Patient Perspectives on AI for Mental Health - With Great [Computing] Power, Comes Great Responsibility: A Cross-sectional Public Survey

Natalie Benda, Pooja M. Desai, Zayan Reza, Anna Zhang, Shiveen Kumar, Sarah Harkins, Alison Hermann, Yiye Zhang, Rochelle Joly, Jessica Kim, Jyotishman Pathak, Meghan Reading Turchioe

Submitted to: JMIR Mental Health
on: March 15, 2024

Table of Contents

Patient Perspectives on AI for Mental Health - With Great [Computing] Power, Comes Great Responsibility: A Cross-sectional Public Survey

Natalie Benda[1] PhD; Pooja M. Desai[2] MA; Zayan Reza[3]; Anna Zhang[4]; Shiveen Kumar[5]; Sarah Harkins[1]; Alison Hermann[6]; Yiye Zhang[7]; Rochelle Joly[8]; Jessica Kim[9]; Jyotishman Pathak[7]; Meghan Reading Turchioe[1]

[1]Columbia University School of Nursing New York US
[2]Columbia University Department of Biomedical Informatics New York US
[3]Columbia University Mailman School of Public Health New York US
[4]Stuyvestant High School New York US
[5]Cornell University Ithaca US
[6]Weill Cornell Medicine Department of Psychiatry New York US
[7]Weill Cornell Medicine Department of Population Health Sciences Division of Health Informatics New York US
[8]Weill Cornell Medicine Department of Obstetrics and Gynecology New York US
[9]Weill Cornell Medicine Department of Population Health Sciences New York US

Corresponding Author:
Natalie Benda PhD
Columbia University
School of Nursing
560 168th Street
New York
US

Abstract

Background: The application of artificial intelligence to health and healthcare is rapidly increasing. Several studies have assessed the attitudes of health professionals but far fewer have explored perspectives of patients or the general public. Studies investigating patient perspectives have focused on somatic issues including radiology, perinatal health, and general applications. Patient feedback has been elicited in the development of specific mental health solutions, but broader perspectives towards AI for mental health have been under-explored.

Objective: To understand public perceptions regarding potential benefits of AI, concerns, comfort with AI accomplishing various tasks, and values related to AI, all pertaining to mental health.

Methods: We conducted a one-time cross-section survey with a nationally representative sample of 500 United States-based adults. Participants provided structured responses on their perceived benefits, concerns, comfortability, and values on AI related to mental health. They could also add free text responses to elaborate on their concerns and values.

Results: A plurality of participants (49.3%) believed AI may be beneficial for mental healthcare, but this perspective differed based on socio-demographic variables ($p<0.05$). Specifically, Black participants (OR = 1.76) and those with lower health literacy (OR=2.16), perceived AI to be more beneficial, and females (OR=0.68) perceived AI to be less beneficial. Participants endorsed concerns related to the use of AI for mental health regarding its accuracy, possible unintended consequences such as misdiagnosis, confidentiality of their information, and loss of connection with their health professional. Over 80% of participants also valued being able to understand individual factors driving their risk, confidentiality, and autonomy as it pertained to the use of AI for their mental health. When asked about who was responsible for misdiagnosis of mental health conditions using AI, 81.6% of participants found the health professional to be responsible. Qualitative results revealed similar concerns related to the accuracy of AI and how its use may impact the confidentiality of their information.

Conclusions: Future work involving the use of AI for mental health should investigate strategies for conveying the level of AI's accuracy, factors that drive risk, and how data are used confidentially so that patients may work with their health professionals to determine when AI may be beneficial. It will also be important in a mental health context to ensure the patient-health professional relationship is preserved when AI is utilized. Clinical Trial: Not applicable

(JMIR Preprints 15/03/2024:58462)

DOI: https://doi.org/10.2196/preprints.58462

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Patient Perspectives on AI for Mental Health: With Great [Computing] Power, Comes Great Responsibility

Natalie C. Benda*

Columbia University, School of Nursing, nb3115@cumc.columbia.edu

Pooja M. Desai

Columbia University, Department of Biomedical Informatics, pmd4002@cumc.columbia.edu

Zayan Reza

Columbia University, Mailman School of Public Health, zsr2108@cumc.columbia.edu

Anna Zhang

Stuyvesant High School, annashuzheng@gmail.com

Shiveen Kumar

Cornell University, sak347@cornell.edu

Sarah Harkins

Columbia University, School of Nursing, seh2238@cumc.columbia.edu

Alison Hermann

Weill Cornell Medicine, Department of Psychiatry, alh9039@med.cornell.edu

Yiye Zhang

Weill Cornell Medicine, Department of Population Health Sciences, yiz2014@med.cornell.edu

Rochelle Joly

Weill Cornell Medicine, Department of Obstetrics and Gynecology, roj9069@med.cornell.edu

Jessica Kim

Weill Cornell Medicine, Department of Population Health Sciences, jek4015@med.cornell.edu

Jyotishman Pathak

Weill Cornell Medicine, Department of Population Health Sciences, jyp2001@med.cornell.edu

Meghan Reading Turchioe

Columbia University, School of Nursing, mr3554@cumc.columbia.edu

**Corresponding author

Patient Perspectives on AI for Mental Health: With Great [Computing] Power, Comes Great Responsibility

Abstract

Background: The application of artificial intelligence to health and healthcare is rapidly increasing. Several studies have assessed the attitudes of health professionals but far fewer have explored perspectives of patients or the general public. Studies investigating patient perspectives have focused on somatic issues including radiology, perinatal health, and general applications. Patient feedback has been elicited in the development of specific mental health solutions, but broader perspectives towards AI for mental health have been under-explored.

Objective: To understand public perceptions regarding potential benefits of AI, concerns, comfort with AI accomplishing various tasks, and values related to AI, all pertaining to mental health.

Methods: We conducted a one-time cross-section survey with a nationally representative sample of 500 United States-based adults. Participants provided structured responses on their perceived benefits, concerns, comfortability, and values on AI related to mental health. They could also add free text responses to elaborate on their concerns and values.

Results: A plurality of participants (49.3%) believed AI may be beneficial for mental healthcare, but this perspective differed based on socio-demographic variables ($p<0.05$). Specifically, Black participants (OR = 1.76) and those with lower health literacy (OR=2.16), perceived AI to be more beneficial, and females (OR=0.68) perceived AI to be less beneficial. Participants endorsed concerns related to the use of AI for mental health regarding its accuracy, possible unintended consequences such as misdiagnosis, confidentiality of their information, and loss of connection with their health professional. Over 80% of participants also valued being able to understand individual factors driving their risk, confidentiality, and autonomy as it pertained to the use of AI for their mental health. When asked about who was responsible for misdiagnosis of mental health conditions using AI, 81.6% of participants found the health professional to be responsible. Qualitative results revealed similar concerns related to the accuracy of AI and how its use may impact the confidentiality of their information.

Conclusions: Future work involving the use of AI for mental health should investigate strategies for conveying the level of AI's accuracy, factors that drive risk, and how data are used confidentially so that patients may work with their health professionals to determine when AI may be beneficial. It will also be important in a mental health context to ensure the patient-health professional relationship is preserved when AI is utilized.

Background

The potential of artificial intelligence (AI) to transform healthcare has been touted since the early 2010s.[1–4] In healthcare applications, practitioners commonly operationalize AI by training machine learning (ML) algorithms using large retrospective data sets to perform human reasoning tasks, such as identifying issues (e.g. anomalies in medical images), predicting events (e.g., disease incidence), recommending treatment (e.g., pharmacogenomics), detecting patterns (e.g., finding symptom clusters), and generating text (e.g., for clinical decision support rules).

AI has already made significant strides in the field of medical imaging, aiding health professionals at various process stages, including improving image quality,[6] guiding image acquisition,[7] risk stratification of images to be reviewed by a specialist (i.e., radiologist),[8,9] and image interpretation.[10,11] More recently, predictive AI has been leveraged to detect mental health related issues including major depressive disorders,[12,13] stress/anxiety,[14] bi-polar disorder,[15] and even suicide.[16,17] ML has also commonly been used for treatment selection in fields such as oncology and mental health.[18,19] Despite the demonstrated predictive accuracy of AI, outside of the field of medical imaging, relatively few of the predictive AI tools created are ever implemented in everyday clinical care,[20] and even fewer have demonstrated a positive clinical impact compared to current standards of care.[21] Furthermore, Of the 63 articles in a systematic review of clinical trials evaluating predictive AI, none focused on mental health-related conditions.

