

REVIEW

Algorithmic biases in mental health diagnoses and their impact on vulnerable populations: a documentary review of advances and challenges

Sesgos algorítmicos en diagnósticos de salud mental y su impacto en poblaciones vulnerables: una revisión documental de avances y desafíos

Ariadna Matos Matos¹  

¹Universidad Técnica de Ambato. Facultad de Ciencias de la Salud, Carrera de Licenciatura en Enfermería. Latacunga, Ecuador.

Cite as: Matos Matos A. Algorithmic biases in mental health diagnoses and their impact on vulnerable populations: a documentary review of advances and challenges. EthAlca. 2022; 1:20. <https://doi.org/10.56294/ai202220>


Submitted: 10-01-2022

Revised: 22-03-2022

Accepted: 15-05-2022

Published: 16-05-2022

Editor: PhD. Rubén González Vallejo 

Corresponding author: Ariadna Matos Matos 

ABSTRACT

Algorithmic biases in mental health diagnostic systems represent a critical challenge, particularly for vulnerable populations, as they perpetuate inequities in access to and quality of care. This article aims to analyze advances and challenges in identifying and mitigating these biases through a documentary review of Spanish and English articles indexed in Scopus between 2018 and 2022. The methodology involved a systematic analysis of 50 selected studies, classified into four thematic areas: types of algorithmic biases, clinical impact on vulnerable populations, technical limitations in algorithm development, and proposed mitigation strategies. The results demonstrate that biases are deeply rooted in training data and the unequal representation of marginalized groups, leading to less accurate diagnoses for women, racialized communities, and low-income individuals. Although technical and ethical approaches have been proposed, gaps persist in their practical implementation. The study concludes that without multidisciplinary intervention integrating public health, ethics, and data science perspectives, algorithms will continue to reproduce structural inequalities. This research underscores the urgency of inclusive policies and robust regulatory frameworks to ensure equity in digital mental health.

Keywords: Algorithmic Biases; Mental Health; Vulnerable Populations; Artificial Intelligence; Health Equity.

RESUMEN

Los sesgos algorítmicos en los sistemas de diagnóstico de salud mental representan un desafío crítico, especialmente para poblaciones vulnerables, al perpetuar inequidades en el acceso y la calidad de la atención. Este artículo tiene como objetivo analizar los avances y desafíos en la identificación y mitigación de estos sesgos, mediante una revisión documental de artículos en español e inglés indexados en Scopus entre 2018 y 2022. La metodología consistió en un análisis sistemático de 50 estudios seleccionados, clasificados en cuatro ejes temáticos: tipos de sesgos algorítmicos, impacto clínico en poblaciones vulnerables, limitaciones técnicas en el desarrollo de algoritmos y estrategias de mitigación propuestas. Los resultados evidencian que los sesgos están profundamente arraigados en los datos de entrenamiento y en la representación desigual de grupos minorizados, lo que deriva en diagnósticos menos precisos para mujeres, comunidades racializadas y personas de bajos ingresos. Aunque se han propuesto enfoques técnicos y éticos, persisten brechas en su implementación práctica. Se concluye que, sin una intervención multidisciplinaria que integre perspectivas de salud pública, ética y ciencia de datos, los algoritmos reproducirán desigualdades estructurales. Este estudio subraya la urgencia de políticas inclusivas y marcos regulatorios robustos para garantizar equidad en la salud mental digital.

Palabras clave: Sesgos Algorítmicos; Salud Mental; Poblaciones Vulnerables; Inteligencia Artificial; Equidad en Salud.

INTRODUCTION

Artificial intelligence (AI) advances have revolutionized the mental health field by offering promising tools for early diagnosis and treatment personalization.^(1,2) However, these systems are not without critical limitations, particularly about algorithmic biases.⁽³⁾ These biases arise because machine learning models reflect and amplify inequalities present in training data, which can lead to misdiagnosis or depersonalized diagnoses.^(4,5) In the context of mental health, where conditions are highly subjective and culturally mediated, the impact of these biases can be especially harmful.⁽⁶⁾

Algorithmic biases in mental health often manifest in multiple dimensions, including discrimination based on gender, ethnicity, socioeconomic status, and geographic location.⁽⁷⁾ Authors such as Straw & Callison-Burch⁽⁷⁾ and Timmons et al.⁽⁸⁾ have shown that algorithms trained with predominantly Western data are less accurate in diagnosing mental disorders in non-white populations due to differences in symptomatic expression and linguistic patterns. Similarly, women and low-income individuals may be misclassified due to stereotypes embedded in data sets.⁽⁹⁾ These errors perpetuate inequities in health care and reinforce systemic barriers to access to appropriate treatment.^(10,11)

