



Published in final edited form as:

AIDS. 2025 July 15; 39(9): 1254–1261. doi:10.1097/QAD.0000000000004193.

Systematic Review of Infodemiology Studies Using Artificial Intelligence: Social Media Posts on HIV Pre-exposure Prophylaxis

Emiko Kamitani, PhD¹, Julia B. DeLuca, MLIS¹, Yuko Mizuno, PhD¹

¹Division of HIV Prevention, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia, United States 30329-4027

Abstract

Objectives. To explore how artificial intelligence (AI) can enhance infodemiology, which distributes and scans information in the electronic medium, to process social media posts for HIV pre-exposure prophylaxis (PrEP).

Design: Systematic Review

Methods. We searched in the U.S. Centers for Disease Control and Prevention's Prevention Research Synthesis database through June 2024 (PROSPERO: CRD42023458870). We included infodemiology studies published in English and reported using AI to process social media posts on PrEP. Two reviewers independently screened citations, extracted data, and conducted a risk of bias assessment using the Joanna Briggs Institute Critical Appraisal Checklist for Prevalence Studies. Findings are narratively summarized.

Results. Of the 135 citations screened, eight infodemiology studies were identified, analyzing over 58.9 million posts. Infodemiology studies found the PrEP topics commonly discussed in communities (e.g., barriers of uptake), rumors that may raise public health concerns (e.g., PrEP is a prevention method against COVID-19 infection), geographic locations where concerns regarding risk of acquiring HIV were raised (e.g., most HIV-related posts were from the 10 states with the highest numbers of new HIV diagnoses), and predicted HIV trends (e.g., HIV-related tweets were negatively correlated with the county-level HIV incidence rate in the following year).

Conclusions. Despite the limitations of this review including a small number of studies reviewed, our review suggests social media posts may provide information on real-time PrEP-related concerns, and AI can accelerate and enhance the processing of mass data to identify the information that communities need and the areas/locations that may need HIV prevention intervention.

*Corresponding author: Emiko Kamitani, ekamitani@cdc.gov.

Author Contributions:

All authors on this article meet the four criteria for authorship as identified by the International Committee of Medical Journal Editors (ICMJE); all authors have contributed to the conception and design of the study, drafted or have been involved in revising this manuscript, reviewed the final version of this manuscript before submission, and agree to be accountable for all aspects of the work. Specifically, using the CRediT taxonomy, the contribution of each author is as follows: Conceptualization: E. Kamitani, Y. Mizuno, and J.B. DeLuca; Original Draft: E. Kamitani; Revision of manuscript: E. Kamitani; Y. Mizuno, and J.B. DeLuca.

Disclosure: There are no conflicts of interest to report or financial disclosures.

Keywords

HIV; pre-exposure prophylaxis; artificial intelligence; infodemiology; social media; machine learning

Social media has been one of the most popular sources for retrieving information, and its use is constantly expanding [1]. In 2023, an estimated 4.9 billion people around the world used social media to share information and engage with news content [2, 3]. Social media platforms are well used by all race/ethnicity groups and sexual orientations, and users frequently share and openly engage in discussions about their health practices, including those related to HIV [4–7].

Studies that report findings from infodemiology (the distribution of health information in electronic media to inform public health and public policy) have been on the rise [8]. Infodemiology allows us to obtain real-time, user-generated data from multiple social media platforms (e.g., a microblogging site [X], a news aggregation and discussion site [Reddit], photo and video sharing platform [Instagram]) [9, 10]. These data are difficult to collect with traditional, time-consuming survey methods [9, 10]. Infodemiology is especially advantageous for getting behavioral responses (e.g., acceptance of vaccinations, prevention and control practice) to health-related issues [11] and information on sensitive topics, such as HIV and pre-exposure prophylaxis (PrEP) knowledge and behaviors, especially among disproportionately affected groups [12].

Social media content expands every second. Scanning and analyzing big data, such as social media posts, are time-consuming and require significant efforts in filtering and removing unwanted information [13]. Artificial intelligence (AI), such as machine learning (ML) or deep learning (DL), may help process social media posts [14]. AI is defined as a multidisciplinary field that focuses on developing intelligent systems capable of performing tasks that typically require human intelligence [15]. ML enables computers to learn from historical data and improve their performance without explicit programming [16], while DL simulates the complex decision-making processes of the human brain through multilayered neural networks [17]. Because of its real-world applications and ability to mimic human decision-making and judgment, AI has rapidly been adopted in medical care, especially during the COVID-19 pandemic to obtain or provide COVID-19-related information [18, 19]. This study is the first systematic review, to our knowledge, that describes infodemiology studies that used AI to process mass social media data to understand what communities are saying about HIV PrEP, to identify the advantages and disadvantages of using AI for this purpose, and to determine the next steps.