Due to the gap between the predictive AI's accuracy and its lack of observed impact on health outcomes, many researchers have studied health professional perceptions of ML-based tools and related implementation challenges.[17,22–27] However, patient perspectives of AI have been understudied.[28–30] While some predictive AI developers may not intend patients to view the AI's output on their own, it has become more likely that patients may have access to predictive AI output, due to recent advances in patient data ownership and access. The U.S. 21st Century Cures Act, for example, prevents information blocking from patients, requiring health organizations and insurance providers to give patients access to their electronic health information without delay or expense.[31] This may result in a patient seeing a predictive AI risk score prior to a discussion with their healthcare team. In a 2020 predictive AI pre-implementation study, health professionals stressed the importance of keeping the patient in the information loop when the AI predicted a risk or a treatment to justify to the patient why they may require further support.[23] In addition to practical considerations, there is also an ethical imperative to ensure patients understand how their data is being used, what predictive AI may show, and what this means, especially for sensitive issues, such as mental health concerns.[32] Before we can design solutions for communicating AI information to patients, it is important to understand the public's perceived benefits, comfort, concerns, and values related to AI use, particularly for mental healthcare.[33]

To address the deficit in knowledge regarding patient perspectives on AI, Khullar et al. conducted a survey of a nationally representative panel of the United States based population. While the majority of respondents reported a perceived benefit of using AI in healthcare, comfort with AI unsurprisingly varied based on accuracy, transparency, and the clinical application of the AI (e.g., reading a chest x-ray vs. making a cancer diagnosis).[34] The Khullar et al. survey focused on somatic applications of AI, leaving questions regarding public perceptions of AI for mental health applications unanswered. Although, others have explored narrower issues related to feedback on specific mental health apps and specific prediction tasks (i.e., suicide).[35,36]

With the simultaneous increase in AI applications for mental health, patient access and ownership to their data, ethical concerns regarding the creation and use of AI, and the stigmas surrounding mental health, understanding patient perceptions of if and how AI may be appropriately used for mental health is critical.[37,38] This study adapts and extends the Khullar et al. survey to evaluate patient perspectives on the utilization for AI for mental health applications. We specifically surveyed members of the public to gain patient perspectives on AI applications for mental health. Khulllar et al. did not explore values regarding AI, in other words, what are patients' priorities for effective, appropriate AI use for mental health? We also explore these issues in this study using a bioethics-informed framework. The specific research questions guiding our work were:

- RQ1: Do the public perceive AI as beneficial for mental health?
    - RQ1E: Do perceived benefits differ by socio-demographic factors?
- RQ2: How concerned is the public about common issues related to AI use in mental health?
- RQ3: What types of predictive tasks are the public comfortable with AI executing in mental health applications?
- RQ4: What are the publics' values related to AI use for mental health?

We also elicited open-ended responses from participants to add to their quantitative feedback.

Methods

Study Design

Our study consisted of a one-time, cross-sectional survey of United States (U.S.)-based adults in September of 2022. We sampled a general U.S. adult population to elicit the public's perspectives on AI for mental health. We partnered with Prolific, an online survey sampling platform, to recruit participants. Prolific provides access to an international sample of verified users (over 100 000 users residing in the U.S.) who are willing to be involved in survey research studies. Prolific matches eligible participants with research studies streamline the recruitment, data collection, and compensation process. Prospective participants had to be verified Prolific users aged 18 or older who were fluent in written and spoken English to be eligible. This study was approved by the BRANY Institutional Review Board.

Questionnaire Design

We designed our questionnaire to mimic items asked in a previous study by Khullar et al. but applied to perspectives related to AI for mental health, instead of healthcare broadly. Question categories related to AI for mental health involved: 1) perceived benefits, 2) concerns, 3) comfort with AI for specific predictive tasks. Adapting questions related to perceived benefits and concerns predominantly involved updating language referring to "health" or "healthcare" to "mental health" or "mental healthcare". In the Khullar et al. questionnaire, questions regarding predictive tasks included reading a screening (i.e., chest x-ray); making a diagnosis for two different conditions, one more severe (pneumonia and cancer); telling a patient they had either of the two aforementioned conditions and making a treatment recommendation. Our team worked with a trained psychiatrist to construct tasks following similar patterns but pertaining to mental health treatment, adding two more tasks (seven total) to explore more sensitive concepts relating to mental health. Appendix A.1 presents the questions in each of the aforementioned categories along with the question from which they were adapted as applicable.

We also extended the questionnaire to understand participant's values pertaining to AI design and implementation for mental health to facilitate more patient-centered design of future AI applications for mental health. This section asked patients to rate their level of importance regarding various statements pertaining to AI for mental health informed by the constructs of MITRE's bioethical framework. Appendix A.2 displays the values statements presented to participants based on the relevant bioethics construct.

In addition to the perspectives and values questions, participants also provided socio-demographic information regarding personal characteristics, health literacy, subjective numeracy, previous mental health care experience, and pregnancy history (results reported in a separate manuscript). The full battery of socio-demographic questions may be found in Appendix A.3.

Lastly, following the sections regarding concerns and values, the survey contained open-ended questions to allow people to free-text responses with additional concerns or values (included in the relevant appendices).

We designed the battery of questions with input from experts in ML, human-centered design, psychiatry, and the author of the original survey from which the questions were adapted. The survey questions also underwent two rounds of pilot testing to improve the comprehensibility of the questions and understand the amount of time needed to complete the questionnaire. The question/answer design was optimized, and pilot tested for both desktop and mobile (i.e., smartphone) completion to ensure those with different device access or preferences could participate in the study.

Participants

All participants were recruited from Prolific's survey sampling panel and were verified users who have agreed to participate in research studies via the Prolific website. Our sample included those 18 years or older, residing in the U.S., with the ability to speak and read English. We recruited a sample representative of the adult U.S. population in terms of age, race, and gender, according to the U.S. Census. We initially recruited 530 survey respondents, of whom 30 did not begin the survey after reading the informed consent document, resulting in a total of 500 respondents. All 500 respondents finished the survey (zero incomplete responses) over a median time of 15 minutes and 24 seconds.

Data Collection

Our team designed and programmed the questionnaire using the Qualtrics XM platform. Participants received an invitation to participate in the questionnaire through Prolific then began the questionnaire by clicking on a secure, anonymous link to Qualtrics. Participants could complete the survey using any smartphone, tablet, or computer, provided they had an Internet connection. Prior to beginning the survey, participants read an information sheet and consented to participate in this specific study. Participants then completed the questionnaire.

There was no time limit for how long participants had to finish the questionnaire, and they had the ability to stop and come back to complete the questionnaire at a later time. Participants had the ability to discontinue the survey at any time. Participants who completed the full questionnaire were compensated at an hourly rate of $13.60 based on Prolific's policies.

Data Analysis

Quantitative Analysis

The first level of analysis involved assessing descriptive statistics to understand trends in participant perceptions and values. We also selected an outcome of interest (perceived benefit of AI for mental health), and calculated a logistic regression model to better understand if perceived benefits may differ by socio-demographic factors, specifically age, gender, race, education, financial resources, mental health history, and self-rated health literacy.[40]. The alpha value for all analyses was set at 0.05, and R version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria, 2022) was utilized. An analysis of a subset of this data (only those reporting female sex at birth) related to differences in perspectives based on pregnancy history has been reported in a separate manuscript.[61]

Qualitative Analysis

We analyzed free text responses using inductive thematic analysis and the constant comparative process. One analyst initially reviewed the codes and created a draft codebook. Free text responses to the two open-ended questions were analyzed using a singular coding scheme. A second analyst then used the coding scheme to independently dual code each free text response. The analysts met with a third team member to resolve discrepancies, coding via consensus, and updating the codebook throughout the discussion. Once detailed codes had been developed, and 50% of the initial coding was completed, the team completed axial coding, coming up with higher level summary themes to describe themes patterns in the detailed codes.

Results

Participant Characteristics

Table 1 describes the demographic makeup of the 500 adult, U.S.-based survey respondents, sampled using the Prolific platform.[39] Respondents were nationally representative based on race, age, and gender.

Table 1: Participant Demographics

Participant characteristic	Category	N =500 (%)
Age	Median (IQR)	46 (31, 59)
	Mean (SD)	46 (16)
	Range	18-93
Gender	Female	249 (50%)
	Male	238 (48%)
	Transgender	1 (0.20%)
	Something else	9 (1.8%)

	Prefer not to answer	3 (0.60%)
Race	Asian	25 (5.0%)
	Black or African American	66 (13%)
	White	388 (78%)
	Other/Prefer not to answer[a]	21 (4.2%)
Perceived financial resources	More than enough	65 (13%)
	Enough	271 (54%)
	Not enough	156 (31%)
	Prefer not to answer	8 (1.6%)
Mental health history[b]	Yes	215 (43%)
	No	271 (54%)
	Prefer not to answer	14 (2.8%)
Health literacy[c]	Adequate	369 (74%)
	Inadequate	131 (26%)

[a] Answer options included: American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, Prefer not to Answer, or the ability to select multiple options.
[b] Question asked, "Have you ever been told have mental illness?"
[c] Measured using the Chew et al.'s Brief Health Literacy Screener[40]

RQ1 - Perceived Benefits of AI for Mental Health

Participants were first asked, "Overall, in the next 5 years, do you think AI will make mental healthcare in the United States…" Answer options included "much better", "somewhat better", "minimal change", "somewhat worse", "much worse", and "don't know". We computed a logistic regression model such that "much better" and "somewhat better" were classified as 1, and the other responses were classified as 0, excluding the three participants answering, "don't know". Among 497 included respondents, 245 (49.3%) respondents believed that AI would make mental health care better or much better. Table 2 reveals that participants of Black/African American race (p=0.04, OR = 1.76) and those with lower health literacy (p = 0.004, OR = 2.19) were significantly more likely to endorse that AI would make mental healthcare somewhat or much better. Female gender participants, on the other hand, were significantly less likely to endorse this statement (p = 0.046, OR = 0.62).