The problem is exacerbated by the fact that mental health diagnostic algorithms are often developed without sufficient representation of vulnerable populations, such as migrants, indigenous communities, or people with cognitive disabilities.⁽¹²⁾ The lack of diversity in the data leads to these groups receiving less accurate or even stigmatizing clinical recommendations.⁽¹³⁾ In addition, many models do not incorporate intercultural perspectives, ignoring how sociocultural factors influence the perception and manifestation of mental disorders.⁽¹⁴⁾ This limits the effectiveness of AI tools and exacerbates mistrust in digitized health systems.^(7,10)

Despite recent efforts to develop fairer frameworks, such as fairness-aware machine learning techniques and algorithmic audits, significant challenges remain in practical implementation. Many technical solutions lack adaptability to local contexts or do not consider the power dynamics underlying data collection.⁽¹⁵⁾ Furthermore, regulation in this area is nascent, allowing potentially discriminatory algorithms to be implemented without sufficient oversight.^(16,17) This raises ethical questions about who bears responsibility when an algorithmic diagnosis fails and harms a vulnerable patient.

Given the growing adoption of AI in mental health, it is urgent to critically examine the advances and challenges in mitigating algorithmic biases, with a special focus on their impact on vulnerable populations. This article seeks to contribute to this debate through a literature review of studies published between 2018 and 2022, analyzing current limitations and proposed strategies to ensure equity in automated diagnosis. The objective is to synthesize recent evidence, identify critical gaps, and suggest future directions for developing more inclusive and ethical algorithms in mental health.

METHOD

This study is based on a systematic review of the scientific literature on algorithmic biases in mental health diagnoses and their impact on vulnerable populations. The review followed a structured approach to ensure thoroughness and rigor in the selection, analysis, and synthesis of sources and to identify advances, limitations, and mitigation strategies reported in recent studies.⁽¹⁸⁾ The methodological process was carried out in four clearly defined stages, which ensured transparent and reproducible data collection.

Definition of search criteria and selection of sources

Specific parameters were established for the collection of literature, including articles published between 2018 and 2022 in the Scopus and PubMed databases, due to their relevance to biomedical and technological research. The search terms combined descriptors such as “algorithmic bias,” “mental health diagnosis,” “health disparities,” and “vulnerable populations” in both English and Spanish. Empirical studies, systematic reviews, and theoretical articles were included, while non-peer-reviewed works or those without solid scientific evidence were excluded.

Filtering and evaluation of study quality

After an initial search yielded 120 documents, filters were applied based on inclusion and exclusion criteria. Studies with transparent methodologies, representative samples, and data-supported conclusions were prioritized. After evaluating titles, abstracts, and full content, 50 articles that met scientific quality standards and thematic relevance were selected.

Thematic analysis and categorization

The selected documents were analyzed using a content analysis approach, identifying patterns and divergences around four main themes: types of algorithmic biases, impact on vulnerable populations, technical limitations, and proposed mitigation strategies. This process allowed for a structured organization of the evidence and facilitated the identification of trends and gaps in the literature.

Synthesis and critical interpretation

Finally, the findings were integrated into a coherent discussion contrasting theoretical, technical, and ethical perspectives. The practical implications of the identified biases were evaluated, and recommendations for future research and public policy interventions were proposed.

This methodological approach provided a comprehensive understanding of the current research on algorithmic biases in mental health, offering a solid foundation for critical analysis and identifying remaining challenges.^(19,20) The rigor at each stage ensured that the results reflected established trends in the literature while highlighting priority areas for action.

RESULTS

An initial literature review showed that artificial intelligence systems applied to mental health diagnosis have demonstrated the potential to improve the accessibility and efficiency of clinical care. However, it was also identified that these systems reproduce and amplify structural inequalities, especially in vulnerable populations, due to intrinsic biases in their designs and training data. The studies analyzed highlight four critical dimensions: the most prevalent types of algorithmic biases, their disproportionate impact on marginalized groups, the technical limitations perpetuating these problems, and the strategies to mitigate them. These themes allowed us to organize the qualitative analysis by articulating theoretical and empirical findings to offer a comprehensive view of the challenge. Each of these themes is discussed below.

Types of Algorithmic Biases in Mental Health

Algorithmic biases in mental health diagnoses manifest in multiple ways, affecting the accuracy and fairness of AI tools (see figure 1).⁽²¹⁾ One of the most documented is ethnic-racial bias, where algorithms trained with data mostly from white Western populations are less accurate when assessing symptoms in racialized groups.^(22,23) In this regard, Hooker et al.⁽²⁴⁾ demonstrate that models for detecting depression underestimate its prevalence in African American and Hispanic communities due to cultural differences in the expression of psychological distress.

