# *Culture by Design*: A Sociotechnical Framework for Culturally Grounded AI for Mental Health

**Sunny Rai**[a] **Elizabeth C. Stade**[b] **Graise Zhou**[a] **Neil K.R. Sehgal**[a]
**Simone H. Schriger**[a] **Sara Gerke**[c] **Lyle H. Ungar**[a] **Sharath Chandra Guntuku**[a]

[a]University of Pennsylvania, Philadelphia, PA
[b]Stanford University, Stanford, CA
[c]University of Illinois Urbana-Champaign, Champaign, IL
{sunnyrai, sharathg}@seas.upenn.edu

## Abstract

AI systems for mental health are developed predominantly using data from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations, raising concerns about their validity, fairness, and generalizability across diverse cultural and geographic contexts. This limitation is especially consequential in mental health, where linguistic expression, symptom presentation, help-seeking behavior, and access to care vary substantially across populations. We argue that culture must be treated as a first-class design requirement throughout the AI development lifecycle, a principle we term *culture by design*. Drawing on evidence from NLP, clinical psychology, HCI, and global mental health, we present ten practical recommendations spanning data collection, modeling, evaluation, deployment, and governance, providing researchers and practitioners with a concrete roadmap for building culturally grounded, equitable, and contextually appropriate mental health AI systems.

## 1 Introduction

AI-powered mental health systems are increasingly positioned as tools for expanding access to care, especially in settings where mental health resources are scarce (World Health Organization, 2024; Cho et al., 2023). However, existing AI mental health systems are primarily developed using data from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (Henrich et al., 2010), often collected from a small number of English-dominant social media platforms (Harrigian et al., 2021; Cho et al., 2023) (see Fig 2). This raises concerns about whether such systems generalize across cultures (Rai et al., 2025a; Aguirre et al., 2021).

In the mental health literature, *culture* has been defined as the shared, learned behaviors, meanings, values, and beliefs that are transmitted socially across generations and that shape how individuals perceive, communicate, and make sense of psychological experience (Marsella and Yamada, 2000; Triandis, 1996). Critically, culture is not a static attribute but a dynamic, evolving system that mediates the etiology, expression, and treatment of psychopathology (Marsella and Yamada, 2010). This view is reflected in DSM-5's Cultural Formulation framework, which recognizes that cultural identity, cultural idioms of distress, cultural explanations of illness, and psychosocial stressors all interact to shape how mental disorders are experienced and communicated across communities (Lewis-Fernández et al., 2014).

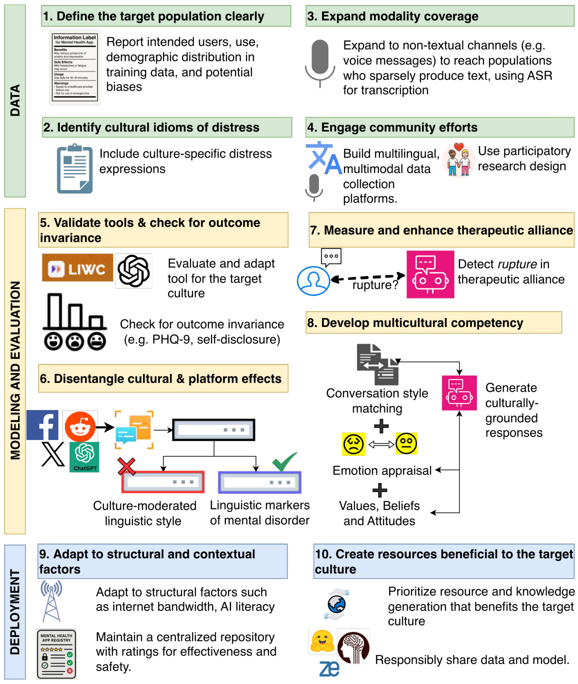


Figure 1: Ten cross-cultural recommendations for designing AI mental health applications. Recommendations are grouped by data design, modeling and evaluation, and deployment.