METHODS

The study protocol for the parent review [20] was registered in PROSPERO (CRD42023458870) [21]. Our report followed the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRSIMA) Statement [22].

Search

A search was conducted in CDC's Prevention Research Synthesis (PRS) Project database that had amassed more than 130,000 HIV prevention citations by the end of March 2025. Details of search strategies were published elsewhere [20]. For this review, a PRS Project team librarian searched the PRS database in August 2023, with a second query conducted in March 2024, for literature published from 2012 to the present. Two components were used in the search. The first component looked at AI with phrases centered on ML and natural language. The second component focused on pre-exposure prophylaxis and PrEP. The search query looked specifically at the title, abstract, and indexing terms for citations focused on PrEP and AI. We also searched for any newly published literature in PubMed, Scopus, and Google Scholar by using the same search terms (June 2024). See the Appendix for full PrEP annual searches, PRS database queries, and supplementary searches for this review.

Inclusion/Exclusion Criteria

Studies published in English and reported using AI, such as ML or natural language processing (NLP), to assess social media posts related to PrEP were included. We excluded studies that used AI to assess HIV information in general and did not focus on PrEP.

Screening and Data Abstraction

A two-step approach was applied to select studies. First, two reviewers independently screened the citations by title and abstract. Second, two reviewers independently reviewed the full text of included citations to confirm study eligibility. Disagreements were resolved through discussion. If the two reviewers failed to reach an agreement, a third reviewer resolved the discrepancy. Reviewers were trained, and all screening forms were pilot-tested and revised as necessary. Identified citations were exported to DistilleSR (a systematic review software, Evidence Partners, Ottawa, Canada) for data management to screen and identify eligible studies. For all eligible studies, two reviewers extracted data on population characteristics and AI program-related information and outcomes with a standard data abstraction form.

Quality Assessment/Data Synthesis

Study quality assessment of included studies was conducted by using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Prevalence Studies [23, 24]. The JBI checklist consisted of 9 questions with "yes" (1 point), "no" (0 points), "unclear" (0 points), or not applicable" (0 points). A total score of nine was possible, with four or less being considered as "high," five or six being considered as "moderate," and seven to nine considered as "low" [25–27]. We narratively synthesized the characteristics and findings of the included studies.

RESULTS

Of the 142 identified citations, 53 were duplications and 32 were excluded by title screening (Figure 1). Of the remaining 57 citations, 49 were excluded by full-report screening because they were modeling studies (n=21), interventional studies (n=13), non-AI studies (n=8), reviews (n=6), or non-PrEP-relevant studies (n=1). The remaining 8 studies were eligible for this review.

The study years included were from 2010 through 2022 (Table 1). Because social media posts could not be validated and may not be reliable, all studies scored low on our study quality assessment for measurement, classification, and other biases. Additionally, because social media posts were voluntary and may not represent the voices of those who did not use social media, the study's validity was diminished; thus, all studies had moderate risk of bias.

Studies used ML in various social media platforms to identify HIV or PrEP-related social media posts (e.g., text, still images, video footage) with the hashtag #HIV or #PrEP on Instagram [28–31], Reddit [30, 31], Tumblr [31], X (formerly Twitter) [30–35], or YouTube [31]. Three studies used NLP [29, 33, 34], which mimicked human decision-making, while one study [28] used convolutional neural networks (CNNs) for analyzing visual imagery.

The total number of social media posts screened by the included studies was more than 58.9 million. The highest number screened for one study was by Adrover [36], who purchased from an official X data reseller almost 40 million tweets that represented the full and unbiased stream of data posted between September 2010 and August 2013. The elements in social media posts that AI analyzed were categorized as either language content (e.g., tweets, texts) or visual content (e.g., still images, video footage).

Language Content

The included infodemiology studies scanned more than 55.8 million tweets and texts with ML or NLP and identified 15,217 relevant social media posts (0.03%). The tweets were posted mainly on X between September 2010 and December 2021 and represented a variety of real-time conversations in communities.

Erdengaslieng [35] randomly selected 1,200 of 3,700 PrEP-related tweets posted between January 2018 - December 2021 to understand the perceptions, awareness, and barriers to using PrEP. They reported that one of the major PrEP-related tweet topics advocated for better access to PrEP with insurance coverage and stated that pricing was the major reason for not using PrEP [35].

Adrover [36] screened X around July 16, 2012, to identify tweets discussing the adverse effects of antiretroviral therapy and associated sentiments around that date when the U.S. Food and Drug Administration (FDA) approved Truvada for PrEP. Out of approximately 40 million tweets, they identified 1,642 tweets and 512 unique users who described personal experiences with HIV drug treatment between September 2010 to August 2013 [37]. In the summer of 2012, around the time of the FDA approval, they observed a peak in tweeting about Truvada with increased media attention. The tweets included a significant number of negative posts, such as adverse effects of the medication (e.g., renal toxicity, nausea, vomiting) [36]. The tweets represented novel events, such as the availability of PrEP to prevent HIV infection and the adverse effects or negative sentiments, that may have been widely discussed in communities.

