Leveraging AI to Deliver Culturally Responsive Mental Health Care at Scale

Alison Cerezo, David Cooper, Vijaykumar Palat, Amber Jolley, Sarah Peregrine Lord

Submitted to: JMIR Mental Health
on: January 31, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Leveraging AI to Deliver Culturally Responsive Mental Health Care at Scale

Alison Cerezo[1*] PhD; David Cooper[1*] PsyD; Vijaykumar Palat[1] MS; Amber Jolley[1*] PhD; Sarah Peregrine Lord[1*] PsyD

[1]mpathic Bellevue US
[*]these authors contributed equally

Corresponding Author:
Alison Cerezo PhD
mpathic
14655 Bel-Red Rd.
Bellevue
US

Abstract

There has been exponential growth in digital health technologies in the last few years, including the use of artificial intelligence (AI) in care delivery. This rapid growth has happened alongside growing disparity in mental health care for racial and ethnic minority (REM) patients. In this paper, we discuss how AI can champion REM mental health equity when developed within a culturally responsive framework. We describe how AI can serve as a layer of provider support that considers a patient's symptoms within their cultural context—specifically how culture shapes beliefs and behaviors towards mental health. We will address common challenges such as how to develop and employ robust, culturally responsive training data, how to center REM lived experience in product development and refinement, and introduce an AI Ethics framework for the use of AI in mental health care delivery.

(JMIR Preprints 31/01/2024:56965)
DOI: https://doi.org/10.2196/preprints.56965

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

Leveraging AI to Deliver Culturally Responsive Mental Health Care at Scale

Alison Cerezo, Ph.D.[1, 2]

David Cooper, Psy.D.[2]

Vijaykumar Palat, M.S.[2]

Amber Jolley, Ph.D.[2]

Sarah Peregrine Lord, Psy.D., ABPP, [2, 3]

1. University of California Santa Barbara, Department of Counseling, Clinical & School Psychology, Santa Barbara, CA; 2137 EDUC, UCSB, Santa Barbara, CA 93106
2. Empathy Rocks, Inc. d/b/a mpathic, Bellevue, WA; 14655 Bel-Red Road; Suite 203; Bellevue, WA 98007
3. University of Washington, Department of Psychiatry and Behavioral Sciences, Seattle, WA; 1959 NE Pacific Street (Box 356560), Seattle, WA 98195-6560

Abstract

There has been exponential growth in digital health technologies in the last few years, including the use of artificial intelligence (AI) in care delivery. This rapid growth has happened alongside growing disparity in mental health care for racial and ethnic minority (REM) patients. In this paper, we discuss how AI can champion REM mental health equity when developed within a culturally responsive framework. We describe how AI can serve as a layer of provider support that considers a patient's symptoms within their cultural context—specifically how culture shapes beliefs and behaviors towards mental health. We will address common challenges such as how to develop and employ robust, culturally responsive training data, how to center REM lived experience in product development and refinement, and introduce an AI Ethics framework for the use of AI in mental health care delivery.

Keywords (5): AI, Culturally Responsive AI, Health Equity, Racial and Ethnic Minorities, Machine Learning

Introduction to the Problem

It is estimated that 57.8 million, approximately one in five U.S. adults, live with a mental illness and more than half do not receive care [1]. The United States spent $280 billion on mental health care in 2020 alone, demonstrating a growing health need [2]. These costs were confirmed in the “Stress in America” report with one in four Americans reporting feeling “too stressed to function on most days.” [3]. Many racial and ethnic minorities (REM) groups report a higher risk of persistence and disability from mental illness than their White counterparts [4]. Research has shown that among REM patients who seek mental health care, 30% terminate prematurely, which contributes to less effective treatment outcomes [5]. Primary drivers of drop-out among REM include diminished client involvement and weak patient-provider alliance [6, 7], with some REM expressing concern that providers are ill-equipped to respond to their mental health needs in an empathic, affirming, and culturally responsive manner [8-10].

Psychologists have argued that Americans are “experiencing the psychological impacts of a collective trauma [11].” Several high-profile killings of Black Americans in the summer of 2020 brought the topic of racial bias to the forefront of collective consciousness. This social reckoning intersected with the COVID-19 pandemic to further unmask social, financial, and health disparities facing REM. In fact, Black and Hispanic Americans were more likely to report mental health distress during the pandemic and at the same time were significantly less likely to initiate mental health service use [12]. Over the last decade–preceding contemporary stressors–REM were already facing significant increases in drug overdose and death by suicide. Black and Hispanic suicide was more than two times greater than Non-Hispanic White adults (43% and 27% compared to 12%, respectively) [13]. Between ongoing social unrest and the disproportionate impact of the pandemic on REM [14], there is a dire need for innovative solutions that can help to address the mental health equity crisis in the United States.

Racial bias in healthcare is commonplace. REM patients regularly describe patient-provider interactions as involving less collaborative communication, fewer positive emotions, and reduced patient-centered care practices [15-18]. Outside of general concerns of cultural mismatch between patients and providers, a national survey of 2,212 REM psychotherapy patients found that 81% reported experiencing at least one racial microaggression in therapy [19]. These microaggressions were described by REM patients as involving bias, avoidance of discussing cultural issues, and denial or lack of awareness of stereotypes [19]. Although mental health providers often look to the Multicultural Orientation Framework (MCO) as a means to embed cultural responsiveness in service delivery, there is a clear gap in the translation of learning to practice. The MCO links the American Psychological Association (APA)’s Multicultural Guidelines (originally introduced in 1990, revised in 2017) to psychotherapy training by focusing on *“how the cultural worldviews, values, and beliefs of the client and the therapist interact and influence one another to cocreate a relational experience that is in the spirit of healing*” [20]. Delivering culturally responsive mental health care demands that the provider is aware of and responsive to the intersections of societal context, culture, and power in the client’s mental health experience [21-23] and is a key ingredient for retaining REM in care [24]. Coaching providers to attend to the MCO is more important than ever. Innovative solutions must be employed to support the needed upscale in providers' cultural responsiveness to improve REM patient outcomes.

