



## OPEN ACCESS

## EDITED BY

Takashi Nakano,  
Fujita Health University, Japan

## REVIEWED BY

Alexander J. Smith,  
University of Bern, Switzerland  
Beate Bokhof,  
Governmental healthy cites network  
Germany, Germany

## \*CORRESPONDENCE

Venkat Bhat

✉ venkat.bhat@utoronto.ca

RECEIVED 05 December 2024

ACCEPTED 23 January 2025

PUBLISHED 18 February 2025

## CITATION

Perivolaris A, Rueda A, Parkington K,  
Soni A, Rambhatla S, Samavi R, Jetly R,  
Greenshaw A, Zhang Y, Cao B, Krishnan S  
and Bhat V (2025) Opinion: Mental health  
research: to augment or not to augment.  
*Front. Psychiatry* 16:1539157.  
doi: 10.3389/fpsy.2025.1539157

## COPYRIGHT

© 2025 Perivolaris, Rueda, Parkington, Soni,  
Rambhatla, Samavi, Jetly, Greenshaw, Zhang,  
Cao, Krishnan and Bhat. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Opinion: Mental health research: to augment or not to augment

Argyrios Perivolaris<sup>1</sup>, Alice Rueda<sup>1</sup>, Karisa Parkington<sup>1</sup>,  
Achint Soni<sup>2</sup>, Sirisha Rambhatla<sup>2,3</sup>, Reza Samavi<sup>4</sup>, Rakesh Jetly<sup>5</sup>,  
Andrew Greenshaw<sup>6</sup>, Yanbo Zhang<sup>6</sup>, Bo Cao<sup>6,7,8</sup>, Sri Krishnan<sup>9</sup>  
and Venkat Bhat<sup>1,10\*</sup>

<sup>1</sup>Interventional Psychiatry Program, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada, <sup>2</sup>Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada, <sup>3</sup>Department of Management Science & Engineering, University of Waterloo, Waterloo, ON, Canada, <sup>4</sup>Vector Institute for Artificial Intelligence, Toronto, ON, Canada, <sup>5</sup>Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada, <sup>6</sup>Department of Psychiatry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada, <sup>7</sup>Ministry of Health, Government of Alberta, Edmonton, AB, Canada, <sup>8</sup>Department of Computing Science, University of Alberta, Edmonton, AB, Canada, <sup>9</sup>Department of Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada, <sup>10</sup>Department of Psychiatry, University of Toronto, Toronto, ON, Canada

## KEYWORDS

augmentation, mental health, data scarcity, synthetic data, artificial intelligence

## Introduction

In recent years, there has been significant growth in the use of artificial intelligence (AI) and machine learning (ML) in healthcare. AI-based tools are increasingly used to predict diagnoses, personalize treatment plans, and assess risk factors, aiming to enable more scalable mental health care solutions. However, mental health research is often limited by the availability of high-quality and large sample datasets and confounded by the multifaceted complexities of human behaviors and emotions (1). To address this gap, researchers have begun to utilize data augmentation techniques to expand available datasets. Generating new data artificially enables models to use larger and more complete training datasets. Consider the medical imaging field, where AI has become a prominent fixture in practice. Data augmentation has demonstrated benefits across all organs and modalities to help promote medical imaging training without investing time and resources into collecting new samples (2). However, mental health research presents many unique barriers to integrating data augmentation. Biases inherent in the original set of mental health data remain and can result in overfitting where a model is unable to make accurate predictions from any other data other than the training data. This article explores the unique challenges researchers must overcome due to the lack of representative mental health data and how these challenges interact with AI and ML advancements. We explore data augmentation as a tool to bridge this gap, offering an integrative perspective on the ethical and practical challenges. As researchers consider data augmentation in mental health research, it is critical to evaluate the promise through rigorous methodologies and research and decide whether 'to augment or not to augment'.