Language technologies are shaped by the populations represented in their training data and by the assumptions embedded in data collection, labeling,



*Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2026)*, pages 128–137
July 4, 2026 ©2026 Association for Computational Linguistics

and evaluation pipelines (Bender and Friedman, 2018; Blodgett et al., 2020; Gebru et al., 2021). Mental health is a particularly challenging domain for cross-cultural AI because the relationship between language and psychopathology is not culture-invariant: linguistic expressions of distress, symptom presentation, help-seeking behavior, and willingness to self-disclose vary substantially across communities (Cork et al., 2019; Zhao et al., 2012; Hall et al., 2016; Rai et al., 2024). Research on cultural adaptation suggests that effective mental health support must account for local norms, values, communication styles, and structural constraints (Hall et al., 2016; Ge et al., 2024). Yet these insights have not been systematically translated into the AI development lifecycle. We argue that culture must be treated as a first-class design requirement throughout the AI development lifecycle, a principle we term *culture by design*, following the proactive embedding logic of established frameworks such as *ethics by design* (Brey and Dainow, 2024) and *privacy by design* (Cavoukian, 2009). This paper addresses this gap by presenting a sociotechnical framework instantiating this principle across ten practical recommendations spanning data collection, model development, evaluation, and deployment, drawing on evidence from NLP, ML, HCI, clinical psychology, and global mental health.

## 2 Related Work

**Cultural differences in mental health expressions** The manner in which symptoms manifest (e.g., psychological versus somatic (Kaiser and Jo Weaver, 2019) and emotion appraised (Girju and Girju, 2022; Li et al., 2023)) differs from culture to culture. Psychological suffering is communicated through culture-specific concepts and expressions that may not align with standardized Western clinical language (Cork et al., 2019). Culture shapes what people disclose, how they seek support, and what they expect from AI systems (Zhao et al., 2012; Ge et al., 2024). Even within social media-based mental health research, language markers (Rai et al., 2024; Aguirre et al., 2021) and motivations (Pendse et al., 2019) can differ substantially across demographic groups. Recent NLP work further suggests that multilingual models fail to interpret emotion or intent in culturally appropriate ways (Havaldar et al., 2023). As a result, AI mental health systems trained on WEIRD populations may be biased not only because of dataset imbalance,