Xu used social media posts to explore the disruptions in PrEP care [31] and barriers to PrEP uptake during the COVID-19 pandemic [30]. The authors screened, from October 2020 to December 2020, 522,430 posts on X, YouTube, Tumblr, Instagram, and Reddit. Terms

linked to PrEP and HIV were examined to explore PrEP experiences, attitudes, and barriers among racial/ethnic minorities and sexual minorities [31]. Based on 785 identified posts, the impact of COVID-19 included access issues due to limited PrEP care services and lack of insurance coverage, referral, and transportation services [31]. The authors also scanned 267,689 posts from March 13, 2020 to December 11, 2020, a time of emerging concern for COVID-19 [30]. They identified 317 posts on X, Reddit, and Instagram discussing PrEP and COVID-19. The major themes identified were concerns that resources were being taken away from PrEP due to COVID-19 and that the pandemic contributed to disruption in PrEP uptake due to the inaccessibility of clinics. False information and myths regarding PrEP uptake included that PrEP could be used as a prevention method or precaution against COVID-19 infections [30]. Authors also found reduced self-reported sexual behavior and perceived lower HIV risk due to social distancing. This led to self-reported termination of PrEP and other HIV prevention methods [30].

Two studies [33, 34] used geolocation of HIV-related tweets and assessed the association with incidence rates of HIV infection. Klein [34] scanned over 3 million tweets containing keywords of gay, bisexual, and other men who have sex with men (collectively referred to as MSM) posted between September 2020 and January 2021 and identified 8,756 unique users with U.S. state-level geolocations and their ages. The authors found that most HIV-related posts (the residence of MSM users) were from the 10 states with the highest numbers of new HIV diagnoses in the United States. Moreover, over two-thirds (68%) of users with county-level geolocations were in counties or states considered priority jurisdictions by the Ending the HIV Epidemic in the U.S. initiative. The age distribution reflected that of persons who received a new diagnosis of HIV in the United States, with the 25–34 years age group being the most common, followed by the 13–24 years age group [34].

The second study by Stevens [33] explored associations between the number of geolocated HIV-related tweets and the future incidence rates of HIV infection. The authors examined how X activities among young men (aged 13–24 years) were related to the incidence of HIV infection in the United States. Reviewing 15.1 million tweets (2,517 were relevant) posted during all of 2016 (January 1–December 31, 2016), the study found that HIV-specific tweets were more likely to appear in counties where HIV incidence rates were high. Furthermore, geolocated HIV tweets were negatively associated with the county-level HIV incidence rate in the following year; for each additional 100 HIV-specific tweets per person in a particular county, there was a 3% reduction in the HIV incidence rate in the following year [33].

Visual Content

Included infodemiology studies scanned 65,999 image-based posts with CNN or NLP and identified 36,802 relevant social media posts (56%). All of these were posted on Instagram from January 2017 through July 2018.

Nobles screened the social media posts to explore the demographic characteristics of human faces posted with the hashtag #HIV [28] and concepts of these images [29]. For the former study, the authors screened 26,766 Instagram posts to explore HIV risk perception and communication in public health and found 10,036 with detectable human faces [28]. Social media posts with the hashtag #HIV on Instagram often feature faces of individuals from

groups that tend to have a lower prevalence of HIV infections, such as older adults, women, and white persons. This discrepancy might lead Instagram users to potentially misjudge their HIV risk [28]. For the latter study, the authors examined 39,233 posts and identified 26,766 that were relevant. They found that the visual content of posts with the hashtag #HIV primarily focused on presenting infographics about clinical interventions (like PrEP) presented images of persons aimed at reducing the stigma of sexual minorities and persons with HIV, and shared their anonymized HIV test results [29].

DISCUSSION

This systematic review identified 8 infodemiology studies that used AI to screen more than 58.9 million social media posts. Analyzing language elements (e.g., tweets, texts) in social media enabled us to analyze PrEP-related experiences or concerns (e.g., barriers to uptake [30, 31, 35], adverse effects of PrEP [36]), and rumors (e.g., PrEP can prevent COVID-19 infection [30]). Analysis also allowed for the geolocation of MSM who may be PrEP candidates [34] and to predict trends of HIV incidence [33]. Meanwhile, analyzing visual elements (e.g., still images, video footage) on video platforms (e.g., Instagram) revealed demographic profiles of pictures of human faces posted [28] and images of #HIV posts [29]. Here, we discuss how AI can enhance infodemiology to process social media posts, the advantages and disadvantages of using AI in the field, and future uses of AI for analysis.