Without an intentional focus on building technology within a culturally responsive framework, new mental health tools and applications hold the potential to exacerbate mental health disparities. Artificial intelligence (AI) that underlies digital mental health products must therefore be designed to promote and meet the cultural needs of REM patients. This begins with a cultural orientation towards diverse symptom presentation among some REM groups for accurate diagnosis [25], intervention preferences among certain REM patients [26], and building a strong relational alliance that is considerate of the client's cultural experience in the world [27]. To build culturally responsive AI, researchers recommend devising clear solutions to addressing algorithmic justice and data diversity [28]. This includes a strategic approach in the AI development lifecycle that accounts for AI bias and health equity across data management, model building, training, and deployment [28].

When developed within a robust cultural framework with attention to balanced curation and mitigation of AI-bias, AI can detect and correct for empathic communication in real-time, including improving providers' capacity to be culturally responsive to their patients. In this definition of empathy, the provider demonstrates efforts to attempt to accurately understand the patient, from within a cultural framework, and *the patient* determines whether they feel understood. This is in contrast to models of empathy wherein an outside observer determines whether empathy is expressed by the provider rather than felt by the patient (e.g., definitions of empathy in Motivational Interviewing,

[29]). AI can help to support the MCO framework and promote accurate understanding by detecting and responding to conversational cues linked to mental health symptomatology including cultural cues that a given provider may overlook or even minimize [30]. For example, in a recent study, researchers analyzed six years of retrospective data and found that Black/African American mothers with depression commonly reported somatic complaints and self-critical behaviors as part of their health experience [31]. These kinds of symptom cues, when more commonly expressed by a non-dominant cultural group, can be better recognized by AI systems to ensure that they are not minimized or overlooked in diagnostic precision.

AI is especially well suited for addressing cultural factors because it can, in principle, be trained to understand all languages with robust cultural idioms of mental health expressions and frameworks for decision-making – gaining more exposure to specific cultural narratives and expressions of mental health than a single provider could experience in their lifetime. In particular, AI holds the potential to help providers be alerted to the ways culture shapes a patient’s beliefs about their mental health, behaviors related to seeking care, and ultimately may hold the promise to reduce REM patient distress [12, 32]. Further, when trained by cultural experts, AI can do more than simply be “good enough” for REM but can specifically be built to promote or embody cultural orientations like Afrocentric [33] or Indigenous [34] approaches to understanding and treating mental health challenges. This can include how an Indigenous patient may be more amenable to seeking support from a spiritual and/or traditional healer than from a medical professional for a mental health concern [35]. Thus, an AI system can coach a provider to recommend

Indigenous forms of healing in addition to standard care practices.

Challenges to Addressing REM Mental Health Disparities at Scale

While there are training methodologies to help therapists become more culturally sensitive and comfortable in addressing racial and cultural issues within therapy [36], there are several challenges to training therapists to meet the mental health needs of REM patients. First, there is an overall shortage of mental health providers who can meet Americans' current and projected mental health needs [37]. In fact, recent estimates show that one-third of Americans do not have access to a mental health provider in their local community. Second, the majority of the mental health provider population is Non-Hispanic White (85% of psychologists; 71% of Marriage and Family Therapists; 59% of Social Workers), demonstrating that the mental health workforce is not representative of the racial and ethnic makeup of the U.S. population [38]. The last few years demonstrated the importance of having a provider workforce that understands the collective stress experienced among REM–from the high-profile killings of Black Americans [39], higher rates of COVID-19 incidence and mortality amongst REM groups [40], and the steep rise in Anti-Asian violence related to COVID-19 [41]. By virtue of their own lived experiences, the majority of mental health providers are likely at a disadvantage when it comes to understanding the nuanced role of racism at individual, collective, and structural levels of lived experience, as well as how racism is embedded within the standard practice of mental health care. It is therefore imperative that novel approaches to training the mental health workforce at scale be adopted.

The emergence of telehealth and digital mental health applications, like therapy chatbots and user-initiated mental health exercises, have helped to improve access to immediate mental health care for many. However, there has been little to no emphasis on using technology to improve culturally-responsive, high-quality care delivery. Rather, the primary focus in digital health applications has been on repeating manualized cognitive-behavioral therapies which have been documented to be efficacious for non-Hispanic White populations but less effective for REM [42]. Given that AI is growing at an exponential rate, health equity must be considered on the front end of design as a means to address health equity at scale. This includes building AI that can coach providers to attend to cultural cues, consider trends in symptom presentation for certain REM groups in diagnosis and treatment, and even understand the client's cultural frame of reference to mental health (e.g., Indigenous worldview).

Addressing Disparities by Leveraging AI

Generative AI, powered by large language models (LLMs) that utilize natural language processing (NLP), is making it possible to conceive novel approaches to augmenting mental health care. With the release of ChatGPT, the market for AI-assisted communications has grown exponentially with continued expected growth in healthcare [43]. This rapid growth requires transparent, intentional efforts to reduce health care disparities, lest we build problematic patterns that contribute to health disparities to be indelibly repeated in AI models. For example, if we train an AI model on general historical data, as in transcripts of psychotherapy sessions, these transcripts are likely to include providers' microaggressions towards REM patients, which research shows is rampant [44]. AI models are only as good as the data they are trained on and will inevitably repeat problematic patterns in patient care. Culturally responsive

AI refers to artificial intelligence systems that are designed and developed with a deep understanding of culturally diverse lived experiences and as such, are trained using data that captures the unique needs, value systems, and beliefs of culturally diverse groups. The goal of building culturally responsive AI is to ensure that first, these systems are inclusive and respectful of different groups, and moreover, can be effective in meeting the needs of diverse peoples across various cultural contexts because they have been trained to consider cultural cues linked to symptom presentation and in the uptake of treatment recommendations.

