms:** As we present in Section , none of the YouTube channels and TikTok accounts we analyzed were verified by the platforms as authorized medical information providers. Independently verifying their credentials required following links to other platforms. We verified 5 medical doctors on YouTube and could not verify the credibility of any of the creators on TikTok.

**Out-Of-Context Platform Use:** As we have discussed in Section 1, while social media platforms are primarily designed to foster social communication, they are increasingly used as search platforms. As a result, features that are optimized for social engagement may interfere with the quality of search results returned; for instance, we find that hashtags with unrelated, non-medical terms result in videos that are not related to the query being returned as search results (Section 4).

**Shift in Cultural Context:** In Section 4, our findings show that salient terms for health queries in low-web data languages include nutritional advice as well as religious

and spiritual terms. Further, from our analysis of the top-viewed videos, we find that videos from religious leaders are among the top-viewed videos for mental health-related queries. Additionally, we find that searching for health information about taboo topics—such as STDs and reproductive health—may result in search results in other languages. These could be the results of the cultural context in which the two low-web data languages are spoken.

The interaction of the five phenomena described above creates a data horizon. The data horizon is the stage right before we go into a data void, where the active manipulation has not begun. As Section 4 shows, the data horizon is mainly manipulated through (1) algorithmic failures in handling borrowed medical terms, similar medical terms, and queries about sensitive medical conditions, (2) (un)intentionally by unverified content creators who post natural cures for chronic and critical medical conditions, and (3) active manipulation by content creators who use popular but non-medical terms as hashtags to boost their content's ranking in retrieval.

In Table 3, we show how the five axes of a data horizon interact with the five sources of search results that diverge from their queries. As discussed in Section 4, the sources of search results that diverge from their queries as listed in order of degree of human manipulation. We find that all five sources of search results that diverge from their queries are affected by the lack of digital resources, the out-of-context use of social media platforms for health information seeking, and the shift in cultural context. For instance, in all five cases, the lack of digital resources in low-web data languages means the algorithm has no documents related to the query. As a result, the retrieval system returns what is “close enough”—be it in another language or about unrelated content. Similarly, the out-of-context use of the social media platforms means that what is returned is not necessarily ranked based on the quality and credibility it has to the query. In the case of non-medical terms, the out-of-context use becomes a breeding ground for manipulation—affordances such as hashtags that were designed for social connection are instead used to manipulate retrieval ranking. Further, the shift in cultural context requires algorithms that handle linguistic features such as borrowed medical terminology and similar medical terms. Cultural context is also important in determining what types of non-medical terms might be used in a given language.

Conversely, the critical domain affects all but non-medical terms in the five sources we identified as non-medical terms can be used to manipulate the algorithm in any domain. The lack of verification, however, only affects non-medical terms; even if the content returned for borrowed medical terms or sensitive medical conditions is coming from verified sources, it will still fail in providing the information the user is seeking. Take the case of searching for content about anemia. As we have demonstrated above, the search results for such queries may be about blood pressure due to the similarity of the medical terms in the Amharic language. Even if the content about blood pressure is coming from a verified medical professional, it fails to provide the information that was requested in the query. Conversely, for non-medical ter-

minology, having verification for content creators will allow people to distinguish between what content to trust, regardless of how it is ranked in the search results.

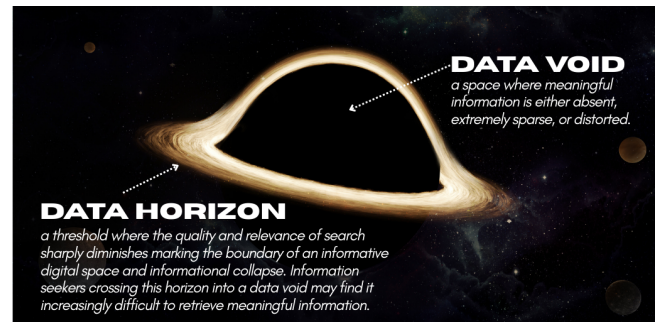


Figure 2: Conceptual illustration of a data void and its data horizon representing the boundary beyond which relevant, credible, or reliable information becomes unavailable

## 6 Discussion

As the use of social media platforms like YouTube and TikTok evolved from their intended design of being platforms for social connection to serving as substitutes for search engines, users are at higher risk of exposure to harm (Sylvia Chou, Gaysynsky, and Cappella 2020). One such harm is due to data voids—information vacuums that are ripe for exploitation by malicious actors (Golebiewski and boyd 2019).