First, infodemiology studies using AI can help gather mass data. This includes real-time, health-related information and how people consume such information (e.g., adverse effects of medications [36] or accessibility of care during public health emergencies [30, 31]), from a wide variety of communities; specifically from those who may not be included in traditional surveys. One of the biggest questions is the reliability and accuracy of capturing the actual communities' experiences and conversations.

The Infodemiology studies in this review found that the reasons for changes in PrEP use during the COVID-19 pandemic were the inaccessibility of PrEP care [31] and reduced sexual behaviors [30]. These were the 2 most common impacts of COVID-19 found by a recent systematic review examining the influences of the pandemic on PrEP care from 46 studies [38]. Additionally, renal toxicity, nausea, and vomiting reported on X as experiences of the communities [36] are well-recognized potential side effects of PrEP [39]. These findings indicated that scanning social media posts with AI can capture information typically collected by using traditional time-consuming and labor-intensive surveys, but with the advantage of collecting mass data quickly in a timely manner.

Second, social media helps spread public health information quickly and widely. Relative to traditional media (e.g., newspapers, TV, radio broadcasts), social media accelerates the recirculating of information [36]. As an example, tweeting about the first antiretroviral treatment approved for PrEP (i.e., Truvada) hit a peak when the FDA approval spurred coverage by the traditional media [36]. 512 unique users who tweeted about their HIV experiences on X had an average of 2,300 followers between 2010-2013 [36]. This suggests that a post by one person can spread broadly and influence thousands of followers and communities.

Meanwhile, one of the major PrEP-related tweet topics was to advocate better access, stating that insurance coverage and pricing were the major concerns in accessing care and using PrEP [35]. As a matter of fact, financial support, including changes in medication options (e.g., availability of generic drugs), Medicaid expansion, insurance coverage policies and regulations (e.g., Affordable Care Act or ACA), and PrEP assistance programs from manufacturers, are available [40]. A significant barrier may be low PrEP awareness, including a lack of information on the availability of PrEP financial support. Thus, when we introduce or promote preventive or public health strategies, we can take advantage of the influential factors of social networks and social media posts to promote awareness of the strategies [36].

On the other hand, information posts on social media may not always be appropriate or accurate. Negative sentiment may increase and influence communities to raise concerns [36]. For example, tweets associating undesirable medical experiences, such as adverse effects of medications, may negatively impact persons interested in the medications. False information can also be widely shared. During the COVID-19 pandemic, the rumor that PrEP can be used as a prevention method or precaution against COVID-19 infection was widely spread on X [30]. Evidence that PrEP can prevent COVID-19 infection, or lead to a less severe course of COVID-19 disease, has not been presented or verified [41].

Our current analyses do not explicitly differentiate between the sources of social media posts. However, distinguishing between unsolicited or unregulated posts and those posted by authoritative sources (e.g., government agencies, community organizations) or professionals and experts could provide deeper insights into the sources of information shared on social media. Future studies in this area could explore this differentiation in greater detail to better understand the impact of different sources on public perceptions and behaviors.

Social media can serve as both a powerful tool for promoting awareness and a platform that may contribute to the spread of misinformation. By targeting key influencers, addressing misinformation, or providing additional information to communities on social media platforms, we can enhance PrEP awareness and uptake.

Next, geolocated social media posts can identify populations with concerns and where those issues are across the country. Our review found that HIV-related tweets were more likely to be posted from the counties that had high HIV incidence rates or the priority jurisdictions of the Ending the HIV Epidemic in the U.S. initiative [33, 34]. Meanwhile, the HIV-related tweets were negatively associated with the county-level HIV incidence rate in the following year [33]. These findings could suggest that higher numbers of HIV-related tweets may be indicators of increased risk awareness, support, and community involvement [33]. Additionally, faces appearing in social media posts with the #HIV hashtag are more likely to belong to individuals who have a lower risk of HIV infection, especially older persons, women, and White persons. This could suggest that HIV-related social media posts may not resonate with persons at high risk for HIV and they might underestimate their own risk [28].

Finally, noise problems in ML could be tricky and impact the performance of scanning and analyzing social media posts. “Noise” means tweets about unrelated topics. By dropping the search terms that could be noise [36], analysis problems can be reduced but the ability to identify relevant data may also be reduced. The strategies of noise reduction, such as data cleaning, algorithm selection, and validation, should be well planned and executed. Beyond the use of AI, researchers still have to go through several validation steps, crowdsourcing, and human interventions to declare the validity of filters [32]. Additionally, further refining AI algorithms to identify relevant, high-quality, and influential social media posts related to PrEP is necessary. The integration of such factors as context, source credibility, and legitimacy into AI models is crucial to ensure that analyses capture not only the volume of posts but also evaluate their quality and influence.

