Original Paper

Capturing Emerging Experiential Knowledge for Vaccination Guidelines Through Natural Language Processing: Proof-of-Concept Study

Lea Lösch[1], MSc; Teun Zuiderent-Jerak[1], PhD; Florian Kunneman[2], PhD; Elena Syurina[1], PhD; Marloes Bongers[3], MD, PhD; Mart L Stein[3], PhD; Michelle Chan[2], MA, MSc; Willemine Willems[1], PhD; Aura Timen[1,3,4], Prof Dr, MD, PhD

[1]Athena Institute, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Department of Computer Science, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[3]Centre for Infectious Disease Control (CIb), National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands

[4]Department of Primary and Community Care, Radboud University Medical Centre, Nijmegen, Netherlands

Corresponding Author:
Lea Lösch, MSc
Athena Institute, Faculty of Science
Vrije Universiteit Amsterdam
De Boelelaan 1105
Amsterdam, 1081 HV
Netherlands
Phone: 31 205987031
Email: lea.loesch@vu.nl

Abstract

Background: Experience-based knowledge and value considerations of health professionals, citizens, and patients are essential to formulate public health and clinical guidelines that are relevant and applicable to medical practice. Conventional methods for incorporating such knowledge into guideline development often involve a limited number of representatives and are considered to be time-consuming. Including experiential knowledge can be crucial during rapid guidance production in response to a pandemic but it is difficult to accomplish.

Objective: This proof-of-concept study explored the potential of artificial intelligence (AI)–based methods to capture experiential knowledge and value considerations from existing data channels to make these insights available for public health guideline development.

Methods: We developed and examined AI-based methods in relation to the COVID-19 vaccination guideline development in the Netherlands. We analyzed Dutch messages shared between December 2020 and June 2021 on social media and on 2 databases from the Dutch National Institute for Public Health and the Environment (RIVM), where experiences and questions regarding COVID-19 vaccination are reported. First, natural language processing (NLP) filtering techniques and an initial supervised machine learning model were developed to identify this type of knowledge in a large data set. Subsequently, structural topic modeling was performed to discern thematic patterns related to experiences with COVID-19 vaccination.

Results: NLP methods proved to be able to identify and analyze experience-based knowledge and value considerations in large data sets. They provide insights into a variety of experiential knowledge that is difficult to obtain otherwise for rapid guideline development. Some topics addressed by citizens, patients, and professionals can serve as direct feedback to recommendations in the guideline. For example, a topic pointed out that although *travel* was not considered as a reason warranting prioritization for vaccination in the national vaccination campaign, there was a considerable need for vaccines for indispensable travel, such as cross-border informal caregiving, work or study, or accessing specialized care abroad. Another example is the ambiguity regarding the definition of medical risk groups prioritized for vaccination, with many citizens not meeting the formal priority criteria while being equally at risk. Such experiential knowledge may help the early identification of problems with the guideline's application and point to frequently occurring exceptions that might initiate a revision of the guideline text.

Conclusions: This proof-of-concept study presents NLP methods as viable tools to access and use experience-based knowledge and value considerations, possibly contributing to robust, equitable, and applicable guidelines. They offer a way for guideline

developers to gain insights into health professionals, citizens, and patients' experience-based knowledge, especially when conventional methods are difficult to implement. AI-based methods can thus broaden the evidence and knowledge base available for rapid guideline development and may therefore be considered as an important addition to the toolbox of pandemic preparedness.

(J Med Internet Res 2023;25:e44461) doi: 10.2196/44461

KEYWORDS

guidelines as topic; COVID-19; public health; natural language processing; NLP; social media; stakeholder engagement; vaccine; vaccination; health policy; coronavirus; SARS-CoV-2

Introduction

Expert opinions and patient experiences have been deemed an essential part of evidence-based medicine right from the outset [1], as these have proved to contribute to high-quality and more applicable public health and clinical practice guidelines [2,3]. Guidelines are developed to systematically synthesize the best available evidence on a given condition, disease, or procedure, to provide recommendations that support health professionals in (clinical) decision-making. Guideline recommendations that consider experiential knowledge and patient preferences more closely reflect the needs and experiences of patients and health care professionals and thereby improve guideline adherence and patient care [4].

However, methods to achieve this, such as surveys, focus groups, and commentary on guideline drafts, vary widely, and incorporating this type of knowledge on a regular basis is not yet common practice [5-7]. Thus, there is substantial underrepresentation of experience-based knowledge and value considerations in most guidelines for a wide range of clinical topics [8].

Integrating this type of knowledge into guideline development is even more important, yet more difficult to achieve, when developing public health guidelines to respond to an ongoing pandemic. Given the urgency and high time pressure to produce the best guidance available, most guideline developers stated that prevailing methods—including those for involving end users—were largely unsuited for developing guidance during the COVID-19 pandemic [9]. In the absence of evidence from randomized clinical trials, experiential knowledge becomes one of the few sources of rapidly evolving knowledge available in the early stages of a health crisis [10,11]. The methodological challenge of its inclusion in outbreak guidance is thus an acute problem for response strategies.

The limited ability to incorporate this type of knowledge is not owing to its absence in public debate. The COVID-19 pandemic has been characterized by the extensive exchange of concerns, experiences, and value deliberations among individuals and groups, fueled by social media and the high turnover of news reports. Experiential knowledge about the subject is thus available, albeit scattered throughout media platforms and obfuscated by echo chamber characteristics of other posts on social media [12,13]. Its sheer volume and unstructured nature make it nearly impossible for guideline developers, without specific tools and methodologies, to use the experiential knowledge and value considerations of patients and professionals: out of 188 guidelines related to COVID-19 analyzed by Stamm et al [14], only 1 had involved patient knowledge.

Computational methods may offer innovative opportunities in guidance production to analyze and use existing data sources that contain experiential knowledge and value consideration systematically and on a large scale, not only after but also alongside the process of guideline development and appraisal of new information. Over the past 2 decades, studies under the heading of infodemiology have explored the use of various artificial intelligence (AI)–based methods to gain insights into disease patterns and health dynamics from digital data to inform public health and public policy [15,16]. These methods have also been further developed and used in a variety of ways across the globe for public health responses during the COVID-19 pandemic (for comprehensive reviews, refer to the papers by Syrowatka et al [17], Tsao et al [18], Chen et al [19], and Gunasekeran et al [20]). For instance, studies have aimed to provide timely and effective support to health authorities with respect to surveillance [21,22], dissemination of health information [23,24], disease detection and prediction [25,26], and monitoring of public opinion and sentiment [27,28]. Although attitudes and opinions, especially positive and negative sentiments toward COVID-19 vaccination and policies, have been analyzed extensively, exploring experiential knowledge using computational methods has received less attention (for notable exceptions, refer to the papers by Chiavi et al [29] and Bacsu et al [30]). Despite the wealth of studies exploring the applications of AI-based methods in various public health and health care contexts, automated approaches have only been explored to a limited extent in the field of guideline development, for example, to support and accelerate literature screening [31,32] but have not yet been leveraged to harness experiences as evidence for clinical or public health guidelines.