Ensuring the Ethical Use of AI in Mental Health Care

As the market for AI in healthcare rapidly grows, so too does the risk of furthering status quo practices that have historically harmed REM and other minoritized groups. Building systems on legacy data without recognizing harms present in the history or the practices used to create that data will continue to perpetuate those harms. Our team has developed a framework for responsible, ethical use of our products (see Table 1) to ensure that we remain vigilant about following ethical principles in the development of our AI tools and their use in the larger world. Below are the key tenants we consider in our work, drawn from principles promoted by industry and academia [45-47]

Table 1: mpathic's AI Ethics Framework

Principle	**Commitment**	**Policies & Practices**
Practice **fairness** in every interaction.	Fairness of an AI system assures that the model treats people and scenarios equitably, free from any discrimination.	We ensure patients have consented to the use of their data.
Build **robust**	Robustness of an AI system verifies	We build with leading

systems.	the model’s accuracy across all potential use cases, while showcasing resilience against malicious attacks.	practices including testing to make sure our models are robust for customer data and model monitoring to validate predictions for customer API users.
Foster **transparency and explainability** in all decisions and actions.	Transparency of an AI system refers to the ability to explain and replicate the decisions made by the system. To achieve transparency, predictions should be accessible, understandable, and occur promptly for users, developers, and all other key stakeholders.	We onboard all stakeholders to the basics of our products in addition to ML/AI. We also update our stakeholders regularly to ensure ongoing explainability as our API advances.
Incorporate **privacy** in all handling of personal information.	Privacy in an AI system relates to protecting sensitive information used in training, validating, testing, and/or in the ongoing use of the model. An AI system should be designed so that it cannot divulge private information about individuals in the training data, nor should it be manipulable to reveal sensitive information about a person through malicious inputs like “jailbreaking” an LLM-based chat.	We use leading practices to protect privacy and adhere to privacy-related regulations ● HIPAA ● GDPR In addition the team has completed GCP Certification for all clinical trial participation.
Strive for **accountability** in all actions.	Accountability of an AI system indicates the readiness of the system’s designers or operators to respond to feedback and appeals and to put remediation mechanisms in place when issues arise. Operators can explain system results and decisions, if their functionality and decisions can be explained, to users, governing bodies, and other key stakeholders to ensure compliance with laws and ethical standards.	Leaders in AI/ML (Palat), Research and Health Equity (Cerezo) and Clinical Product (Jolley-Paige) keep our company aligned on AI Ethics best practices. We also regularly publish our policies and commitments for public view, including our AI Ethics framework.
Prioritize **safety** in all	The safety of an AI system emphasizes that it does not cause	We institute safety protocols to report

environments and situations.	harm to individuals, environments, or societies. This includes preventing the deployment of an AI system in inappropriate contexts and/or among populations for which it was not adequately designed. This could occur when the AI system has not been meticulously calibrated to align with clinical and/or legal standards.	instances of bias and imminent risk detected by our products and/or human annotators to customers.

It is important to recognize that the rapid pace of AI innovation requires consistent re-visitation of this framework. We come together as a larger team every six months to update and refine this framework as new developments in ML/AI emerge, including new methods for the responsible, ethical use of AI.

Curating Data to Produce an Inclusive and Equitable Environment

LLMs, such as those that power ChatGPT, are from a branch of AI known as Machine Learning (ML) and focus specifically on NLP. The processes for building deep learning models from annotated textual data have been described in detail in the literature [48-50]. In this section, we highlight the aspects of this process that are particularly prone to AI bias and how we have mitigated this concern via data curation, expert-level annotation, and thorough assessments of our models.

One of the main challenges in AI model building is that the content these systems generate is only drawn from the data on which they are trained. This means that if the developer chooses to expose an AI model to biased training data, the model will produce biased content. For instance, if AI is trained on tens of thousands of White faces but only a small sample of Black and Brown faces, it is likely to be better at recognizing White faces and make more generalization errors with Black and Brown faces during inference. Biased AI outputs like this have directly impacted REM in detrimental ways. Porcha Woodruff, a pregnant woman in Detroit, was wrongfully arrested and jailed for robbery and carjacking due to a faulty AI facial recognition program used by the Detroit police. Six officers showed up to her home on the morning of February 16, 2023 while she was getting her children ready for school. In this particular case, the AI system demonstrated AI bias with significant real-life impacts [51].

If a developer wants an AI system to demonstrate cultural representation, the system must be trained on robust, culturally representative data. Thus, it stands to reason that one of the formative challenges of building these systems is the curation of diverse, representative datasets. One of the most commonly used datasets to train LLMs is The Common Crawl dataset, a large collection of web pages scraped from the internet since 2014. These pages contain the gamut from high-quality news, science, and literature to the rantings and ravings of hate groups, conspiracy theorists, and propagandists. To make this and other datasets acceptable after training their models, OpenAI paid workers in Kenya approximately $2 USD an hour to remove harmful content and curate the data to various levels of success [52]. Even with this, there are still instances where biased, harmful content passes through the curation process [53, 54].

The challenges described above will be replicated in AI in the healthcare space if AI systems are trained on general health data. As demonstrated in one study, REM are faced with biased interactions from providers frequently; it should be expected that standard health datasets will therefore include instances of AI bias [19]. Thus, attention to algorithmic justice and diversity in data must be at the forefront of every AI health tool [28, 53]. And although current AI tools on the market demonstrate the potential for AI bias, they also demonstrate how well AI, like the transformer architecture used in large language models, interprets human language. NLP applications can be used to analyze and comprehend complex sentences, extract information, and generate meaningful responses, thus making these applications ideal for clinical settings. The challenge in building these tools is access to robust training data free of racial and other cultural biases so that they do not unintentionally demonstrate AI bias. Ensuring that standard psychotherapy practices are “good enough” for REM patients is not the end goal. Rather, AI systems should be built in collaboration with key cultural experts, specifically providers who practice from diverse cultural approaches to mental health, and who can aid in the building of AI systems that are responsive to the breadth of mental health experience and that consider cultural cues in mental health care.