## What is data augmentation?

As AI and ML algorithms have advanced exponentially in recent years, one of the most prominent limiting factors remains the availability of representative training data that determines model performance (3). In contrast to synthetic data that creates data from scratch, data augmentation is an ML technique used to create new data based on existing data points, thereby artificially expanding a dataset. There are several data augmentation methods, with some incorporating simple transformations to text data (rotating images by random degrees, flipping images horizontally, and back-translation of data to a new language). Expanding on this, generative adversarial network (GAN) based augmentation is a more sophisticated strategy that uses neural networks to create novel samples from a pre-existing dataset. For example, GANs can augment data for chest X-rays that not only improve classification accuracy but perform better than other simple transformation methods (4). Large language models (LLMs), such as GPT-4o, have also been used for clinical transcript data augmentation (5). With various strategies available, data augmentation can be a potential tool for all fields of medical research moving forward. Augmentation can address class imbalance while preserving anonymity, facilitating cross-lingual and robust mental health research with available data. While data augmentation may enhance model generalizability and facilitate new research, mental health research introduces concerns about augmentation because of unique challenges in balancing realism and mitigating biases.

## To augment: overcoming data scarcity

Data scarcity remains a significant challenge in mental health research. Unlike other areas of medicine that can evaluate objective data from available biomarkers and imaging, mental health research relies on qualitative interviews, self-reported surveys, questionnaires, and clinical notes. The subjective nature of mental health concepts, such as emotional well-being (6), also makes developing universally accepted definitions challenging. Despite self-reported measurements being cost-efficient, flexible, and valuable for uncovering personal perceptions (7), many datasets do not provide the comprehensive, diverse, and sufficient data necessary for generalizable and reliable research. Furthermore, data collection is hindered by high costs, privacy concerns, stigma, and recruitment difficulties.

Augmented data presents a promising opportunity to address these issues. By artificially generating new data, such as augmented text or audio, researchers can increase usable data, mitigate the concerns of dependency on subjective reports of experiences, and enhance the scalability of mental health studies (8). Data augmentation is a cost-effective alternative to collecting new clinical data, reducing the reliance on expensive longitudinal studies. By using augmented data, researchers may limit the reliance on personally identifiable information, enhancing privacy protection. As well, researchers can instead focus efforts on

generating new insights and testing hypotheses using readily available datasets.

Mental health datasets are often highly imbalanced, with certain conditions underrepresented (e.g., borderline personality disorder) compared to others (e.g., depression), and gender disparities in diagnosis, treatment, and research. Rare mental health disorders can present with uncommon symptoms, which can complicate diagnoses (9). Moreover, certain populations—such as children, seniors, racial minorities, LGBTQ2+, and marginalized groups—are also underrepresented in datasets. This imbalance can lead to biased conclusions and unreliable predictive models, which can perpetuate disparities and further marginalize underserved populations.

Addressing these issues, data augmentation can create more balanced datasets by artificially increasing the representation of minority classes, allowing ML models to better detect and treat underrepresented conditions and populations. AI-generated data can impute missing information and ensure datasets are more diverse, leading to more inclusive and equitable models. For instance, psychiatric symptoms often manifest differently across age groups and genders, with adolescents and adults experiencing distinct presentations of similar conditions (10). Augmented data allows for a better representation of subgroups, which can enhance diagnostic accuracy and treatment outcomes.

Augmentation is also crucial in scaling AI models. Introducing synthetic variations, such as noise injection, makes models more robust and less prone to overfitting. This increased variability enables models to learn general patterns rather than memorizing specific instances, thus improving their generalizability across different patient groups. This is particularly beneficial in mental health research, where there is significant variability in behavior and emotions. For example, consider a research team studying depressive disorders in a population skewed towards high symptom severity levels. An ML model trained on this real-world data may not predict accurate outcomes when applied to patient groups with lower symptom severity levels (11). However, researchers can achieve more accurate and generalizable predictions by generating augmented data that mimics these underrepresented cases (12). Incorporating data augmentation could improve research and clinical practice outcomes, allowing decision-support tools to be developed, and offering more equitable recommendations.