The objective of this proof-of-concept study was to develop AI methods from the field of natural language processing (NLP), for harvesting experience-based knowledge and value considerations to make guidelines more inclusive and representative and, ultimately, improve their performance in the field. Thus, our fundamental question was the following: How can AI-based methods be used to identify and analyze the experiential knowledge of health care professionals, patients, and citizens that is being shared on the web, to contribute to the development of public health guidelines? We examined these methods in the case of the development of the COVID-19 vaccination guideline in the Netherlands, which was first published in December 2020 and has since been updated in >50 guideline development and 20 user feedback meetings (as of November 2022). The guideline supports health professionals

in the implementation and administration of COVID-19 vaccination [33]. It contains, for example, information about contraindications, available vaccines, and organizational and practical aspects.

Methods

Data Sources

We focused on 2 types of existing textual data sources: social media platforms and internal databases from the Dutch National Institute for Public Health and the Environment (RIVM). Social media were considered to be valuable for sourcing a wide range of experiences and values, as these tend to be expressed through the comment-oriented nature of posts. Facebook comments to Dutch news articles about COVID-19 vaccination were selected as a promising open data source because comments were extensive, diverse, and accessible owing to the public status of these Facebook pages through the Facebook API. In total, we collected 230,863 Facebook comments about news articles posted by the 4 popular Dutch news outlets, *Nederlandse Omroep Stichting*, *NU.nl*, *Telegraaf*, and *Nieuwe Rotterdamsche Courant*, between December 1, 2020, and June 8, 2021.

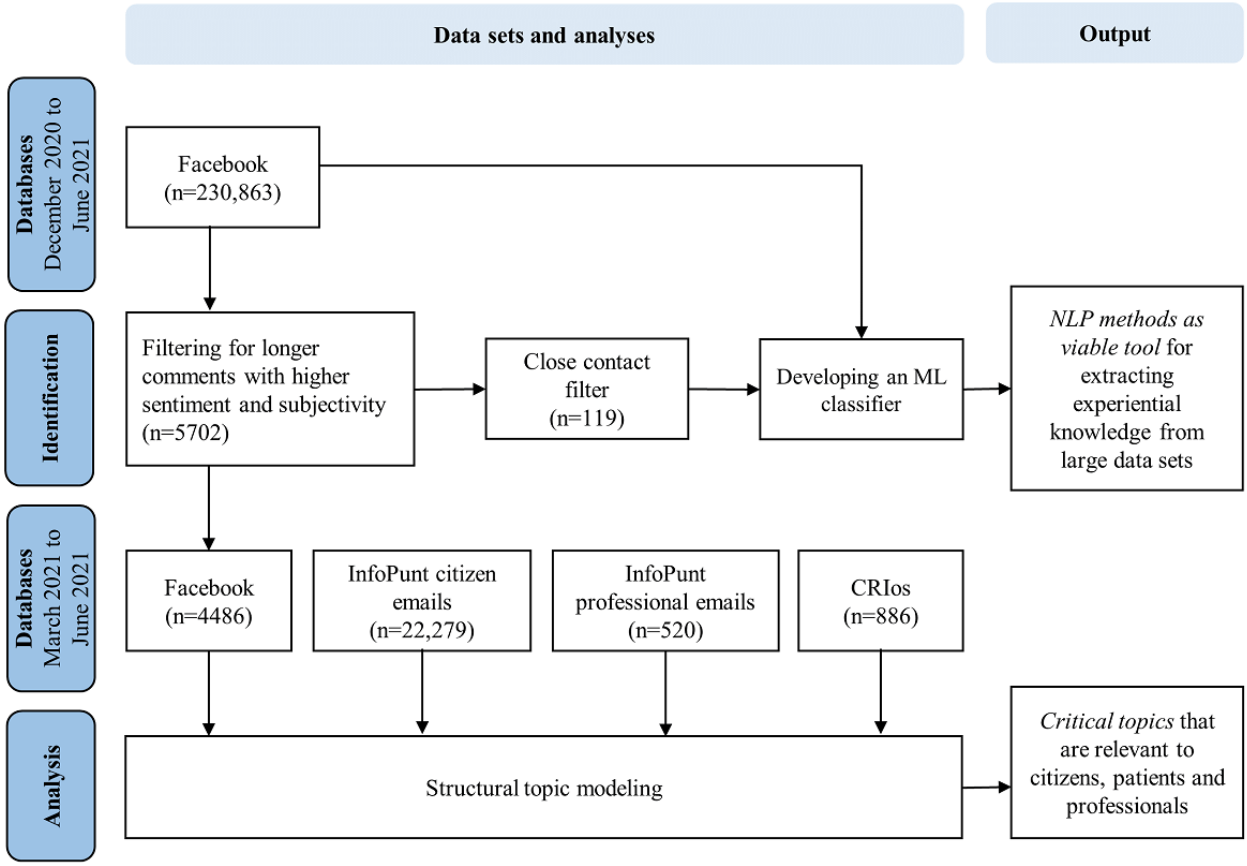
To also enable the sourcing of more targeted questions and concerns, we used 2 databases of RIVM: *InfoPunt*, the telephone and email support service for COVID-19 vaccination, and Casuïstiek Registratie Infectiezieken–operating system (*CRIos*), a case registry related to infectious diseases where professionals can report any challenges, questions, medical complications, and so on. From both databases, reports were extracted from January 2021 to June 2021 on the topic of "COVID-19 vaccination": 34,243 anonymized emails sent to InfoPunt by citizens and professionals and 1408 from CRIos. To safeguard protected health information and to guarantee the anonymity of all senders, a data protection protocol and anonymization script were developed and applied.

Analyses

Overview

An impeding factor for analyzing COVID-19 vaccination experiences from social media is the predominance of expressions of attitudes and opinions in the data and that only a fraction contains descriptions of experiences. Advancing the methodology for filtering out the health information needed and making it available to responders has been identified as a major research challenge to leveraging social media for public health emergencies [34]. Therefore, to identify experiential knowledge amidst the huge volume of Facebook data, we first developed a rudimentary filter and trained a machine learning (ML) classifier. Subsequently, the selected texts and the RIVM data sets were analyzed using structural topic modeling (STM) to discover the content of people's experiences and values related to COVID-19 vaccination. Figure 1 illustrates these data analysis streams.