Building Culturally Relevant LLM at *mpathic*

A patient’s mental health is inextricably linked to their cultural experience in the world. It is for this reason that cultural responsiveness is paramount to the assessment of mental health symptomatology and in the delivery of quality mental health care. Given the challenges of meeting the diverse mental health needs of patients at scale, we built an empathy skills training platform (mpathic’s mConsult, www.mpathic.ai) to detect and provide actionable behavioral adjustments in the form of tips (e.g., “Try repeating back what you just heard”) and generative suggested text (e.g., “Now say, am I understanding you correctly?”). mpathic’s products also coach in common factors therapy skills (i.e., empathy, collaboration, rapport building) in provider-patient communication in real-time. Our mission is to improve communication in the healthcare and life science spaces by not only identifying behaviors that weaken collaboration and relational alliance–including the minimization of a patient's cultural experience–but also by providing actionable guidance to improve providers' communication behaviors. mConsult detects hundreds of different therapist and patient behaviors in real-time with the same reliability as a gold-standard clinician and has been applied in commercial environments for English-speaking clinical interactions globally. In one application, the combination of mConsult and gold-standard clinicians detected 34 instances of suicidal ideation that were missed by the clinicians alone. To be clear, the application was not built for the replacement of human providers but as a co-pilot. We aimed to improve providers' communications and thus, increase access to and retention of quality care for diverse patients. By providing AI-powered support of providers' communication in real-time, rooted in robust, culturally representative training data, we are helping to ensure that all patients receive high-quality mental health care that is free of racial bias. There are many challenges to delivering this technology, and we haven't solved all of them. What follows are the practices we've adopted to manage and improve our quality and mitigate the risk of AI bias.

The Potential of AI for Health Equity

AI is well suited for supporting advances in mental health treatment at scale because it can analyze the content of treatment and quality of conversations, allowing researchers to understand the key conversational moments, predictors of success, and mechanisms of change in mental health care. Going a step further, AI can be trained to recognize patterns in mental health symptomatology, including beliefs and behaviors expressed to providers by cultural lived experience (e.g., racial identity, gender identity). These factors are key in training an AI system to recognize markers of key moments in therapy, including cultural elements across various groups that one sole provider is likely to miss and, in some cases, even minimize. Via AI-powered empathy support, a provider can improve their capacity to attend to critical mental health symptoms, cultural beliefs, and behaviors as opposed to solely relying on their professional judgment to attend to patient data that is perceived as important.

Our overarching goal was to build a tool that would foster provider effectiveness by encompassing and being responsive to the needs of culturally diverse populations (see Table 2). This included an acknowledgment that most AI is built on unrepresentative training data that not only fosters models rooted in dominant lived experience (i.e., White patients and providers), but that also ignores the significant role of culture in how mental health is both experienced and understood and what drives patients' engagement with treatment. Thus, having robust training data, in addition to a diverse Clinical AI team, has been fundamental to building an API that can redress mental health care disparities for REM patients.

Table 2: Building Culturally Responsive AI

Step 1	Step 2	Step 3	Step 4
Collection of Robust Training Data	Curation of Representative Training Data	Benchmarking AI Outputs against Robust Clinical Thresholds	Assess and Correct for AI Bias

Step 1: Collection of Robust Training Data. Given the amount of time

and data involved in the training of high-quality AI, it is advantageous to use an existing dataset or model rather than make an existing dataset more culturally relevant via curation. When relying on curation to prevent AI bias, the original dataset is already rooted in cultural experiences, assumptions, and values that are from the dominant culture and that have been proven to show bias against minoritized groups – many of whom carry the disproportionate burden of health disparities. Our team avoided a cold-start problem in AI by building models initially using labeled data derived from a coaching/therapy training game (www.empathy.rocks) [55]. This game involved a data acquisition and labeling flywheel where providers responded to clinical vignettes and practiced empathy-grounded communication to earn continuing education credits. Providers/players ranked other players' responses; every response with three positive rankings was added to our dataset. Players included providers from a range of professional and cultural backgrounds including digital mental health companies, a state-level crisis text line, and a state-level Native American health service network. Our goal was to have clinicians from diverse healthcare settings and with diverse lived experiences build our initial training dataset so that our API could detect and respond to clinical scenarios in an empathetic manner across cultural lived experiences.

Step 2: Curation of Representative Training Data. Upon ingesting data into our API from our data flywheel, we built an expert clinical annotation team that first annotated transcripts of our enterprise customers in the Health/Life Sciences space from a range of settings including clinical trials, healthcare coaching, therapy, medical appointments, and insurance claims. Specifically, we employed AI and NLP processing on conversational data from real-life mental health care sessions. For each new conversational behavior, we aim to train our models on anywhere from 2000 utterances to 300+ sessions with our Clinical AI team providing expert annotation of culturally-specific behaviors. We also aim to collect metadata (patients' age, gender, race, etc.) and to have our U.S. training data be at least 60% REM to account for diverse lived experience while allowing for comparison to White patients when appropriate (i.e., assessing differences in assessment of minimization). Our data is not limited to English or the United States locations; however, our first models are largely built on only English-speaking data with translational layers to other languages using LLMs as needed when deployed in the real world. Our goal is to ensure that our AI accurately captures cultural factors and that it does so equitably across REM and White patients for U.S. settings and beyond. Leveraging high-quality health session training data that is annotated and created by a diverse Clinical AI team is paramount to successfully shaping the precision and cultural relevance of our API.

After building the first version of these models, we then can create synthetic "twins" of the data and models that replicate similar situations created by our gold-standard clinicians. Synthetic data, also called algorithmically generated, can be used to stand in for real-world data when data is missing or incomplete, or in this case where concern for privacy in the training data is paramount. By making synthetic data with similar properties to the real, diverse clinical data we can commercialize the models broadly and without concern for private information being included in the models in perpetuity. Our clinical annotation team also builds in-house synthetic data consisting of noteworthy communications designated as "rare" such as an egregious positive or negative provider response to further train our models (e.g., provider demonstrating anger and/or using derogatory words or phrases in session). This data is developed by providers who have the clinical and cultural expertise to capture rare cases. Synthetic data are necessary to ensure that the real-life distribution of mental health and therapy experience is represented—something that is often lacking in open, public data sets. Our clinical annotation team is charged with bridging examples of behaviors not well represented in the training data, ensuring that the synthetic data are realistic and high quality.

In the construction of synthetic data for model development, we consider data to be high quality when they are (1) coherent, (2) demonstrate clear construct validity to the behavior, (3) are rich in structure and diversity, and (4) include realistic vocabularies, including vocabularies for different cultural groups. First, it is important to note that not all data generated by synthetic systems are coherent. While LLMs add new use cases for the commercial and academic use of AI, there is still the risk of models generating poorly constructed sentences. Second, clear sentiment describes examples in the data (e.g., conflict) that clearly illustrate the behavior we are seeking to train the model on. Third, data should be rich, demonstrating a range of diversity in style, content and representation. The data should also represent different types of patient experiences. Fourth, realistic vocabularies are important with respect to culture and domain knowledge. It is easy to identify simple examples of behavior, but real people in real conversations often construct more complex and varied utterances. There must be a balance in the training data when it comes to including realistic vocabularies given that realistic conversations may contain incoherent phrases. Raw, real data are much more diverse than what synthetic data typically represents. By testing our models against real data, we are better able to evaluate where our models are weakest and can then synthesize new data points that more accurately represent real-world distributions.

