Responsible AI to Support Mental Health in Black Communities

A discussion about biases, health care access, and the potential of technology

Over the last year, intertwined crises have impacted people throughout the world. A global pandemic caused millions to rely on technology to work and learn from home. Given the digital divide, this meant millions missed out on support and opportunities where technological access was essential, especially tied to health.

In fact, the number of people looking for help with anxiety and depression skyrocketed from 2020 to 2021 [1]. For many African Americans and Black people in the U.S., police violence, especially the video-captured murder of George Floyd compounds racial trauma.

It is at this intersection of technology, mental health, and societal bias that we held a two-day workshop exploring the benefits and harm artificial intelligence (AI) can have on mental health. Our Microsoft AI for Accessibility team worked with individuals from across Microsoft and the Columbia University School of Social Work to learn from experts in mental health, data science, AI, and Black communities.

In this research paper, we're sharing takeaways from a two-day workshop regarding correcting biases, creating responsible AI, and increasing access to culturally responsive mental health care. Takeaway themes include:

- Overcoming a history of trauma
- Collecting inclusive and representative data
- Co-designing responsible AI solutions made by and with affected communities
- Addressing barriers to accessing services
- Taking next steps for technology to support mental health

There were many clinicians, researchers, data scientists, and people with lived experience of mental health conditions who contributed their expertise and ideas. We cannot fully capture the richness of discussions that took place during the workshop, but this document summarizes what we learned—and the questions we have yet to answer as part of our collective journey to develop and leverage technology for equitable mental health care in Black communities.

[1] COVID-19 and Mental Health: A Growing Crisis, Mental Health America, 2021

Overcoming a history of trauma

Racism—via individual, institutional, and systemic bias—has a direct impact on the mental health and wellbeing of people in Black communities. Inequities of health care and access, as well as historical medical abuses[2], also cause many to be wary of seeking care.

For example, Black patients are less likely to receive adequate pain management strategies[3], more likely to be diagnosed with higher-stigma conditions such as schizophrenia and psychotic disorders[4], and less likely to receive appropriate treatment for substance use disorders[5].

It's not race that's a social determinant of health; it's racism.

- Dr. Danielle Hairston

When it comes to mental health, risks are even higher in Black communities[6]. As Gloria Washington, PhD, highlighted during the workshop, "Until we acknowledge the constant battle Black persons fight internally to overcome the current and historical trauma created by racism, sexism, and every other '-ism' that is uniquely applied to Black people, we will not have an effective strategy to deal with societal bias that comes along with treating Black mental health patients."

The potential of technology to improve mental health care

The need for culturally responsive care has never been higher: A quarter of the world's population will likely face a mental health condition in their lifetime, according to the United Nations[7]. Technology is far from being the only solution to meet this need, especially because it cannot understand a person's internal state or wellbeing. That said, technology can expand the reach and complement the person-to-person care of mental health providers, which can help meet high demand for support.

Responsible AI does not make decisions for or about people; it helps empower them to make better decisions for themselves.

We already see inspiring examples. Young people are using the social media platform TikTok to de-stigmatize mental health conditions. A national bill will create a 988 number for mental health services, which will offer an alternative to 911 calls. And our AI for Accessibility grantees Mental Health America, University of Toronto, and Northwest University are developing an AI-powered text messaging platform to reach youth who are less likely to seek formal supports.

While these are promising solutions, all AI must be developed with extreme care in order to create inclusive technologies for Black communities and other vulnerable groups.

[2] Mental Illness in Black Community, 1700-2019: A Shorth History, Black Past, 2019

[3] Black and African American Communities and Mental Health, Mental Health America, 2020

[4] A Naturalistic Study of Racial Disparities in Diagnoses at an Outpatient Behavioral Health Clinic, Gara et al, 2018

[5] Layisett et al, 2019 [NEED THIS STUDY]

[6] Black and African American Communities and Mental Health, Mental Health America, accessed 2021

[7] Mental Health and Development, United Nations, accessed 2021

Developing responsible AI

AI is one tool to address persistent, widespread, and complex issues of mental health, yet it can only help if it is inclusive.

To create equitable AI solutions, technology must use inclusive and representative data, and it must be co-designed with diverse stakeholders, including people with personal experience of mental health conditions.

Driving inclusion in data sets

Artificial intelligence is technology designed to mimic human intelligence and perception and, through machine learning (ML), refine its "decision making" processes and adapt to novel situations. AI can enhance tools and services by personalizing the experience to an individual's needs and strengths.

AI needs data. Equitable AI needs equitable data.

Today, lack of diversity is a common problem in datasets used to develop machine learning models for mental health applications, which creates a dangerous feedback loop. Technologies created without the input of vulnerable communities may result in recommendations that further marginalize them and expand the disparity of their care[8].

As workshop presenter Dr. Courtney Cogburn said, to create more equitable health care for people in Black communities, we need to "move [the conversation] beyond individual bias to focus on systems, patterns, and structures." Given that AI is, at its core, based on patterns, we need to be vigilant to avoid perpetuating biases.

AI Solutions for Wellbeing

Workshop presenter Ehi Nosakhare, PhD, explained that mental health technologies are being developed in four main areas.

- **Behavioral intervention technologies** can support habit change and treatment plan adherence.
- **Mobile and wearable apps** can monitor symptoms and assess health risks.
- **Computerized treatments** can support treatment in partnership with human health providers and help expand clinical knowledge.
- **Platforms for peer or community support** match people with similar lived experiences to facilitate support relationships and crisis care.