Figure 1. Flowchart visualizing the data analysis streams and respective outputs. CRIos: Casuïstiek Registratie Infectiezieken–operating system; ML: machine learning; NLP: natural language processing.



Identifying Experience-Based Knowledge and Value Considerations

A rudimental filter was developed for the Facebook data set to narrow down the parts of the data that were likely to include experiences. First, we selected comments exceeding 250 characters, as experiences tend to be lengthy to describe. Second, posts with higher degree of subjectivity, typical for accounts of experiences and, third, posts of higher sentiment, that is, expressing stronger positive or negative emotions, were retained. To select such comments, sentiment analysis was conducted using the Python package, *Pattern* [35]. All comments were thereby assigned a value between −1 (negative) and +1 (positive) for their respective sentiment and a value between 0 (objective) and 1 (subjective) for subjectivity. Selected comments had an above-average subjectivity score (≥0.4) and a sentiment score ≤−0.25 or ≥0.25. This resulted in 5702 comments, from which a random sample of 500 (8.77%) comments was coded independently by 2 researchers for the presence of experiential knowledge (Cohen κ=0.66). The definition of experience-based knowledge related to COVID-19 vaccination in comments was kept broad, including first-hand and second-hand experiences. Our understanding is closest to the definition of *experience* from The Oxford Pocket Dictionary of Current English as "practical contact with and observation of facts or events." We then analyzed the comments coded as experience to find further features characteristic of these comments. These descriptions often contained references to a close contact, such as a family member. This was incorporated into the filter by additionally filtering for the combination of the words "my" and a close contact, for example, "my grandma." Whether the resulting comments exhibited experiential knowledge was assessed by another annotation round by the same 2 annotators at a moderate interannotator agreement (Cohen κ=0.60).

To detect a more diverse set of experiences beyond the rule-based filters, we set out to train a ML classifier on part of the initial data set. The "close contact" filter described previously supplied examples labeled as experiences for training the classifier. To obtain more experience posts to train on, this filter was expanded with 2 additional filter rules: the pattern, "have...had," and a first-level or second-level hypernym to the verb, "feel" (extracted from Open Dutch Wordnet [36]).

We sampled 70,830 Facebook posts as training data from the initial set, of which 1258 (1.78%) were labeled as experience based on the 3 conditions. The posts were first cleaned of URLs, emojis, punctuation, numbers, and symbols, and the remaining tokens were lowercased. The cleaned and normalized texts were tokenized and lemmatized using the Dutch Stanza pipeline (Stanford NLP Group) [37]. Comments matching any of the 3 filters formed the baseline against which we tested the performance of 2 different algorithms, logistic regression and extreme gradient boosting [38], and 2 different feature weightings, *tf*idf* [39] and binary.

The best-performing model was applied to identify experiences in a heldout set of 49,034 posts. We then evaluated the number of identified experiences by inspecting 2 samples of 250 posts each, 1 with high (>0.90) and 1 with low (0.50-0.90) classifier confidence.

Analyzing Experience-Based Knowledge and Value Considerations

We analyzed all data sets using STM to discern thematic patterns that may relate to experiential knowledge. Topic modeling algorithms are "unsupervised" ML methods for discovering manifest and latent topics in large collections of texts [40]. "Topics" are formed based on the co-occurrence of certain words. The underlying linguistic assumption is that words that systematically appear together across multiple texts are also associated thematically [41]. Topic modeling is particularly suitable for analyzing data sets after initial relevance filtering, owing to its exploratory perspective and ability to provide rich insights into the nature of a corpus [42].

Overall, 4 different topic models were estimated. To learn about health professionals' experiences and values related to COVID-19 vaccination, a topic model was run with 520 emails from health professionals received by InfoPunt and a second model was run with the 886 requests submitted to CRIos. To explore the experiences of citizens and patients, a third model with 22,279 emails from citizens to InfoPunt and a fourth model with 4486 long (>200 characters) comments from the high sentiment and subjectivity Facebook subset were run. Although we used only Facebook comments that resulted from the rudimentary filter, the RIVM databases did not require initial filtering; approximately all entries were regarding COVID-19 vaccination.

In all models, we included data from March 2021 onward, when vaccination of the wide population got underway in the Netherlands [43]. All non-Dutch contributions were excluded, and all texts were subjected to common preprocessing steps such as conversion to lowercase and removal of duplicates, numbers, punctuation, symbols, and web links. Words were stemmed, and a list of common Dutch "stop words," supplied by the *quanteda* R package, was removed. In the CRIos data set, staff-specific abbreviations such as "pat" for "patient" were also omitted as certain spellings distorted the topic formation process. Finally, words that appear in very few (eg, <0.3%) or almost all (eg, >95%) documents were screened out as they are unlikely to be discriminating.

Following Roberts et al [44] we have analyzed the measures of semantic coherence and exclusivity of different models to inform the selection of an adequate number of topics (K). Although semantic coherence is maximized when the most probable words of a given topic occur frequently together within documents, exclusivity measures the share of top topic words that are distinct to a given topic. For the Facebook data set, the highest values for both measures were achieved at K=13. Similarly, this was achieved at K=17 for InfoPunt (citizens), K=8 for InfoPunt (professionals), and K=10 for CRIos. All topic models were conducted with the *stm* package in R [45].

To understand the resulting topics and to check the model's validity, 20 documents highest associated with each topic were analyzed. On the basis of examining the most probable and most exclusive terms in conjunction with a close reading of exemplary documents, labels were assigned to the topics. We will provide an overview of all semantically meaningful topics; topics that

are not interpretable and without substantive meaning are not presented and not displayed in the tables [46].

Ethical Considerations

Ethics approval was not required for the analysis of the Facebook data set as it comprises publicly available posts from public pages on the platform. We did not publish specific comments that could be used to identify the original user and only share comment IDs in our Facebook data set [47] to preserve users' privacy. The analysis of the 2 RIVM internal databases, *InfoPunt* and *CRIos*, followed a strict data protection protocol, which was compliant with the General Data Protection Regulation and approved by the RIVM privacy coordinators.

Results

Identifying Experience-Based Knowledge and Value Considerations

To identify experiential knowledge in Facebook data, we developed a rudimentary filter that retained long comments with high levels of sentiment and subjectivity. On the basis of the annotated set that was sampled from the filtered data, where 10% (50/500) of the messages were found to be an experience, we can assume that the same percentage holds for the entire filtered set of 5702 messages. Extending the filter by searching for comments referring to a close, personal contact resulted in 2.09% (119/5702) of comments, of which 77.3% (92/119) featured experiential knowledge.

This shows that simple but carefully selected NLP filters can identify some instances of even a fairly complex concept such as "experience" and thus greatly speed up the search for such comments. The filters also supplied examples for training an ML classifier, with the aim to capture more diverse experiences than the rule-based filters would yield.