Table 2: Logistic regression results – impact of socio-demographic variables on perceived benefit of AI for mental health

Variable	β Estimate	Odds ratio	P-value
(Intercept)	0.378	--	0.367
Age	-0.006	0.99	0.325
Gender (Female)	-0.388	0.68	0.046[a]
Race (Black/African American)	0.567	1.76	0.040[a]
Perceived financial Resources (Not Enough)	0.003	1.00	0.989
Mental Illness History (Yes)	-0.044	0.96	0.827
Health Literacy (Inadequate)	0.782	2.16	0.004[a]

a Statistically significant based on α <0.05

Concerns, Comfort with Predictive Tasks, and Values (Quantitative)

RQ2 - Concerns regarding AI for mental health

Based on the Khullar et al. survey, we asked participants their level of concern (very concerned, somewhat concerned, not concerned, don't know) related to six potential challenges of using AI for mental health (Figure 1). Participants reported being somewhat or very concerned about AI making the wrong diagnosis (89.9%), leading to inappropriate treatment (87.0%), or leading to them not knowing their mental health provider as well (81.8%) Participants had some concern (very or somewhat) regarding spending less time with their mental health professional (69.2%) and their confidentiality (60.4%), but expressed relatively less concern regarding increased costs (43.2%).

[INSERT FIGURE 1 HERE]

RQ3 - Comfort with AI accomplishing mental health tasks

We next asked patients their level of comfort with AI performing various tasks instead of their mental health professional. We assessed a range of tasks (i.e., assessment, diagnosis, diagnosis delivery, and treatment recommendation) and mental health issues of varied levels of perceived severity (i.e., depression, bi-polar disorder, suicide). Participants were most comfortable (reporting being very or somewhat comfortable) with recommendations of non-pharmacological interventions, including general wellness management strategies (71.5%) or talk therapy (65.5%). Participants expressed moderate comfort with AI performing a mental health assessment but were less comfortable (only 20-32% selecting very or somewhat comfortable) with various prediction, diagnosis, and diagnosis delivery tasks, as well as a medication recommendation task. Participants were least comfortable with diagnosis delivery tasks (i.e., telling someone directly they have a mental health condition) including for clinical depression (24.3%) and bi-polar disorder (20.3%).

[INSERT FIGURE 2 HERE]

We also asked participants their level of comfort sharing mental health information with a (human) mental health professional, an AI chatbot, or to improve an AI program that treats disease. We chose these categories to understand perspectives of common use of patient data for AI, such as using patient data to build models that make predictions or help treat diseases as compared to those who use patient information directly for patient support (e.g., and AI chatbot). Patients were most comfortable (very or somewhat) sharing with the mental health professional (77.8%), followed by helping to improve an AI program that treats disease (60.0%), then the AI chatbot (47.4%).

RQ4 - Values Related to AI for Mental Health

We replicated a series of questions from Khullar et al. assessing patient values of various aspects of AI, including transparency, explainability and performance, responsibility, and the effect of AI on trust in health professionals (Table 3).

Transparency of AI use: Participants were first asked how important it was to know when AI played a 1) small or 2) big role in their mental health treatment or diagnosis. The vast majority of participants found it somewhat or very important to know if AI played a small (90.2%) or big (95.0%) role in mental health treatment or diagnosis, although participants tended to report it was very important based on whether AI played a big (73.0%) vs. small role (50.6%). This pattern remained consistent with a specific scenario regarding use of an AI program in prescribing antidepressants, with many of the participants stating it was very important (71.0%) or somewhat important (22.6%) that their mental health professional informed them regarding the AI's involvement in this decision.

Explainability and performance: In comparing AI that was not explainable (i.e., it could not describe why it made a given diagnosis), participants were generally uncomfortable even with stated AI performance accuracies of 90% (76.0% somewhat or very uncomfortable) and 98% (57.8% somewhat or very uncomfortable).

Responsibility and AI: Participants answered a series of questions regarding who was responsible in the event of a medical error when AI was used in conjunction with their mental health treatment; answer options included: the mental health professional who made the decision, the company that made the computer program, the hospital or clinic that bought the computer program, the government agency that approved the computer program, someone else, no one, and don't know. In the case where AI was used in collaboration with a single mental health professional, the majority of participants reported that the mental health professional (over 80% of participants) would be the one responsible if a medical error (e.g. wrong diagnosis, unnecessary treatment) occurred for both a specific (i.e.., sleep disorder) and a general scenario.

Participants were more divided on who had responsibility for ensuring an AI program for mental health was safe with the plurality stating the company created the program was responsible (45.8%), followed by the mental health professional (27.0%), then the health system (14.6%), with fewer than 10% of participants selecting the each of the remainder of the answer options.

Effect of AI on trust in mental health professionals: The majority of participants (53.0%) said that if AI that was accurate 80% of the time in detecting health issues related to sleeping, eating, and concentrating disagreed with their mental health professional, it would make them question the health professional's assessment. Notably, 30.0% said they "did not know" how such information would change their view of their mental health professional's assessment, with the remaining 17% stating it would not change.

Table 3: Summary of questions regarding values related to AI for mental health.

Value	Question	Answer Choice	N =500 (%)
Transparency of AI use	How important do you think it is that you are told when an AI program has played a big role in your mental health diagnosis or treatment?	Not important	13 (2.6%)
		Somewhat important	111 (22.2%)
		Very important	364 (72.8%)

		Don't know	12 (2.4%)
	How important do you think it is that you are told when an AI program has played a small role in your mental health diagnosis or treatment?	Not important	39 (7.8%)
		Somewhat important	128 (25.6%)
		Very important	253 (50.6%)
		Don't know	10 (2.0%)
	[Clinical scenario][a] How important is it that your doctor tells you that the computer program helped make this decision?	Not important	25 (5.0%)
		Somewhat important	66 (13%)
		Very important	388 (78%)
		Don't know	21 (4.2%)
Explainability and performance	How comfortable would you be receiving a mental health diagnosis from a computer program that made the right diagnosis **<u>90%</u>** of the time but could not explain why it made the diagnosis?	Very comfortable	15 (3.0%)
		Somewhat comfortable	94 (18.8%)
		Somewhat uncomfortable	184 (36.8%)
		Very uncomfortable	199 (39.8%)
		Don't know	8 (1.6%)
	How comfortable would you be receiving a mental health diagnosis from a computer program that made the right diagnosis **<u>90%</u>** of the time but could not explain why it made the diagnosis?	Very comfortable	63 (12.6%)
		Somewhat comfortable	138 (27.6%)
		Somewhat uncomfortable	172 (34.4%)
		Very uncomfortable	117 (23.4%)
		Very comfortable	10 (2.0%)
		Don't know	63 (12.6%)
Responsibility and AI	Imagine that your mental health professional and a computer program work together to treat your mental illness and a medical error occurs. An example of a medical error is getting a diagnosis that was wrong, or a treatment that was not needed. Who is responsible? (Select all that apply.)[b]	The mental health professional	412 (82.4%)
		The company that made the computer program	30 (6.0%)
		The hospital or clinic that bought the computer program	20 (4.0%)
		The government agency that approved the computer program	6 (1.2%)
		Someone else	12 (2.4%)
		No one	1 (0.2%)

		Don't know	19 (3.8%)
	Imagine that you have a sleeping disorder that might be due to a mental health issue. You have a test done. Your doctor uses a computer program that says the sleeping disorder might be mental health-related, so you start medication to treat it. The medication leads to bad side effects. After another doctor evaluates your sleeping disorder, it turns out it was NOT mental health related. Who, if anyone, is to blame? (Select all that apply.)[b]	The mental health professional	408 (81.6%)
		The company that made the computer program	35 (7.0%)
		The hospital or clinic that bought the computer program	14 (2.8%)
		The government agency that approved the computer program	5 (1.0%)
		Someone else	6 (1.2%)
		No one	19 (3.8%)
		Don't know	13 (2.6%)
	Imagine that your hospital recently started using a computer program to help diagnose mental health problems. Who do you think has checked to make sure the computer program is safe before it is rolled out? (Select all that apply.)[b]	The mental health professional	135 (27.0%)
		The company that made the computer program	229 (45.8%)
		The hospital or clinic that bought the computer program	73 (14.6%)
		The government agency that approved the computer program	31 (6.2%)
		Someone else	4 (0.8%)
		No one	16 (3.2%)
		Don't know	12 (2.4%)
Effect of AI on trust in mental health professionals	Imagine that you have some symptoms that have been bothering you for a while, such as difficulty sleeping, eating, and focusing on work. You visit a doctor who runs some tests and he says he does NOT think you have any mental health issue. He also puts your symptoms into a computer program that can make the right diagnosis about 80% of the time, but can't say why it chose the diagnoses. It says you DO have mental health issue. How does the computer program affect your view?	It would not affect my trust of the mental health professional's assessment	85 (17.0%)
		It would make me question the mental health professional's assessment	265 (53.0%)
		I do not know if it would change my view of the mental health professional's assessment.	137 (27.4%)
		Don't know	13 (2.6%)

[a] Scenario wording: "Imagine that you have been told that you have been diagnosed with depression, a common mental illness that affects your mood, thoughts, and behavior. In the past, your doctor would have decided whether to prescribe a medication or refer you for psychotherapy depending on the type of symptoms you have and how severe they are. //Your doctor now has a computer program that uses many other factors. This computer program says you should start an antidepressant."