Step 3: Benchmark AI Outputs Against Robust Clinical Thresholds

Building tools for applications in real-world settings is a complex endeavor, not easily captured by measuring accuracy or any other single metric alone. Our team uses a variety of measures to benchmark our solutions. We start during model development via numerous tests. First, we measure the precision of our model which considers the quality of the true predictions (out of all the predictions where the model identified an utterance as belonging to a class, how many of them were actually in the class [true positives/ true positives + false positives]). Second, we examine the recall or sensitivity of our model, meaning the ability of the model to identify positive cases in a given evaluation set (a separate subset of the dataset that is not used during the training phase; rather, it is employed to assess how well the model generalizes to new, unseen data). By charting how the model does across several different thresholds, we can determine where the model provides the best performance. We strive to make our validation sets as representative of real-world data as possible, including samples from customer usage to assess how far their language and examples are from our training data, and how accurate our system remains.

While we work to make our models as robust as possible to work across many different customers and environments, we find that different customers may use language in ways that our model may be more or less sensitive to. At the software or API infrastructure level, our team includes controls that allow us to increase or decrease the sensitivity of the model to different inputs. Our annotation team will also create additional qualitative assessments that allow them to develop a better understanding of the strengths or weaknesses of the model for application in specific customers' real-world scenarios. Leveraging the Clinical AI team's training, varied lived experiences, and understanding of our customers allows us to better identify shifts in the behaviors of our customers and recognize when our API is starting to lose accuracy for a given customer. For example, the Clinical AI team has been able to identify instances where models are technically accurate based on linguistic patterns but were not contextually accurate for the customer (e.g. a model may be identifying confrontation but in the customer context it was a different behavior).

In each instance of a provider disagreement with mConsult feedback, our annotation team reviews the data to either refine the models or provide feedback to the customer (i.e., our annotation team deems mConsult's detections to be correct). It may be the case that disagreement between automated feedback and the provider is rooted in a cultural lens of our annotation team, and hence the model can detect a problem that is not recognized as a communication behavior to be corrected by the provider. Again, each instance of disagreement is reviewed with the end goal of improving the precision of our behavioral detection models.

One of the challenges common to qualitative textual analysis and ML data annotation is keeping in alignment over time (drift) and across the team (interrater reliability). To ensure high agreement over time we have multiple measures and checkpoints for inter-rater reliability (IRR) in the workflow for our annotation team. During labeling, we sample 20% of the content for IRR and verify how close we remain to our gold standards of annotation (comparison reference for annotation) using pairwise Krippendorf's alpha at the sentence or utterance level. Our team has previously written about IRR for building mental health ML applications [56, 57]. As we see drift in our annotation, we immediately initiate internal evaluation and discussion. In some cases, this may be a teaching moment for an annotator who can be given more specific feedback on how to address a certain label (e.g., confrontation). It can also be a moment to discuss with the team ambiguities in the criteria for a particular label. This discussion allows the team to disagree and raise concerns if a label is too restrictive, or vague, or may highlight a weakness with the current definition, thus avoiding groupthink and honoring diverse lived experiences of a particular phenomenon (i.e., confrontation). These practices have allowed our team to keep alignment scores higher than many found in the clinical literature (e.g., internally, our IRR ranges from .73 to .86 for sentence-level IRR for behaviors). In some cases, we find the definition of a particular phenomenon has changed enough that we will go back and relabel existing material to better reflect the new, shared understanding.

Step 4: Assessing and Correcting for AI Bias

Delivering high-quality predictions that serve all REM groups, in addition to other cultural lived experiences (singular; intersectional), is a challenge we continue to address. After curating our data sets and augmenting them with synthetic data, we then train our models to detect behaviors in speech and text. When developing models for cultural responsiveness, we first compare the model outputs for all automatically annotated patient data against the performance of human annotators. Second, we assess whether our models demonstrate bias, meaning whether they demonstrate stronger predictive validity on key outcomes (i.e., relational alliance, confrontation) for White patients than REM patients by comparing our model outputs against our human annotators. It is our goal that our API can assess empathic communication in a consistent manner that is free of bias by demonstrating no differences in performance between ethnic and racial groups, in addition to other historically minoritized communities. A challenge we have faced is finding ground truth in how our models are being used in real-world contexts. For example, most customers in the commercial health space do not collect or share metadata (i.e., demographic information about patients) when submitting data to our platform. To circumvent this challenge, our team develops data partnerships outside of our customers to ensure that we (a) have culturally representative provider and patient data, and (b) have the kinds of data that allow us to build models from a particular cultural frame of reference (e.g., Indigenous healing). We continue to source examples of natural conversations to ensure that our training data is representative, thereby increasing the likelihood that the models are bias-free.

Auditing. Beginning with clinical evaluation, upon obtaining ground truth annotations of a given model (e.g., relational alliance), our team continuously works towards extension and refinement of the said model via testing its effectiveness in new contexts (i.e., with new customers). In some cases, our team must refine a model to a target population or context if the guidelines and/or requirements for implementation in the new context are different from when the model was first developed and validated. For example, our AI model’s ability to correctly detect and coach for improved relational alliance in a standard medical health visit (between provider and patient) will be different from relational alliance in a surgical training room–between the training doctor and resident, for example. To audit needed changes for horizontal implementation in different health settings, our AI Clinical Products team must have a thorough understanding of the training data, the ML algorithm employed to develop the models, and evaluation metrics that are specific for their intended use.