The optimal ML setup was an extreme gradient boosting classifier using a binary weighted lemma representation, yielding an F_1-score of 0.47 on predicting experiences in the annotated sample of 500 comments, considerably outperforming a rule-based filter (F_1-score of 0.20). The substantial difference in recall (0.59 vs 0.14) shows the generalizability of the ML approach. After this classifier was applied to previously unseen data, 4 coders annotated samples of 250 posts with low and high classifier confidence, to evaluate the number of comments classified as experience. The coders reached a slightly weak agreement (Cohen κ=0.59). An analysis of the disagreements revealed several factors complicating the interpretation, which included nonexperiences ("My mother did not take an inoculation"), distant experiences ("I heard about people who"), and lack of context ("I did that").

Our analysis has been a crucial step toward capturing the diversity of manifestations of experiential knowledge relevant to guideline developers, but further refinement of the ML classifier is needed before it can be integrated into the workflow of guideline developers.

Analyzing Experience-Based Knowledge and Value Considerations

We performed STM to gain insight into the main themes addressed in different data sources by citizens (InfoPunt and Facebook) and health professionals (InfoPunt and CRIos).

Health Professionals' Experiences and Values Related to COVID-19 Vaccination

The emails that health professionals directed to the InfoPunt help service revolved largely around 2 topic clusters: practical and organizational matters and questions about the vaccination of health professionals themselves (refer to Table 1 for an overview of all topics).

Table 1. Topics identified in health professionals' inquiries at InfoPunt, their assigned labels, proportions, and the most frequent and simultaneously most exclusive (FREX) words (n=520).

Topic and label	Corpus, n (%)	FREX terms
1—Vaccinating other health care providers	116 (22.3)	zeneca, astra, patient, nurse, birthday, caregiver, employer, elderly, staff, and ggz
2—Business requests	100 (19.2)	registrati, zeeland, perform, ms, registered, brba, already, register, quick test, and europe
3—Administrative problems	97 (18.6)	letter, appointment, mr, client, invite, jab, mail, call, adr, and called
4—GPs[a] and vaccination	60 (11.5)	bmi, envelopes, forwarding, forward, GPs, sir, transport, meet, online, and in advance
5—Tests	49 (9.4)	buildings, positive, tests, tested, test, symptoms, days, quarantine, scientolog, and self-test
6—Organization of vaccinations for care workers	41 (7.8)	laboratory technicians, radiological, care worker, acute, occupational group, care, contact person, infections, for example, and receives

[a]GP: general practitioner.

A topic cluster addresses administrative difficulties such as incorrect registrations of administered vaccines (topic 3) and challenges around vaccination in general practitioner (GP) practices (topic 4). Some GPs experience great strain having to organize COVID-19 vaccination in their practices in addition to their regular workload, including, for example, ordering vaccinations and prioritizing, inviting, and vaccinating patients. A second topic cluster concerns coordinating vaccination of nurses and GP practice staff who are prioritized for vaccination (topic 6). Besides, other health care professionals and employees

who could not keep the recommended safe distance (1.5 m) enquire when they will get their vaccination (topic 1). This second topic cluster demonstrates how the boundary between the roles of "health professional" and "citizen" or even "patient" as adopted in our study design and in the guideline becomes blurred, as a person can fall into all 3 categories.

The questions of health professionals logged in CRIos are more technical ones (Table 2). For instance, topic 1, with the highest prevalence (150/886, 17%) in this data set, comprises reports of symptoms occurring shortly after vaccination or lasting only briefly, for example, itching skin. Another prominent subject is vaccinating patients with serious health conditions (topics 2 and 4). The guideline specifies which groups of people are prioritized for vaccination because of their health condition. However, health professionals' questions indicate that there remains some uncertainty regarding the implementation of how (eg, with which vaccine) and by whom (eg, specialist, GP, or public health service) patients with certain conditions should be vaccinated. Moreover, professionals submit requests because some of their patients are not yet prioritized but are nevertheless deemed considerably more susceptible. This illustrates the tensions that professionals experience when they have to translate a policy that sharply defines priority groups for clear selection of patients to be prioritized for vaccination during a nationwide campaign to individualize patient care in medical practice. Another topic concerns mishaps at the vaccination location such as interrupted cold chains and vaccination with wrong needles or expired vaccines (topic 7). The analysis of topics from the case registry, owing to its focus on challenges in clinical practice, directly points out issues that arise in the application of the guideline that need to be covered and clearly explained in the guideline.

Table 2. Topics identified in health professionals' inquiries at Casuïstiek Registratie Infectiezieken–operating system, their assigned labels, proportions, and the most frequent and simultaneously most exclusive (FREX) words (n=886).

Topic and label	Corpus, n (%)	FREX terms
1—Immediate vaccination reactions	150 (17)	swollen, spots, lips, feeling, itching, oedema, body, red, minutes, and itching
2—Vaccination for medical risk groups	113 (12.8)	note, friendly, hear, deemed, unfortunately, GPs[a], want, watchman, employee, and medical
3—Pfizer vaccination	92 (10.4)	sum, other, summarize, dose, assist, birthday, madam, series, contraindication, and allergies
4—Medical conditions risk group	90 (10.2)	priority, parent, patients, quarantine, group, obesity, morbid, bmi, turn, and ml
5—Thrombosis	85 (9.6)	zeneca, 3e, thrombosis service, inr, called, agreement, astra, factor, children, and contra
6—Known allergies and other health conditions	75 (8.5)	person, jabbed, covid19, flu, reason, component, swallows, observation period, asap, and develops
7—Errors at vaccination site	74 (8.35)	resident, answer, know, hour, sure, work, guideline, logistics, two, and of course
8—Vaccination and previous COVID-19 infection	73 (8.2)	test, patient, positive, infection, scheduled, negative, antibodies, tested, past, and shorter
9—Allergic reactions	70 (7.9)	mobil, man, mg, tavegil, vaccination doctor, dd, old, diverse, day shift, and got
10—Questions and answers regarding diverse cases	64 (7.2)	account, holiday, use, website, message, url, in consultation, absolute, contraindication, and side effects

[a]GP: general practitioner.