One of the review limitations was the limited number of studies identified and reviewed. This is likely due to the evolving nature of research in this field, as well as the relatively novel application of AI in PrEP and infodemiology studies. Some of the evidence discussed in this review was concluded from single studies rather than from a synthesis of multiple studies. It is important to avoid making overly broad conclusions based on the limited number of studies. A synthesis of evidence across multiple studies is needed to draw more definitive conclusions and guide policy and practice effectively. Additionally, the variability in platforms used, study years, and data collection methods presented challenges in generalizing the findings across studies. Our included studies did not provide detailed information on several aspects, such as content type, engagement metrics, tone, and target audiences, which are crucial for understanding the reach and impact of these posts. To strengthen the evidence of PrEP use and behavioral trends, we need additional high-quality studies that explore these research areas in greater depth and consistency.

In conclusion, this review found that infodemiology studies provide real-time, public health trends and behaviors. AI, including ML and DL, helps scan mass, real-time, user-generated data on social media platforms quickly and accurately to capture simultaneous public health concerns and the locations/communities of those discussing the concerns. Social media data may serve as a proxy of people’s behaviors in communities, and analyzing big data has the potential to provide the information needed to improve patient outcomes and gain new insights into various human behaviors [42–44]. Our review explored whether AI can be used to identify populations who may face public health concerns and the areas/locations that may need public health interventions the most, as well as messages that the public seeks. However, we cannot ignore the issue of infodemic, which refers to an excessive amount of information, including false information in online or offline settings [45]. AI may further help identify specific public health topics that are unnecessary in certain locations. We should take advantage of developing technology and use it appropriately and wisely to improve and promote good health for all.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the U.S. Centers for Disease Control and Prevention. This work was entirely funded by the U.S. Government. All authors are federal government employees and this report is not subject to copyright in the United States.

References

1. Pew Research Center. Social Media Fact Sheet 2022 [updated January 11. Available from: <https://www.pewresearch.org/internet/fact-sheet/social-media/>.
2. Pew Research Center. Social Media Fact Sheet 2024 [updated January 31, 2024. Available from: <https://www.pewresearch.org/internet/fact-sheet/social-media/>.
3. Forbes. Top Social Media Statistics and Trends of 2024 2023 [updated May 18, 2023. Available from: <https://www.forbes.com/advisor/business/social-media-statistics/#source>.
4. Cai M, Shah N, Li J, Chen WH, Cuomo RE, Obradovich N, et al. Identification and characterization of tweets related to the 2015 Indiana HIV outbreak: A retrospective infoveillance study. *PLoS One*. 2020;15(8):e0235150. [PubMed: 32845882]
5. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman Isler M. Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. *J Med Internet Res*. 2015;17(11):e248. [PubMed: 26525289]
6. van Heerden A, Young S. Use of social media big data as a novel HIV surveillance tool in South Africa. *PLoS One*. 2020;15(10):e0239304. [PubMed: 33006979]
7. Kudrati SZ, Hayashi K, Taggart T. Social Media & PrEP: A Systematic Review of Social Media Campaigns to Increase PrEP Awareness & Uptake Among Young Black and Latinx MSM and Women. *AIDS Behav*. 2021;25(12):4225–34. [PubMed: 33939035]
8. Klimiuk KB, Balwicki ŁW. What is infodemiology? An overview and its role in public health. *Przegl Epidemiol*. 2024;78(1):81–9. [PubMed: 38904314]
9. 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;11(1):e11. [PubMed: 19329408]
10. Eysenbach G. Infodemiology: The epidemiology of (mis)information. *Am J Med*. 2002;113(9):763–5. [PubMed: 12517369]
11. Chen J, Wang Y. Social Media Use for Health Purposes: Systematic Review. *J Med Internet Res*. 2021;23(5):e17917. [PubMed: 33978589]
12. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. *Journal of Big Data*. 2018;5(1):17.
13. Bazzaz Abkenar S, Haghi Kashani M, Mahdipour E, Jameii SM. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telemat Inform*. 2021;57:101517. [PubMed: 34887614]
14. Saheb T, Sidaoui M, Schmarzo B. Convergence of artificial intelligence with social media: A bibliometric & qualitative analysis. *Telematics and Informatics Reports*. 2024;14:100146.
15. IBM. What is Artificial Intelligence (AI)? 2024 [Available from: <https://www.ibm.com/topics/artificial-intelligence>.
16. U.S. Centers for Disease Control and Prevention. Artificial Intelligence and Machine Learning [updated July 3, 2023. Available from: <https://www.cdc.gov/surveillance/data-modernization/technologies/ai-ml.html>.
17. IBM. What is computer vision? [Available from: <https://www.ibm.com/topics/computer-vision>.
18. Piccialli F, di Cola VS, Giampaolo F, Cuomo S. The Role of Artificial Intelligence in Fighting the COVID-19 Pandemic. *Inf Syst Front*. 2021;23(6):1467–97. [PubMed: 33935585]
19. Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S, et al. Artificial Intelligence for COVID-19: A Systematic Review. *Front Med (Lausanne)*. 2021;8:704256. [PubMed: 34660623]