For technical validation, our team routinely carries out precision and recall. An F1 score provides a numerical assessment of a given model’s accuracy or precision. Specifically, the F1 score computes how many times an AI model made an accurate prediction across a specific dataset. An F1 score of 0.75, which roughly equates to 75% accuracy, is an acceptable F1 score by most ML researchers. Relatedly, recall involves the proportion of true positives predicted by a model in proportion to actual positives in the dataset. In other words, how often a model correctly predicts depression for Blacks patients when depression is the correct diagnosis (as determined by clinical experts). Thus, if an ML model correctly predicts depression in 79 of 100 Black patients in the dataset–missing 21 patients–then the F1 score is 0.79, meaning 79% of Black patients were accurately diagnosed with depression. Most ML researchers would deem an F1 score of 0.79 as acceptable. However, in clinical practice, the F1 score may need to be higher in critical situations, as in the case of schizophrenia or even suicide. In these particular scenarios, a clear model performance threshold should be established for a given clinical context (i.e., emergency room care; inpatient care).

Sharing Model Cards. In ML, model cards are often shared in the spirit of transparency and integrity in AI model development and validation. Model cards are the equivalent to a codebook and/or manualized treatment handbook in the health sciences wherein the details of the methods employed to build and validate a model are available for other ML researchers and interested parties to understand the details of the process [58]. In the case of mitigating AI bias in model development and refinement, we recommend the following details to be included: (1) performance metric[1]s (precision, recall, F1 score), (2) differential performance across demographic groups, which requires the collection of robust metadata, (3) bias evaluations (specific methods used to evaluate bias in addition to the identification of potential sources of bias in the training data), (4) bias mitigation efforts via clinical auditing by cultural experts, and (5) ethical and social impact considerations, as in the case of the deployment of our models in clinical settings that are not appropriate and/or may inadvertently cause harm. Model cards must be shared on a regular, ongoing basis to retain model precision.

Leveraging AI for Mental Health Care Equity

A major focus in the use of AI for minoritized groups has been on the protection of human rights and the reduction of AI bias. Clinicians and researchers must continuously engage in practices that ensure Justice, Accountability, and Equity (see Table 1) in model development, refinement, and use. And while these considerations are paramount, they are just the start–not the end solution. Our team has intentionally built an API trained on robust, culturally representative data - and annotated by an expert, culturally diverse team. These foundational steps are requisites to building AI tools that are powered to coach providers in the reduction of biased language and missed cultural opportunities for the understanding of mental health presentation; and

[1]

in the uptake of empathic, culturally attuned language that has been shown to improve REM treatment outcomes, including retention in care.

AI holds significant potential to revolutionize mental health care delivery, including the capacity to scale health equity and social justice for REM patients. But in order for this innovation to take place, REM patients and providers must be in leadership positions that have the power to drive innovation. This includes the adoption of human-centered design [59] and community advisory boards [60] where REM voices are centered in product development, refinement and dissemination. REM clinicians also hold immense power in the AI space given their expert training in clinical practices and research that not only consider culture, but more so leverages cultural strengths as a means for patient healing and growth. Just as the common idiom of "garbage in, garbage out" denotes that an AI system is only as good as the data it is trained on, the same holds true for the potential of AI when robust training data are available. In other words, "quality in, quality out" can be championed to train AI systems to leverage culturally responsive care at scale. Our team is realizing a vision wherein an AI system can be trained on the expert skills of REM clinicians to ensure that all clinicians can adopt AI provider support that helps them to provide empathic, culturally attuned care.

This paper offers an introduction to our team's approach to developing AI for health equity. As the AI industry continues to grow, so will our efforts to innovate with the overarching goal of centering the lived experiences and needs of minoritized groups so that AI solutions are bias-free and well-positioned to foster equitable care for all.

Conclusion

This review of mpathic’s Ethical AI framework and approach to model building demonstrates that AI in the mental health space should be considered as a mechanism to develop robust, culturally responsive products. Further, orienting to an Ethical AI framework fosters consideration of how these products are used in healthcare settings–ensuring they are employed to redress versus further exacerbate extant health disparities. Many of the problems facing the mental health field, such as the massive gap in available providers [61] and the continued presence of provider bias against REM patients [62] require large-scale innovative approaches. AI is one such solution. We encourage providers and researchers to imagine how AI can help to identify patterns, connect patients to immediate care, and guide providers to provide care that is free of racial bias. By leveraging the power of AI, the mental health field is better able to address health disparities at scale.

References

1. National Institute of Mental Health. Mental illness. National Institute of Mental Health. Published March 2023. https://www.nimh.nih.gov/health/statistics/mental-illness
2. Reinert M, Fritze D, Nguyen T. The State of Mental Health in America 2023. Mental Health America, Alexandria VA. 2022. https://mhanational.org/sites/default/files/2023-State-of-Mental-Health-in-America-Report.pdf/. [Accessed 2024/January 18]
3. American Psychological Association. More than a quarter of U.S. Adults say they’re so stressed they can’t function. Published 2022/October 19. https://www.apa.org/news/press/releases/2022/10/multiple-stressors-no-

function#:~:text=Washington%20%E2%80%94%20Americans%20are%20struggling%20with,for%20the%20American%20Psychological%20Association. [Accessed 2023/Ocotober 31]

4. Breslau J, Aguilar-Gaxiola S, Kendler KS, Su M, Williams D, Kessler RC. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. Psychol Med. 2006 Jan;36(1):57-68. doi: 10.1017/S0033291705006161. Epub 2005 Oct 5. PMID: 16202191; PMCID: PMC1924605.
5. Green JG, McLaughlin KA, Fillbrunn M, Fukuda M, Jackson JS, Kessler RC, Sadikova E, Sampson NA, Vilsaint C, Williams DR, Cruz-Gonzalez M, Alegría M. Barriers to Mental Health Service Use and Predictors of Treatment Drop Out: Racial/Ethnic Variation in a Population-Based Study. Adm Policy Ment Health. 2020 Jul;47(4):606-616. doi: 10.1007/s10488-020-01021-6. PMID: 32076886; PMCID: PMC7260099.
6. Thomas L. Psychotherapy dropout: The influence of ethnic identity and stigma on early termination. Electronic Theses and Dissertations. 2016; https://doi.org/10.18297/etd/2406
7. Owen J, Imel Z, Tao KW, Wampold B, Smith A, Rodolfa E. Cultural ruptures in short-term therapy: Working alliance as a mediator between clients' perceptions of microaggressions and therapy outcomes. Counseling and Psychotherapy Research. 2011. 11(3), 204–212. https://doi.org/10.1080/14733145.2010.491551
8. Mongelli F, Georgakopoulos P, Pato MT. Challenges and Opportunities to Meet the Mental Health Needs of Underserved and Disenfranchised Populations in the United States. Focus (Am Psychiatr Publ). 2020

Jan;18(1):16-24. doi: 10.1176/appi.focus.20190028. Epub 2020 Jan 24. PMID: 32047393; PMCID: PMC7011222.

