General Psychiatry

Public perspectives on AI diagnosis of mental illness

Cliodhna O'Connor

To cite: O'Connor C. Public perspectives on AI diagnosis of mental illness. *General Psychiatry* 2024;**37**:e101370. doi:10.1136/gpsych-2023-101370

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/gpsych-2023-101370).

Received 28 September 2023
Accepted 23 April 2024

University College Dublin, Dublin, Ireland

Correspondence to
Dr Cliodhna O'Connor; cliodhna.oconnor1@ucd.ie

To the editor:

Psychiatric theory, policy and practice are currently grappling with the risks and opportunities presented by artificial intelligence (AI) applications in mental healthcare. Synthesising data to generate diagnosis is an aspect of mental healthcare where AI is anticipated to have the greatest and soonest impact.[1–4] While such technologies remain some distance from clinical application, preliminary evidence suggests AI-derived classifications may predict certain treatment outcomes and clinical trajectories, and could soon become available to supplement or replace traditional manual-based diagnostic assessment.[5]

The use of AI algorithms to diagnose mental illness raises many ethical challenges. These include the potential for security breaches or misuse of private mental health data, the risk that AI trained on biased data sets will reinforce societal inequalities, the risk of false-positive diagnoses that expose patients to stress and discrimination, and issues with the interpretability of 'black box' AI decisions.[6–13] For any emerging technology, evidence on how the lay public views its ethical challenges (eg, the risks that most concern end users) is vital to ensuring socially responsible application. Moreover, to optimise the value for policy and practice, this analysis should occur prospectively rather than retrospectively, while technological development can still be adjusted in line with societal values and priorities. With AI-informed diagnosis likely approaching implementation in clinical settings, minimal data reveal societal perspectives on this technology or the ethical issues it raises.[14]

To enlighten these issues, an online survey study was recently conducted, with ethical approval from the Research Ethics Committee of University College Dublin. A research company was contracted to recruit samples in the USA (n = 1060) and UK (n = 1000) that were nationally representative on gender, age and region. Using Qualtrics software, participants were randomly assigned to read one of four vignettes (online supplemental material). All vignettes described a person ('Morgan'; gender unspecified) undergoing clinical assessment for the same mental health difficulties (eg, flat mood, sleep difficulties, paranoia), but differed in whether they were diagnosed either using an AI or a standard Diagnostic and Statistical Manual of Mental Disorders (DSM) approach. Furthermore, to ensure the

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Artificial intelligence (AI)-supported diagnosis of mental illness is a viable prospect for future clinical practice but raises many ethical challenges.
⇒ Minimal empirical evidence enlightens how such concerns are viewed by the general public.
⇒ Availability of data elucidating lay ethical concerns about AI diagnosis is crucial to ensure socially responsible development and application of new technology.

WHAT THIS STUDY ADDS

⇒ This paper reports a large-scale representative survey (n=2060) of the US and UK populations, which explored lay perspectives on the ethical issues raised by AI diagnosis in psychiatry, compared with standard Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnostic approaches.
⇒ Results identify the specific ethical issues that cause the greatest public concern and suggest that the lay public are less concerned overall about AI compared with DSM diagnosis.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings alert researchers, practitioners and policymakers to the specific ethical concerns that should be prioritised in developing and implementing new approaches to psychiatric diagnosis.
⇒ Understanding the opinions and preferences of the lay public, who represent the users and potential users of mental health services, will help ensure AI diagnostic technologies can be steered towards maximal benefit and minimal harm.

Table 1 Ethical concerns (means and SD) within the DSM and AI conditions and across the total sample