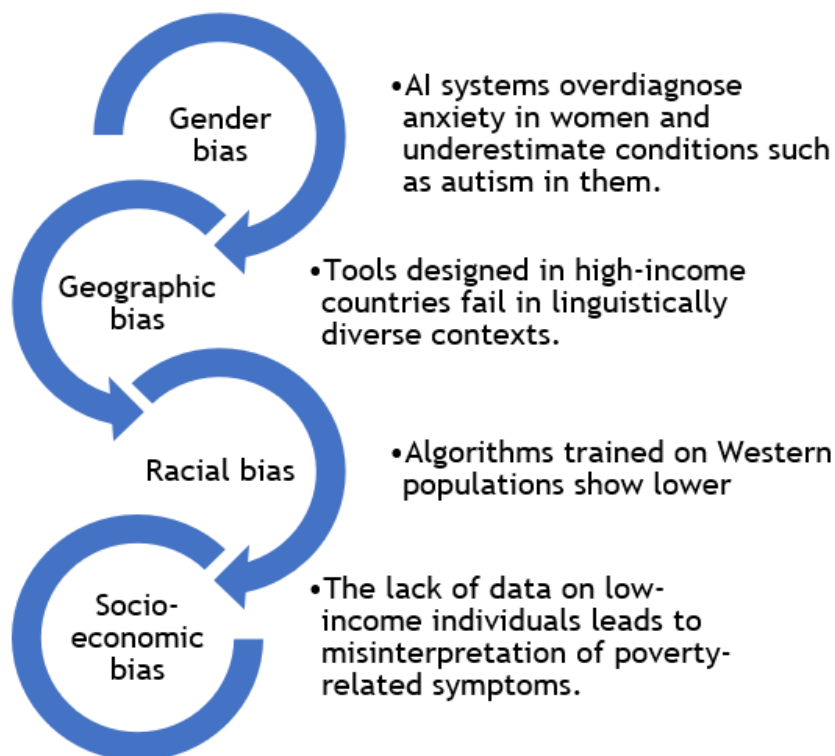


Figure 1. Impact of algorithmic biases in medicine Impact on Vulnerable Populations

Another critical type is gender bias, which arises when algorithms associate stereotypes with diagnoses.⁽¹³⁾ Sedgewick et al.⁽²⁵⁾ reveal that AI systems tend to overdiagnose anxiety disorders in women and underestimate conditions such as autism in them. Similarly, socioeconomic bias is reflected in the lack of representative data on low-income individuals, leading to symptoms associated with poverty (such as chronic stress) being misinterpreted as individual pathologies.⁽²⁶⁾

In addition, geographical biases persist, where tools designed in high-income countries fail when applied in contexts with linguistic diversity or limited access to health services.⁽²⁷⁾ A case in point is algorithms that analyze natural language. When trained with English texts, they ignore idiomatic expressions or grammatical constructions specific to other languages, which affects their cross-cultural validity.⁽²⁸⁾

Impact on Vulnerable Populations

Algorithmic biases have serious consequences for historically marginalized groups (see figure 2), exacerbating mental health inequalities.^(29,30) In indigenous communities, AI tools often overlook cultural manifestations of psychological distress, leading to misdiagnosis or the invisibility of real needs.⁽³¹⁾