[b] Multi-select, so the sum of proportions will be greater than 100%

To better understand what participants valued most related to AI for their mental health, we asked the importance of various ethical constructs, based on MITRE's ethical framework for consumer-generated health information,[41] as they pertained to an example AI program used to support treatment for depression. Over 80% (range 80.2%-96.0%) of participants found each of the constructs somewhat or very important. Notably, the highest proportion of participants (80.3%) viewed explainability and transparency, "understanding individual risk factors" as very important. Participants tended to perceive decreasing the risk of negative outcomes as slightly more important than improving symptoms. Participants found AI **not** reducing trust in their mental health professional as least important by comparison, though 37.2% and 43.0% rated this trait as very and somewhat important, respectively.

[INSERT FIGURE 3 HERE]

Qualitative Results

Participants provided free text responses describing themes related to nuanced aspects of AI's performance; human-AI dynamics; and further values or concerns pertaining to AI. Free text responses were mandatory, but some participants simply stated they had no additional concerns (33.1%), or they did not provide sufficient detail for their responses to be categorized (1.3%). Ninety-seven of the 1,000 (two per each participant) responses involved more than one code, so percentages listed below reflect the proportion of total codes observed.

AI Performance

Table 4 provides the detailed codes, proportion of occurrence, and examples related to AI performance. Participants described issues related to AI's accuracy, biases in AI data or that may occur in the use of AI, concerns regarding potential errors AI may commit, and how this may affect quality of care. Participants expressed mixed opinions regarding if AI would improve or degrade the quality of mental health care.

Table 4. Detailed qualitative codes related to AI performance.

Detailed code	Count	%	Example quotes
AI Performance & Error - Accuracy	96	8.8%	I think the main concern would be potential for getting an inaccurate diagnosis or the wrong treatment. It would definitely take time to trust the reliability. Programming errors, for example, could potentially lead to fatal outcomes for patients.
Bias in Data & Use	50	4. 6%	I am concerned that the algorithms/data set that was utilized to train AI would be biased. For instance, if more white people seek mental health care, and AI is trained on their data, would AI be as good at diagnosing mental health conditions in people of color? It will be biased, sexist and racist. It will rely on old ideas of mental health care and not use current information. It will be used to ignore or bully patients.
Risk of Harm	29	2.6%	I think there is far too much on the line when it comes to mental health that it's risky to rely on AI for it. I think my main concern would be being over-diagnosed and having to be put in a psychiatric ward. I think that would be horrific.
Care Quality	13	1.2%	I think that it might lessen the quality standard of hired healthcare professionals since the expertise of AI system could become more important. It would improve the quality of work.
Total	235	21.4% [a]	

[a] Percentages do not add to 100%, as "no additional concerns" and "indeterminable" codes are included in the count.

AI and Humans - superior, inferior, or simply better together

Table 5 presents participant feedback related to human-AI dynamics. Participants described worry that AI may not be able to replicate things done by humans (AI capabilities, human reasoning and communication, importance of human connection). On the other hand, some pointed out ways in which AI may offer advantages to human cognition (AI capabilities). They also provided feedback on how AI and humans may (or may not) work together (human-AI collaboration, overreliance on AI), with many noting that AI should be overseen by humans and not work autonomously in mental health applications. Lastly, a few participants expressed concerns regarding how AI may take away jobs from humans.

Table 5. Detailed qualitative codes related to human-AI dynamics.

Detailed code	Count	%	Example quote
AI Capabilities	78	7.1%	A software program's understanding of mental health will never be as nuanced as a real humans. There will always be less common variables that AI systems aren't programmed to take into account. I am concerned that with AI driven healthcare, patient[s] with more unusual backgrounds, experiences and symptoms will not have access to human professional[s] who can more fully consider their circumstances. AI is better at playing human games than humans are - it's already World Go Champion and World Chess Champion…
Human-AI Collaboration	39	3.6%	While I think that AI will be useful in mental health scenarios, I think that oversight should still be done, just like how I would prefer a few doctors to confirm a diagnosis. I think that it will be good at detecting some trends that can help move people towards better help, but that the help itself should be a joint effort and more personalized. Overall, the use of AI is to assist the doctor in providing an accurate diagnosis. It improves the reliability of the diagnosis.
Overreliance on AI	30	2.7%	The biggest ethical concern I can think of is a mental health professional being completely reliant on AI without taking a closer look into how the program works.
Human Reasoning & Communication	54	4.9%	That it can't pick up on the subtleties of some symptoms that a human can.
Importance of Human Connection	90	8.2%	Human connection and understanding are crucial in mental health diagnosis and treatment. I finally found a doctor that made me feel understood, heard, and cared for. It resulted in an effective treatment for my major depression and suicidal ideology after many years. AI can't do that.
Jobs	6	0.6%	As someone who will be working in healthcare in the new future, I am concerned that AI in health scenarios will take away jobs from real people who put in all the work to be working there.
Total	297	27.1% [a]	

[a] Percentages do not add to 100%, as "no additional concerns" and "indeterminable" codes are included in the count.

Additional values and concerns

Respondents expressed further values and concerns beyond performance and human-AI dynamics, many of which were also covered in the closed-ended survey responses, including trust, transparency, privacy, responsibility, and cost (Table 6).

Table 6. Detailed qualitative codes related to additional values and concerns regarding the use of AI for mental health.

Detailed code	Count	%	Example quote
Privacy	98	8.9%	I would be concerned about the company that owns the AI and if they could share the data with third parties. I'd also be concerned about what happens if someone admitted suicidal thoughts.
Transparency	22	2.0%	I would worry about how my data is used to make AI based decisions. I would also wonder about the type of data being used.
Ethics	23	2.1%	One worry is that AI might be used to diagnose and treat mental health conditions without a person's consent. This could lead to people being treated for conditions they do not have, or not receiving treatment for conditions they do have…
Trust	26	2.4%	I am just not comfortable with any machine diagnosing and treating any symptoms of mine, bottom line.
Appropriate Use	19	1.7%	Yes, that a manipulative enough person could sway the machine into getting what they want rather than what they need.
Responsibility	14	1.3%	I wonder who would be held liable in the event a patient dies or experiences bad side effects due to the diagnosis or advice of AI. Would it be the AI itself or would the specialist also be held accountable?
Cost	20	1.8%	I think if the purpose is to provide better and more thorough health care, then it is a good endeavor. If the purpose is to decrease the costs of providing health care while maximizing profits, the project is specious. That's why I can see it as a diagnostic tool to help healthcare providers come to a more accurate and thorough diagnosis. But I think it's about maximizing profits and all stages of the healthcare process.
Total	222	20.2%[a]	

[a] Percentages do not add to 100%, as "no additional concerns" and "indeterminable" codes included in the count.

Discussion

This is one of the first studies to explore public perspectives related to the use of AI for mental health-related applications. Our results expand upon other work studying public perceptions of AI for non-mental health applications and raise important considerations regarding patient involvement in AI use for their mental health.[30,42] We also focused on various applications of AI to mental health, differentiating our results from previous user-centered design studies that have elicited participant perceptions of a single, specific AI tool under development. Our study highlights the nuances of patient perspectives regarding AI for mental healthcare such that their comfort with its use depends on the purpose of the AI (tasks it performs), process for use (when it is used, what factors drive predictions), and its performance (how well it works, and what happens when it is wrong).[43]

Perceived benefits of AI for mental health (RQ1)

Just under half of the respondents in our study reported that they thought AI would make mental healthcare better or somewhat better. This is similar to a study conducted in Germany where 53% of patients reported positive or very positive attitudes towards AI, but not specifically for mental health.[44] Participants in other studies asking perceptions regarding more specific applications of AI (e.g., radiology image interpretation, pregnancy and postpartum) had stronger positive attitudes towards AI.[45,46] This is also consistent with qualitative studies that have found patients with specific health challenges more readily connected with AI's potential benefits. Our study also found that female gender was associated with lower perceived benefits of AI for mental health, while lower self-rated health literacy and Black/African American race were each associated with more positive perceptions. The previously cited study conducted in Germany had similar findings related to lower perceived benefit among females.[44] Although, it is interesting that this result remained consistent for mental health applications, in light of the fact that females have reported lower levels of stigma regarding mental health than males.[47,48] Our study also detected an interesting paradox such that those having lower self-reported health literacy had more positive perceptions towards AI for mental health. While this finding warrants further replication and investigation, it highlights the importance of inclusivity of patient-facing information regarding AI, ensuring those of various levels of literacy and numeracy may equitably comprehend its functions. Those of African American and Black race in the United States consistently report greater stigma and lower levels of trust towards mental (and other) health institutions due to biases, discrimination, and systemic racism.[49,50] The greater perceived benefit of AI for mental health may represent a view that AI can be more just and lacking the biases of humans, but this notion also requires further exploration, particularly given that the biases of human can often be embedded into the AI since training data embodies previous human behavior.

Concerns regarding AI for mental health (RQ2)

Participants in our study cited concerns consistent with previous work related to AI accuracy, risk of harm (e.g., wrong diagnosis, inappropriate treatment),[42,51,52] decreased human communication/connection,[34,42,46,51] and issues pertaining to confidentiality.[34,42,46,52–55] Issues related to privacy were also the most commonly mentioned concern in the qualitative feedback. Participants also qualitatively described concerns about the performance of AI and doubts in AI's ability to truly replicate human reasoning. In our study, participants expressed some concern related to rising costs, though, as found in other studies, this worry was less pronounced.[34,56] These results continue to stress the importance of contextualizing for patients: the accuracy of AI, harms and how they are mitigated, and data use and protections. While it has also been mentioned as a concern in previous studies, continuing to support human connection is particularly needed in mental health applications given the importance of the patient - mental health professional therapeutic relationship.