		DSM		AI		Total	
Ethical issue	**I would be concerned…**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Discrimination	about facing discrimination due to my diagnosis	5.03	1.71	4.91	1.78	4.97	1.75
Communicability	that it will be difficult to explain my diagnosis to others	4.74	1.70	4.50	1.78	4.61	1.78
Stress	about the stress caused by receiving the diagnosis	4.70	1.64	4.51	1.76	4.60	1.70
Self-concept	that the diagnosis may change how I see myself (my self-concept)	4.66	1.71	4.51	1.79	4.58	1.75
Understanding	about not fully understanding the clinical assessment procedures that led to my diagnosis	4.47	1.75	4.42	1.72	4.44	1.74
Treatability	that my diagnosis will not lead to effective treatment	4.40	1.74	4.31	1.72	4.35	1.73
Accuracy	that my diagnosis may be inaccurate	4.31	1.68	4.36	1.69	4.34	1.69
Medicalisation	about everyday human functions, like sleep and language, becoming targets of medical attention	4.44	1.69	4.22	1.79	4.33	1.75
Inequality	about inequalities in access to this type of clinical assessment	4.14	1.72	4.13	1.73	4.14	1.72
Security	about the security of my private mental health data	4.02	1.93	4.06	1.95	4.04	1.94
Accountability	about who could be held accountable for any errors in the clinical assessment process	3.93	1.73	4.04	1.73	3.99	1.73
Impersonality	about the impersonal nature of the clinical assessment process	3.83	1.71	3.94	1.82	3.89	1.77
Obsolescence	that my diagnosis may become outdated as clinical knowledge and assessment techniques evolve	3.91	1.70	3.85	1.68	3.88	1.69
Bias	that the clinical assessment process was biased	3.54	1.69	3.56	1.76	3.55	1.72
Intrusion	that the clinical assessment process was too intrusive	3.54	1.77	3.44	1.79	3.49	1.78
Clinical competence	about Dr Smith's competence as a clinician	3.26	1.77	3.14	1.77	3.20	1.77

AI, artificial intelligence; DSM, Diagnostic and Statistical Manual of Mental Disorders.

generalisability of results given that different diagnostic labels trigger different associations regarding severity and stigma,[15] half of the participants (evenly broken down across the AI/DSM groups) read that 'Morgan' had been diagnosed with major depressive disorder (MDD) and half with schizophrenia spectrum disorder (SSD). After reading the vignettes and completing a brief attention check, participants were asked to imagine they were in Morgan's position themselves and to rate their degree of concern (on a 7-point Likert Scale) about 16 issues after receiving their diagnosis. Given the study's interest in mapping contemporary responses to emerging diagnostic technologies, these 16 issues were derived from a prestudy review of the literature on ethical challenges of AI diagnosis in mental health. The ethical issues were carefully phrased so that they could, in principle, apply to both AI and DSM diagnosis (for instance, the question on 'bias' could be equally interpreted as connoting algorithmic or human bias, while 'intrusion' could be interpreted with reference to an individual clinician asking personal questions or technology that tracks one's daily activity and speech). Table 1 displays the range of ethical issues queried, in order of their average levels of concern within the total sample. Self-reported demographic identifications indicated participants were 51.6% female; aged 18–89 (mean = 48.41) years; 25.4% identified as an ethnic minority; 71.5% had tertiary education; and 30.7% had previously received a psychiatric diagnosis.