## Not to augment: bias and clinician fidelity

Mental health data is nuanced and profoundly contextual, with small variations in symptoms or patient perceptions potentially leading to different clinical outcomes. This type of data has multifactorial and complicated biological, psychological, and social components. One of the most significant risks of data augmentation is replicating and potentially amplifying biases present in the original datasets. If the original dataset underrepresents certain cultural, gender, or ethnic groups, these biases may be further embedded into the model. Poorly designed augmented data, if not inspected by

mental health professionals, may fail to respect the nuanced interplay of different symptomatology and can mistakenly intensify biases present in the original dataset, which may also introduce new biases (13, 14). In mental health research, historical biases regarding race, gender, and socioeconomic status are well-documented and must be mitigated (15).

Creating augmented data may risk the loss of meaning, especially when nuanced cultural and individual differences are simplified. This could lead to generalized stereotypes or poor representation of complex mental health experiences (16). Augmented data may fail to consider the complexity of identities intersecting such factors, which may result in inaccurate predictions, leading to inconsistencies in treatment recommendations. This is especially problematic in mental health, where symptoms and coping mechanisms can vary greatly across cultures due to differences in language, values, and stigma around mental illness. Augmented data generated without consideration of cultural contexts might promote the development of AI models that misinterpret the mental health challenges of underrepresented populations. Traditional augmentation techniques may treat diverse groups as homogeneous, reducing cultural and ethnic variability to a few representative data points, thus risking generalization and misrepresentation.

Moreover, augmented data may lack clinical expertise and the ability to reproduce real-world patient behavior and presentation (17). Augmented data may oversimplify the variability that clinicians rely on for diagnosis and treatment. Adding synthetic noise or random data augmentation may alter key data features, causing a loss of context crucial to understanding mental health conditions. For example, AI-generated text transcripts of patient interviews might lack the subtle linguistic cues and emotional context necessary for a clinician's judgment (18). This disconnect could result in models that appear highly accurate in theory but fail to translate into reliable real-world clinical support. Evaluating the quality of augmented data is particularly challenging in mental health due to the subjective nature of psychological assessments and a lack of consistent validation benchmarks.

## Discussion: harmonizing novelty with caution

Given the different perspectives in this argument, how should mental health research proceed with augmented data? The key is cautious optimism. While augmented data should not be dismissed outright, it must be integrated with real-world data in a way that preserves transparency and mitigates bias.

One approach is to utilize augmented data to supplement rather than replace real-world data. Combining both traditional and augmented datasets can enhance the dataset diversity without over-relying on synthetic information. Models trained on augmented data should also be evaluated by mental health professionals with standard accuracy metrics and qualitative appraisals. This will help ensure that augmented data-driven predictions align with clinical expertise and judgment. To integrate augmented data into mental health research effectively, researchers should prioritize pilot and feasibility studies to

assess its practicality and ensure alignment with clinical expertise. Collaborative efforts based on these findings can also address challenges related to bias, equity, and implementation.

Societal and cultural norms heavily influence how mental health symptoms are expressed, understood, and treated. For example, some cultures may emphasize physical symptoms like headaches, while others focus on emotional or behavioural aspects (19). Data augmentation preserves the distributional properties of the original dataset, including those with small sample sizes, imbalances, or underrepresented features. By enhancing the diversity and representation within the dataset, models trained on well-augmented data are more likely to generalize effectively and exhibit reduced bias. Importantly, if cultural nuances are present in the original data but captured unevenly, data augmentation can help balance representation. This allows the model to better generalize across cultural subtleties, improving its fairness and applicability. Incorporating ethnographic insights or consulting cultural experts during data creation can further improve augmented data's realism and applicability (20).