Comfort with AI accomplishing mental health tasks (RQ3)

Our study provided further evidence that patient comfort with AI varies based on what the AI is doing. People were least comfortable with diagnosis delivery tasks, which lends further support to the importance of continuing to keep health professionals in the loop related to AI.[34] Similar to a previous study of pregnant people,[45] patients were most comfortable with tasks that recommended general wellness strategies or talk therapy. These results also suggest that AI for tasks patients are less comfortable with may require greater care to explain and also emphasize how the AI works with the health professional.

Despite the rapid proliferation of chatbots,[36,57] less than half of participants were comfortable sharing mental health information with a chatbot, which may simply signify that these types of tools should be usable on an opt-in basis. Previous studies have suggested that it may be easier for someone to share these sensitive feelings with a computer or AI,[58] and this may be true for certain people, but our findings did not universally support this assertion. It was also notable that almost a quarter of the sample was not comfortable sharing mental health information with a mental health professional. We acknowledge this may have been impacted by the types of mental health information listed in our survey, but it may also represent a continued stigma related to sharing mental health concerns.

Values related to AI for Mental Health (RQ4)

Findings in our study related to patient values for AI for mental healthcare revealed challenges AI integration may present to the patient-health professional relationship. In individual scenarios, patients overwhelmingly found mental health professionals responsible for AI-related errors. While this does reflect similarity to the current standard of practice (i.e., that health professionals are held responsible for medical errors even when computer systems are involved), future work should consider how this affects health professional well-being given the challenges related to burnout and shortages in trained mental health workers. It also supports programs, such as AI falling under the U.S. Food and Drug Administration's purview and the European Union's AI Act of 2023, where algorithms may be reviewed prior to use and subject to regulations.[59] Participants also noted that their trust in their mental health professional would decrease if their assessment disagreed with AI. Although, this was somewhat at odds with relatively fewer patients viewing issues with AI decreasing trust in their mental health professional as "very important". It was also notable that almost a third of participants said they "did not know" how this scenario would affect trust in their mental health professional, which seems to highlight that patients may still be wrapping their heads around feelings regarding emerging technologies, such as AI.

Patients desired a high degree of transparency related to AI use as over 90% of participants also found it important they be told when AI played even a small role in their care. Participants also valued explainability, such that the majority of participants were not comfortable with highly accurate (98%) AI that could not explain how it made its predictions and that "understanding individual risk factors" was rated as the most important provided value related to AI for a mental health application. It is at best unclear what patients are typically told regarding when AI is used for their care, how explainable it is, and to what extent, if at all, they are informed what factors drive predictions regarding their care. These results suggest that patient values may be at odds with the current standard of practice for patient communication. Similar to the concerns previously described, participants also rated highly the importance of AI not leading to errors and that it helps with their mental health symptoms.

Implications and future challenges

Navigating the patient - health professional relationship

The therapeutic relationship between patient and health professional is crucial in mental health settings. Our study revealed issues that will need to be reconciled if AI is to be safely, transparently, and acceptably utilized for mental healthcare. From our results, it is clear that patients want mental health professionals to be the ultimate decision makers, using AI to support (but not make) decisions, when the AI is deemed safe and effective.[35] Participants also overwhelmingly viewed mental health professionals as the people responsible if an error occurred related to treatment where AI had been utilized. While this is important for mental health professionals to be aware of this, future regulations may consider some level of shared responsibility between the health professional, AI developer, and relevant regulating bodies to ensure mental health professionals are adequately protected. It may also be helpful to explore what type of competencies health professionals may need when interacting with AI so they may use it safely, effectively, and efficiently, if there is evidence that it can improve patient outcomes.

Given these conflicting issues, it will be important to consider how AI workflows and interfaces may be designed to support collaborative patient-health professional decision making in a way that fosters trust instead of degrades it, while also not creating undue burden for the health professional. Previous studies have described how clinicians should be able to contest AI, such as ignoring it (when it is not relevant or appropriate), trusting it when it is appropriate, or being able to uncover explanations to negotiate in borderline cases.[60] Such systems should be able to track health professional decisions in relation to the AI, possibly allowing health professionals to provide brief rationale that they may use in conversations with patients. In creating such systems usability, and model explainability will be critical.

Communicating AI-related information to patients

Participants in this study desired various information regarding the use of AI for their mental healthcare – to know when it was used, for what, why, how accurate it was, and the risk factors that drove a decision. Even with highly predictive AI, patients were still not comfortable with AI that could not explain how it arrived at a result, and they qualitatively expressed concerns related to misdiagnosis and improper treatment that could result from AI. They also reported "understanding individual risk factors" as the most important value. It is unclear as to how much if any patients currently receive regarding AI's role in their treatment, and there seems to be a mismatch between the current deployment of AI for mental health and patients' desires. At minimum, we need to promote transparency in AI's use, communication of its accuracy, and possibly include individual risk factors to help patients and clinicians decide when AI may be appropriate for use in health-related applications.

Communicating the desired information to patients, however, is not straightforward as concepts such as AI performance and process involve complex mathematical concepts. Furthermore, this desire for additional information regarding AI is also at odds with how patient communication has traditionally been practiced. When we consider other non-AI based diagnostic or decision support tools (e.g., MRI's, blood tests, screening assessments), communicating information regarding how they work (i.e., their process) or their performance, is far from standard practice. AI seems to be held to a higher standard than other diagnostic tools related to transparency in performance. Future work may consider not only communicating the process and performance of AI but providing it in the context with the performance of the existing approach to a given task. This would allow health professionals and patients to determine if the potential benefits from AI outweigh their concerns.

Providing patients with explanations of AI performance and factors driving prediction will require extensive study involving the experts in human-centered, inclusive design working alongside AI developers, mental health professionals, and patients. There is a need for lab-based studies to understand what information regarding AI balances recognizing patient values, but also supports comprehension of important concepts and fostering appropriate trust. These issues raise many questions for numeracy and data visualization experts regarding how information may be conveyed inclusively to patients with different needs, so that the benefits of AI may be equitably realized.

Individuality and autonomy

Our study detected differences in who may find AI beneficial, and for what tasks they may be comfortable using it. Future work should explore how we may respect individual autonomy with regards to the utilization of AI for mental health applications.

Limitations

Our study was limited in that the sample was recruited using an online platform, which may not generalize to those with technology, literacy, and other barriers to online survey completion. The sample was representative of the U.S. based on age, gender, and race distributions,15 leading to a majority of survey respondents being White due to over 70% due to the U.S. demographic makeup.30 Therefore, important perspectives from other racial/ethnic groups may be limited.

Conclusions

Our study found that approximately half of U.S. adults surveyed perceived some benefit for the use of AI in mental healthcare applications. These perceived benefits were lower among women but higher among Black/African American participants and those with lower self-rated health literacy. Participants also expressed nuanced differences in the types of tasks they would be comfortable with AI completing, showing the greatest discomfort regarding clinical diagnosis, diagnosis delivery, and recommendation of medication. Those surveyed valued high-performing AI that could explain individual risk factors driving predictions. In general, participants were also concerned that AI may mean loss of human connection, and they perceived humans as the ultimate decision makers, using AI as an extra data point when appropriate. Qualitative feedback also revealed participants' deeply seeded fears regarding the use of AI for their mental healthcare. These findings stress the importance of working with patients and mental health professionals to understand if and how AI may be safely, ethically, and acceptably implemented for mental health applications.

Acknowledgements

This work was supported by the National Institute for Mental Health (R41MH124581-S1; PI: Pathak)

Author contributions

NCB and MRT jointly conceptualized the study design and questionnaire materials with input from all other authors. JK, PMD, and ZR completed the data analysis advised by NCB and MRT. NCB drafted the manuscript with substantive input from all other authors.

Competing interests statement

The authors have no competing interests to report.

Data availability statement

Data are available upon reasonable request to the corresponding author.

Abbreviations

OR – Odds Ratio
AI – Artificial Intelligence
ML – Machine Learning
U.S. – United States

References

1. Science N, Technology Council. *Preparing for the Future of Artificial Intelligence*. Executive Office of the President; 2016. Accessed January 3, 2023. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

2. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. *Deep medicine: how artificial intelligence can make healthcare human again*. Published online 2019.

3. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. *J Med Internet Res*. 2019;21(4):e12286.

4. Ting DSW, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. *Nat Med*. 2018;24(5):539-540.

5. Bruce G. Predictive or generative AI: Which will change healthcare the most? Accessed

September 11, 2023. https://www.beckershospitalreview.com/innovation/predictive-or-generative-ai-which-will-change-healthcare-the-most.html

6. TrueFidelity CT \textbar GE Healthcare (United States). Accessed January 3, 2023. https://www.gehealthcare.com/products/truefidelity

7. Caption Health. Caption AI. Accessed January 3, 2023. https://captionhealth.com/technology

8. Viz.ai. Accessed January 3, 2023. https://www.viz.ai/

9. Aidoc. Accessed January 3, 2023. https://www.aidoc.com/

10. IBM Watson Health. Enterprise imaging solutions. Accessed January 3, 2023. https://www.ibm.com/watson-health/solutions/enterprise-imaging

11. Siemens. AI-Rad Companion. Accessed January 3, 2023. https://www.siemens-healthineers.com/en-us/digital-health-solutions/digital-solutions-overview/clinical-decision-support/ai-rad-companion

12. Seal A, Bajpai R, Agnihotri J, Yazidi A, Herrera-Viedma E, Krejcar O. Deprnet: A deep convolution neural network framework for detecting depression using EEG. *IEEE Trans Instrum Meas*. 2021;70:1-13.

13. Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *J Affect Disord*. 2021;279:1-8.

14. Flesia L, Monaro M, Mazza C, et al. Predicting Perceived Stress Related to the Covid-19 Outbreak through Stable Psychological Traits and Machine Learning Models. *J Clin Med Res*. 2020;9(10). doi:10.3390/jcm9103350

15. Sundaram L, Bhat RR, Viswanath V, Li X. DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. *Hum Mutat*. 2017;38(9):1217-1224.

16. Sheikh M, Qassem M, Kyriacou PA. Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring. *Frontiers in digital health*. 2021;3:662811.

17. Reale C, Novak LL, Robinson K, et al. User-Centered Design of a Machine Learning Intervention for Suicide Risk Prediction in a Military Setting. *AMIA Annu Symp Proc*. 2020;2020:1050-1058.

18. Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis Oncol*. 2020;4:19.

19. Lin E, Lin CH, Lane HY. Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *Int J Mol Sci*. 2020;21(3). doi:10.3390/ijms21030969

20. Girosi. *Narrative Review and Evidence Mapping: Artificial Intelligence in Clinical Care*. Patient-Centered Outcomes Research Institute; 2021.

21. Zhou Q, Chen ZH, Cao YH, Peng S. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. *npj Digital Medicine*. 2021;4(1):154.

22. Lee TC, Shah NU, Haack A, Baxter SL. Clinical Implementation of Predictive Models Embedded within Electronic Health Record Systems: A Systematic Review. *Informatics (MDPI)*. 2020;7(3). doi:10.3390/informatics7030025

23. Benda NC, Das LT, Abramson EL, et al. "How did you get to this number?" Stakeholder needs for implementing predictive analytics: a pre-implementation qualitative study. *J Am Med Inform Assoc*. 2020;27(5):709-716.

24. Matthiesen S, Diederichsen SZ, Hansen MKH, et al. Clinician Preimplementation Perspectives of a Decision-Support Tool for the Prediction of Cardiac Arrhythmia Based on Machine Learning: Near-Live Feasibility and Qualitative Study. *JMIR human factors*. 2021;8(4):e26964.

25. Baxter SL, Bass JS, Sitapati AM. Barriers to implementing an artificial intelligence model for unplanned readmissions. *ACI Open*. 2020;4(2):e108-e113.

26. Nong P, Adler-Milstein J. Socially situated risk: challenges and strategies for implementing algorithmic risk scoring for care management. *JAMIA open*. 2021;4(3):ooab076.

27. Salwei ME, Anders S, Novak L, et al. Preventing clinical deterioration in cancer outpatients: Human centered design of a predictive model and response system. *J Clin Orthod*. 2022;40(16_suppl):e13567-e13567.

28. Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, et al. Application of Artificial Intelligence in Shared Decision Making: Scoping Review. *JMIR Med Inform*. 2022;10(8):e36199.

29. Walsh CG, McKillop MM, Lee P, Harris JW, Simpson C, Novak LL. Risky business: a scoping review for communicating results of predictive models between providers and patients. *JAMIA open*. 2021;4(4):ooab092.

30. Khullar D, Casalino LP, Qian Y, Lu Y, Chang E, Aneja S. Public vs physician views of liability for artificial intelligence in health care. *J Am Med Inform Assoc*. 2021;28(7):1574-1577.

31. Act CC. Public Law 114--255. 114 th Congress. Published online 2021.

32. Dai T, Tayur S. Designing AI-augmented healthcare delivery systems for physician buy-in and patient acceptance. *Prod Oper Manag*. Published online September 9, 2022. doi:10.1111/poms.13850

33. Ghassemi M, Mohamed S. Machine learning and health need better values. *NPJ Digit Med*. 2022;5(1):51.

34. Khullar D, Casalino LP, Qian Y, Lu Y, Krumholz HM, Aneja S. Perspectives of patients about artificial intelligence in health care. *JAMA network open*. 2022;5(5):e2210309.

35. Yarborough BJH, Stumbo SP. Patient perspectives on acceptability of, and implementation preferences for, use of electronic health records and machine learning to identify suicide risk. *Gen Hosp Psychiatry*. 2021;70:31-37.

36. Shan Y, Ji M, Xie W, Lam KY, Chow CY. Public Trust in Artificial Intelligence Applications in Mental Health Care: Topic Modeling Analysis. *JMIR Hum Factors*. 2022;9(4):e38799.

37. Frankel J. How Artificial Intelligence Could Help Diagnose Mental Disorders. *The Atlantic*. Published online August 23, 2016. Accessed September 12, 2023. https://www.theatlantic.com/health/archive/2016/08/could-artificial-intelligence-improve-psychiatry/496964/

38. Hudson KL, Collins FS. The 21st century cures act - A view from the NIH. *N Engl J Med*. 2017;376(2):111-113.

39. Prolific · academic researchers. Accessed December 8, 2023. https://www.prolific.com/academic-researchers

40. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med*. 2004;36(8):588-594.

41. Granger E, Skopac J. An ethical framework for the use of consumer generated data in health care. In: *2021 Health Datapalooza and National Health Policy Conference*. AcademyHealth; 2021. Accessed December 8, 2023. https://academyhealth.confex.com/academyhealth/2021hdpnhpc/meetingapp.cgi/Paper/44315

42. Esmaeilzadeh P, Mirzaei T, Dharanikota S. Patients' Perceptions Toward Human–Artificial Intelligence Interaction in Health Care: Experimental Study. *J Med Internet Res*. 2021;23(11):e25856.

43. Benda NC, Novak LL, Reale C. Trust in AI: why we should be designing for APPROPRIATE reliance. *Journal of the American*. Published online 2022. https://academic.oup.com/jamia/article-abstract/29/1/207/6415464

44. Fritsch SJ, Blankenheim A, Wahl A, et al. Attitudes and perception of artificial intelligence in healthcare: A cross-sectional survey among patients. *Digit Health*. 2022;8:20552076221116772.

45. Oprescu AM, Miró-Amarante G, García-Díaz L, et al. Towards a data collection methodology for Responsible Artificial Intelligence in health: A prospective and qualitative study in pregnancy. *Inf Fusion*. 2022;83-84:53-78.

46. Bahakeem BH, Alobaidi SF, Alzahrani AS, et al. The General Population's Perspectives on Implementation of Artificial Intelligence in Radiology in the Western Region of Saudi Arabia. *Cureus*. 2023;15(4):e37391.

47. Corrigan PW, Watson AC. The stigma of psychiatric disorders and the gender, ethnicity, and education of the perceiver. *Community Ment Health J*. 2007;43(5):439-458.

48. Chatmon BN. Males and Mental Health Stigma. *Am J Mens Health*. 2020;14(4):1557988320949322.

49. Novacek DM, Hampton-Anderson JN, Ebor MT, Loeb TB, Wyatt GE. Mental health ramifications of the COVID-19 pandemic for Black Americans: Clinical and research recommendations. *Psychol Trauma*. 2020;12(5):449-451.

50. Eylem O, de Wit L, van Straten A, et al. Stigma for common mental disorders in racial minorities and majorities a systematic review and meta-analysis. *BMC Public Health*. 2020;20(1):879.

51. Wu C, Xu H, Bai D, Chen X, Gao J, Jiang X. Public perceptions on the application of artificial intelligence in healthcare: a qualitative meta-synthesis. *BMJ Open*. 2023;13(1):e066322.

52. Richardson JP, Curtis S, Smith C, et al. A framework for examining patient attitudes regarding applications of artificial intelligence in healthcare. *Digit Health*. 2022;8:20552076221089084.

53. Nelson CA, Pérez-Chada LM, Creadore A, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA Dermatol*. 2020;156(5):501-512.

54. Paprica PA, de Melo MN, Schull MJ. Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study. *CMAJ Open*. 2019;7(1):E40-E46.

55. Thenral M, Annamalai A. Challenges of Building, Deploying, and Using AI-Enabled Telepsychiatry Platforms for Clinical Practice Among Urban Indians: A Qualitative Study. *Indian J Psychol Med*. 2021;43(4):336-342.

56. Harnkham N. *Artificial Intelligence in Medicine: Measuring Respondents' Trust in Artificial Intelligence for Better Patient Care*. William Howard Taft University; 2023. http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/artificial-intelligence-medicine-measuring/docview/2800671825/se-2

57. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *J Med Internet Res*. 2021;23(1):e17828.

58. Miner A, Chow A, Adler S, et al. Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect. In: *Proceedings of the Fourth International Conference on Human Agent Interaction*. HAI '16. Association for Computing Machinery; 2016:123-130.

59. EU AI Act: first regulation on artificial intelligence. Published August 6, 2023. Accessed September 12, 2023. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

60. Sivaraman V, Bukowski LA, Levin J, Kahn JM, Perer A. Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Association for Computing Machinery; 2023:1-18.