20. Kamitani E, Mizuno Y, Khalil GM, Viguerie A, DeLuca JB, Mishra N. Improving HIV Pre-exposure Prophylaxis Uptake with Artificial Intelligence and Automation: A Systematic Review. *AIDS* (London, England). 2024.
21. Kamitani E, Mizuno Y, Mishra N, Khalil G, Viguerie A, DeLuca JB. Artificial Intelligence Used in Interventions or Programs to Promote HIV Pre-exposure Prophylaxis (PrEP) Clinical Care: A Systematic Review. *PROSPERO* [Internet]. 2023; CRD42023458870. Available from: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=458870.
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*. 2021;372:n71. [PubMed: 33782057]
23. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Chapter 5: Systematic reviews of prevalence and incidence. 2020 [cited January 19, 2024]. In: *JBIManual for Evidence Synthesis* [Internet]. JBI, [cited January 19, 2024]. Available from: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fjbi.global%2Fsites%2Fdefault%2Ffiles%2F2021-10%2FChecklist_for_Prevalence_Studies.docx&wdOrigin=BROWSELINK.
24. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *Int J Evid Based Healthc*. 2015;13(3):147–53. [PubMed: 26317388]
25. Lu C, Dasgupta P, Cameron J, Fritschi L, Baade P. A systematic review and meta-analysis on international studies of prevalence, mortality and survival due to coal mine dust lung disease. *PLoS One*. 2021;16(8):e0255617. [PubMed: 34343220]
26. van der Kruk S, Harrison NJ, Bartram A, Newton S, Miller C, Room R, et al. Prevalence of parental supply of alcohol to minors: a systematic review. *Health Promotion International*. 2023;38(5).
27. Aonso-Diego G, Krotter A, García-Pérez Á. Prevalence of energy drink consumption world-wide: A systematic review and meta-analysis. *Addiction*.n/a(n/a).
28. Nobles AL, Leas EC, Noar S, Dredze M, Latkin CA, Strathdee SA, et al. Automated image analysis of instagram posts: Implications for risk perception and communication in public health using a case study of #HIV. *PLoS One*. 2020;15(5):e0231155. [PubMed: 32365124]
29. Nobles AL, Leas EC, Latkin CA, Dredze M, Strathdee SA, Ayers JW. #HIV: Alignment of HIV-Related Visual Content on Instagram with Public Health Priorities in the US. *AIDS Behav*. 2020;24(7):2045–53. [PubMed: 31916098]
30. Xu Q, McMann T, Godinez H, Nali MC, Li J, Cai M, et al. Impact of COVID-19 on HIV Prevention Access: A Multi-platform Social Media Infodemiology Study. *AIDS Behav*. 2023;27(6):1886–96. [PubMed: 36471205]
31. Xu Q, Nali MC, McMann T, Godinez H, Li J, He Y, et al. Unsupervised Machine Learning to Detect and Characterize Barriers to Pre-exposure Prophylaxis Therapy: Multiplatform Social Media Study. *JMIR Infodemiology*. 2022;2(1):e35446. [PubMed: 37113799]
32. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter. *JMIR Public Health Surveill*. 2015;1(2):e7. [PubMed: 27227141]
33. Stevens R, Bonett S, Bannon J, Chittamuru D, Slaff B, Browne SK, et al. Association Between HIV-Related Tweets and HIV Incidence in the United States: Infodemiology Study. *J Med Internet Res*. 2020;22(6):e17196. [PubMed: 32579119]
34. Klein AZ, Meanley S, O'Connor K, Bauermeister JA, Gonzalez-Hernandez G. Toward Using Twitter for PrEP-Related Interventions: An Automated Natural Language Processing Pipeline for Identifying Gay or Bisexual Men in the United States. *JMIR Public Health Surveill*. 2022;8(4):e32405. [PubMed: 35468092]
35. Erdengasileng A, Tian S, Green SS, Naar S, He Z. Using Twitter Data Analysis to Understand the Perceptions, Awareness, and Barriers to the Wide Use of Pre-Exposure Prophylaxis in the United States. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 2022;2022:3000–7. [PubMed: 36818418]