Citizens' Experiences and Values Related to COVID-19 Vaccination

Similar to professionals' inquiries, citizens' emails to the InfoPunt support service mainly revolved around practical questions about their COVID-19 vaccination. In total, we identified 13 technically meaningful, large topics (Table 3). A few topics were about the pandemic more broadly, for example, questions about COVID-19 preventive measures (topic 3) and criticism of policies (topic 4). However, most questions were related to practical matters. This comprises organizational, administrative questions about incorrectly addressed vaccination invitations (topic 2); vaccination certificates (topic 8); and receiving vaccination when living abroad and traveling (topic 10). Travel was not considered as a relevant reason for vaccination according to the stated national objectives of the COVID-19 vaccination campaign (prevention of disease, hospitalization, and death) and thus not addressed initially in the guideline. However, our analysis revealed that the reasons for travel were more nuanced than just for recreational purposes, resulting in a considerable need for vaccines for indispensable travel for cross-border informal caregiving, accessing specialized care in a neighboring country, or cross-border work or study.

Table 3. Topics identified in citizens' inquiries at InfoPunt, their assigned labels, proportions, and the most frequent and simultaneously most exclusive (FREX) words (n=22,279).

Topic and label	Corpus, n (%)	FREX terms
1—Heterologous vaccination (AstraZeneca)	1849 (8.3)	astra, zeneca, zenica, 2nd, shot, pfizer, 1st, get, and moderna
2—Vaccination invitations	1715 (7.7)	invite, address, letter, receive, call, mail, send, present, receive, and born
3—Implementation of COVID-19 rules	1626 (7.3)	open, distance, infections, meter, measures, mask, keep, children, shop, and rule
4—Criticism of and recommendations for corona polices	1604 (7.2)	citizen, real, everyone, government, let, choice, ministry, policy, trust, and life
5—Scheduling a vaccination appointment	1603 (7.2)	appointment, call, online, download, called, make, location, succeed, and telephone
6—Thrombosis risk with AstraZeneca	1514 (6.8)	thrombosis, mother, women, birthday, afraid, father, birthday, 60-64, group, and can
7—Complaints and attachments	1470 (6.6)	sir, madam, esteemed, hereby, sir, request, awaits, organization, and dr
8—Request for vaccination certificate	1403 (6.3)	booklet, yellow, vaccination booklet, certificate, registration card, proof, registration, application, and registered
9—Definition of risk groups	1314 (5.9)	hospital, medical, patients, fall, risk group, indication, operation, asthma, priority, and care
10—Abroad	1292 (5.8)	germany, netherlands, belgium, dutch, abroad, travel, spanish, holiday, travel, and italy
11—Tests and results	1225 (5.5)	test, tested, positive, testing, pcr, result, negative, antibodies, and symptoms
12—Efficacy of different vaccines	1223 (5.5)	janssen, research, protect, variant, mrna, cases, less, effectiveness, studies, and offers
13—Vaccination side effects and second shot	1158 (5.2)	weeks, injection, two, burden, week, three, first, sensible, pain, and couple

Furthermore, questions were asked with respect to individual situations, for example, about heterologous vaccination, especially a messenger RNA vaccine following an initial shot of AstraZeneca (topic 1), and about how to proceed with the second vaccination after COVID-19 infection or adverse reactions to the first vaccination (topic 13). People are also uncertain whether they belong to the prioritized groups defined in the guideline, for example, owing to an accumulation of various mild risk factors in them (topic 9). Other people included in this prioritized group experience difficulties in receiving their vaccination, for example, because they no longer undergo active treatment but are nonetheless at risk or because their condition is not known to health professionals or authorities at all (eg, very high BMI).

What stands out in this analysis is that people who contact InfoPunt form a selection of citizens who do not seem to doubt the purpose of vaccination but already have practical questions about its implementation. As with professionals, these questions reflect the tension that might occur when applying a population-wide guideline to individual care. The RIVM InfoPunt contact center could be seen as an effort to assist in bridging this tension by providing answers to professionals' questions when providing care for individuals. It could, with the methods presented in this paper, additionally be used for analyzing which questions arise so often that an adjusted formulation of the guideline text may be required, with more focus on such frequently occurring exceptions. The topics resulting from this analysis can thus provide insights into the issues related to individualizing guidance [3] and can indicate the possibly required guideline changes.

Citizens' questions to InfoPunt mostly focused on the practical implementation of vaccination, whereas the Facebook comments to news articles on COVID-19 vaccination come from a more diverse group of citizens and span a wide range of values and experiences. We identified 12 technically meaningful topics (Table 4). As with InfoPunt, some topics relate to the pandemic more generally (topics 1 and 6). Among the more vaccine-specific topics, only the topic on the risk of thrombosis from vector vaccines, particularly from AstraZeneca, overlaps with those found in InfoPunt. Other topics identified in Facebook discussions revolve around uncertainties about processes and techniques that people first learned about in the wake of COVID-19 vaccination. Topic 5, for instance, assembles comments discussing the messenger RNA technique and if and how it affects the body's DNA. Discussions also focus on how reliably vaccinations have been tested and what a vaccine's provisional approval means (topic 11), which is often a reason for citizens to be skeptical about vaccines and consider them as actually not approved. Comments associated with topic 2 are about the purpose of vaccination, its effect on contagiousness, and the chance and severity of the COVID-19 disease.

Table 4. Topics identified in Facebook comments, their assigned labels, proportions, and the most frequent and simultaneously most exclusive (FREX) words (n=4486).

Topic and label	Corpus, n (%)	FREX terms
1—COVID-19 measures	489 (10.9)	freedom, mask, hear, rule, distance, holiday, oh, back, home, and shop
2—Effect of vaccination on contagiousness and disease	453 (10.1)	vaccinated, sick, infected, elderly, infecting, less, contagious, child, and serious
3—Own free choice	444 (9.9)	choice, sir, everyone, respect, own, free, fine, pressure, and feels
4—Meta-discussions	435 (9.7)	best, fine, reaction, via, message, government, latest, things, fun, and do
5—mRNA[a] technique and consequences	354 (7.9)	term, mrna, gene, long, dna, plan, gate, compare, and effects
6—Criticism of the government	345 (7.7)	rutte, hugo, jonge, measures, cabinet, numbers, deaths, minister, bring, and ic
7—Healthy body	336 (7.5)	healthy, syringe, dead, put, inject, couple, stay, quite, junk, and trust
8—Risk of thrombosis	319 (7.1)	thrombosis, astra, pfizer, group, astrazeneca, so many, zeneca, two, pill, and janssen
9—Research	301 (6.7)	research, article, namely, israel, indeed, strong, seems, some, information, and know
10—Dangers and risks	287 (6.4)	flu, covid, dangerous, flu shot, dying, fever, scared, deadly, exists, and side effects
11—Approval and testing of the vaccination	274 (6.1)	test phase, ema, medicine, tested, 2023, package leaflet, medication, approved, error, and safe
12—Jurisdictions and costs for tests	183 (4.1)	jab, GP[b], test, free, ggd, pcr, testing, pay, advice, and info

[a]mRNA: messenger RNA.

[b]GP: general practitioner.