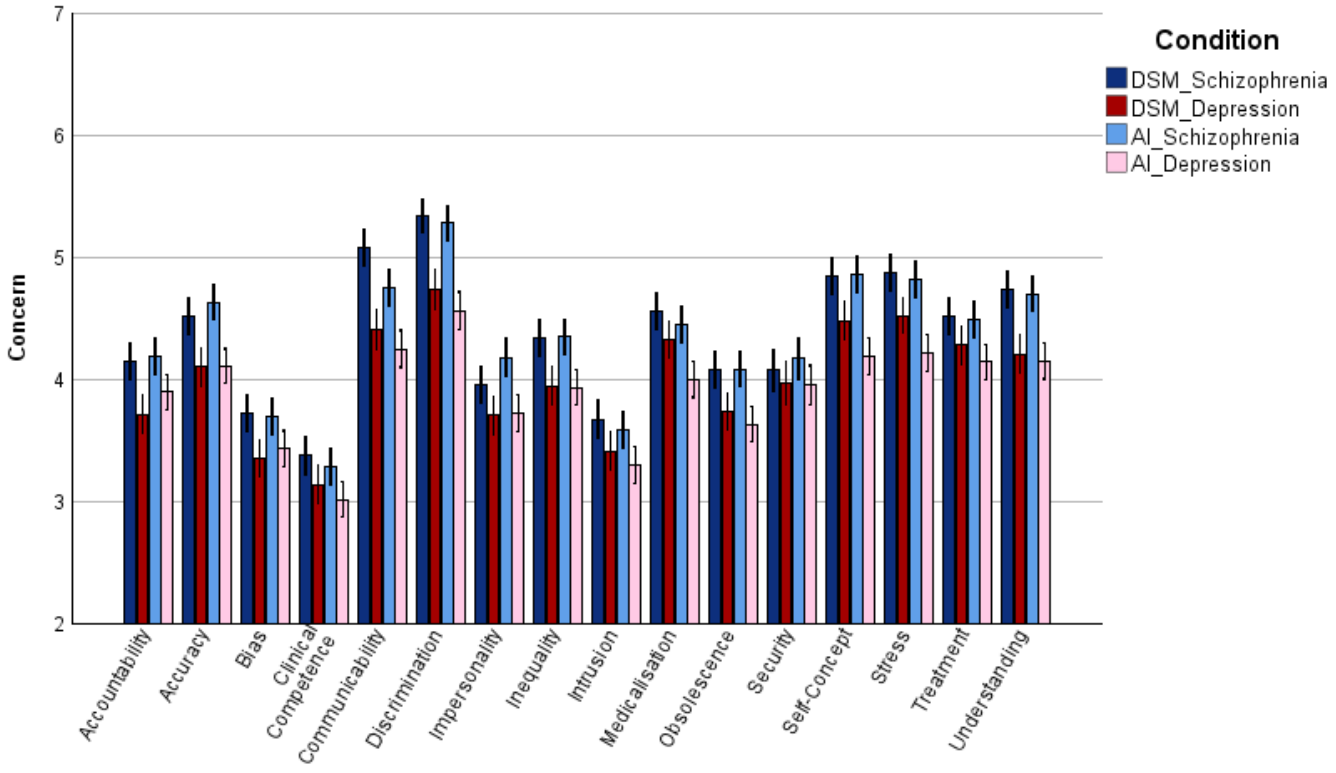

Figure 1 Mean levels of concern across vignette conditions. AI, artificial intelligence; DSM, Diagnostic and Statistical Manual of Mental Disorders.

Data were analysed using SPSS V.27. Participants who failed attention checks (n = 84), completed the survey implausibly quickly (n = 25), or had suspicious response patterns (eg, selecting the same button for every question, n = 10) were removed from the final data set. A two-way multivariate analysis of variance (MANOVA) using Pillai's trace assessed the impact of vignette condition (Diagnostic Method: DSM vs AI; and Diagnostic Category: MDD vs SSD) on ethical concerns, controlling for country, gender, age, ethnicity, education and personal diagnosis experience. The MANOVA showed no significant interaction between Diagnostic Method and Diagnostic Category. A main effect of Diagnostic Category indicated that people had significantly greater concern about the implications of a diagnosis of SSD than MDD, $F(16,2017) = 7.32$, $p < 0.001$, $\eta_p^2 = 0.06$. Most interestingly for the present purposes, a main effect of Diagnostic Method suggested that the DSM vignettes elicited significantly more concerns than the AI vignettes, $F(16,2017) = 2.94$, $p < 0.001$, $\eta_p^2 = 0.02$. Tests of between-subjects effects indicated that the DSM vignettes prompted significantly greater concern on the dimensions of communicability, $F(1,2032) = 10.84$, $p = 0.001$, $\eta_p^2 = 0.005$, stress, $F(1,2032) = 5.90$, $p = 0.015$, $\eta_p^2 = 0.003$, and medicalisation, $F(1,2032) = 7.76$, $p = 0.005$, $\eta_p^2 = 0.04$. Figure 1 displays mean levels of concern across Diagnostic Category and Diagnostic Method conditions.

These results raise multiple important points. First, regarding diagnosis overall, the ethical issues that most concern the lay public relate to the personal and social impacts of a diagnosis (eg, its implications for discrimination, communicability, stress and self-concept). These data suggest the public is relatively unconcerned about the possibility of clinical assessment being biased, intrusive or conducted by an incompetent clinician. The hierarchy of ethical concerns illustrated in Table 1 can inform the development of diagnostic technologies that align with lay priorities. For example, addressing diagnoses' potential to trigger discrimination and stress, and difficulty explaining a diagnosis to others, is imperative to ensure the acceptability of new diagnostic approaches. While the lower-ranked concerns may reflect genuine indifference, they could also indicate a need to raise public awareness regarding certain risks; for example, the possibility for bias in both clinician judgements and AI algorithms.

