




REGISTERED REPORT PROTOCOL

Artificial intelligence applications in social media for depression screening: A systematic review protocol for content validity processes

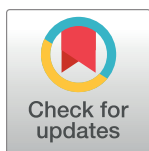
Priscilla N. Owusu¹^{*}, Ulrich Reininghaus^{2‡}, Georgia Koppe^{2‡}, Irene Dankwa-Mullan^{3‡}, Till Bärnighausen^{1,4}

1 Institute of Global Health, University Hospital Heidelberg, Heidelberg, Germany, **2** Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany, **3** IBM Watson Health, Maryland, Bethesda, MD, United States of America, **4** Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States of America

 These authors contributed equally to this work.

 These authors also contributed equally to this work.

* priscilla.owusu@uni-heidelberg.de



This is a Registered Report and may have an associated publication; please check the article page on the journal site for any related articles.

OPEN ACCESS

Citation: Owusu PN, Reininghaus U, Koppe G, Dankwa-Mullan I, Bärnighausen T (2021) Artificial intelligence applications in social media for depression screening: A systematic review protocol for content validity processes. PLoS ONE 16(11): e0259499. <https://doi.org/10.1371/journal.pone.0259499>

Editor: Astrid M. Kamperman, Erasmus Medical Center, NETHERLANDS

Received: January 15, 2021

Accepted: October 20, 2021

Published: November 8, 2021

Copyright: © 2021 Owusu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The background evidence supporting this protocol discussion are fully stated in the reference section of the paper. The systematic review is forthcoming, and authors confirm that all data underlying the findings of the anticipated study outlined in this Registered Report

Abstract

Background

The popularization of social media has led to the coalescing of user groups around mental health conditions; in particular, depression. Social media offers a rich environment for contextualizing and predicting users' self-reported burden of depression. Modern artificial intelligence (AI) methods are commonly employed in analyzing user-generated sentiment on social media. In the forthcoming systematic review, we will examine the content validity of these computer-based health surveillance models with respect to standard diagnostic frameworks. Drawing from a clinical perspective, we will attempt to establish a normative judgment about the strengths of these modern AI applications in the detection of depression.

Methods

We will perform a systematic review of English and German language publications from 2010 to 2020 in PubMed, APA PsychInfo, Science Direct, EMBASE Psych, Google Scholar, and Web of Science. The inclusion criteria span cohort, case-control, cross-sectional studies, randomized controlled studies, in addition to reports on conference proceedings. The systematic review will exclude some gray source materials, specifically editorials, newspaper articles, and blog posts. Our primary outcome is self-reported depression, as expressed on social media. Secondary outcomes will be the types of AI methods used for social media depression screen, and the clinical validation procedures accompanying these methods. In a second step, we will utilize the evidence-strengthening **Population, Intervention, Comparison, Outcomes, Study type (PICOS)** tool to refine our inclusion and exclusion criteria. Following the independent assessment of the evidence sources by two authors for the risk of bias, the data extraction process will culminate in a thematic synthesis of reviewed studies.

Protocol will be made fully available without restriction upon completion.

Funding: The author(s) received no specific funding for this work.

Competing interests: Priscilla N. Owusu is the founder of Oogtech AI UG, a startup company in Germany, developing digital health technology solutions for diabetic retinopathy. The products incorporate artificial intelligence applications to promote wellbeing including mental health. Dr. Irene Dankwa-Mullan is employed by the commercial company IBM Watson Health, IBM Corporation as the chief health equity officer and deputy chief health officer. IBM Watson Health develops computational technology products for clinical use purposes. No other authors have competing interests. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; AI, Artificial Intelligence; COSMIN, Consensus-based Standards for the selection of health status Measurement Instruments; GRADE, Grading of Recommendations, Assessment, Development and Evaluations framework; LGBTQ+, Lesbian, gay, bisexual, transsexual, queer, and others; ML, Machine learning; PICOS, Problem, Intervention, Comparison, Outcomes, Study type; SRDR, Systematic Review of Data Repository.

Discussion

We present the protocol for a systematic review which will consider all existing literature from peer reviewed publication sources relevant to the primary and secondary outcomes. The completed review will discuss depression as a self-reported health outcome in social media material. We will examine the computational methods, including AI and machine learning techniques which are commonly used for online depression surveillance. Furthermore, we will focus on standard clinical assessments, as indicating content validity, in the design of the algorithms. The methodological quality of the clinical construct of the algorithms will be evaluated with the COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) framework. We conclude the study with a normative judgment about the current application of AI to screen for depression on social media.

Systematic review registration

International Prospective Register of Systematic Reviews PROSPERO (registration number [CRD42020187874](https://doi.org/10.1186/CRD42020187874)).