Nosakhare explained during the workshop that inclusive technologies must avoid or overcome these types of biases:

Selection bias is a systemic difference between the characteristics of those selected for a study and those who are not. When people of color are not sufficiently represented, it limits the ability to personalize AI for a user's needs.	**Algorithmic bias** occurs when an ML model produces unintended results, often by optimizing for the majority group to boost "accuracy."	**Human bias** results from misunderstanding mental health data such as clinical notes, self-assessments, and unstructured text such as social media posts. Interpreting these data is subjective, unlike other health data, such as heart rate or blood sugar levels. A mix of data sources (e.g., video, audio, text, telemetry) and type (e.g., from lab and "real world" contexts) can help counterbalance human bias.

[8] Benjamin, 2019

The scarcity of Black and culturally responsive providers contributes to human bias because "Black folks, while not a monolith, experience and express mental health differently than their white counterparts," said Kelechi Ubozoh, author and workshop attendee. People who are used to conforming to white-dominant scenarios (such as therapy) may rely on "covering" (not accurately communicating their symptoms) or "masking" (expressing symptoms in different ways than their white counterparts) to protect themselves from microaggressions. When not interpreted correctly, this can skew data and the resulting solutions.

AI solutions trained on non-representative data can fail to serve —or even harm—people not included in the data sets.

Co-designing solutions

To create equitable AI solutions, we need to design with, and not just for, the communities we intend to serve.

The commitment to co-designing solutions requires inclusion from top to bottom—from ideation and funding to testing and rollout—rather than just setting aside a "seat at the table." As Theresa Nguyen, LCSW, from Mental Health America said, "Equitable AI research needs to include not just inclusive samples; it needs to include people of color as researchers and not sidelined as advisors or testers."

Similarly, transdisciplinary teams must include people with direct, personal experience of mental health conditions. This way, teams ensure they ask the right questions, address the most pressing problems, and collect the relevant data—from the beginning, not after the fact.

The 3 questions of Design Justice

Design Justice is an approach that helps counteract the biases described above and create solutions that disrupt inequitable systems.

Every AI developer should take an anti-racist approach by asking themselves a series of questions at the outset of any project. These questions, which center people who are typically marginalized, aren't meant to elicit the "right" answers but rather spark deep consideration and conversation.

1. **Who participates?** Are affected communities involved? Are stakeholders representative and diverse?
2. **Who benefits?** Consider the many people who interact in the scenario the technology addresses, including patient, practitioner, insurer, researcher, and others.
3. **Who is harmed?** Any AI in mental health solution must be evaluated for potential harm caused to Black communities.

> **"To tackle systemic change, it is important to ask if or how any new development in AI helps to dismantle the legacy of discrimination and racism faced by the Black community."**
>
> - Theresa Nguyen

Design Justice aligns with Microsoft's Responsible AI principles to guide an organization's approach to the design, development, and delivery of human-centered AI technologies. However, principles alone are not enough. We must translate them into engineering and business practices and continually investigate AI for biases.

Access to culturally responsive care

The previous sections addressed the creation of AI technology that can improve mental health treatment and support. Access considers the format, platform, or delivery of that care. Meeting people in Black communities where they are, and with the technology they use, will increase access to mental health services.

Equitable care is only possible when people can access what they need in a way that is helpful to them.

We see three main limits to accessing much-needed services.

1. **Insufficient infrastructure includes** the location of clinics and medical offices, public transportation needed to get to appointments, insurance, childcare, and more.
2. **The technology gap**. More than a third of Black adults in the U.S. do not have broadband at home[9], and they are more at risk of losing mobile services[10]. Some COVID-19 safety precautions, such as closing libraries to in-person services, worsened the gap.
3. **Scarcity of culturally responsive providers.** Less than two percent of American Psychological Association-affiliated providers identify as Black or African American, and many non-Black therapists lack the tools or background to help Black patients' concerns.

While AI and other technologies cannot, and should not, replace human interaction and the need for culturally responsive care, we believe that addressing these inadequacies through equitable and human-centered technology can help people and their providers monitor, manage, and improve their wellbeing.

Where we go from here

We are committed to continuing this conversation, engaging in the work, and making investments to build and support equitable AI solutions for mental health.

No matter your background—as a social worker, mental health provider, technologist, disability rights advocate, person with direct experience with mental health conditions—you have something to contribute. By combining our respective strengths, we can do more to advance AI for better mental health in Black communities.

[9] Internet/Broadband Fact Sheet, Pew Research Center, 2019

[10] Smartphones help blacks, Hispanics bridge some – but not all – digital gap with whites, Pew Research Center, 2019

Ways you can contribute

1. **Become an advocate.** Amplify diverse voices and invest in a wide range of projects and research.
2. **Broaden who participates.** Invite diverse and representative perspectives in every stage of developing a project.
3. **Expand your sources.** Seek out varying viewpoints, including those with lived experience of mental health conditions.
4. **Ask Design Justice questions.** Make these questions a routine part of every project.
5. **Apply for a grant.** If you have an idea that can use AI to empower people with disabilities, we encourage you to learn more about our AI for Accessibility program.

We invite you to join us.

While we deeply appreciate the contributions of more people than we can name individually here, we would particularly like to thank Dr. Desmond Patton, Dr. Courtney Cogburn, Dr. Danielle Hairston, Leigh Felton, Ehi Nosakhare, PhD, Theresa Nguyen, Harold Javid, Jina Suh, Megan Lawrence, Heather Dowdy, Dr. Gloria Washington, Dr. Christine Crawford, Kelechi Ubozoh, Robin Stevens, Jennifer Zhang, Mary Czerwinski, Andre Brock, Lex Barbosa, Catherine Ryan Gregory, Kate Madrid, Noyonima Hassan, Katy Snyder, and Darren Ward.