36. Adrover C, Bodnar T, Huang Z, Telenti A, Salathe M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. *JMIR Public Health and Surveillance*. 2015;1(2):e7.1–e7.10. [PubMed: 27227141]
37. U.S. Food and Drug Administration. Truvada for PrEP Fact Sheet: Ensuring Safe and Proper Use 2012 October 28, 2024 [cited 2022 November 30th]. Available from: <https://www.fda.gov/media/83586/download>.
38. Hong C. Characterizing the Impact of the COVID-19 Pandemic on HIV PrEP care: A Review and Synthesis of the Literature. *AIDS Behav*. 2023;27(7):2089–102. [PubMed: 36459263]
39. Gilead. Medication Guide: TRUVADA® (tru-VAH-dah) (emtricitabine and tenofovir disoproxil fumarate) tablets. What is TRUVADA®? [Internet]. May 9, 2024. Available from: https://www.gilead.com/~media/Files/pdfs/medicines/hiv/truvada/truvada_medication_guide.pdf.
40. Bonacci RA, Van Handel M, Huggins R, Inusah S, Smith DK. Estimated Uncovered Costs For HIV Preexposure Prophylaxis In The US, 2018. *Health Aff (Millwood)*. 2023;42(4):546–55. [PubMed: 37011310]
41. Ayerdi O, Puerta T, Clavo P, Vera M, Ballesteros J, Fuentes ME, et al. Preventive Efficacy of Tenofovir/Emtricitabine Against Severe Acute Respiratory Syndrome Coronavirus 2 Among Pre-Exposure Prophylaxis Users. *Open Forum Infectious Diseases*. 2020;7(11).
42. SAS. Riverside County relies on data integration and analytics from SAS to improve the health and well-being of vulnerable Californians [Available from: https://www.sas.com/en_us/customers/riverside-county.html].
43. SAS. Big Data: What it is and Why it Matters [Available from: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html].
44. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: Opportunities and challenges. *Neurocomputing*. 2017;237:350–61.
45. World Health Organization. Infodemic [Available from: https://www.who.int/health-topics/infodemic#tab=tab_1].

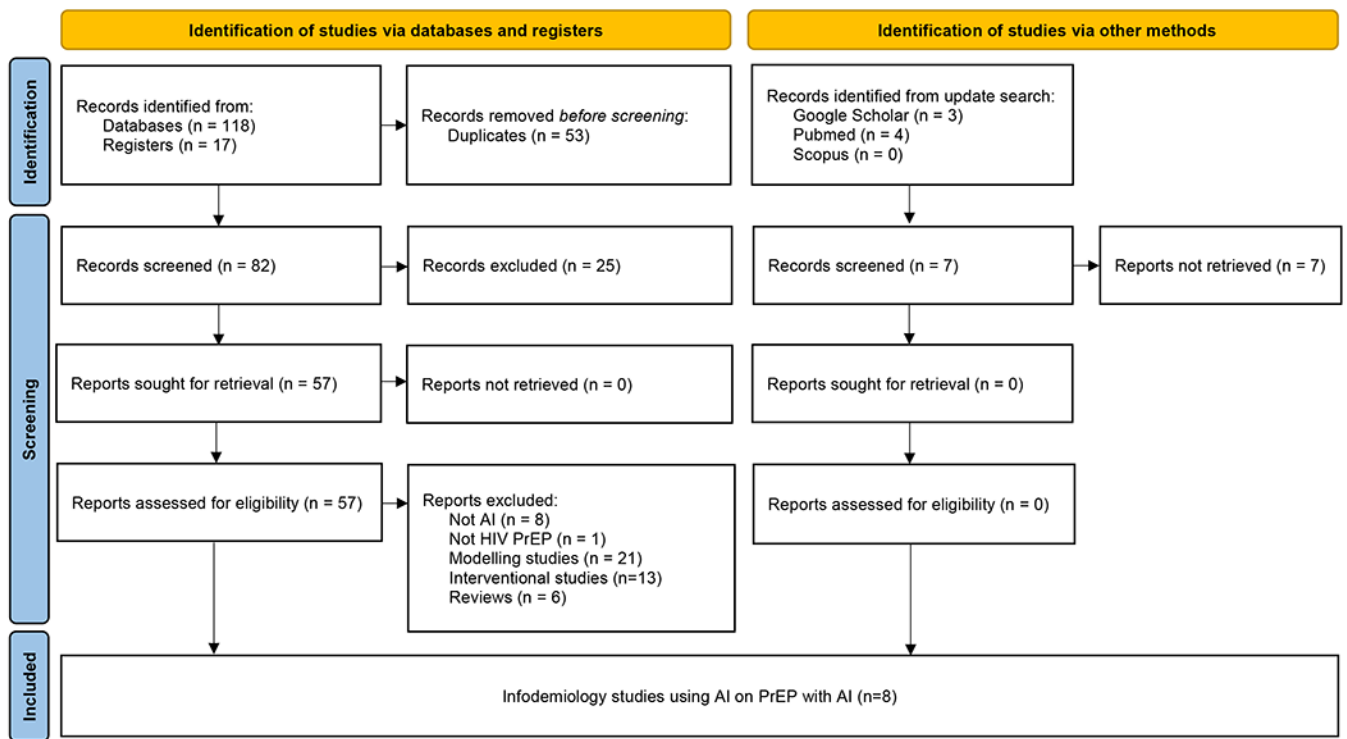


Figure 1:
Flow diagram of included studies

Table 1: Study characteristics of pre-exposure prophylaxis (PrEP) infodemiology studies using artificial intelligence (AI) published from 2012 to 2024 (N=8)