Background

Worldwide phenomenon of depression

Globally, depression is one of the leading causes of disability burdens affecting almost 300 million of the world's population, with growing prevalence among children and adolescents [1, 2]. When left untreated, severe depression increases the risk of suicide, claiming 800,000 lives yearly, counting as the second leading mortality cause among 15–29 year-olds [1]. As many nations remain severely impacted by the SARS-CoV-2 pandemic, both clinical and subsyndromal levels of depression are expected to increase by unprecedented proportions [2–4]. Depression is associated with a complex interplay of biological, emotional, and social impairments on the affected individual [5]. Biologically, chronic exposure to depression-induced neurotransmitters is suspected to induce epigenetic changes which can alter brain functioning in the long term [6]. Kupferberg and colleagues (2016) [5] summarize the interpersonal deficits of depression into a framework with the subconstructs of “Affiliation and Attachment”, “Social Communication”, “Perception and Understanding of Self” and “Perception and Understanding of Others.” In addition, the economic impact of depression manifests as reduced work productivity and unemployment, contributing to a cumulative loss of \$1 trillion each year to the global economy [7].

Despite the increasing global burden of depression, it remains underdiagnosed and poorly managed [8]. Literature evidence indicates that the multidimensional barriers to adequate care and improved outcomes for depression center around individual-level challenges including resistance to care-seeking and psychosocial limitations. At the clinical-level, diagnostic and treatment incongruity involving physician-oriented interventions are known barriers, while reduced access to formal mental health care services constitute structural-level problems linked to unmet care [9].

Internet-based surveillance of depression

Internet-based sources including search queries, blogs, web encyclopedias, and social media websites provide proxy indicators for public health surveillance. This practice, termed

“infoveillance” or “infodemiology,” aims to understand health trends, outbreaks, and disease prevalence from semi-structured data reported by internet users [10]. Social media networks such as Twitter, Facebook, TikTok, Reddit, Instagram, and blog sites each offer a virtual community network where people of various demographic backgrounds share sentiments, exchange information, and provide mutual support for common disease conditions. Remarkably, these platforms serve as channels for members of stigmatized populations such as LGBTQ+ communities, and ethnic and religious minorities to coalesce anonymously [11].

Numerous studies have demonstrated the link between social media material and mental health status, including depression [12–15]. The surveillance of online content and users’ posting activity has been proposed as a complementary or even alternative precision tool for the early detection of depression markers. The strong appeal toward the analysis of social media content, as a form of patient self-reporting, arises from its potential as a measurement modality which is unconstrained by recall bias or selective disclosure. Moreover, risk factors including patient behavior, demographics, and socioeconomic status can be directly deduced from publicly accessible user profiles [16].

Artificial intelligence (AI) is a term that is broadly applied to the computerized automation of tasks normally requiring human intelligence [17, 18]. Through a subset of AI, known as machine learning (ML), algorithms or computer-generated operations can be developed to train large quantities of data obtained from social media postings to detect the presence of depression. With supervised learning ML methods, these algorithms gain the ability to recognize psychometric outcomes based on the degree of matching with the predefined indicator variables. For example, Coppersmith et al. (2015) [19] used linguistic analyses to predict depression and PTSD as well as demographic identities of Twitter users with age- and gender-matched controls. Yazdavar et al. (2020) [20] have similarly designed a multimodal framework to process visual imagery, textual data, and user interactions to detect depression among Twitter users. Further advanced applications of AI such as deep neural networks enable the matching of depressed individuals following in-depth evaluations of images based on characteristics of “colorfulness, facial presentation, sharpness, and naturalness.”

Addressing the content validity of computational diagnoses for depression

Diagnostic manuals including the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (DSM-V) and *International Classification of Disease* version 11 (ICD-11) are considered the gold standard for clinical practice. Following the administration of standardized psychiatric interviews, these frameworks provide a “behaviorally descriptive and systematic diagnostic system” [21] for identifying mental disorders, including depression. The psychopathology of a patient is determined by following a dimensionalized approach [21, 22] where a diagnosis is made on the basis of the degree of matching between his, her or their self-reported symptoms and the criteria description in the DSM-V manual. This “classical view” of diagnostic decision-making is predicated on strict adherence to the defined categories.

Even so, the DSM system is not without its unique set of limitations [23]. Conventionally, clinicians adopt a “probabilistic” approach to diagnosis, meaning that the most likely diagnosis is offered following the consideration of information relevant to a patient’s emotional state [22]. Given that a similar empirical methodology is used in computational psychiatry, there is a potential convergence point for the two disciplines to develop a modern, sophisticated standard for diagnosing mental disorders.

While previous studies [24–27] discuss the computational prediction of self-reported depression and other mental health disorders on social media, a remaining gap in the literature is the evidence of the scope of their clinical validity. Indeed, in the domain of psychiatry,

Jablensky [28] notes that the nosology of neurological and psychological conditions remains unresolved. A valid diagnosis, however, is conceptualized as the level of correlation to the “idealized” norm, pathology, demographic and cultural variation of a reference population [28–30].

In their critical review of computational predictive techniques to determine mental health status on social media, Chancellor and De Choudhury [31] delve into the frameworks and practices underlying the working definition of a mental condition. Their review examines the methodology involved in such computational approaches, including the social media data sources, in addition to the selection of key variables, including linguistic style, topic content, and status disclosure. Of critical note is the question of ground truth approaches for establishing construct validity, *i.e.* a thorough enquiry into the “theoretical construct of knowledge [pertaining] to the observed phenomenon [of mental health status] within the dataset” [31]. The authors expound on the potential value of “digital psychiatry” in complementing the traditional clinical approaches to mental health treatment; meanwhile, admonishing that “stronger connections to traditional psychiatry” are essential for the purpose of accurate diagnoses.