9. Takeshita J, Wang S, Loren AW, Mitra N, Shults J, Shin DB, Sawinski DL. Association of Racial/Ethnic and Gender Concordance Between Patients and Physicians With Patient Experience Ratings. JAMA Netw Open. 2020 Nov 2;3(11):e2024583. doi: 10.1001/jamanetworkopen.2020.24583. PMID: 33165609; PMCID: PMC7653497.

10. Rebuilding trust in health care. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/health-care/trust-in-health-care-system.html

11. American Psychological Association. Stress in America 2023. Published 2023/November. https://www.apa.org/news/press/releases/stress/2023/collective-trauma-recovery [Accessed 2024/January 16]

12. Panchal N, Saunders H, Ndugga N. Five Key Findings on Mental Health and Substance Use Disorders by Race/Ethnicity. KFF. Published September 22, 2022. https://www.kff.org/racial-equity-and-health-policy/issue-brief/five-key-findings-on-mental-health-and-substance-use-disorders-by-race-ethnicity/

13. Sue DW, Sue D, Neville HA, Smith L. Counseling the culturally diverse: Theory and practice. John Wiley & Sons; 2022 Mar 22.

14. Miu AS, Moore JR. Behind the Masks: Experiences of Mental Health Practitioners of Color During the COVID-19 Pandemic. *Acad Psychiatry*. 2021;45(5):539-544. doi:10.1007/s40596-021-01427-w

15. Merced K, Imel ZE, Baldwin SA, Fischer H, Yoon T, Stewart C, Simon

G, Ahmedani B, Beck A, Daida Y, Hubley S. Provider contributions to disparities in mental health care. Psychiatric services. 2020 Aug 1;71(8):765-71.PMCID: PMC7590958

16. Bird ST, Bogart LM. Perceived race-based and socioeconomic status(SES)-based discrimination in interactions with health care providers. Ethn Dis. 2001 Autumn;11(3):554-63. PMID: 11572421.

17. Sorkin DH, Ngo-Metzger Q, De Alba I. Racial/ethnic discrimination in health care: impact on perceived quality of care. J Gen Intern Med. 2010 May;25(5):390-6. doi: 10.1007/s11606-010-1257-5. Epub 2010 Feb 10. PMID: 20146022; PMCID: PMC2855001.

18. Lee C, Ayers SL, Kronenfeld JJ. The association between perceived provider discrimination, healthcare utilization and health status in racial and ethnic minorities. Ethn Dis. 2009 Summer;19(3):330-7. PMID: 19769017; PMCID: PMC2750098.

19. Hook JN, Farrell JE, Davis DE, DeBlaere C, Van Tongeren DR, Utsey SO. Cultural humility and racial microaggressions in counseling. *J Couns Psychol*. 2016;63(3):269-277. doi:10.1037/cou0000114

20. American Psychological Association. Guidelines on multicultural education, training, research, practice, and organizational change for psychologists. The American Psychologist. 2003 May;58(5):377-402.

21. Knudson-Martin C, McDowell T, Bermudez JM. From Knowing to Doing: Guidelines for Socioculturally Attuned Family Therapy. J Marital Fam Ther. 2019 Jan;45(1):47-60. doi: 10.1111/jmft.12299. Epub 2017 Nov 10. PMID: 29125887.

22. Betancourt JR, Green AR, Carrillo JE, Ananeh-Firempong O 2nd.

Defining cultural competence: a practical framework for addressing racial/ ethnic disparities in health and health care. Public Health Rep. 2003 Jul-Aug;118(4):293-302. doi: 10.1093/phr/118.4.293. PMID: 12815076; PMCID: PMC1497553.

23. Ranjbar N, Erb M, Mohammad O, Moreno FA. Trauma-Informed Care and Cultural Humility in the Mental Health Care of People From Minoritized Communities. Focus (Am Psychiatr Publ). 2020 Jan;18(1):8-15. doi: 10.1176/appi.focus.20190027. Epub 2020 Jan 24. PMID: 32047392; PMCID: PMC7011220.

24. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Current Treatment Options in Psychiatry. 2019 Sep 15;6:243-55.

25. Liang J, Matheson BE, Douglas JM. Mental Health Diagnostic Considerations in Racial/Ethnic Minority Youth. *J Child Fam Stud*. 2016;25(6):1926-1940. doi:10.1007/s10826-015-0351-z

26. Jull J, Fairman K, Oliver S, Hesmer B, Pullattayil AK; Not Deciding Alone Team. Interventions for Indigenous Peoples making health decisions: a systematic review. *Arch Public Health*. 2023;81(1):174. Published 2023 Sep 27. doi:10.1186/s13690-023-01177-1

27. Escovar EL, Craske M, Roy-Byrne P, et al. Cultural influences on mental health symptoms in a primary care sample of Latinx patients. *J Anxiety Disord*. 2018;55:39-47. doi:10.1016/j.janxdis.2018.03.005

28. Dankwa-Mullan I, Scheufele EL, Matheny ME, Quintana Y, Chapman WW, Jackson G, South BR. A proposed framework on integrating health equity and racial justice into the artificial intelligence development

lifecycle. Journal of Health Care for the Poor and Underserved. 2021;32(2):300-17.

29. Miller WR, Rollnick S. Motivational interviewing: Helping people change. Guilford press; 2012 Sep 1.

30. Denecke K, Gabarron E. How Artificial Intelligence for Healthcare Look Like in the Future?. *Stud Health Technol Inform*. 2021;281:860-864. doi:10.3233/SHTI210301

31. Perez NB, D'Eramo Melkus G, Wright F, et al. Latent Class Analysis of Depressive Symptom Phenotypes Among Black/African American Mothers. *Nurs Res*. 2023;72(2):93-102. doi:10.1097/NNR.0000000000000635

32. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behaviour research and therapy. 2021 Dec 1;147:103990.