We introduce data horizons, which are areas in the information landscape where the fabric of information is stretched by systemic constraints on access to information, causing it to wrap and act like an invisible boundary around a data void. In Figure 2, we show this phenomenon. At the data horizon, the quality of information is spread thin due to sociotechnical constraints (Section 5). As a result, individuals searching for information while at the horizon may not be able to access the information they seek. However, they are also not always targets of active manipulation—instead, their experiences are largely shaped by the degrading algorithmic performance and the shift in socio-cultural context. To understand the characteristics of data horizons and identify how they are formed, we focused on health queries in two low-web data languages, Amharic and Tigrinya. We set out to (1) characterize search results, (2) understand the quality of health information, and (3) investigate causes for search results that diverge from their queries.

We find that the health information that is available online in the two languages of study leans towards providing nutritional and natural remedies, includes religious content, especially for mental health queries (Section 4.1). Specifically for mental health queries, we find that the most viewed videos are religious content. This indicates a tendency by consumers of health information to seek information that may not be classified as authoritative medical content. However, as prior work shows, the lack of authoritative content about topics that are not mainstream may lead to unverified and at times harmful content reaching consumers (Abebe

Source	Digital Re-source	Out-of-Context Use	Culture Shift	Critical Domain	Lack of Verification
Borrowed Medical Terms	✓	✓	✓	✓	
Similar Medical Terms	✓	✓	✓	✓	
Sensitive Medical Conditions	✓	✓	✓	✓	
Critical Medical Conditions	✓	✓	✓	✓	
Non Medical Terms	✓	✓	✓		✓

Table 3: Table showing which axes of a data horizon influence the sources of search results that diverge from their queries.

et al. 2019). We want to emphasize that we are not making claims that religious content is unequivocally invalid for mental health support; many religious societies rely on their spiritual practices for mental health support (Oewel, Aream, and Agapie 2024). Rather, we argue that knowing what type of content is most effective in reaching this particular community would allow us to target tailored interventions. One of the religious content creators states in their video that they are “not opposing seeking medical help.” Hence, there is an opportunity to, for instance, collaborate with religious leaders to create awareness about mental health issues.

Golebiewski and boyd (2019) state, for queries that are associated with authoritative content, such as medical queries, search engines are more likely to return authoritative medical content in their top results. However, medical resources are not equally available in all languages (Hu et al. 2025). Additionally, it may be more difficult for accredited health providers from certain communities to be labeled “authoritative” (Section 2). Our results also show that it is difficult to verify the credentials of content creators who post health information in low-web data languages (Section 4). It was particularly challenging to verify the credentials for TikTok content creators. Further, while content creators who share health information on online platforms in the two languages of study post frequently, they may not always post health-related content (Section 4). This might indicate a relatable angle for social media health information, as has been shown by prior work (Milton et al. 2023). However, as we have described in Section 2, relatability may result in mistrust by users who consume health information on social media platforms. This factor might be confounded by how content creators rarely cite their sources (Section 4). Hence, low-web data language speakers may have limited access to trusted, verified health information on social media platforms.

As for social media platforms, there are a few avenues to protect users from falling into data voids in low-web data language queries. First, while it might be difficult to verify medical content creators from countries that do not have the proper mechanisms set up, platforms can require content creators to provide their own verifications (e.g. what hospital they work at, certifications they might have e.t.c) on their channel descriptions. While this does not serve as a replacement for the proper accreditation for medical professionals verified by the platforms, it will make it easier for users to make their own verifications. Platforms could also consider adding links to websites of recognized organizations

like the WHO on videos that are posting medical content, so users have options to search for further information. Platforms could also encourage content creators to add citations and provide additional sources when posting health-related videos, regardless of them being verified by the platform as a medical professional. However, we caution that this should not take the form of punitive measures against existing content creators who are providing medical information with the limited resources available to them.