Second, results suggest that AI-based assessments do not heighten lay concern relative to traditional DSM diagnosis. On the contrary, accounts of DSM diagnosis elicited more concern about issues such as the diagnosis' communicability to others, stress to self and medicalisation. This unanticipated result suggests that prevailing manual-based diagnostic methods may not have strong residual acceptability among the general population. In considering the implementation of new diagnostic technology, it is equally important to critically appraise the diagnostic methods it proposes to replace or supplement; while AI diagnosis may raise specific ethical challenges, the public may deem these less risky than the known

limitations of traditional diagnostic methods. However, it remains unclear whether the public's relative comfort regarding AI diagnosis authentically reflects lay priorities, or results from unfamiliarity with a still-hypothetical clinical technique. Moreover, the fictional clinical cases described in the vignettes all resulted in classification into a traditional diagnostic category (MDD or SSD). Since one anticipated outcome of AI diagnosis is the elimination or subdivision of traditional diagnoses by algorithmically derived diagnoses reflecting intricate biological and behavioural profiles,[6 10] public and service-user responses to unfamiliar precision diagnoses represent a further unknown that requires clarification.

This preliminary study is subject to numerous limitations, particularly pertaining to the reliance on hypothetical vignettes, the superficial nature of the online survey method and the unavailability of previously validated measures of ethical concern. Nevertheless, it represents the first data internationally on how lay publics evaluate the ethical challenges of AI-enabled diagnostic technologies. As public opinion will likely evolve in parallel with technological developments, continuing to track lay perspectives as AI diagnosis comes onstream is crucial to ensuring it can be steered towards the maximal benefit and minimal harm.

Contributors The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding This research was funded by a University College Dublin Career Development Award (ref. SF1881).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. This study was approved by the University College Dublin Human Research Ethics Committee (ref. HS-23-12-OConnor). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Cliodhna O'Connor http://orcid.org/0000-0001-8134-075X

REFERENCES

1 Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. *Artif Intell Med* 2020;102:101753.
2 D'Alfonso S. AI in mental health. *Curr Opin Psychol* 2020;36:112–7.
3 Zhou Z, Wu TC, Wang B, *et al*. Machine learning methods in psychiatry: a brief introduction. *Gen Psychiatr* 2020;33:e100171.
4 Rocheteau E. On the role of artificial intelligence in psychiatry. *Br J Psychiatry* 2023;222:54–7.
5 Abd-Alrazaq A, Alhuwail D, Schneider J, *et al*. The performance of artificial intelligence-driven Technologies in diagnosing mental disorders: an umbrella review. *NPJ Digit Med* 2022;5:87.
6 Lee EE, Torous J, De Choudhury M, *et al*. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2021;6:856–64.
7 Carr S. AI gone mental': engagement and ethics in data-driven technology for mental health. *J Ment Health* 2020;29:125–30.
8 Graham S, Depp C, Lee EE, *et al*. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep* 2019;21:116.
9 Loch AA, Lopes-Rocha AC, Ara A, *et al*. Ethical implications of the use of language analysis technologies for the diagnosis and prediction of psychiatric disorders. *JMIR Ment Health* 2022;9:e41014.
10 Montag C, Sindermann C, Baumeister H. Digital phenotyping in psychological and medical sciences: a reflection about necessary prerequisites to reduce harm and increase benefits. *Curr Opin Psychol* 2020;36:19–24.
11 Koutsouleris N, Hauser TU, Skvortsova V, *et al*. From promise to practice: towards the realisation of AI-informed mental health care. *Lancet Digit Health* 2022;4:e829–40.
12 Lennon JC. Machine learning Algorithms for suicide risk: a premature arms race *Gen Psychiatr* 2020;33:e100269.
13 McCradden M, Hui K, Buchman DZ. Evidence, ethics and the promise of artificial intelligence in psychiatry. *J Med Ethics* 2023;49:573–9.
14 Higgins O, Short BL, Chalup SK, *et al*. Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: an integrative review. *Int J Ment Health Nurs* 2023;32:966–78.
15 O'Connor C, Brassil M, O'Sullivan S, *et al*. How does diagnostic labelling affect social responses to people with mental illness? A systematic review of experimental studies using vignette-based designs. *J Ment Health* 2022;31:115–30.

Cliodhna O'Connor holds degrees from Trinity College Dublin (BA in Psychology, 2009) in Ireland, the London School of Economics & Political Science (MSc in Social & Cultural Psychology, 2010) and University College London (PhD in Psychology, 2014) in the UK. She currently is an Associate Professor in the School of Psychology, University College Dublin, where she has been employed since 2017. She leads the UCD Classification & Attribution Lab, which forms a hub for research exploring the social and psychological processes through which we classify individuals into social categories and attribute causal explanations for those groups' characteristics. Her research interests focus on how scientific and clinical classifications influence self-concept and social identity, and she has explored these links in a range of social contexts including gender stereotypes and psychiatric diagnosis.