Whereas Chancellor and De Choudhury give an overview of the practice of using automated predictive techniques on social media to determine mental health status, in addition to the construct validity of these techniques, our forthcoming systematic review will analyze content validity, with respect to the contribution of clinical diagnostic frameworks in computational methods used to identify depression on social media. Of vital note, we consider social media content on depression as a *self-reported* health status indicator. Through the lens of the COSMIN framework [32], we aim to critique the performance of these computational instruments in measuring the clinical construct of depression. Likewise, we will pose a normative judgment about the observed practices in the use of predictive techniques on social media for self-reported depression against current clinical diagnostic norms.

Methods

Objectives

The purpose of this impending systematic review is to present the structures and processes employed in the design of AI methods including ML algorithms for detecting depression on social media. Primarily, this research will focus on whether clinical depression screening instruments are integrated into the design of algorithms for diagnostic assessments of social media user-generated content. In this study, social media content depression will be considered as patient self-reported health outcomes. Using the COSMIN checklist [32], we will appraise the measurement properties of the computational methods with respect to content validity against clinical diagnostic instruments. A flowchart summary of the steps to be followed in the preparation of this systematic review is illustrated in Fig 1.

Protocol and registration

The upcoming systematic review will be prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] guidelines for evidence-based reporting. The review protocol is registered with the University of York’s online international prospective register of systematic reviews (PROSPERO; registration ID CRD42020187874). The authors will document any amendments to the protocol in the PROSPERO record.

Eligibility criteria

Participants. The Pew Research Center [34] has studied the demographic profile of heavy users of social media. Per its findings, young people between the ages of 18 and 29 comprise

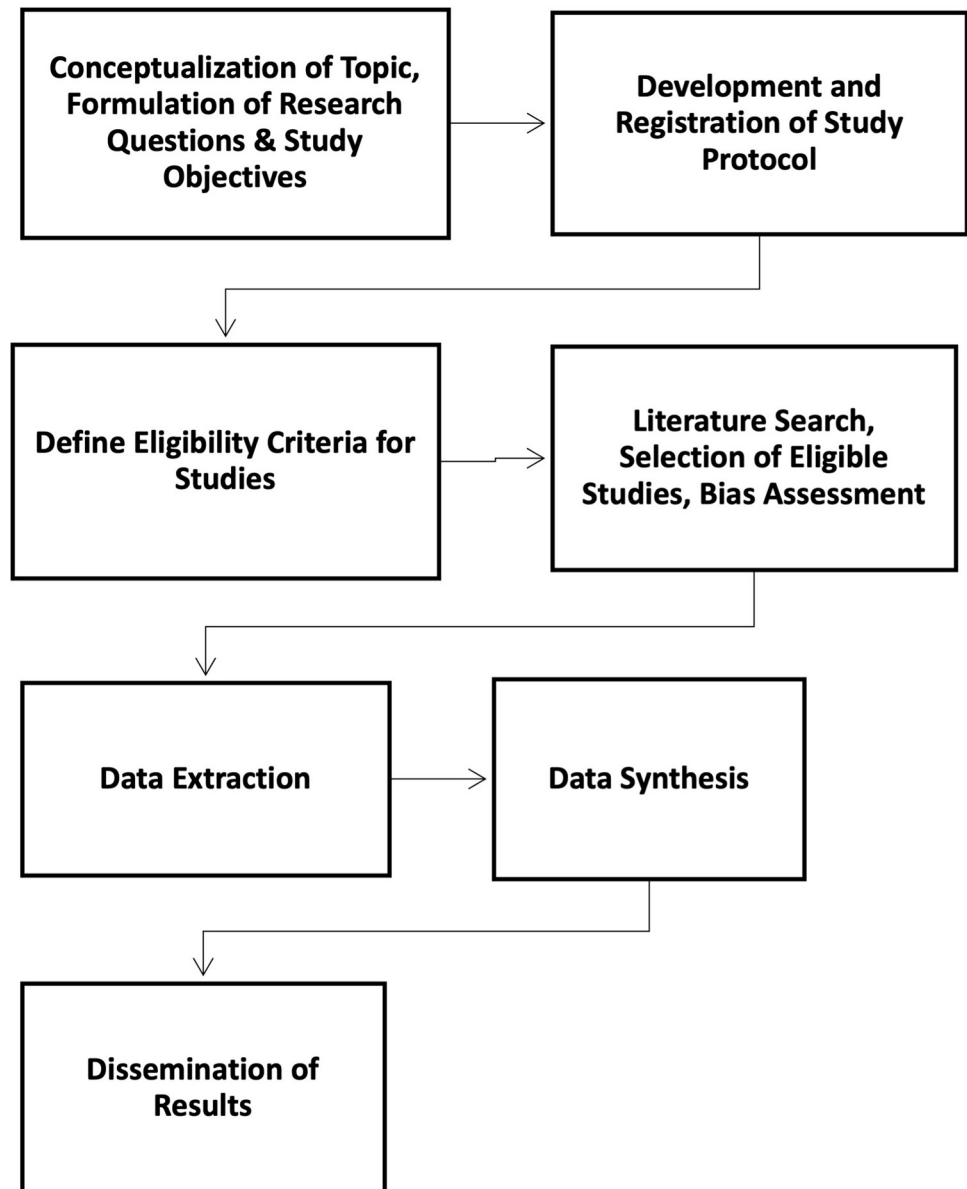


Fig 1.