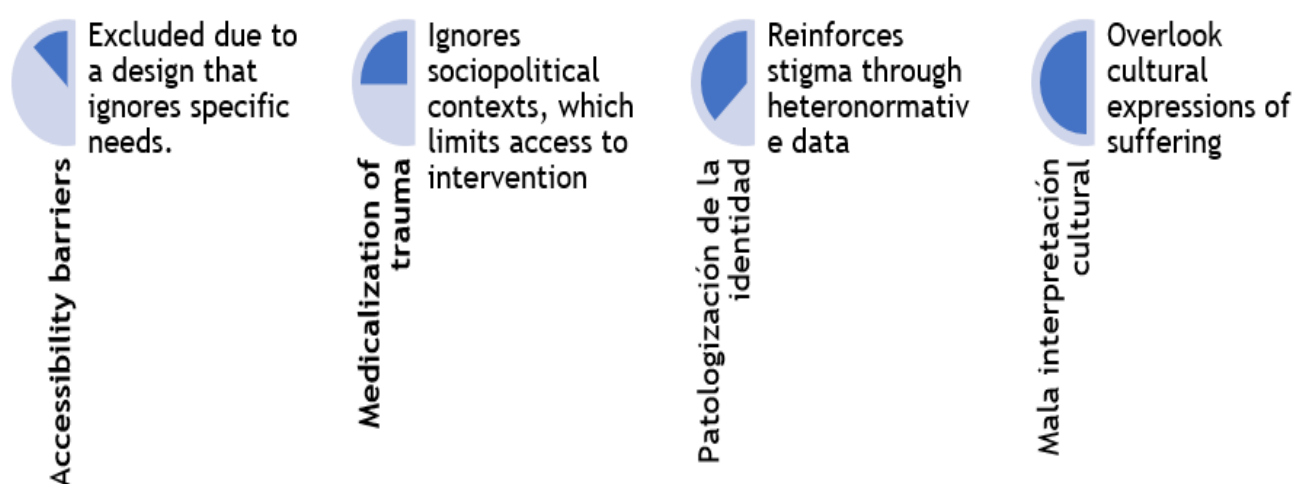


Figure 2. Impact of algorithmic biases on vulnerable populations

LGBTQ+ people also face unique risks. Algorithms trained with heteronormative data can pathologize non-binary identities or interpret gender dysphoria as psychotic disorders.⁽³²⁾ One study found that mental health chatbots displayed invalidating responses to users who mentioned their sexual orientation, reinforcing stigmas.⁽³³⁾

Another affected group is migrants and refugees, whose experiences of trauma are often medicalized by algorithms without considering sociopolitical contexts.⁽³⁴⁾ Finally, older adults and people with cognitive disabilities face barriers due to designs that do not incorporate their needs.^(35,36) Complex interfaces or assessments based on written language exclude those with visual or cognitive limitations, further marginalizing them.⁽³⁷⁾

Technical and Structural Limitations

The perpetuation of biases is not only a problem of insufficient data, but also of deep limitations in the development and deployment of algorithms (see figure 3).⁽³⁷⁾ A key barrier is the homogeneity of data sets, where vulnerable populations are underrepresented.^(8,38)

Another limitation is the lack of transparency in proprietary models. Many medical technology companies do not disclose how their algorithms are trained, making independent audits impossible.⁽³⁹⁾

Moreover, challenges persist in clinical interpretation. Mental health professionals often lack training to question algorithmic results, accepting them as objective.⁽²¹⁾ This is dangerous because algorithms reinforce stereotypes, such as associating poverty with lower treatment adherence.⁽⁴⁰⁾ These limitations, in the author's opinion, reveal that biases are systemic, requiring technical adjustments as well as changes in regulatory frameworks and data governance.