This analysis of Facebook comments provides a sense of some key uncertainties that people—possibly also health professionals—express on social media about COVID-19 vaccination. Guideline users at the vaccination front line likely get confronted with these questions, to which the guideline currently does not provide answers.

The guideline provides guidance about the implementation of vaccination, such as instructions for injection, information about contraindications, vaccine combinations, and intervals [33]. An extension of the guideline by considering the uncertainties around the topics described previously could provide health professionals with possible answers to some of the public's biggest concerns. These straightforward clarifications would potentially render the guideline more relevant and applicable for guideline users and support them more comprehensively in their day-to-day practice.

Discussion

Principal Findings

This paper reveals the potential of AI methods to capture experience-based knowledge and value considerations of patients, professionals, and citizens, in our case, regarding COVID-19 vaccination. We first developed strategies to filter for this type of knowledge in a large Facebook data set. We subsequently applied STM to map the landscape of questions and discussion surrounding COVID-19 vaccination in this data set and in 2 RIVM databases. Our results indicate that it is indeed possible to extract experience-based knowledge and analyze value considerations, some of which have a direct relationship to the recommendations formulated in vaccination guidelines.

By using 2 different types of data sources, we were able to identify the unique focus and added value of each. We found that people contact InfoPunt, that is, the national public health institute, for advice on practical questions about organizing their COVID-19 vaccination. The user group posting on Facebook is differently composed and has different types of questions and concerns, mainly expressed in the form of comments and prompted by the topics of the news articles. Analyzing plural data channels is hence crucial for the inclusion of a wide range of citizens and professionals' experience-based knowledge and value considerations.

The topics revealed in the internal CRIos and InfoPunt databases are more similar to those in the guideline and thereby provide a more immediate way to incorporate value considerations and experiential knowledge in guideline development. Our finding that social media discourses are very different from what guideline developers consider relevant could have more profound implications for guideline development practices. In the initial phases of guideline development, the Population of interest, Intervention, Comparison, and Outcome framework is often consulted to select clinical questions and to search for clinical evidence. However, our analysis demonstrates that if people's experiences and value considerations are given a more central place, one also arrives at new starting points and

questions for discussion that do not fit this more clinical framing [8]. This approach also departs from public and patient involvement techniques, where these stakeholders are consulted only after scientific evidence has been gathered and priorities have been established. To improve the performance of vaccination guidelines in the field, addressing topics that are central concerns to citizens and professionals may be equally important and may need to occur during all stages of guideline development.

Limitations and Future Studies

This analysis can be seen as a proof of concept for the analysis of experiential knowledge and values using automated text analysis methods, which yield important initial results but should be developed further. The development of our ML classifier was a crucial first step in successfully uncovering experience descriptions in a vast data set. However, the diverse manifestations of experiences make it a difficult task for an ML classifier, which is why further refinement is crucial to be able to capture experiences in large data sets with more accuracy and reliability and to yield a more complete set of experiences.

Moreover, the approach should be tested with other data sets. Regarding social media platforms, we only analyzed Facebook and did not succeed in accessing, for example, WhatsApp groups, which were used intensively to rapidly exchange information. It is thus likely that we have missed discourses on other platforms. The issue of COVID-19 vaccination presented a particular case marked by a tremendous exchange of information and experience on the web. Although this amount of data from different channels has been useful for developing our methodology, a drawback of this extent of public attention is that one must expect the deliberate spread of misinformation on the web, possibly aided by bots [48]. We tried to mitigate the impact of bots in our analysis by following a human-in-the-loop (HITL) approach, where automated analyses constantly alternated with close reading and qualitative analysis of the data. Nevertheless, our approach could be further enhanced by integrating methods specifically designed to filter out misinformation and bot-written texts [49,50].

Following this proof-of-concept study, this approach should also be extended to other (vaccine) guidelines, which are developed in less dynamic contexts. Apart from other, more widely debated topics (eg, human papillomavirus vaccination in the Netherlands), diseases that have dedicated web-based forums and those where people seek the anonymity of the internet to share their experiences could be other promising opportunities for the application of our methods in guideline development. Another challenge in method advancement is to maintain the HITL aspect while standardizing the approach. For successfully developing our approach, it was vital to constantly combine automated methods with close qualitative analysis and human assessment of the data. The HITL approach has yielded the highest diagnostic performance when using AI to support clinical practice [51]. It seems to be equally important when trying to process complex concepts such as experiential knowledge and values using automated methods. Automated methods must be flexible and sensitive to formal, linguistic, or content-related peculiarities of a given data set.

Conclusions

Our study has shown how AI-based methods can be leveraged to extract and analyze experiential knowledge and value considerations on a large scale, which might open up new opportunities for the integration of this type of knowledge into public health guideline development, potentially improving the performance of the guidance produced. Using NLP methods, our rudimentary filter and ML classifier identified experiential knowledge in a large data set and subsequently allowed for performing STM to analyze health professionals, citizens, and patients' experiences with COVID-19 vaccination. The methods presented offer a novel approach for guideline developers to access and gain insights into experience-based knowledge from a broad range of people, especially in cases where conventional methods of incorporating such knowledge become impractical. They thereby provide a way to broaden the evidence and knowledge base available for public health guideline development, which is particularly valuable for rapid decision-making about pandemic response strategies. Despite the limitations, this proof-of-concept study has shown that AI-based methods developed in this study may need to be considered as important additions to the toolbox of guideline development and pandemic preparedness.

Acknowledgments

This project was funded by ZonMw, the Dutch Organization for Health Research and Care Innovation, under grant agreement 516022526. The authors would like to thank the Dutch National Institute for Public Health and the Environment (RIVM) and especially the database administrators of InfoPunt and Casuïstiek Registratie Infectiezieken–operating system (CRIos) for providing the data and technical assistance. Furthermore, the authors thank Suleika El Fassi for her support with the initial exploratory analysis of various data channels and annotation of Facebook comments.

Data Availability

Data from the internal databases (InfoPunt and Casuïstiek Registratie Infectiezieken–operating system; CRIos) from the Dutch National Institute for Public Health and the Environment (RIVM) were supplied after anonymization under strict data protection protocols agreed between the Vrije Universiteit Amsterdam and RIVM. Owing to the sensitive nature of the data, data cannot be redistributed to researchers other than those approved in the data protection protocol, which means that the data sets are not publicly available. The data collected from Facebook are publicly available in an open data repository [47].