61. Reading Turchioe M, Harkins S, Desai P, et al. Women's perspectives on the use of artificial intelligence (AI)-based technologies in mental healthcare. *JAMIA Open*. 2023;6(3):ooad048.

Figures

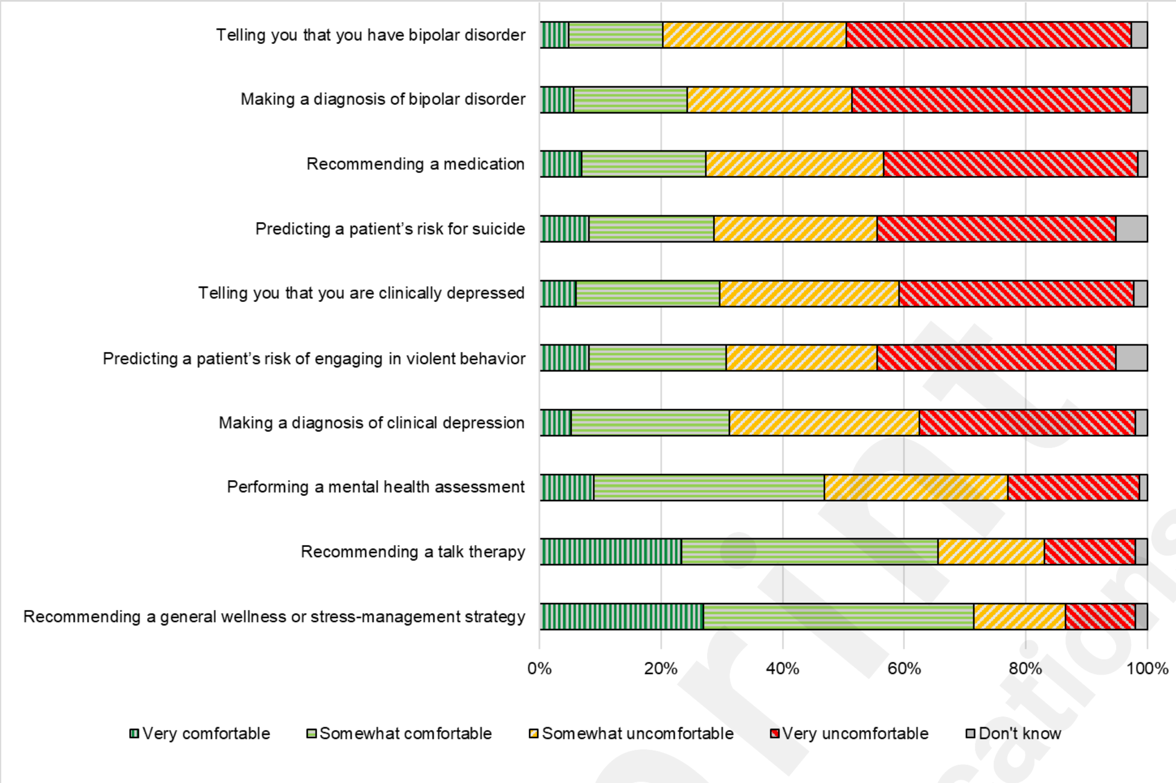


Figure 1: Reported levels of perceived concerns regarding AI use for mental health.

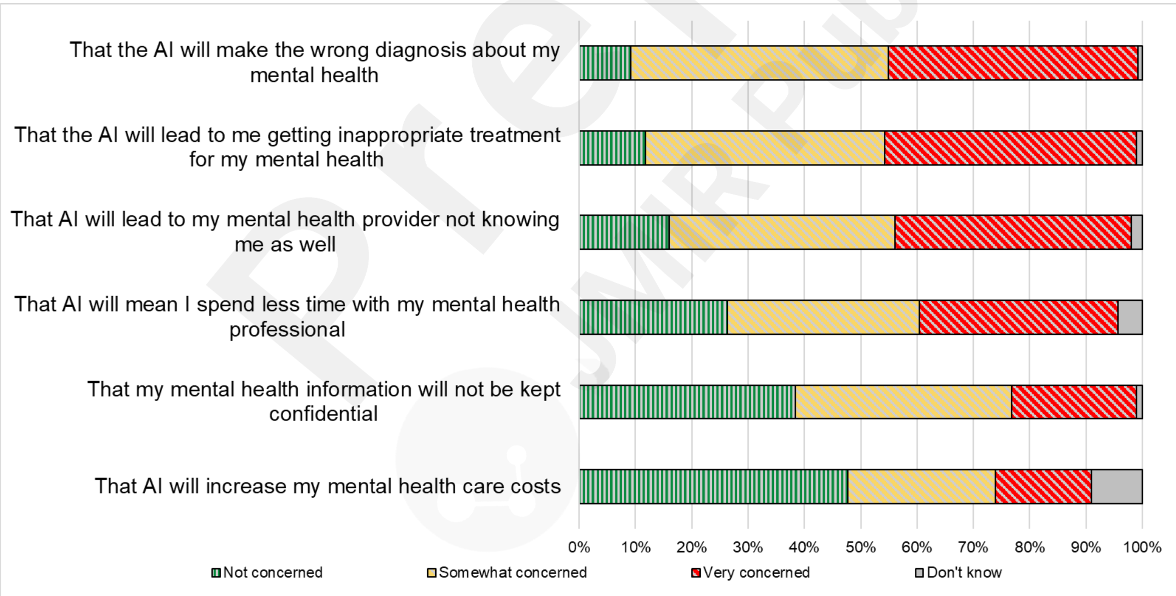


Figure 2: Reported level of comfort with AI instead of a mental health professional conducting various tasks.

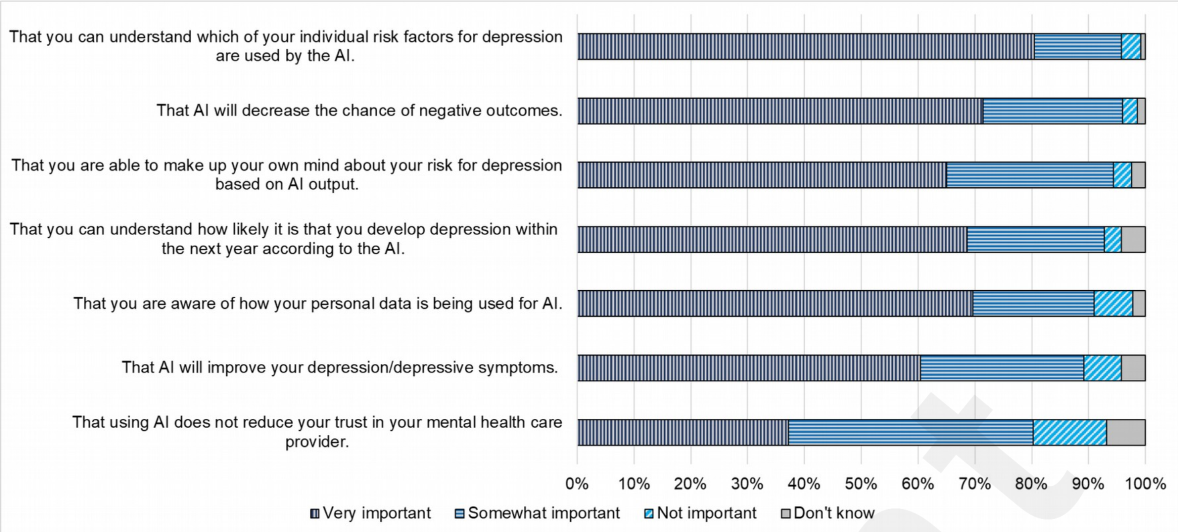
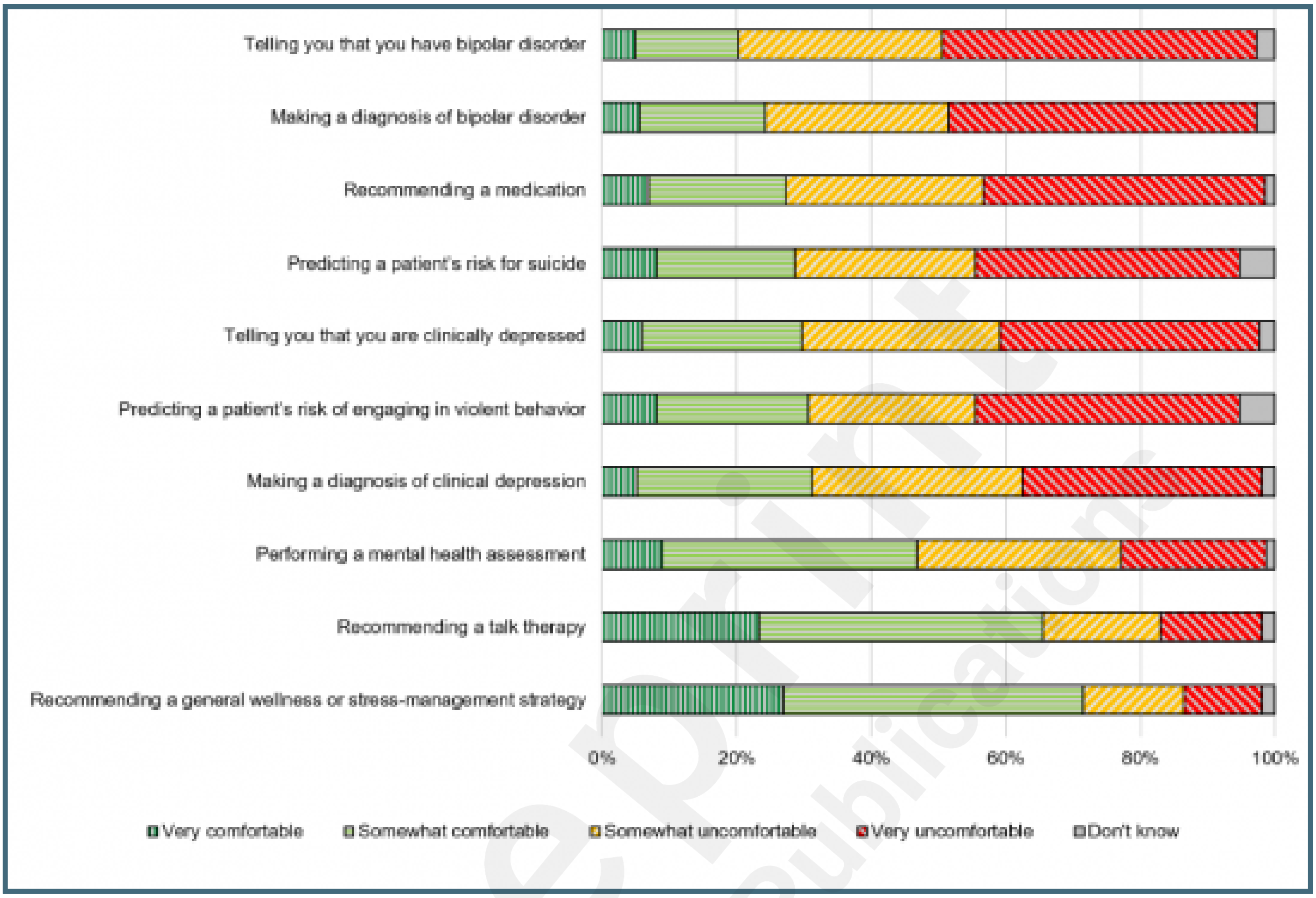