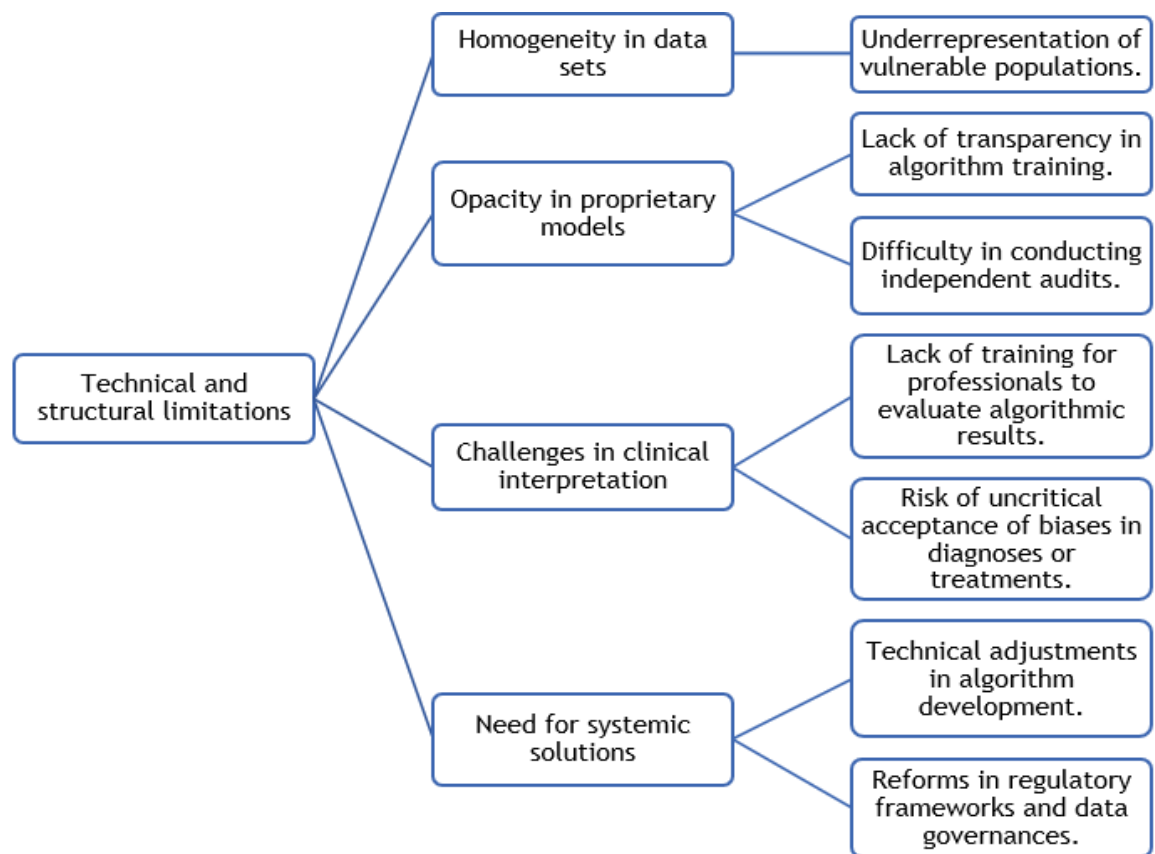


Figure 3. Ethical and structural limitations

Proposed Mitigation Strategies

The reviewed literature suggests multiple approaches to reduce biases, although their implementation remains in early stages (see figure 4).^(41,42) One promising direction is *fairness-aware machine learning*, which incorporates equity metrics during model training.⁽⁴³⁾ Techniques such as *reweighting* and *adversarial debiasing* have improved accuracy for underrepresented groups.⁽⁴⁴⁾

Mandatory algorithmic audits stand out at the institutional level, requiring assessments of their impact on fundamental rights.^(45,46) Existing ethical frameworks can also be adapted to address the new challenges posed by AI in mental health diagnosis, and practical recommendations are provided for health professionals.⁽⁴⁷⁾

However, the author’s opinion is that challenges remain in the scalability of these solutions and the political will to adopt them. In addition to technical innovation, algorithmic justice in mental health will require a redistribution of power in producing medical knowledge.

Technical approaches to reducing bias	Institutional and regulatory measures	Remaining challenges and critical considerations
<ul style="list-style-type: none">•Fairness-aware machine learning: integrating fairness metrics into model training.•Techniques such as reweighting and adversarial debiasing to improve accuracy in underrepresented groups.	<ul style="list-style-type: none">•Implementation of mandatory algorithmic audits with a focus on fundamental rights.•Adapting existing ethical frameworks to address AI challenges in mental health.•Practical recommendations for professionals in the sector.	<ul style="list-style-type: none">•Limitations in the scalability of the proposed solutions.•Lack of political will to adopt mitigation measures.•Need for redistribution of power in the production of medical knowledge.

Figure 4. Strategies for mitigating bias

DISCUSSION

The findings of this review reveal that biases in AI-based diagnostic systems are not mere technical flaws but manifestations of structural inequalities deeply rooted in data and algorithmic designs.^(23,26) Evidence shows that these biases operate differentially, affecting historically marginalized groups more severely and reproducing patterns of exclusion in access to mental health care.⁽⁴⁸⁾ This poses an urgent ethical challenge: the need to recognize that the objectivity of algorithms is, in reality, a mirror of the prejudices that exist in the societies that create them.⁽²⁵⁾ The solution cannot be limited to superficial technical adjustments but requires a fundamental questioning of who participates in developing these technologies and whose voices are systematically silenced in the process.

A critical issue that emerges from the literature is the tension between the democratizing potential of AI and its ability to amplify inequalities.⁽⁴⁹⁾ While these tools promise to expand access to diagnostics in regions with a shortage of specialists, their implementation without adequate safeguards can perpetuate forms of digital colonialism, where vulnerable populations are subject to systems they do not understand or control.⁽³¹⁾ This problem is exacerbated by the commercialization of opaque algorithms, whose internal mechanisms are inaccessible to healthcare professionals and patients.^(12,33) The lack of transparency limits accountability and erodes trust in interventions that, paradoxically, seek to improve mental health care.⁽⁵⁰⁾

The mitigation strategies analyzed, while promising, face significant barriers to real-world implementation. Initiatives such as fairness-by-design frameworks and participatory audits represent significant advances, but they clash with commercial interests, budgetary constraints, and the absence of robust regulatory frameworks.^(38,46) In addition, there remains a disconnect between the technical solutions proposed and the specific needs of local contexts, particularly in low- and middle-income countries. This underscores the importance of developing glocal approaches that combine international standards with culturally situated adaptations, avoiding universal solutions that ignore the particularities of the health systems and communities they serve.⁽³⁷⁾

Finally, this review highlights that the fight against algorithmic bias in mental health requires collective and multidisciplinary action. Researchers, clinicians, policymakers, and affected communities must collaborate to create ethical AI ecosystems prioritizing justice over efficiency.^(46,47) This involves improving algorithms and transforming the power structures that determine what knowledge is validated and which populations are considered a priority.⁽⁴⁾ The path to truly equitable algorithmic diagnoses will require, above all, recognizing that mental health technology is not neutral: it is a battleground where competing visions of who deserves to be heard and what forms of suffering are legitimate are being fought out.