<https://doi.org/10.1371/journal.pone.0259499.g001>

the highest proportion of adults subscribed to at least one social media site. However, as depression is not only limited to this age demographic, the systematic review will consider reports about male, female and gender diverse social media users worldwide, of all ages. Any existing social media websites will be considered.

Outcomes

Primary outcome.

- Self-reported depression on social media

Secondary outcomes.

- The wide-ranging application of AI in detecting depression on social media websites.
- Demographic profile of social media users screened for depression.
- Distinct subsets of AI methods employed.
- Structures and processes involved in the selection of eligible postings for depression screening.
- Clinical assessments and standards used in identifying depression cases on social media.

Our reviews will report on the absence or presence of clinical frameworks in the design of ML algorithms. Since the review focuses on the clinical diagnosis of depression, we will not consider studies that refer to the broad spectrum of mental health disorders.

Types of studies

The impending systematic review will consider research publications discussing the use of AI and ML in the context of social media for the detection of treatment of depression. We will consider all existing literature from peer reviewed and some gray literature sources published from 2010 to present. Study designs will be inclusive of cross-sectional, cohort, and case-control studies, in addition to randomized controlled trials available in English or German languages. Publications from both high-income and developing country settings will be considered. Gray literature sources will be limited to conference reports. We will not consider newspaper, magazine articles, opinions/commentaries, dissertations, editorials, blog posts, and policy briefs. The number of excluded studies at each stage will be documented and shown in the review manuscript.

Information sources and search strategy. English and German language studies from 2010 to 2020 meeting eligibility criteria will be searched from PubMed, APA PsychInfo, Science Direct, EMBASE Psych, Google Scholar, and Web of Science databases. The review will follow a rigorous search process from both scientific publications and selected gray literature sources. It will be performed in four phases as shown in [Fig 2](#).

In phase 2 of the search process ([Fig 2](#)), the main keywords will be entered into the PICOS [35] tool available in PubMed, where Population is “social media users”; Intervention is “artificial intelligence” or “machine learning” or “Natural Language Processing (NLP)” use for sentiment screening; Comparison is standard “psychiatric diagnostic tools” used in clinical settings; Outcomes are “depression,” “AI applications,” “structures and processes used in social media depression screenings,” and “clinically-derived standards applied in these processes”; Study design is “cross-sectional,” “case-control,” “case series,” “randomized controlled trial.”

Preliminary search keywords to be used are “Artificial intelligence”, “social media”, “depression”, “Artificial intelligence AND social media AND depression”, “Machine learning AND social media AND behavioral health”, “social media AND depression”, “Natural Language Processing (NLP) AND social media AND depression.”

Study selection. After applying search terms and keywords, all identified records from our databases will be exported to the Endnote® research information management system. The total number of records will be noted, with duplicates record numbers deducted from this amount. The selection of eligible studies will involve a preliminary skim through titles and abstracts, with non-relevant records eliminated. This process will be independently conducted by two reviewers on the authorship team, and final selections will be compared. A third reviewer will be consulted to reconcile any disagreements with study selections. Selection processes and tallies will be documented in the PRISMA flow chart.