Figure 3: Importance of various values related to AI use for mental health.

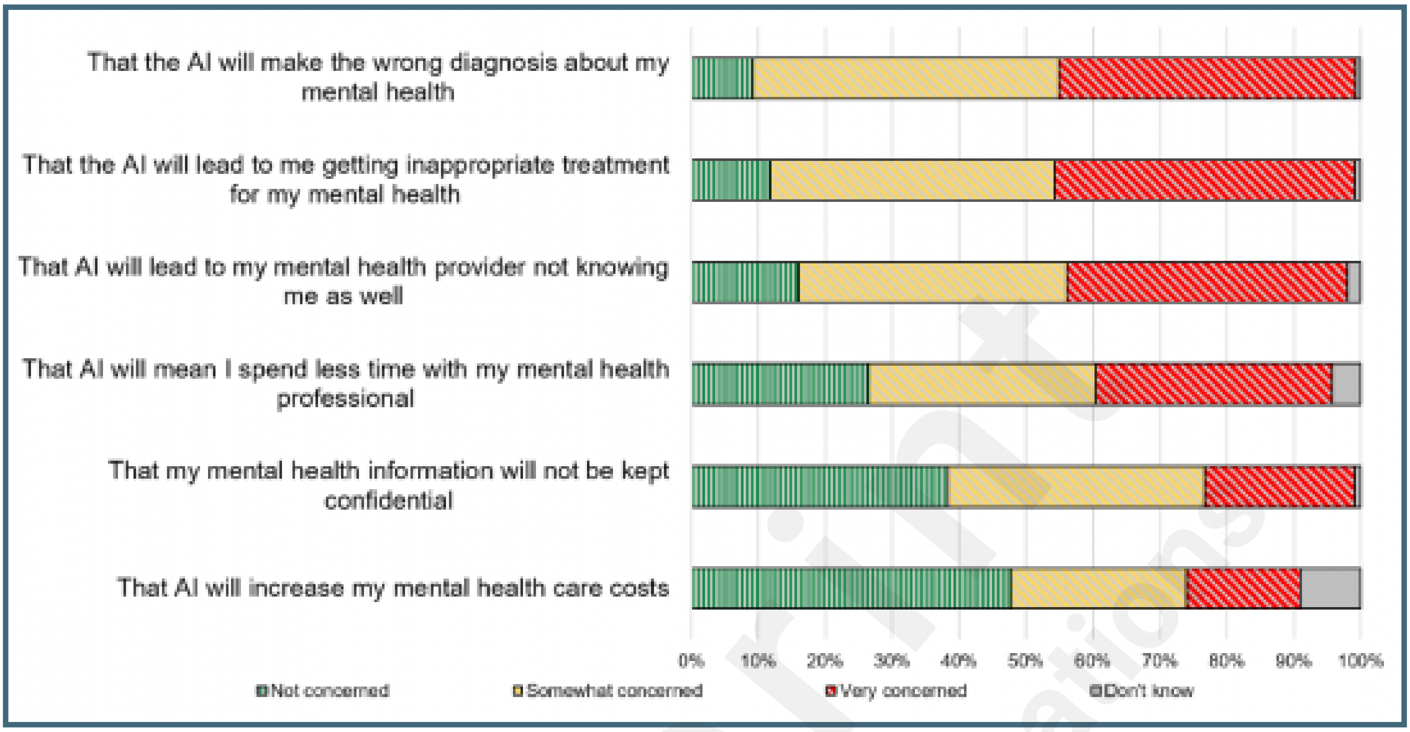
Supplementary Files

Figures

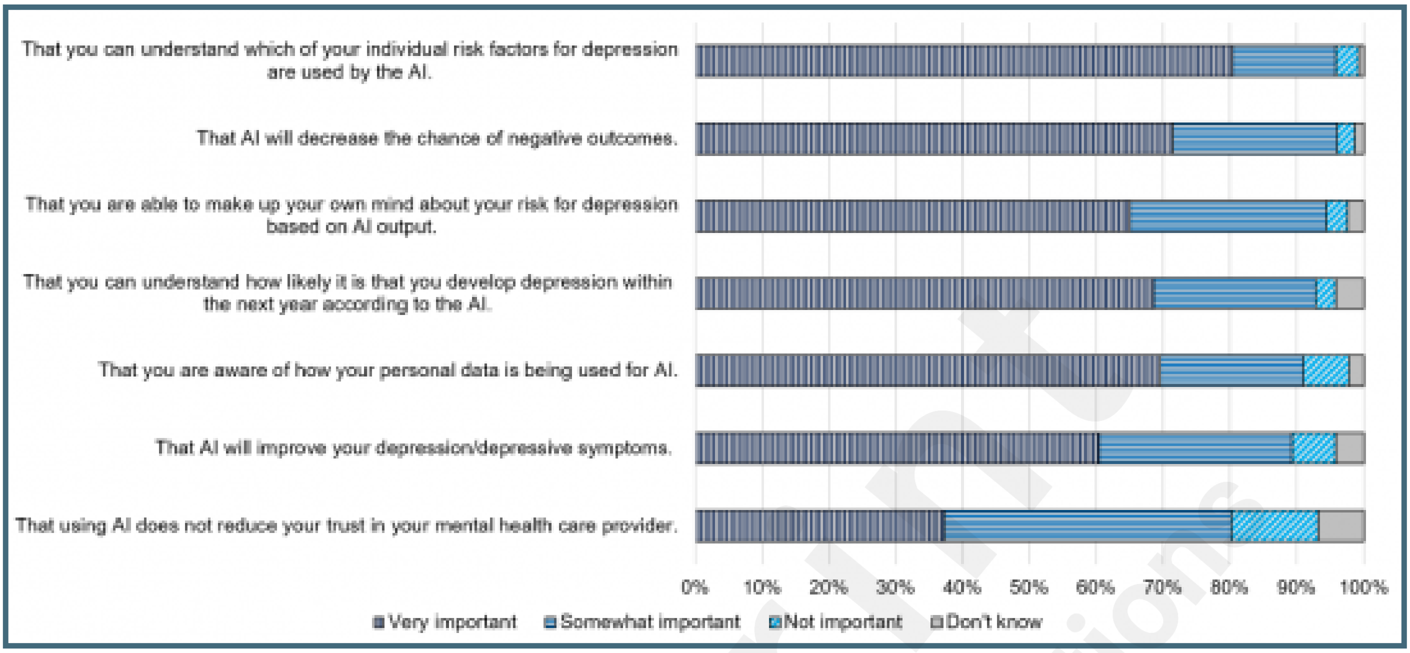
Reported levels of perceived concerns regarding AI use for mental health.



Reported level of comfort with AI instead of a mental health professional conducting various tasks.



Importance of various values related to AI use for mental health.



Multimedia Appendixes

Full survey battery.
URL: http://asset.jmir.pub/assets/842efe5e34fea84b33ac87152f93455b.docx

Related publication(s) - for reviewers eyes onlies

Publication of a subset of results only related to those with female sex at birth.
URL: http://asset.jmir.pub/assets/60ef514e9493916fe5baba58336d39aa.pdf