CONCLUSIONS

This literature review shows that algorithmic biases in mental health diagnostic systems reproduce and amplify structural inequalities by disproportionately affecting vulnerable populations. While technical and ethical strategies have been identified to mitigate these biases, their practical implementation requires a multidisciplinary approach combining artificial intelligence advances with inclusive public policies, community participation, and robust regulatory frameworks. Ensuring equity in automated diagnosis is not only a technological challenge but an ethical imperative that requires transforming mental health systems to prioritize social justice over algorithmic efficiency.

Overcoming these challenges requires active collaboration between developers, health professionals, policymakers, and affected communities to create tools that are not only accurate but also culturally sensitive and socially responsible. Only through this collective commitment can the transformative potential of AI in mental health be realized, ensuring that its application benefits all groups equally without exacerbating existing inequalities.

REFERENCES

1. Graham S, Depp C, Lee E, Nebeker C, Tu X, Kim H, Jeste D. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current Psychiatry Reports* 2019;21:116. <https://doi.org/10.1007/s11920-019-1094-0>
2. Grzenda A. Artificial Intelligence in Mental Health. In: *Convergence Mental Health*. Oxford University Press; 2021. <https://doi.org/10.1093/MED/9780197506271.003.0011>
3. Pérez Gamboa AJ, Gómez Cano CA, Sánchez Castillo V. Decision making in university contexts based on knowledge management systems. *Data and Metadata* 2022;1:92. <https://doi.org/10.56294/dm202292>
4. Ntoutsi E, Fafalios P, Gadiraju U, Iosifidis V, Nejdl W, Vidal M, et al. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2020;10(3):e1356. <https://doi.org/10.1002/widm.1356>

5. Peters U. Algorithmic Political Bias in Artificial Intelligence Systems. *Philosophy & Technology* 2022;35:25. <https://doi.org/10.1007/s13347-022-00512-8>
6. Cano CAG, Castillo VS. Mobbing en una institución de educación superior en Colombia. *Aglala* 2021;12(2):100-116. <https://dialnet.unirioja.es/servlet/articulo?codigo=8453105>
7. Straw I, Callison-Burch C. Artificial Intelligence in mental health and the biases of language based models. *PLoS ONE* 2020;15(12):e0240376. <https://doi.org/10.1371/journal.pone.0240376>
8. Timmons A, Duong J, Fiallo N, Lee T, Vo H, Ahle M, et al. A Call to Action on Assessing and Mitigating Bias in Artificial Intelligence Applications for Mental Health. *Perspectives on Psychological Science* 2022;18(5):1062-1096. <https://doi.org/10.1177/17456916221134490>
9. Park J, Arunachalam R, Silenzio V, Singh V. Fairness in Mobile Phone-Based Mental Health Assessment Algorithms: Exploratory Study. *JMIR Formative Research* 2022;6:e34366. <https://doi.org/10.2196/34366>
10. Estrada-Cely GE, Sánchez-Castillo V, Gómez-Cano CA. Bioética y desarrollo sostenible: entre el biocentrismo y el antropocentrismo y su sesgo economicista. *Cilo América* 2018;12(24):255-267. <http://dx.doi.org/10.21676/23897848.2818>
11. Embí PJ. Algorithmic vigilance—Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity. *JAMA Network Open* 2021;4(4):e214622. <https://doi.org/10.1001/jamanetworkopen.2021.4622>
12. Gooding P, Kariotis T. Ethics and Law in Research on Algorithmic and Data-Driven Technology in Mental Health Care: Scoping Review. *JMIR Mental Health* 2021;8(6):e24668. <https://doi.org/10.2196/24668>
13. Gómez-Cano CA, Sánchez V, Tovar G. Factores endógenos causantes de la permanencia irregular: una lectura desde el actuar docente. *Educación y Humanismo* 2018;20(35):96-112. <https://revistas.unisimon.edu.co/index.php/educacion/article/view/3030>
14. Schouler-Ocak M. Transcultural Aspect of Mental Health Care. *European Psychiatry* 2022;65(S1):S3. <https://doi.org/10.1192/j.eurpsy.2022.37>
15. Schick A, Rauschenberg C, Ader L, Daemen M, Wieland L, Paetzold I, et al. Novel digital methods for gathering intensive time series data in mental health research: scoping review of a rapidly evolving field. *Psychological Medicine* 2023;53(1):55-65. <https://doi.org/10.1017/S0033291722003336>
16. Guayara Cuéllar CT, Millán Rojas EE, Gómez Cano CA. Diseño de un curso virtual de alfabetización digital para docentes de la Universidad de la Amazonia. *Revista científica* 2019;(34):34-48. <https://doi.org/10.14483/23448350.13314>
17. Gómez Cano CA, Sánchez Castillo V, Millán Rojas EE. Capitalismo y ética: una relación de tensiones. *Económicas CUC* 2019;40(2):31-42. <https://doi.org/10.17981/econcuc.40.2.2019.02>
18. Siddaway AP, Wood AM, Hedges LV. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology* 2019;70:747-770. <https://doi.org/10.1146/annurev-psych-010418-102803>
19. Hiebl MRW. Sample Selection in Systematic Literature Reviews of Management Research. *Organizational Research Methods* 2023;26(2):229-261. <https://doi.org/10.1177/1094428120986851>
20. Pérez Gamboa AJ, García Acevedo Y, García Batán J. Proyecto de vida y proceso formativo universitario: un estudio exploratorio en la Universidad de Camagüey. *Transformación* 2019;15(3):280-296. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-29552019000300280
21. Richter T, Fishbain B, Fruchter E, Richter-Levin G, Okon-Singer H. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research* 2021;141:199-205. <https://doi.org/10.1016/j.jpsychires.2021.06.044>

22. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366(6464):447-453. <https://doi.org/10.1126/science.aax2342>
23. Rai T. Racial bias in health algorithms. *Science* 2019;366(6464):440. <https://doi.org/10.1126/science.366.6464.440-e>
24. Hooker K, Phibbs S, Irvin V, Mendez-Luck C, Doan L, Li T, et al. Depression Among Older Adults in the United States by Disaggregated Race and Ethnicity. *The Gerontologist* 2019;59(5):886-891. <https://doi.org/10.1093/geront/gny159>
25. Sedgewick F, Kerr-Gaffney J, Leppanen J, Tchanturia K. Anorexia nervosa, autism, and the ADOS: How appropriate is the new algorithm for identifying cases? *Frontiers in Psychiatry* 2019;10:507. <https://doi.org/10.3389/fpsyt.2019.00507>
26. Berthonnet I. Where Exactly Does the Sexist Bias in the Official Measurement of Monetary Poverty in Europe Come From? *Review of Radical Political Economics* 2021;55(1):132-146. <https://doi.org/10.1177/0486613420981785>
27. Ciecierski-Holmes T, Singh R, Axt M, Brenner S, Barteit S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *NPJ Digital Medicine* 2022;5:162. <https://doi.org/10.1038/s41746-022-00700-y>
28. Gomez Cano CA, Sánchez Castillo V, Clavijo Gallego TA. English teaching in undergraduate programs: A reading of the challenges at Uniamazonia from teaching practice. *Horizontes Pedagógicos* 2018;20(1):55-62. <https://doi.org/10.33881/0123-8264.hop.20107>
29. Buda T, Guerreiro J, Iglesias J, Castillo C, Smith O, Matic A. Foundations for fairness in digital health apps. *Frontiers in Digital Health* 2022;4:943514. <https://doi.org/10.3389/fdgth.2022.943514>
30. Xu J, Xiao Y, Wang W, Ning Y, Shenkman E, Bian J, Wang F. Algorithmic fairness in computational medicine. *EBioMedicine* 2022;84:104250. <https://doi.org/10.1016/j.ebiom.2022.104250>
31. Gloria K, Rastogi N, DeGroff S. Bias Impact Analysis of AI in Consumer Mobile Health Technologies: Legal, Technical, and Policy. *arXiv* 2022;abs/2209.05440. <https://doi.org/10.48550/arXiv.2209.05440>
32. Simpson E, Semaan B. For You, or For "You"? *Proceedings of the ACM on Human-Computer Interaction* 2021;4(CSCW3):1-34. <https://doi.org/10.1145/3432951>
33. Oliva T, Antonialli D, Gomes A. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture* 2021;25(2):700-732. <https://doi.org/10.1007/s12119-020-09790-w>
34. Wylie L, Van Meyel R, Harder H, Sukhera J, Luc C, Ganjavi H, Elfakhani M, Wardrop N. Assessing trauma in a transcultural context: challenges in mental health care with immigrants and refugees. *Public Health Reviews* 2018;39:22. <https://doi.org/10.1186/s40985-018-0102-y>
35. Huq S, Maskeliūnas R, Damaševičius R. Dialogue agents for artificial intelligence-based conversational systems for cognitively disabled: a systematic review. *Disability and Rehabilitation: Assistive Technology* 2022;19(5):1059-1078. <https://doi.org/10.1080/17483107.2022.2146768>
36. Lee-Cheong S, Amanullah S, Jardine M. New assistive technologies in dementia and mild cognitive impairment care: A PubMed review. *Asian Journal of Psychiatry* 2022;73:103135. <https://doi.org/10.1016/j.ajp.2022.103135>
37. Wangmo T, Lipps M, Kressig R, Ienca M. Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. *BMC Medical Ethics* 2019;20:98. <https://doi.org/10.1186/s12910-019-0437-z>
38. Pérez A, Echerri D, García Y. Life project as a category of Higher Education pedagogy: Approaching