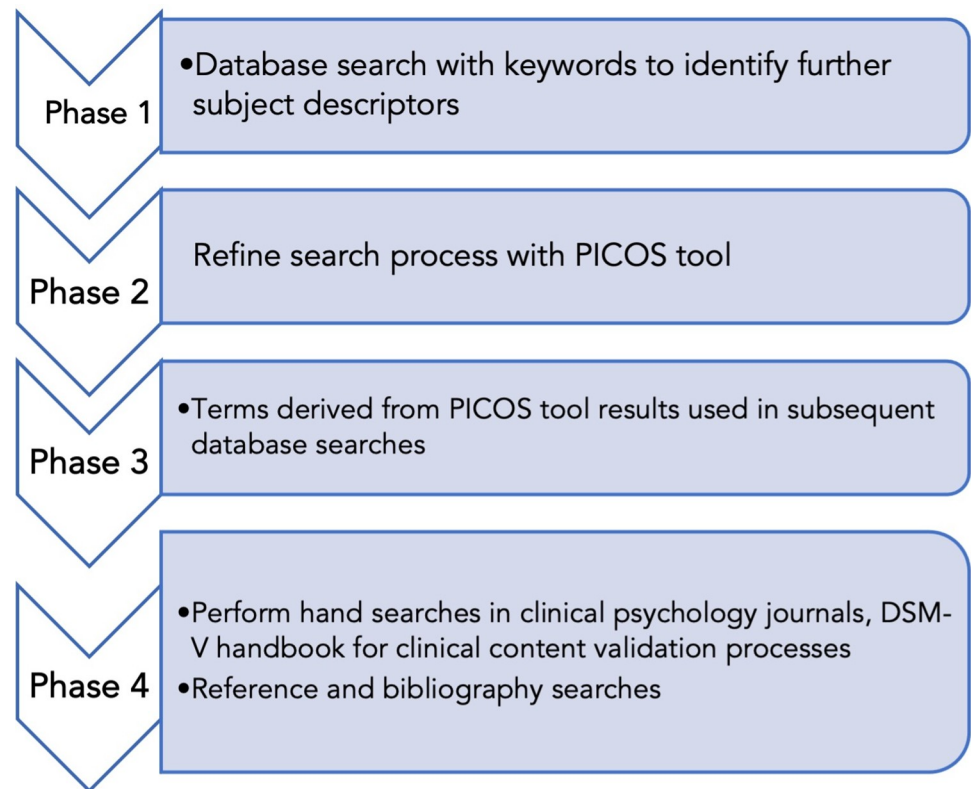


Fig 2. Key phases in literature searches.

<https://doi.org/10.1371/journal.pone.0259499.g002>

Data management and extraction

The data extraction process will be facilitated by the Systematic Review Data Repository (SRDR) web interface offered by the U.S. Department of Health and Human Services Agency for Healthcare Research and Quality (AHRQ). This free online tool allows project managers to define the outcomes of the systematic review, and to import records with information pertaining to the outcomes. This will include demographic characteristics of target population, study type, settings, interventions and comparator studies. The SRDR tool has the capacity to produce graphical and statistical summaries of extracted data. Data extraction will be performed by two independent reviewers, and results discussed with a third reviewer, in case of divergencies.

Risk of bias

Siemieniuk and Guyatt [36] describe bias as occurring when there is the failure to report the complete truth to a phenomenon, due “inherent limitations in the design or conduct of a study.” Therefore, to avoid the risk of bias, publications meeting the eligibility criteria for this systematic review will be evaluated through the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework [37]. The GRADE is a widely used tool to rate the quality of evidence for the desired outcome to be presented in the review. At every stage of the review process, information regarding the reasons for exclusion will be documented. Expert opinion will be sought for publications whose eligibility warrants further debate.

Patient and public involvement

This prospective study will not directly involve either patients or the general public. We will utilize secondary data reporting obtained from published peer review and selected gray literature sources.

Data synthesis

For this report, evidence will be qualitatively synthesized according to themes. Thematic synthesis derives from the grounded theory method used to contextualize health-related data by overarching concepts [37]. Previous work describes this approach as involving iterative collection and analysis of data resulting in the emergence of theories from the evidence [38].

Where possible, meta-analyses will be conducted following considerations about selected publications' study design, reported findings, and outcomes.

In evaluating the content validity of selected studies, we will apply the COSMIN checklist in its binary form for evaluating the properties of measurement instruments for patient-reported health outcomes. It will be accompanied by a four-item checklist (see supplementary information) for the type of evidence presented in the literature, with composite scores out of 20 assigned for each selected publication. Additional quantitative analyses will involve sensitivity analyses to determine robustness of our outcome variables.

Discussion

Mental illness persists as a highly stigmatized condition. The double burden is especially pronounced among ethnic and sexual minority groups, whose self-stigmatization results in feelings of embarrassment, shame, and poor health seeking behavior [39]. Cross-cultural barriers in patient-provider communication also contribute to lower patient satisfaction, misdiagnoses, and poor retention of minorities in mental health care settings [40]. Social media functions as a valuable outlet for individuals with mental illness to find validation, hope and peer-to-peer support to cope with their condition [41, 42]. Digitalized records of user engagement in online peer networks contain rich insight into the peculiar burdens faced by people suffering from mental disability.

On social media, everyday language is effected in describing sentiments, as well as to chronicle life events. User characteristics including social connectedness, demographics, and socio-economic status may be inferred from materials often voluntarily shared by individuals [42]. In some instances, users provide direct evidence of mental illness by self-reporting or through membership in support groups on social media such as reddit's *r/depression*, *r/anxiety*, *r/ADHD*, and various Facebook groups for mental health. Contemporary advances in big data analytics propose the analysis of language use and linguistic styles in social media content to predict ongoing or future incidence of mental illness. For example, De Choudhury and coworkers [43, 44] have that found that prior Facebook postings can serve as indicators of postpartum depression in new mothers. Computational algorithms have been trained to not only identify the presence of mental illness, but also to enable the multiclassification of psychiatric conditions with over 70% accuracy [44].

In 2001, a blueprint for effective depression care management was developed using the template of the Chronic Care Model by Wagner and colleagues [45–50]. This framework begins with the “systematic identification of patients at increased risk for depression,” and recommends the use of an assessment tool that allows patients to be categorized by “episode” as well as “severity.” Within the domain of psychometrics, validity, which is a property of instrument quality for measuring patient-reported health outcomes, is necessary to reduce biased estimates and false diagnoses [50]. Content validity focuses on the instrument's accuracy in

“representing” the clinical domain in question [30, 50]. From the available evidence, the instrument or algorithm performs qualitative or quantitative assessments of the indicators and formulates a conclusion in accordance with established clinical benchmarks. Thus, if AI applications are being implemented as alternative precision tools for depression screening, then the core requirements of this blueprint for effective depression care management can similarly be applied toward the systematic identification and stratification of high-risk individuals.