Authors' Contributions

TZ-J, AT, ES, FK, MB, and MLS conceived the study. LL, supported by MLS, collected the data, developed and conducted the topic modeling analysis, contributed to data analysis and interpretation, and wrote the manuscript together with TZ-J and FK, which was supported by AT's comments about the manuscript at all stages. TZ-J, AT, and FK provided conceptual advice and contributed to data analysis and interpretation. FK developed and trained the machine learning classifier together with MC. MLS provided technical support and commented about the manuscript. ES conducted an initial exploratory analysis of the various data channels and contributed to the main data analysis and interpretation. WW contributed to data analysis and interpretation and commented about the manuscript. All authors reviewed the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
2. Cowl J. "Community members brought real life experience": An evaluation of lay people's contribution to public health guidelines. Otolaryngol Head Neck Surg 2010 Jul;143(S1):48-48 [FREE Full text] [doi: 10.1016/j.otohns.2010.04.191]
3. Wieringa S, Dreesens D, Forland F, Hulshof C, Lukersmith S, Macbeth F, AID Knowledge Working Group of the Guidelines International Network. Different knowledge, different styles of reasoning: a challenge for guideline development. BMJ Evid Based Med 2018 Jun;23(3):87-91 [FREE Full text] [doi: 10.1136/bmjebm-2017-110844] [Medline: 29615396]
4. den Breejen EM, Hermens RP, Galama WH, Willemsen WN, Kremer JA, Nelen WL. Added value of involving patients in the first step of multidisciplinary guideline development: a qualitative interview study among infertile patients. Int J Qual Health Care 2016 Jun;28(3):299-305 [doi: 10.1093/intqhc/mzw020] [Medline: 26968684]
5. Kim C, Armstrong MJ, Berta WB, Gagliardi AR. How to identify, incorporate and report patient preferences in clinical guidelines: a scoping review. Health Expect 2020 Jul 12;23(5):1028-1036 [FREE Full text] [doi: 10.1111/hex.13099] [Medline: 32656807]
6. Selva A, Sanabria AJ, Pequeño S, Zhang Y, Solà I, Pardo-Hernandez H, et al. Incorporating patients' views in guideline development: a systematic review of guidance documents. J Clin Epidemiol 2017 Aug;88:102-112 [doi: 10.1016/j.jclinepi.2017.05.018] [Medline: 28579379]
7. Zuiderent-Jerak T, Forland F, Macbeth F. Guidelines should reflect all knowledge, not just clinical trials. BMJ 2012 Oct 05;345:e6702 [doi: 10.1136/bmj.e6702] [Medline: 23043093]
8. Moleman M, Jerak-Zuiderent S, van de Bovenkamp H, Bal R, Zuiderent-Jerak T. Evidence-basing for quality improvement; bringing clinical practice guidelines closer to their promise of improving care practices. J Eval Clin Pract 2022 Dec;28(6):1003-1026 [FREE Full text] [doi: 10.1111/jep.13659] [Medline: 35089625]
9. Moleman M, Macbeth F, Wieringa S, Forland F, Shaw B, Zuiderent-Jerak T. From "getting things right" to "getting things right now": developing COVID-19 guidance under time pressure and knowledge uncertainty. J Eval Clin Pract 2022 Feb;28(1):49-56 [FREE Full text] [doi: 10.1111/jep.13625] [Medline: 34617367]
10. Atkinson S, Bradby H, Gadebusch Bondio M, Hallberg A, Macnaugthon J, Söderfeldt Y. Seeing the value of experiential knowledge through COVID-19. Hist Philos Life Sci 2021 Jun 29;43(3):85 [FREE Full text] [doi: 10.1007/s40656-021-00438-y] [Medline: 34185187]
11. Kothari A, Rudman D, Dobbins M, Rouse M, Sibbald S, Edwards N. The use of tacit and explicit knowledge in public health: a qualitative study. Implement Sci 2012 Mar 20;7:20 [FREE Full text] [doi: 10.1186/1748-5908-7-20] [Medline: 22433980]
12. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
13. Jiang J, Ren X, Ferrara E. Social media polarization and echo chambers in the context of COVID-19: case study. JMIRx Med 2021 Aug 05;2(3):e29570 [FREE Full text] [doi: 10.2196/29570] [Medline: 34459833]
14. Stamm TA, Andrews MR, Mosor E, Ritschl V, Li LC, Ma JK, et al. The methodological quality is insufficient in clinical practice guidelines in the context of COVID-19: systematic review. J Clin Epidemiol 2021 Jul;135:125-135 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.005] [Medline: 33691153]
15. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765 [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
16. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