By definition, the lack of data in low-web data languages creates a void. This void can lead users to harm, particularly in use cases like healthcare. International and local organizations can partner with content creators to fill the void. Such organizations can (1) help speed up the verification of content creators by providing the necessary certifications for medical professionals from these communities, and (2) help content creators produce more content in these languages with the right resources for citations. The organizations can also create their own content in various languages.

Data voids are difficult to detect in that they require external triggers to surface—by the time a data void becomes a problem, the harm has already been done (Golebiewski and boyd 2019). With the concept of data horizons, however, we can identify data voids *before* they are fully exploited. To uncover a data horizon, we need only look at the *systemic constraints* to information access. As we have shown in Section 5, we can rely on the axis of a data horizon to understand how it manifests. Hence, we can strategize about ways to protect users.

## Limitations and Conclusion

Our paper proposes data horizons—boundaries in the information landscape created due to the tension between various systemic constraints to information access. We used the context of health information seeking in low-web data languages to characterize data horizons. We show that by studying socio-technical constraints that lead to degradation of information quality, we can identify data horizons before they become full-fledged data voids. Our study is limited in that we study only two languages. This was due to the human resource limits on how many languages we could cover. Future work could explore extending our study to more languages and mapping the data horizon to a different critical domain (e.g. legal domain). We hope that our work informs stakeholders on potential avenues for identifying data horizons and protecting users before they fall into data voids.