First author (publication year)	Used Social Media and AI	Study Objectives	Characteristics of Social Media Posts	Relevant Findings	Risk of Bias ¹
<i>Language Contents (n=6)</i>					
Erdengasileng (2023)	X ML	Understand the perceptions, awareness, and barriers to using PrEP	3,700 tweets posted from January 2018 to December 2021 1,200 tweets were randomly selected.	PrEP issues often discussed in personal tweets include advocacy, risks/benefits, access, pricing, insurance coverage, legislation, stigma, health education, and prevention of HIV.	Moderate
Xu (2023)	X, Reddit, Instagram ML	Identify and characterize key barriers associated with access to PrEP therapy during the COVID-19 pandemic	267,689 posts between March 13, 2020 and December 11, 2020. 317 posts discussing PrEP and/or HIV prevention practices and also mention COVID-19	COVID-19 pandemic contributed to disruptions in HIV prevention services that may impact PrEP uptake and adherence and other HIV prevention services, particularly among ethnic, racial, and sexual minority populations.	Moderate
Xu (2022)	X, YouTube, Tumblr, Instagram, Reddit ML	Identify and characterize barriers associated with PrEP uptake and adherence.	522,430 posts from October 13, 2020, to December 11, 2020. 785 posts were identified that specifically related to barriers.,	The main barriers included a lack of knowledge, access issues (e.g., lack of insurance coverage, no prescription, the impact of COVID-19 pandemic), and low adherence (e.g., side effects, social stigma)	Moderate
Adrover (2015)	X ML	Examine if adverse effects of HIV drug treatment and associated sentiments can be identified using social media data	39,988,306 tweets were posted between September 2010 to August 2013. 1,642 described personal experiences with HIV drug treatment	Concerns for adverse effects expressed in social media about specific HIV drugs or drug combinations, accurately capture well-recognized toxicities.	Moderate
Stevens (2020)	X NLP	Examine how Twitter activity among young (13-24 years) men is associated with the incidence of HIV infection in the U.S.	15.1 million tweets between January 1, 2016, and December 31, 2016. 2,517 were identified as potentially relevant to HIV risk and prevention tags.	Every 100-tweet increase in HIV-specific tweets per capita from the noninstitutional account was associated with a decrease in HIV incidence in the following year ²⁹	Moderate
Klein (2022)	X NLP	Develop an automated NLP pipeline to identify MSM in the U.S. and assess the extent to which they demographically represent MSM in the US with new HIV diagnoses	More than 3 million tweets between September 2020 and January 2021. 8,756 unique users with US state-level geolocations.	Among the 8,756 users with US state-level geolocations, 5096 (58.2%) were in the 10 states with the highest numbers of new HIV diagnoses. Among the 6240 users with county-level geolocations, 4252 (68.1%) were in counties or states considered priority jurisdictions by the Ending the HIV Epidemic initiative	Moderate
<i>Visual Contents (n=2)</i>					
Nobles (2020)	Instagram ML, specifically CNN	Implications for risk perception and communication in public health	26,766 posts from January 2017 through July 2018. 10,036 posts containing #HIV with detectable human face	Compared to demographic characteristics of people with new HIV diagnoses in the U.S. in 2017, faces posted on Instagram containing #HIV were more likely to be white (43% vs. 26%), older (35-39 years old) (47% vs. 11%), and female (41% vs. 19%), and less likely to be black (31% vs 44%), and Hispanic (13% vs 25%). The results were similar among the subset of #HIV posts mentioning PrEP.	Moderate

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

First author (publication year)	Used Social Media and AI	Study Objectives	Characteristics of Social Media Posts	Relevant Findings	Risk of Bias ¹
Nobles (2020)	Instagram NLP	Explore thematic concepts in the image using automated image recognition and topic modeling	39,233 posts containing #HIV from January 2017 through July 2018 26,766 of the posts were authored in English	Most posts containing #HIV focused on awareness (28%), LGBTQ (24.5%) or living with HIV (17.9%). Less than one third cited specific HIV strategies including HIV testing (10.8%), treatment (10.3%), PrEP (6.2%), and condoms (4.1%).	Moderate

CNN: convolutional neural networks; LGBTQ: lesbian, gay, bisexual, transgender, queer or questioning; ML: machine learning; MSM: men who have sex with men; NLP: natural language processing; PrEP: pre-exposure prophylaxis

¹ Risk of Bias was assessed by using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Prevalence Studies.