a grounded theory. *Transformación* 2021;17(3):542-563. <http://scielo.sld.cu/pdf/trf/v17n3/2077-2955-trf-17-03-542.pdf>

39. Bird K, Castleman B, Mabel Z, Song Y. Bringing Transparency to Predictive Analytics: A Systematic Comparison of Predictive Modeling Methods in Higher Education. *AERA Open* 2021;7:23328584211037630. <https://doi.org/10.1177/23328584211037630>

40. Gómez Cano CA, García Acevedo Y, Pérez Gamboa AJ. Intersection between health and entrepreneurship in the context of sustainable development. *Health Leadership and Quality of Life* 2022;1:89. <https://doi.org/10.56294/hl202289>

41. Norori N, Hu Q, Aellen F, Faraci F, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns* 2021;2(10):100347. <https://doi.org/10.1016/j.patter.2021.100347>

42. Fazelpour S, Danks D. Algorithmic bias: Senses, sources, solutions. *Philosophy Compass* 2021;16(8):e12760. <https://doi.org/10.1111/PHC3.12760>

43. Bird S, Kenthapadi K, Kıcıman E, Mitchell M. Fairness-Aware Machine Learning: Practical Challenges and Lessons Learned. *Proceedings of the 12th ACM International Conference on Web Search and Data Mining* 2019:834-835. <https://doi.org/10.1145/3289600.3291383>

44. Petrović A, Nikolić M, Radovanović S, Delibašić B, Jovanović M. FAIR: Fair Adversarial Instance Re-weighting. *Neurocomputing* 2022;476:14-37. <https://doi.org/10.1016/j.neucom.2021.12.082>

45. Mantelero A. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review* 2018;34(4):754-772. <https://doi.org/10.1016/J.CLSR.2018.05.017>

46. Yam J, Skorburg J. From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology* 2021;23(4):611-623. <https://doi.org/10.1007/s10676-021-09599-7>

47. Straw I. Ethical implications of emotion mining in medicine. *Health Policy and Technology* 2021;10(1):167-171. <https://doi.org/10.1016/j.hlpt.2020.11.006>

48. Walsh C, Chaudhry B, Dua P, Goodman K, Kaplan B, Kavuluru R, Solomonides A, Subbian V. Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. *JAMIA Open* 2020;3(1):9-15. <https://doi.org/10.1093/jamiaopen/ooz054>

49. Koster R, Balaguer J, Tacchetti A, Weinstein A, Zhu T, Hauser O, et al. Human-centred mechanism design with Democratic AI. *Nature Human Behaviour* 2022;6(10):1398-1407. <https://doi.org/10.1038/s41562-022-01383-x>

50. D'Alfonso S. AI in mental health. *Current Opinion in Psychology* 2020;36:112-117. <https://doi.org/10.1016/j.copsyc.2020.04.005>

FUNDING

The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION

Conceptualization: Ariadna Matos Matos.

Data curation: Ariadna Matos Matos.

Formal analysis: Ariadna Matos Matos.

Research: Ariadna Matos Matos.

Methodology: Ariadna Matos Matos.

Project management: Ariadna Matos Matos.

Resources: Ariadna Matos Matos.

Software: Ariadna Matos Matos.

Supervision: Ariadna Matos Matos.

Validation: Ariadna Matos Matos.

Visualization: Ariadna Matos Matos.

Writing - original draft: Ariadna Matos Matos.

Writing - revision and editing: Ariadna Matos Matos.