Our forthcoming systematic review aims to investigate the content validity of these computational methods, taken as the clinical structures and processes involved in identifying social media users exhibiting depression symptoms. We will assess the extent to which the depression screening conducted by newly developed, automated instruments reflect the variables measured in clinical practice. In capturing the parallels between these modern computational methods and gold standard clinical frameworks, we seek to establish the content validity of these AI and ML algorithm tools in social media-based depression screening.

Study limitations

Psychiatric epidemiology [51, 52], comprising population-based studies of mental disorders, employs both self-reports and structured diagnostic interviews [53]. Within this practice, weak concordance between self-reported and standard clinical assessments is frequently observed [53]. In psychiatric practice, a known maxim is that a person’s own “claim of mental impairment is alone not enough. . .” [53–55]. As such, this tenet highlights the inherent limitation of online content, taken as self-reports, as a sufficient data source for population-level inferences of mental health status. Considering the need for objective medical examination to triangulate self-disclosed disability, this systematic review will overlook any observed discrepancies by focusing on the clinical frameworks utilized in validating depressive symptoms.

Ethics, privacy, and dissemination

Ethics scholars deliberate on the dilemmas of privacy and user consent for research purposes in this emerging field of “social computing” [56]. Undeniably, this question falls into the wider theme of the societal impacts of AI in consumer products. As early as the 1980s and 1990s, ethics educators have highlighted the challenge for digital computing applications to responsibly protect marginalized communities from harm through the integration of good ethical practices [56, 57]. Ethical and privacy concerns are to be carefully resolved as social media users increasingly include minors who may not fully grasp the implication of digitally traceable personal information. For academics, we still confront the controversy surrounding the norms and best practices for collecting and disseminating public content for research purposes [58, 59].

This ongoing study is being prepared according to the guidelines of the PRISMA checklist. Information will be derived entirely from secondary source material in peer reviewed literature and selected gray sources, as aforementioned. The completed systematic review including the dataset of literature findings will be published and shared at conferences and proceedings related to clinical psychiatry, big data research, and artificial intelligence methods.

Supporting information

S1 File. Four-item checklist to evaluate systematic review literature evidence.
(PDF)

S2 File. PRISMA 2009 flow diagram.
(DOC)

S3 File. PRISMA-P checklist.
(PDF)

Acknowledgments

This review is supported by resources available through the Heidelberg Institute of Global Health and the Heidelberg University Library.

Author Contributions

Conceptualization: Priscilla N. Owusu, Till Bärnighausen.

Investigation: Priscilla N. Owusu.

Methodology: Priscilla N. Owusu, Ulrich Reininghaus, Georgia Koppe, Till Bärnighausen.

Project administration: Priscilla N. Owusu.

Supervision: Till Bärnighausen.

Visualization: Priscilla N. Owusu.

Writing – original draft: Priscilla N. Owusu.

Writing – review & editing: Ulrich Reininghaus, Georgia Koppe, Irene Dankwa-Mullan, Till Bärnighausen.

References

1. World Health Organization. Depression. [Internet] 2020. Accessed 24 August 2020. Available from <https://www.who.int/news-room/fact-sheets/detail/depression>
2. UNICEF. Ensuring mental health and well-being in an adolescent's formative years can foster a better transition from childhood to adulthood. Accessed 17 August 2020. Available from <https://data.unicef.org/topic/child-health/mental-health/>
3. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. *Asian journal of psychiatry*. 2020 Aug 1; 52:102066. <https://doi.org/10.1016/j.ajp.2020.102066> PMID: 32302935
4. Duan L, Shao X, Wang Y, Huang Y, Miao J, Yang X, et al. An investigation of mental health status of children and adolescents in china during the outbreak of COVID-19. *Journal of affective disorders*. 2020 Oct 1; 275:112–8. <https://doi.org/10.1016/j.jad.2020.06.029> PMID: 32658812
5. Kupferberg A, Bicks L, Hasler G. Social functioning in major depressive disorder. *Neuroscience & Biobehavioral Reviews*. 2016 Oct 1; 69:313–32. <https://doi.org/10.1016/j.neubiorev.2016.07.002> PMID: 27395342
6. Lohoff FW. Overview of the genetics of major depressive disorder. *Current psychiatry reports*. 2010 Dec 1; 12(6):539–46. <https://doi.org/10.1007/s11920-010-0150-6> PMID: 20848240
7. Bank World. Mental Health. Accessed 24 August 2020. Available from <https://www.worldbank.org/en/topic/mental-health>
8. Sheehan DV. Depression: underdiagnosed, undertreated, underappreciated. *Managed care (Langhorne, Pa.)*. 2004 Jun 1; 13(6 Suppl Depression):6–8. PMID: 15293765
9. Schumann I, Schneider A, Kantert C, Löwe B, Linde K. Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies. *Family practice*. 2012 Jun 1; 29(3):255–63. <https://doi.org/10.1093/fampra/cmr092> PMID: 22016322
10. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): systematic review. *Journal of medical internet research*. 2020; 22(3): e13680. <https://doi.org/10.2196/13680> PMID: 32167477
11. Cortland CI, Craig MA, Shapiro JR, Richeson JA, Neel R, Goldstein NJ. Solidarity through shared disadvantage: Highlighting shared experiences of discrimination improves relations between stigmatized groups. *Journal of Personality and Social Psychology*. 2017 Oct; 113(4):547. <https://doi.org/10.1037/pspi0000100> PMID: 28581301

12. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions—evidence from Twitter. *Journal of the American Medical Informatics Association*. 2017 May 1; 24(3):496–502. <https://doi.org/10.1093/jamia/ocw133> PMID: 27707822
13. Escobar-Viera CG, Shensa A, Bowman ND, Sidani JE, Knight J, James AE, et al. Passive and active social media use and depressive symptoms among United States adults. *Cyberpsychology, Behavior, and Social Networking*. 2018 Jul 1; 21(7):437–43. <https://doi.org/10.1089/cyber.2017.0668> PMID: 29995530
14. Gkotsis G, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific reports* 2017(7): 45141.
15. Thorstad R, Wolff P. Predicting future mental illness from social media: A big-data approach. *Behavior research methods*. 2019 Aug; 51(4):1586–600. <https://doi.org/10.3758/s13428-019-01235-z> PMID: 31037606
16. Ernala SK, Birnbaum ML, Candan KA, Rizvi AF, Sterling WA, Kane JM, et al. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. In *Proceedings of the 2019 chi conference on human factors in computing systems* 2019 May 2 (pp. 1–16).
17. Krafft PM, Young M, Katell M, Huang K, Bugingo G. Defining AI in policy versus practice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 2020 Feb 7 (pp. 72–78).
18. Wang P. On defining artificial intelligence. *Journal of Artificial General Intelligence*. 2019 May 1; 10(2):1–37.
19. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 shared task: Depression and PTSD on Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* 2015 (pp. 31–39).
20. Yazdavar AH, Mahdavejad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al. Multimodal mental health analysis in social media. *Plos one*. 2020 Apr 10; 15(4):e0226248. <https://doi.org/10.1371/journal.pone.0226248> PMID: 32275658
21. Ortigo KM, Bradley B, Westen D. An empirically based prototype diagnostic system for DSM-V and ICD-11. *Contemporary directions in psychopathology: scientific foundations of the DSM-V and ICD-11*. New York: Guilford. 2010:374–90. Accessed 24 August 2020. Available from <https://pdfs.semanticscholar.org/8fd9/abff2858d532fec9463cd7d0d02cddc44797.pdf>
22. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*. 2014 Feb; 13(1):28–35. <https://doi.org/10.1002/wps.20087> PMID: 24497240
23. Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, et al. Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Archives of general psychiatry*. 1998 Feb 1; 55(2):109–15. <https://doi.org/10.1001/archpsyc.55.2.109> PMID: 9477922
24. Zirikly A, Resnik P, Uzuner O, Hollingshead K. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* 2019 Jun (pp. 24–33).
25. Harrigan K, Aguirre C, Dredze M. On the state of social media data for mental health research. *arXiv preprint arXiv:2011.05233*. 2020 Nov 10.
26. Shing HC, Nair S, Zirikly A, Friedenberg M, Daumé H, III, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* 2018 Jun (pp. 25–36).
27. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*. 2017 Dec 1; 18:43–9.
28. Jablensky A. Psychiatric classifications: validity and utility. *World Psychiatry*. 2016 Feb; 15(1):26–31. <https://doi.org/10.1002/wps.20284> PMID: 26833601
29. Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. *American journal of psychiatry*. 2003 Jan 1; 160(1):4–12. <https://doi.org/10.1176/appi.ajp.160.1.4> PMID: 12505793
30. Zamanzadeh Vahid, Ghahramanian Akram, Rassouli Maryam, Abbaszadeh Abbas, Hamid Alavi-Majd, and Ali-Reza Nikanfar. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of caring sciences* 4, no. 2 2015: 165. <https://doi.org/10.15171/jcs.2015.017> PMID: 26161370
31. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*. 2020 Mar 24; 3(1):1–1. <https://doi.org/10.1038/s41746-020-0233-7> PMID: 32219184

32. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research*. 2010 May; 19(4):539–49. <https://doi.org/10.1007/s11136-010-9606-8> PMID: 20169472
33. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Review*. 2015; 4(1):1. <https://doi.org/10.1186/2046-4053-4-1> PMID: 25554246
34. Pew Research Center, Internet and Technology. Social media fact sheet. Accessed 21 April 2020. Available from <https://www.pewresearch.org/internet/fact-sheet/social-media/>
35. Robinson KA, Saldanha IJ, Mckoy NA. Development of a framework to identify research gaps from systematic reviews. *Journal of clinical epidemiology*. 2011 Dec 1; 64(12):1325–30. <https://doi.org/10.1016/j.jclinepi.2011.06.009> PMID: 21937195
36. Siemieniuk R, Guyatt G. What Is GRADE? | BMJ Best Practice. Accessed 21 April 2020. Available from <https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/>.
37. Barnett-Page Elaine, and Thomas James. Methods for the synthesis of qualitative research: a critical review. *BMC medical research methodology* 9.1 2009: 1–11. <https://doi.org/10.1186/1471-2288-9-59> PMID: 19671152
38. Glaser BG, Strauss AL: *The Discovery of Grounded Theory: Strategies for Qualitative Research* New York: Aldine De Gruyter; 1967.
39. Corrigan PW, Druss BG, Perlick DA. The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest*. 2014 Oct; 15(2):37–70. <https://doi.org/10.1177/1529100614531398> PMID: 26171956
40. Alegría M, Roter DL, Valentine A, Chen CN, Li X, Lin J, et al. Patient–clinician ethnic concordance and communication in mental health intake visits. *Patient education and counseling*. 2013 Nov 1; 93(2):188–96. <https://doi.org/10.1016/j.pec.2013.07.001> PMID: 23896127
41. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. *Epidemiology and psychiatric sciences*. 2016 Apr; 25(2):113–22. <https://doi.org/10.1017/S2045796015001067> PMID: 26744309
42. Jaidka K, Giorgi S, Schwartz HA, Kern ML, Ungar LH, Eichstaedt JC. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*. 2020 May 12; 117(19):10165–71 <https://doi.org/10.1073/pnas.1906364117> PMID: 32341156
43. De Choudhury M, Counts S, Horvitz EJ, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* 2014 Feb 15 (pp. 626–638).
44. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI conference on human factors in computing systems* 2013 Apr 27 (pp. 3267–3276).
45. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. *The Milbank Quarterly*. 1996 Jan 1:511–44. PMID: 8941260
46. Wagner EH, Davis C, Schaefer J, Von Korff M, Austin B. A survey of leading chronic disease management programs: are they consistent with the literature?. *Managed care quarterly*. 1999 Jan 1; 7(3):56–66. PMID: 10620960
47. Pincus HA, Pechura CM, Elinson L, Pettit AR. Depression in primary care: linking clinical and systems strategies. *General hospital psychiatry*. 2001 Nov 1; 23(6):311–8. [https://doi.org/10.1016/s0163-8343\(01\)00165-7](https://doi.org/10.1016/s0163-8343(01)00165-7) PMID: 11738461
48. Belnap BH, Kuebler J, Upshur C, Kerber K, Mockrin DR, Kilbourne AM, et al. Challenges of implementing depression care management in the primary care setting. *Administration and Policy in Mental Health and Mental Health Services Research*. 2006 Jan; 33(1):65–75. <https://doi.org/10.1007/s10488-005-4237-z> PMID: 16215660
49. Jani BD, Purves D, Barry S, Cavanagh J, McLean G, Mair FS. Challenges and implications of routine depression screening for depression in chronic disease and multimorbidity: a cross sectional study. *Plos one*. 2013 Sep 13; 8(9):e74610. <https://doi.org/10.1371/journal.pone.0074610> PMID: 24058602
50. National Institute of Health and Clinical Excellence Depression in people with a chronic physical health problem. CG91. 2009. Accessed; 18 Feb 2020. Available: <http://www.nice.org.uk/guidance/index.jsp?action=download&o=45914>.
51. Weissman MM. Advances in psychiatric epidemiology: rates and risks for major depression. *American journal of public health*. 1987 Apr; 77(4):445–51. <https://doi.org/10.2105/ajph.77.4.445> PMID: 3826462

52. Susser E, Schwartz S, Morabia A, Bromet EJ. *Psychiatric epidemiology: searching for the causes of mental disorders*. Oxford University Press; 2006 Jun 1.
53. Eaton WW, Neufeld K, Chen LS, Cai G. A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. *Arch Gen Psychiatry*. 2000 Mar; 57(3):217–22. <https://doi.org/10.1001/archpsyc.57.3.217> PMID: 10711906
54. Hanssen M, Bak M, Bijl R, Vollebergh W, Van Os J. The incidence and outcome of subclinical psychotic experiences in the general population. *British Journal of Clinical Psychology*. 2005 Jun; 44(2):181–91. <https://doi.org/10.1348/014466505X29611> PMID: 16004653
55. Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations; Board on the Health of Select Populations; Institute of Medicine. *Psychological Testing in the Service of Disability Determination*. Washington (DC): National Academies Press (US); 2015 Jun 29. 4, Self-Report Measures and Symptom Validity Tests. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK305237/>
56. Garrett N, Beard N, Fiesler C. More Than "If Time Allows" The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 2020 Feb 7 (pp. 272–278).
57. Saltz J, Skirpan M, Fiesler C, Gorelick M, Yeh T, Heckman R, et al. Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)*. 2019 Aug 2; 19(4):1–26.
58. Vitak J, Shilton K, Ashktorab Z. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* 2016 Feb 27 (pp. 941–953).
59. Buchanan E.A. and Zimmer M., 2012. Internet research ethics. Accessed on August 20, 2021. Available from <https://plato.stanford.edu/entries/ethics-internet-research/>