

Letter to the Editor

Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on “Machine Learning for Mental Health in Social Media: Bibliometric Study”

Philip Resnik¹, PhD; Munmun De Choudhury², PhD; Katherine Musacchio Schafer³, MEd, MS; Glen Coppersmith⁴, PhD

¹Department of Linguistics and Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States

²School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

³Department of Psychology, Florida State University, Tallahassee, FL, United States

⁴Qntfy, Boston, MA, United States

Corresponding Author:

Philip Resnik, PhD

Department of Linguistics and Institute for Advanced Computer Studies

University of Maryland

1401 Marie Mount Hall

College Park, MD, 20814

United States

Phone: 1 301 405 7002

Email: resnik@umd.edu

Related Articles:

Comment on: <https://www.jmir.org/2021/3/e24870/>

Comment in: <https://www.jmir.org/2021/6/e29549>

(*J Med Internet Res* 2021;23(6):e28990) doi: [10.2196/28990](https://doi.org/10.2196/28990)

KEYWORDS

bibliometric analysis; machine learning; mental health; social media; bibliometrics

Bibliometric studies like the recent article by Kim et al [1] in the *Journal of Medical Internet Research* play an essential part in understanding the evolution of emerging, fast-moving research on machine learning for mental health in social media. However, the intended value of this paper's contribution is tempered by some important lessons it teaches us about the current state of research on this topic.

The first key lesson is that computationally oriented research on mental health remains highly fragmented. Notably, variants on the cover term “mental health” are included in the illustrative search query but, crucially, “clinical psychology” and “psychiatry” are not. The terminological difference here reflects a prevailing technological focus often separated from clinical research and even more distant from clinical practice. Kim et al [1] do discuss a trend toward clinically validated self-report questionnaires to gather clinically relevant information. However, the review's overall approach, from the search terms to the keyword analysis, simultaneously reflects and reinforces a widespread technological disregard for basic considerations in clinical psychology and psychiatry, such as the distinction between the symptoms of the disorders versus the disorders themselves. As technologists, we are often happy just to get our hands on enough data to work with. However, real progress

toward solving these important problems demands a more careful definition of the actual mental health constructs under investigation and greater attention to the question of validity [2,3], with research questions and experimental choices guided by knowledge of the subject domain.

Second, the inclusion terms reflect a widespread narrow focus on methods, such as “neural network” and “hybrid intelligent system,” rather than the problems for which those methods are contributing solutions, such as “screening,” “risk assessment,” or “monitoring.” Even the cover term “natural language processing” focuses narrowly on engineering versus “computational linguistics” as a parent scientific discipline. Further lacking in the methodology-centric perspective are searches based on theoretical frameworks (which guide research, treatment, and intervention) or DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th Edition) diagnoses (eg, major depressive disorder or persistent depressive disorder versus “depression,” which is not a diagnosis). The review reflects and reinforces a general tendency to frame machine learning research in terms of technical “tasks” rather than connecting them more directly with real-world problems, a necessary step toward translating technological progress into

the broader mental health ecosystem within which the technology will ultimately need to be situated [4,5].

Third, the bibliometric approach taken here reflects a traditional top-down view that fails to break down information silos in a rapidly evolving field. It is now standard to cast the net more broadly by searching for citations in resources like Google Scholar and/or looking at papers' references (cf Franklin et al [6]), and then narrow using exclusion criteria. Such practices can illuminate the wider space of relevant search terms and sources—for example, the notable absence of suicidality here among mental conditions, at least in the illustrative search—and uncover unexpected connections. Even within the most rigorous meta-analysis frameworks (Moher et al [7]), studies can miss “gray literature” (eg, conference proceedings, preprints, collected data that have never been analyzed, presented on, or published). For example, the substantially similar prior study by Chancellor and De Choudhury [3] needed to adjust for the limitations of indexing services, which had large gaps for

conferences known to be important in this research area (eg, Association for the Advancement of Artificial Intelligence [AAAI], Association for Computational Linguistics [ACL], Association for Computing Machinery [ACM], Neural Information Processing Systems [NIPS/NeurIPS], American Medical Informatics Association [AMIA])—they were careful in particular to include the Workshop on Computational Linguistics and Clinical Psychology (CLPsych), a key interdisciplinary publication venue for natural language processing, machine learning, and mental health since 2014.

Kim et al [1] are to be commended for undertaking a bibliometric study with the goal of advancing our understanding of machine learning for mental health in social media. However, we would encourage thinking about their article as a different kind of contribution, even if not the intended one: it is an opportunity to draw attention to an increasing need, as the field grows, to approach this research space not only as technologists, but also as partners with clinical researchers and clinicians.

Conflicts of Interest

GC is a stockholder and employee of Qntfy.

References

1. Kim J, Lee D, Park E. Machine Learning for Mental Health in Social Media: Bibliometric Study. *J Med Internet Res* 2021 Mar 08;23(3):e24870 [FREE Full text] [doi: [10.2196/24870](https://doi.org/10.2196/24870)] [Medline: [33683209](https://pubmed.ncbi.nlm.nih.gov/33683209/)]
2. Ernala SK, Birnbaum ML, Michael L, Candan KA, Rizvi AF, Sterling WA, 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. New York, NY: Association for Computing Machinery; 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9; Glasgow, Scotland, UK p. 1-16. [doi: [10.1145/3290605.3300364](https://doi.org/10.1145/3290605.3300364)]
3. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digit Med* 2020;3:43 [FREE Full text] [doi: [10.1038/s41746-020-0233-7](https://doi.org/10.1038/s41746-020-0233-7)] [Medline: [32219184](https://pubmed.ncbi.nlm.nih.gov/32219184/)]
4. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, et al. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2021 Feb 08. [doi: [10.1016/j.bpsc.2021.02.001](https://doi.org/10.1016/j.bpsc.2021.02.001)] [Medline: [33571718](https://pubmed.ncbi.nlm.nih.gov/33571718/)]
5. Resnik P, Foreman A, Kuchuk M, Musacchio Schafer K, Pinkham B. Naturally occurring language as a source of evidence in suicide prevention. *Suicide Life Threat Behav* 2021 Feb;51(1):88-96. [doi: [10.1111/sltb.12674](https://doi.org/10.1111/sltb.12674)] [Medline: [32914479](https://pubmed.ncbi.nlm.nih.gov/32914479/)]
6. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol Bull* 2017 Feb;143(2):187-232. [doi: [10.1037/bul0000084](https://doi.org/10.1037/bul0000084)] [Medline: [27841450](https://pubmed.ncbi.nlm.nih.gov/27841450/)]
7. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]

Abbreviations

- AAAI:** Association for the Advancement of Artificial Intelligence
ACL: Association for Computational Linguistics
ACM: Association for Computing Machinery
AMIA: American Medical Informatics Association
CLPsych: Computational Linguistics and Clinical Psychology
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
NIPS/NeurIPS: Neural Information Processing Systems

Edited by T Derrick; this is a non-peer-reviewed article. Submitted 21.03.21; accepted 13.05.21; published 17.06.21.

Please cite as:

Resnik P, De Choudhury M, Musacchio Schafer K, Coppersmith G

Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on “Machine Learning for Mental Health in Social Media: Bibliometric Study”

J Med Internet Res 2021;23(6):e28990

URL: <https://www.jmir.org/2021/6/e28990>

doi: [10.2196/28990](https://doi.org/10.2196/28990)

PMID:

©Philip Resnik, Munmun De Choudhury, Katherine Musacchio Schafer, Glen Coppersmith. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 17.06.2021. 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 work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.