33. Daugherty PR, Wilson HJ, Chowdhury R. Using artificial intelligence to promote diversity. MIT Sloan Management Review. 2018 Nov 21.

34. O'Keefe VM, Cwik MF, Haroz EE, Barlow A. Increasing culturally responsive care and mental health equity with indigenous community mental health workers. *Psychol Serv*. 2021;18(1):84-92. doi:10.1037/ser0000358

35. Mental Health America. Native and Indigenous Communities and Mental Health. Mental Health America. Published 2020. https://www.mhanational.org/issues/native-and-indigenous-communities-and-mental-health

36. Asnaani A, Hofmann SG. Collaboration in multicultural therapy:

establishing a strong therapeutic alliance across cultural lines. *J Clin Psychol*. 2012;68(2):187-197. doi:10.1002/jclp.21829

37. Mongelli F, Georgakopoulos P, Pato MT. Challenges and Opportunities to Meet the Mental Health Needs of Underserved and Disenfranchised Populations in the United States. *Focus (Am Psychiatr Publ)*. 2020;18(1):16-24. doi:10.1176/appi.focus.20190028

38. Lin L, Stamm K, Christidis P. How diverse is the psychology workforce? *American Psychological Association*. https://www.apa.org/monitor/2018/02/datapoint. Published February 2018.

39. Lett E, Asabor EN, Corbin T, Boatright D. Racial inequity in fatal US police shootings, 2015–2020. J Epidemiol Community Health. 2020 Oct 21.

40. Lundberg DJ, Wrigley-Field E, Cho A, et al. COVID-19 Mortality by Race and Ethnicity in US Metropolitan and Nonmetropolitan Areas, March 2020 to February 2022. *JAMA Netw Open*. 2023;6(5):e2311098. Published 2023 May 1. doi:10.1001/jamanetworkopen.2023.11098

41. Wong-Padoongpatt G, Barrita AM. The fast and slow violence of the COVID-19 pandemic on Asians in the USA. InCOVID-19: Cultural Change and Institutional Adaptations 2022 Dec 30 (pp. 147-158). Routledge.

42. Windsor LC, Jemal A, Alessi EJ. Cognitive behavioral therapy: a meta-analysis of race and substance use outcomes. *Cultur Divers Ethnic Minor Psychol*. 2015;21(2):300-313. doi:10.1037/a0037929

43. Snyder J, Silberschatz G. The Patient's experience of attunement and responsiveness scale. Psychotherapy Research. 2017 Sep 3;27(5):608-19.

44. Kanter JW, Rosen DC, Manbeck KE, Branstetter HM, Kuczynski AM, Corey MD, Maitland DW, Williams MT. Addressing microaggressions in

racially charged patient-provider interactions: a pilot randomized trial. BMC Medical Education. 2020 Dec;20:1-4.

45. Google. Google AI Principles. Google AI. Published 2023. https://ai.google/responsibility/principles/

46. Microsoft. Responsible AI principles from Microsoft. Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai

47. Baker S, Xiang W. *Explainable AI Is Responsible AI: How Explainability Creates Trustworthy and Socially Responsible Artificial Intelligence*. https://arxiv.org/pdf/2312.01555.pdf

48. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. arXiv.org. Published June 12, 2017. https://arxiv.org/abs/1706.03762

49. Radford A, Narasimhan K, Salimans T, Sutskever I. *Improving Language Understanding by Generative Pre-Training*.; 2018. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

50. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*. 2019;1. doi:https://doi.org/10.18653/v1/n19-1423

51. Pregnant woman’s arrest in carjacking case spurs call to end Detroit police facial recognition. AP News. Published August 7, 2023. Accessed January 21, 2024. https://apnews.com/article/detroit-police-facial-recognition-lawsuit-cab0ae44c1671fc30617d301b21b2d13

52. Perrigo B. Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer. Time. Published January 18, 2023.

https://time.com/6247678/openai-chatgpt-kenya-workers/

53. Birhane A, Prabhu VU, Kahembwe E. Multimodal datasets: misogyny, pornography, and malignant stereotypes. arXiv preprint arXiv:2110.01963. 2021 Oct 5.

54. Luccioni A, Viviano J. What’s in the box? an analysis of undesirable content in the Common Crawl corpus. InProceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) 2021 Aug (pp. 182-189).

55. Lord SP, Bertagnolli N, inventors; Empathy Rocks Inc, assignee. System and method for increasing effective communication through evaluation of multimodal data, auto-correction and behavioral suggestions based on models from evidence-based counseling, motivational interviewing, and empathy. United States patent application US 17/731,230. 2022 Oct 27.

56. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 2012;8(1):23-34. doi: 10.20982/tqmp.08.1.p023. PMID: 22833776; PMCID: PMC3402032.

57. Lord SP, Sheng E, Imel ZE, Baer J, Atkins DC. More than reflections: empathy in motivational interviewing includes language style synchrony between therapist and client. Behav Ther. 2015 May;46(3):296-303. doi: 10.1016/j.beth.2014.11.002. Epub 2014 Nov 11. PMID: 25892166; PMCID: PMC5018199.

58. Mitchell M, Wu S, Zaldivar A, et al. Model Cards for Model Reporting.

Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT '19*. Published online 2019. doi:https://doi.org/10.1145/3287560.3287596

59. Stiles-Shields C, Cummings C, Montague E, Plevinsky JM, Psihogios AM, Williams KD. A call to action: using and extending human-centered design methodologies to improve mental and behavioral health equity. Frontiers in Digital Health. 2022 Apr 25;4:848052.

60. Gonzalez-Guarda RM, Jones EJ, Cohn E, Gillespie GL, Bowen F. Advancing nursing science through community advisory boards: working effectively across diverse communities. ANS. Advances in nursing science. 2017 Jul;40(3):278.

61. Kazdin AE. Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behav Res Ther*. 2017;88:7-18. doi:10.1016/j.brat.2016.06.004

62. Hall WJ, Chapman MV, Lee KM, et al. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *Am J Public Health*. 2015;105(12):e60-e76. doi:10.2105/AJPH.2015.302903