## References

- Abebe, R.; Hill, S.; Vaughan, J. W.; Small, P. M.; and Schwartz, H. A. 2019. Using Search Queries to Understand Health Information Needs in Africa. *Proceedings of the International AAAI Conference on Web and Social Media*, 13: 3–14.
- Al Shamsi, H.; Almutairi, A. G.; Al Mashrafi, S.; and Al Kalbani, T. 2020. Implications of Language Barriers for Healthcare: A Systematic Review. *Oman Medical Journal*, 35(2): e122.
- Alkhodair, R.; Alfawzan, A.; Alharthi, S. A.; AlOmar, A.; Alqerafi, A.; and Alfaleh, A. 2024. Assessing the Quality and Accuracy of Syphilis-Related Content on TikTok and YouTube: A Comprehensive Analysis. *Sexually Transmitted Diseases*.
- Basha, S. J.; Veeraiah, D.; Charan, B. V.; Joyce Yeddu, W. S.; and Babu, D. G. 2023. Detection and Comparative Analysis of Handwritten Words of Amharic Language to English using CNN-Based Frameworks. In *2023 International Conference on Inventive Computation Technologies (ICICT)*, 422–427. ISSN: 2767-7788.
- Bautista, J. R.; Zhang, Y.; and Gwizdka, J. 2021. Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics*, 148: 104375.
- Beestrup, M. 2007. The CRAP Test.
- Blei, D. M. 2003. Latent Dirichlet Allocation.
- Brooks, L. A.; Manias, E.; and Bloomer, M. J. 2019. Culturally sensitive communication in healthcare: A concept analysis. *Collegian*, 26(3): 383–391.
- Chirumamilla, S.; and Gulati, M. 2021. Patient Education and Engagement through Social Media. *Current Cardiology Reviews*, 17(2): 137–143.
- Clarke, S.; Jangid, G.; Nasr, S. E.; Atchade, A.; Moody, B. R.; and Narayan, G. 2023. Polycystic Ovary Syndrome (PCOS): A Cross-Sectional Observational Study Analyzing the Quality of Content on YouTube. *Cureus*.
- Dahm, M. R. 2012. Tales of time, terms, and patient information-seeking behavior-an exploratory qualitative study. *Health Communication*, 27(7): 682–689.
- Freeman, J. L.; Caldwell, P. H. Y.; and Scott, K. M. 2023. How Adolescents Trust Health Information on Social Media: A Systematic Review. *Academic Pediatrics*, 23(4): 703–719.
- Freimuth, V. S.; and Quinn, S. C. 2004. The Contributions of Health Communication to Eliminating Health Disparities. *American Journal of Public Health*, 94(12): 2053–2055.
- Ghenai, A.; and Mejova, Y. 2018. Fake Cures: User-centric Modeling of Health Misinformation in Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 1–20. Publisher: Association for Computing Machinery (ACM).
- Golebiewski, M.; and boyd, d. 2019. Data Voids: Where Missing Data Can Easily Be Exploited.
- Haile, N.; Ali, N.; and Hadgu, A. T. 2023. ERROR ANALYSIS OF TIGRINYA – ENGLISH MACHINE TRANSLATION SYSTEMS.
- Hu, S.; Oppong, A.; Mogo, E.; Collins, C.; Occhini, G.; Barford, A.; and Korhonen, A. 2025. Natural Language Processing Technologies for Public Health in Africa: Scoping Review. *Journal of Medical Internet Research*, 27: e68720.
- Jafar, Z.; Quick, J. D.; Larson, H. J.; Venegas-Vera, V.; Napoli, P.; Musuka, G.; Dzinamarira, T.; Meena, K. S.; Kanmani, T. R.; and Rimányi, E. 2023. Social media for public health: Reaping the benefits, mitigating the harms. *Health Promotion Perspectives*, 13(2): 105–112.
- Jeyaraman, M.; Ramasubramanian, S.; Kumar, S.; Jeyaraman, N.; Selvaraj, P.; Nallakumarasamy, A.; Bondili, S. K.; and Yadav, S. 2023. Multifaceted Role of Social Media in Healthcare: Opportunities, Challenges, and the Need for Quality Control. *Cureus*, 15(5): e39111.
- Karusala, N.; and Anderson, R. 2022. Towards Conviviality in Navigating Health Information on Social Media. In *CHI Conference on Human Factors in Computing Systems*, 1–14. New Orleans LA USA: ACM.
- Lee, J. L.; Rawl, S. M.; Dickinson, S.; Teal, E.; Baker, L. B.; Lyu, C.; Tarver, W. L.; and Haggstrom, D. A. 2020. Communication About Health Information Technology Use Between Patients and Providers. *Journal of General Internal Medicine*, 35(9): 2614–2620.
- Milton, A.; Ajmani, L.; DeVito, M. A.; and Chancellor, S. 2023. “I See Me Here”: Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–17. Hamburg Germany: ACM.
- Milton, A.; Maestre, J. F.; Roy, A.; Umbach, R.; and Chancellor, S. 2024. Seeking in Cycles: How Users Leverage Personal Information Ecosystems to Find Mental Health Information. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–16. Honolulu HI USA: ACM.
- Munoz, I. B.; Liu-Zarzuela, J.; Oorjitham, N.; and Jacob, D. 2024. YouTube and TikTok as a source of medical information on dissociative identity disorder. *Journal of Affective Disorders Reports*, 15: 100707.
- Nangsangna, R. D.; and Vroom, F. d.-C. 2019. Factors influencing online health information seeking behavior among patients in Kwahu West Municipal, Nkawkaw, Ghana. *Online Journal of Public Health Informatics*, 11(2): e62588. Company: Online Journal of Public Health Informatics Distributor: Online Journal of Public Health Informatics Institution: Online Journal of Public Health Informatics Label: Online Journal of Public Health Informatics Publisher: JMIR Publications Inc., Toronto, Canada.
- Naslund, J. A.; Grande, S. W.; Aschbrenner, K. A.; and Elwyn, G. 2014. Naturally Occurring Peer Support through Social Media: The Experiences of Individuals with Severe Mental Illness Using YouTube. *PLOS ONE*, 9(10): e110171. Publisher: Public Library of Science.
- Nigatu, H. H.; and Raji, I. D. 2024. “I Searched for a Religious Song in Amharic and Got Sexual Content Instead”:

- Investigating Online Harm in Low-Resourced Languages on YouTube. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 141–160. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.
- Nigatu, H. H.; Tonja, A. L.; Rosman, B.; Solorio, T.; and Choudhury, M. 2024. The Zeno's Paradox of 'Low-Resource' Languages. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 17753–17774. Miami, Florida, USA: Association for Computational Linguistics.
- Njenga, G. G.; Mogambi, H.; and Oriaso, S. 2024. Issues Associated with Media and Health Information Seeking Behavior Among the Elderly in Kenya. *Kabarak Journal of Research & Innovation*, 14(02): 180–190. Number: 02.
- Oewel, B.; Arean, P. A.; and Agapie, E. 2024. Approaches for tailoring between-session mental health therapy activities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–19. Honolulu HI USA: ACM. ISBN 9798400703300.
- Oleribe, O. O.; Momoh, J.; Uzochukwu, B. S.; Mbofana, F.; Adebisi, A.; Barbera, T.; Williams, R.; and Taylor-Robinson, S. D. 2019. Identifying Key Challenges Facing Healthcare Systems In Africa And Potential Solutions. *International Journal of General Medicine*, 12: 395–403.
- Ozdemir, O.; Diren, F.; Boyali, O.; Civelek, E.; and Kabatas, S. 2021. Metric Evaluation of Reliability and Transparency of the Videos About Carpal Tunnel Syndrome Surgery in the Online Platforms: Assessment of YouTube Videos' Content. *Neurospine*, 18(2): 363–368.
- Quintero Johnson, J. M.; Yilmaz, G.; and Najarian, K. 2017. Optimizing the Presentation of Mental Health Information in Social Media: The Effects of Health Testimonials and Platform on Source Perceptions, Message Processing, and Health Outcomes. *Health Communication*, 32(9): 1121–1132.
- Raynard S. Kington; Stacey Arnesen; Wen-Ying Sylvia Chou; Susan J. Curry; David Lazer; and Antonia M. Villaruel. 2021. Identifying Credible Sources of Health Information in Social Media: Principles and Attributes - NAM.
- Roberts, B.; Kobritz, M.; Nofi, C.; Demyan, L.; Guevara, J.; Hansen, L.; and Giangola, M. 2023. Social Media, Misinformation, and Online Patient Education in Emergency General Surgical Procedures. *Journal of Surgical Research*, 287: 16–23.
- Sylvia Chou, W.-Y.; Gaysynsky, A.; and Cappella, J. N. 2020. Where We Go From Here: Health Misinformation on Social Media. *American Journal of Public Health*, 110(Suppl 3): S273–S275.
- Talie Fenta, E.; Bogale, E. K.; and Anagaw, T. F. 2024. The role of social media on COVID-19 preventive behaviors worldwide, systematic review. *PLOS ONE*, 19(7): e0306284.
- Tang, H.; Lenzini, G.; Greiff, S.; Rohles, B.; and Sergeeva, A. 2024. "Who Knows? Maybe it Really Works": Analysing Users' Perceptions of Health Misinformation on Social Media. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, 1499–1517. New York, NY, USA: Association for Computing Machinery. ISBN 9798400705830.
- Trübner, M.; Patzina, A.; Lehmann, J.; Brinkhaus, B.; Kessler, C. S.; and Hoffmann, R. 2025. Health information-seeking behavior among users of traditional, complementary and integrative medicine (TCIM). *BMC Complementary Medicine and Therapies*, 25: 111.
- Ukonu, M. O.; and Ajaebili, N. C. 2021. Socio-Cultural Determinants of Women's Health Information Opportunities in Nsukka, Southeast Nigeria. *Asian Women*.
- Usman, J. E.; Childs, C.; Rogerson, D.; and Klonizakis, M. 2024. Online Health Information-Seeking Behaviour among People of African Descent in the United Kingdom: A Qualitative Study. *Healthcare*, 12(9): 897.
- Viviani, M.; and Pasi, G. 2017. Credibility in social media: opinions, news, and health information—a survey. *WIREs Data Mining and Knowledge Discovery*, 7(5): e1209. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1209>.
- YouTube. 2021. Apply to be a source in YouTube health features - YouTube Help.
- Zhang, A. Q.; Milton, A.; and Chancellor, S. 2023. #Pragmatic or #Clinical: Analyzing TikTok Mental Health Videos. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, 149–153. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701290.