Machine learning for mental health diagnosis

Tackling contributory injustice and epistemic oppression

Pozzi, Giorgia; De Proost, Michiel

DOI
10.1136/jme-2024-110059

Publication date
2024

Document Version
Final published version

Published in
Journal of medical ethics

Machine learning for mental health diagnosis: tackling contributory injustice and epistemic oppression

Giorgia Pozzi [1], Michiel De Proost [2]

INTRODUCTION

In their contribution, Ugar and Malele[1] shed light on an often overlooked but crucial aspect of the ethical development of machine learning (ML) systems to support the diagnosis of mental health disorders. The authors restrain their focus on pointing to the danger of misdiagnosing mental health pathologies that do not qualify as such within sub-Saharan African communities and argue for the need to include population-specific values in these technologies' design. However, an analysis of the nature of the harm caused to said populations once their values remain unrecognised is not offered.

Building on Ugar and Malele's considerations, we add a further perspective to their analysis by showing the need to design considering intended values to avoid the occurrence of epistemic injustices.[2] First, we argue that failing to acknowledge the hermeneutical offerings of the populations interacting with these systems can qualify as *contributory injustice*.[3] Second, we show that this form of injustice paves the way to patterns of epistemic oppression that need scrutiny, particularly given the epistemic authority these systems tend to increasingly acquire.

CONTRIBUTORY INJUSTICE IN ML FOR MENTAL HEALTH SUPPORT

Dotson's concept of *contributory injustice*[3] points out that in the case of blind spots in collectively shared epistemic resources, people in marginalised social positions often develop epistemic resources to make sense of their experiences and reality, which knowers from dominant social positions do not acknowledge. More precisely, this form of injustice arises if knowers in dominant social positions *willfully* ignore alternative epistemic resources despite the awareness that these could be relevant to a certain epistemic goal, thus perpetuating the preponderance of dominant ones. We argue that the examples that Ugar and Malele discuss in their article can be captured as an illustration of Dotson's concept. While Western understandings of mental health are often the standard approach that underlies the design of ML technologies, this is far from being the only approach to make sense of such experiences as 'the South African cases of ukuthwasa, ukufa kwabantu and ufufunyane' indicate.[1] (p 4).

The discussion of Ugar and Malele exemplifies that different epistemic resources, in terms of interpretations, are available to refer to mental diversity beyond the standard biomedical approach. This has important effects not only on epistemic practices within healthcare encounters but also on a structural level. We can consider ML systems as epistemic authoritative entities that, by their presence and use, influence and shape the epistemic resources that are collectively available in society. They have the potential to support specific epistemic attitudes in the general public and mental health professionals that, for instance, favour Western interpretations of psychosis that would be classified as biological dysfunction and neglect others. Alternative interpretations and understandings of mental health conditions may remain unaccounted for.

The use of ML can, thereby, undermine non-Western users' epistemic agency, reinforcing the marginalisation of epistemic resources of non-dominant social groups, thus perpetuating contributory injustice. Let us also point out that the contributory injustice we hint towards does not imply that epistemic agents *lack* the resources to make sense of their experience (as with hermeneutical forms of injustice).[2] The issue is rather in their articulations of lived experience, which 'fail to gain appropriate uptake according to the biased hermeneutical resources used by the perceiver.[3]' (p 34) In the case of interest, the bias amounts to the tacit and wrongful assumption of homogeneity in understanding mental health conditions across populations that often underlie the design of ML systems. This assumption undermines the need to consider and acknowledge the role of individuals as producers of local knowledge and the value of varied epistemic resources, thus excluding pertinent epistemic offerings.

TACKLING EPISTEMIC OPPRESSION

Contributory injustice can become so pervasive that it risks propagating epistemic oppression. In its most general definition, epistemic oppression pertains to practices of *systematic* exclusion that hinder the contribution to central epistemic activities such as knowledge production.[4] The diagnostic process in mental healthcare is intrinsically epistemic since it strongly influences patients' self-understanding and contributes to the pool of knowledge they hold about themselves. Imposing an understanding of health and disease incompatible with patients' collective values pertaining to their embeddedness in a particular social context can hamper these epistemic activities. The introduction of ML systems disregarding the target population's epistemic offerings risks systematising the exclusion of their hermeneutical resources. A misalignment can emerge between the concepts available to the target population to make sense of their lived experiences and those pertaining to the ML system. This means that the problem will not remain at the level of an occasional lack of recognition; this misrecognition rather becomes encoded in the system, which will, by design, not be able to pick up on conceptions of mental health conditions that do not qualify as pathological within a particular community.

These considerations thus hint at complex issues in which power dynamics, social identities and the normative weight of relevant concepts such as health and disease intersect. Addressing and amending epistemic injustice and oppression in ML for mental health diagnosis is far more challenging than one account can accommodate. A catchall theory of epistemic injustice is, hence, an unrealistic expectation within the scope of this commentary. However, effective open conceptual structures can aid in reducing the perpetuation of contributory injustice and the epistemic oppression that often accompanies it. Medina, for instance, describes a connection between social imaginaries, poor epistemic habits, active ignorance and 'meta-blindness'—the lack of actively searching 'for more alternatives than those noticed' and engaging those alternatives.[5] (p 78). This insensitivity to the limits of one's instituted social imaginaries fosters and maintains poor

[1]Delft University of Technology, Delft, Netherlands
[2]Ghent University, Ghent, Belgium

Correspondence to Dr Giorgia Pozzi, Delft University of Technology, Delft, Netherlands; g.pozzi@tudelft.nl

epistemic habits and vices, such as epistemic laziness, closed-mindedness and epistemic arrogance. Medina encourages seeking 'epistemic friction' and 'to remain open to epistemic counterpoints'.[5] (p 78). With our contribution, we hope to have shown the importance of including these critical approaches in the design processes of ML systems for mental health support to avoid sustaining forms of epistemic injustice and oppression.

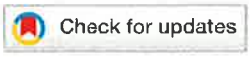
Contributors GP and MDP contributed equally to the manuscript.

Funding MDP's contribution to this work was supported by the H2020 European Research Council (grant number 949841).

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; internally peer reviewed.



To cite Pozzi G, De Proost M. *J Med Ethics* 2024;**50**:596–597.

Received 4 April 2024
Accepted 20 May 2024
Published Online First 30 May 2024

► http://dx.doi.org/10.1136/jme-2023-109711

J Med Ethics 2024;**50**:596–597.
doi:10.1136/jme-2024-110059

ORCID iDs
Giorgia Pozzi http://orcid.org/0000-0001-8928-5513
Michiel De Proost http://orcid.org/0000-0003-0545-8515

REFERENCES

1 Ugar ET, Malele N. Designing AI for mental health diagnosis: challenges from sub-Saharan African value-laden judgements on mental health disorders. *J Med Ethics* 2024;50:592–5.
2 Fricker M. In: *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007. Available: https://academic.oup.com/book/32817
3 Dotson K. A cautionary tale: on limiting epistemic oppression. *Fro* 2012;33:24–47.
4 Dotson K. Conceptualizing epistemic oppression. *Social Epistemology* 2014;28:115–38.
5 Medina J. In: *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford University Press, 2013. Available: https://academic.oup.com/book/9202