17. Syrowatka A, Kuznetsova M, Alsubai A, Beckman AL, Bain PA, Craig KJ, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. NPJ Digit Med 2021 Jun 10;4(1):96 [FREE Full text] [doi: 10.1038/s41746-021-00459-8] [Medline: 34112939]
18. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
19. Chen Q, Leaman R, Allot A, Luo L, Wei CH, Yan S, et al. Artificial intelligence in action: addressing the COVID-19 pandemic with natural language processing. Annu Rev Biomed Data Sci 2021 Jul 20;4:313-339 [doi: 10.1146/annurev-biodatasci-021821-061045] [Medline: 34465169]
20. Gunasekeran DV, Chew A, Chandrasekar EK, Rajendram P, Kandarpa V, Rajendram M, et al. The impact and applications of social media platforms for public health responses before and during the COVID-19 pandemic: systematic literature review. J Med Internet Res 2022 Apr 11;24(4):e33680 [FREE Full text] [doi: 10.2196/33680] [Medline: 35129456]
21. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, et al. Big Data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: infoveillance study on twitter and Instagram. JMIR Public Health Surveill 2020 Aug 25;6(3):e20794 [FREE Full text] [doi: 10.2196/20794] [Medline: 32750006]
22. Andreadis S, Antzoulatos G, Mavropoulos T, Giannakeris P, Tzionis G, Pantelidis N, et al. A social media analytics platform visualising the spread of COVID-19 in Italy via exploitation of automatically geotagged tweets. Online Soc Netw Media 2021 May;23:100134 [FREE Full text] [doi: 10.1016/j.osnem.2021.100134] [Medline: 36570037]
23. Alhassan FM, AlDossary SA. The Saudi Ministry of Health's Twitter Communication Strategies and Public Engagement during the COVID-19 pandemic: content analysis study. JMIR Public Health Surveill 2021 Jul 12;7(7):e27942 [FREE Full text] [doi: 10.2196/27942] [Medline: 34117860]
24. Tang L, Liu W, Thomas B, Tran HT, Zou W, Zhang X, et al. Texas public agencies' tweets and public engagement during the COVID-19 pandemic: natural language processing approach. JMIR Public Health Surveill 2021 Apr 26;7(4):e26720 [FREE Full text] [doi: 10.2196/26720] [Medline: 33847587]
25. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
26. Yousefinaghani S, Dara R, Mubareka S, Sharif S. Prediction of COVID-19 waves using social media and Google search: a case study of the US and Canada. Front Public Health 2021 Apr 16;9:656635 [FREE Full text] [doi: 10.3389/fpubh.2021.656635] [Medline: 33937179]
27. Mishra S, Verma A, Meena K, Kaushal R. Public reactions towards COVID-19 vaccination through Twitter before and after second wave in India. Soc Netw Anal Min 2022;12(1):57 [FREE Full text] [doi: 10.1007/s13278-022-00885-w] [Medline: 35668822]
28. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
29. Chiavi D, Haag C, Chan A, Kamm CP, Sieber C, Stanikić M, et al. The real-world experiences of persons with multiple sclerosis during the first COVID-19 lockdown: application of natural language processing. JMIR Med Inform 2022 Nov 10;10(11):e37945 [FREE Full text] [doi: 10.2196/37945] [Medline: 36252126]
30. Bacsu JD, O'Connell ME, Cammer A, Azizi M, Grewal K, Poole L, et al. Using Twitter to understand the COVID-19 experiences of people with dementia: infodemiology study. J Med Internet Res 2021 Feb 03;23(2):e26254 [FREE Full text] [doi: 10.2196/26254] [Medline: 33468449]
31. Whittington C, Feinman T, Zelman Lewis SZ, Lieberman G, del Aguila M. Clinical practice guidelines: machine learning and natural language processing for automating the rapid identification and annotation of new evidence. J Clin Oncol 2019 Mar 10;37(8_suppl):77 [FREE Full text] [doi: 10.1200/jco.2019.37.8_suppl.77]
32. Harmsen W, Groot JD, Harkema A, Dusseldorp IV, Bruin J. Artificial intelligence supports literature screening in medical guideline development: towards up-to-date medical guidelines. Zenodo. 2021 Jun 25. URL: https://zenodo.org/record/5031907 [accessed 2023-09-01]
33. COVID-19 vaccination. Rijksinstituut voor Volksgezondheid en Milieu. 2022. URL: https://lci.rivm.nl/richtlijnen/covid-19-vaccinatie [accessed 2023-07-03]
34. Chan JL, Purohit H. Challenges to transforming unconventional social media data into actionable knowledge for public health systems during disasters. Disaster Med Public Health Prep 2020 Jun;14(3):352-359 [doi: 10.1017/dmp.2019.92] [Medline: 31610817]
35. De Smedt T, Daelemans W. Pattern for Python. J Mach Learn Res 2012;13:2063-2067 [FREE Full text] [doi: 10.5555/2188385.2343710]
36. Postma M, van Miltenburg E, Segers R, Schoen A, Vossen P. Open Dutch WordNet. In: Proceedings of the 8th Global WordNet Conference. 2016 Presented at: GWC '16; January 27-30, 2016; Bucharest, Romania p. 302-310 URL: https://aclanthology.org/2016.gwc-1.43.pdf

37. Qi P, Zhang Y, Zhang Y, Bolton J, Manning C. Stanza: a Python natural language processing toolkit for many human languages. arXiv. Preprint posted online March 16, 2020 2020 [FREE Full text] [doi: 10.18653/v1/2020.acl-demos.14]
38. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 13-17, 2016; San Francisco, CA p. 785-794 URL: https://dl.acm.org/doi/10.1145/2939672.2939785 [doi: 10.1145/2939672.2939785]
39. Day WH, Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods. J Classif 1984 Dec;1(1):7-24 [FREE Full text] [doi: 10.1007/bf01890115]
40. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84 [FREE Full text] [doi: 10.1145/2133806.2133826]
41. Matthes J, Kohring M. The content analysis of media frames: toward improving reliability and validity. J Commun 2008 Jul 09;58(2):258-279 [FREE Full text] [doi: 10.1111/j.1460-2466.2008.00384.x]
42. Murakami A, Thompson P, Hunston S, Vajn D. 'What is this corpus about?': using topic modelling to explore a specialised corpus. Corpora 2017 Aug;12(2):243-277 [FREE Full text] [doi: 10.3366/cor.2017.0118]
43. Coronavirus dashboard. Rijksoverheid. URL: https://coronadashboard.government.nl/landelijk/vaccinaties [accessed 2022-11-18]
44. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder‐Luis J, Gadarian SK, et al. Structural topic models for open-ended survey responses. Am J Pol Sci 2014 Mar 06;58(4):1064-1082 [FREE Full text] [doi: 10.1111/ajps.12103]
45. Roberts ME, Stewart BM, Tingley D. stm: an R package for structural topic models. J Stat Softw 2019;91(2):1-40 [FREE Full text] [doi: 10.18637/jss.v091.i02]
46. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, et al. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. Commun Methods Meas 2018 Feb 16;12(2-3):93-118 [FREE Full text] [doi: 10.1080/19312458.2018.1430754]
47. Lösch L. Facebook comments dataset IDs. Figshare. URL: https://figshare.com/articles/dataset/Facebook_comments_dataset_IDs_csv/19328762 [accessed 2023-09-01]
48. Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz HA, et al. Bots and misinformation spread on social media: implications for COVID-19. J Med Internet Res 2021 May 20;23(5):e26933 [FREE Full text] [doi: 10.2196/26933] [Medline: 33882014]
49. Kudugunta S, Ferrara E. Deep neural networks for bot detection. Inf Sci 2018 Oct;467:312-322 [FREE Full text] [doi: 10.1016/j.ins.2018.08.019]
50. Kolluri N, Liu Y, Murthy D. COVID-19 misinformation detection: machine-learned solutions to the infodemic. JMIR Infodemiology 2022 Aug 25;2(2):e38756 [FREE Full text] [doi: 10.2196/38756] [Medline: 37113446]
51. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2019 Nov 18;2:111 [FREE Full text] [doi: 10.1038/s41746-019-0189-7] [Medline: 31754637]

Abbreviations

AI: artificial intelligence
CRIos: Casuïstiek Registratie Infectiezieken–operating system
GP: general practitioner
HITL: human-in-the-loop
ML: machine learning
NLP: natural language processing
RIVM: Dutch National Institute for Public Health and the Environment
STM: structural topic modeling

Edited by A Mavragani; submitted 20.11.22; peer-reviewed by M Coccia, E Ferrara; comments to author 20.06.23; revised version received 11.07.23; accepted 27.07.23; published 14.09.23

Please cite as:
Lösch L, Zuiderent-Jerak T, Kunneman F, Syurina E, Bongers M, Stein ML, Chan M, Willems W, Timen A
Capturing Emerging Experiential Knowledge for Vaccination Guidelines Through Natural Language Processing: Proof-of-Concept Study
J Med Internet Res 2023;25:e44461
URL: https://www.jmir.org/2023/1/e44461
doi: 10.2196/44461
PMID: 37610972