The financial implications of data augmentation are an important consideration in promoting global health equity. Researchers in wealthier regions often have greater access to the tools and funding, potentially exacerbating inequalities (21). In contrast, researchers in underserved areas may face significant barriers to adopting these technologies. Open Science initiatives could help promote the sharing of augmented datasets and tools, enabling broader access (21, 22). Publicly available platforms can democratize research opportunities, while transparency protocols requiring researchers to disclose their augmentation methods could foster collaboration and reduce disparities. By addressing these financial and equity concerns, the benefits of augmented data can be distributed more equitably across research communities (22).

Finally, the ethical implications of using augmented data in mental health research must not be overlooked. Augmented data can mitigate privacy concerns however, generating realistic patient data raises questions about consent and transparency. In addition, ethicists will need to develop clear guidelines for using augmented data in healthcare AI models that align with clinicians' preferred practices and optimize patient confidentiality. Frameworks that promote positive clinician-AI interactions can ensure that AI data-driven models undertake the same rigorous inspection as models based on real-world data and be successfully implemented in clinical settings (23, 24).

## Conclusion

The use of augmented data in mental health research is an exciting frontier, contributing to the potential to overcome long-standing challenges of data scarcity and imbalance. Data augmentation has been demonstrated to be a useful tool in other medical fields, such as medical imaging. However, introducing augmented data into the mental health field must be handled with caution. While the promise of enhanced model performance and data diversity is desirable, the risks of bias, unreliability, and ethical concerns may limit feasibility.

## Author contributions

AP: Conceptualization, Writing – original draft, Writing – review & editing. AR: Conceptualization, Writing – review & editing. KP: Writing – review & editing. AS: Writing – review & editing. SR: Writing – review & editing. RS: Writing – review & editing. RJ: Writing – review & editing. AG: Writing – review & editing. YZ: Writing – review & editing. BC: Writing – review & editing. SK: Conceptualization, Writing – review & editing. VB: Conceptualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. KP is supported by a CIHR Post-doctoral Fellowship (2024 - 2026). VB is supported by an Academic Scholar Award from the University of Toronto Department of Psychiatry and has received research funding from the Canadian Institutes of Health Research, Brain & Behavior Foundation, Ontario Ministry of Health Innovation Funds, Royal College of Physicians and Surgeons of Canada, Department of National Defence (Government of Canada), New Frontiers in Research Fund, Associated Medical Services Inc. Healthcare, American Foundation for Suicide Prevention, Roche Canada, Novartis, and Eisai. The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

## Acknowledgments

We extend our gratitude to the researchers and clinicians whose foundational work on data augmentation and mental health

inspired this commentary. Special thanks to the Interventional Psychiatry Program lab members at St. Michael's Hospital for their valuable insights and feedback.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Pratap A, Homiar A, Waninger L, Herd C, Suver C, Volponi J, et al. Real-world behavioral dataset from two fully remote smartphone-based randomized clinical trials for depression. *Sci Data*. (2022) 9:522. doi: 10.1038/s41597-022-01633-7
- Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: A systematic literature review. *Comput Biol Med*. (2023) 152:106391. doi: 10.1016/j.compbiomed.2022.106391
- Mumuni A, Mumuni F. Data augmentation: A comprehensive survey of modern approaches. *Array*. (2022) 16:100258. doi: 10.1016/j.array.2022.100258
- Motamed S, Rogalla P, Khalvati F. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Inform Med Unlocked*. (2021) 27:100779. doi: 10.1016/j.imu.2021.100779
- Wu Y, Mao K, Zhang Y, Chen J. CALLM: enhancing clinical interview analysis through data augmentation with large language models. *IEEE J BioMed Health Inform*. (2024). doi: 10.1109/JBHI.2024.3435085
- Koslowski JB, Wilson-Mendenhall CD, Parsafar P, Goldberg S, Martin MY, Chafouleas SM. Measuring emotional well-being through subjective report: a scoping review of reviews. *BMJ Open*. (2022) 12:e062120. doi: 10.1136/bmjopen-2022-062120
- Kormos C, Gifford R. The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *J Environ Psychol*. (2014) 40:359–71. doi: 10.1016/j.jenvp.2014.09.003
- Harley JM. Chapter 5 - measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In: *Emotions, Technology, Design, and Learning*. United States: Academic Press (2016). p. 89–114. doi: 10.1016/B978-0-12-801856-9.00005-0
- Schaefer JD, Caspi A, Belsky DW, Harrington H, Houts R, Horwood LJ, et al. Enduring mental health: Prevalence and prediction. *J Abnorm Psychol*. (2017) 126:212–24. doi: 10.1037/abn0000232
- Rice F, Riglin L, Lomax T, Souter E, Potter R, Smith DJ, et al. Adolescent and adult differences in major depression symptom profiles. *J Affect Disord*. (2019) 243:175–81. doi: 10.1016/j.jad.2018.09.015
- Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns (N Y)*. (2021) 2:100347. doi: 10.1016/j.patter.2021.100347
- Faryna K, van der Laak J, Litjens G. Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology. *Comput Biol Med*. (2024) 170:108018. doi: 10.1016/j.compbiomed.2024.108018
- Hall M, van der Maaten L, Gustafson L, Jones M, Adcock A. A systematic study of bias amplification. *arXiv*. (2022). doi: 10.48550/arXiv.2201.11706
- Iosifidis V, Ntoutsis E. Dealing with bias via data augmentation in supervised learning scenarios. *Jo Bates Paul D. Clough Robert Jäschke*. (2018) 24.
- Snowden LR. Bias in mental health assessment and intervention: theory and evidence. *Am J Public Health*. (2003) 93:239–43. doi: 10.2105/ajph.93.2.239
- Hovy D, Spruit S. (2016). The social impact of natural language processing, in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Vol. 2. pp. 591–8. doi: 10.18653/v1/P16-2096
- Moodley K. Artificial intelligence (AI) or augmented intelligence? How big data and AI are transforming healthcare: Challenges and opportunities. *South Afr Med J = Suid-Afrikaanse tydskrif vir geneeskunde*. (2023) 114:22–6. doi: 10.7196/SAMJ.2024.v114i1.1631
- Sönmez YÜ, Varol A. In-depth investigation of speech emotion recognition studies from past to present –The importance of emotion recognition from speech signal for AI. *Intelligent Syst Applications*. (2024) 22:200351. doi: 10.1016/j.iswa.2024.200351

19. Jimenez DE, Bartels SJ, Cardenas V, Dhaliwal SS, Alegria M. Cultural beliefs and mental health treatment preferences of ethnically diverse older adult consumers in primary care. *Am J Geriatr Psychiatry*. (2012) 20:533–42. doi: 10.1097/JGP.0b013e318227f876
20. de Seta G, Pohjonen M, Knuutila A. Synthetic ethnography: Field devices for the qualitative study of generative models. *Big Data Society*. (2024). doi: 10.1177/2053951724130307
21. Liebreiz M, Bhugra D, Alibudbud R, Ventriglio A, Smith A. AI in health care and the fragile pursuit of equity and social justice. *Lancet*. (2024) 404:843. doi: 10.1016/S0140-6736(24)01604-0
22. Rubeis G, Dubbala K, Metzler I. Democratizing” artificial intelligence in medicine and healthcare: Mapping the uses of an elusive term. *Front Genet*. (2022) 13:902542. doi: 10.3389/fgene.2022.902542
23. Tikhomirov L, Semmler C, McCradden M, Searston R, Ghassemi M, Oakden-Rayner L. Medical artificial intelligence for clinicians: the lost cognitive perspective. *Lancet Digit Health*. (2024) 6:e589–94. doi: 10.1016/S2589-7500(24)00095-5
24. Perivolaris A, Adams-McGavin C, Madan Y, Kishibe T, Antoniou T, Mamdani M, et al. Quality of interaction between clinicians and artificial intelligence systems. A systematic review. *Future Healthc J*. (2024) 11(3):100172. doi: 10.1016/j.fhj.2024.100172