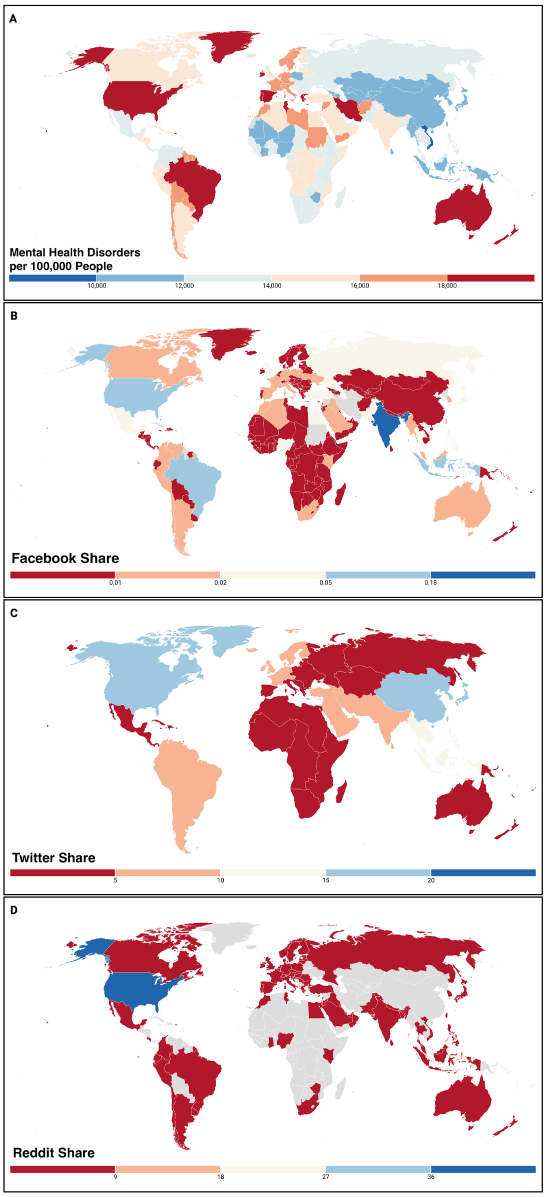


Figure 2: **Geographic mismatch between global mental health burden and the platform ecosystems from which the majority of mental health datasets are drawn**. **Panel A**. Prevalence of Mental Health Disorders - Country Level, 2021. Source: (Global Burden of Disease Collaborative Network, 2021) . **Panel B**. Share of Facebook Users - Country Level, July 2024. Source: (NapoleonCat, 2024) **Panel C**. Share of Twitter Advertising Audience - UN Subregion Level, April 2023. Source: (Kemp, 2023) **Panel D**. Share of Global Reddit Traffic - Country Level, March 2023. Source: (Semrush, 2023). Reddit has no traffic from China; Twitter is more widely used, the audience share is poor for Africa, Oceania, and Eastern Europe. Facebook is most actively used in the United States, India, Southeast Asia, and Brazil. Note that Twitter Audience Share is reported at the United Nations subregion level. As a result, subregions may include countries where platforms are banned (e.g. Twitter is banned in China but is displayed in Panel B due to its belonging in Southern Asia).

but also because the constructs they aim to model are themselves expressed differently across cultural contexts (e.g. shame-related expressions in East vs West (Rai et al., 2025b)).

**Bias, underspecified data, and linguistic diversity in NLP.** A substantial body of work in NLP argues for greater transparency about the populations and conditions under which data are collected (Bender and Friedman, 2018; Gebru et al., 2021; Mitchell et al., 2019). Language technologies can reproduce social biases through data selection, labeling assumptions, and deployment practices (Blodgett et al., 2020), and NLP resources remain heavily concentrated in a small number of high-resource languages (Joshi et al., 2020). Despite the direct relevance of these frameworks to mental health AI, where construct validity is culturally contingent, and data collection practices carry ethical stakes, their uptake in the mental health NLP literature has been limited. Research papers rarely describe population distributions or subgroups in training and test data (Cho et al., 2023), and do not report disparate impact across demographic groups.

**Robustness and evaluation under cultural distribution shift.** ML model performance may degrade substantially under distribution shift, and strong benchmark performance can mask failures under underspecification (Ovadia et al., 2019; D'Amour et al., 2022; Koh et al., 2021). Cultural and linguistic variations that characterize global mental health contexts remain a particularly understudied form of such shift.

In summary, prior work has studied AI-driven mental health and cultural adaptation largely in isolation. To our knowledge, there is limited AI-focused work that integrates these perspectives into a unified framework for culturally grounded AI mental health system development. This paper fills that gap with ten practical recommendations.

## 3 A Sociotechnical Framework for Developing Culturally Grounded AI Mental Health Applications

The ten recommendations are organized across three phases of the AI development lifecycle: **data design**, **modeling and evaluation** , and **deployment and governance**, reflecting the view that cultural grounding must be addressed end-to-end rather than as a post-hoc correction.

**DATA DESIGN**

**1. Define the target population clearly**

The absence of critical demographic information such as age, gender, and race/ethnicity, in training datasets raises concerns about the generalizability of AI models across subpopulations (e.g., how performance differs across racial/ethnic subgroups (Rai et al., 2024)). Researchers should explicitly document demographic details, following established guidelines in medical and psychological research (Kissel and Friedman, 2023). At minimum, a study or model should report data source, language(s), distributions for country/region/race, age, and gender, and source of labels (self-report vs clinical diagnosis vs machine prediction). Additionally, clear, standardized labeling practices for AI mental health applications are essential for communicating the intended user, use, and warnings about potential biases, thereby fostering user trust and informed adoption (Gerke, 2024; Mitchell et al., 2019).

**2. Identify cultural idioms of distress**

Building assessment or intervention applications that rely on self-disclosures about mental health on social media platforms skews the applicability of such applications to demographic groups with existing treatment access or high mental health literacy. This excludes disadvantaged groups that use different cultural idioms of distress (e.g., *tension* in India, *shenjing shuairuo* in China) or cultural narratives to describe their mental health problems, especially on public platforms. When developing AI-based mental health applications, collaborating with cultural psychologists, anthropologists, and local mental health practitioners is crucial to examining how mental health is discussed. Such partnerships can enable the identification, validation, and integration of culturally resonant expressions of psychopathology, ensuring AI applications accurately reflect the diverse ways global communities convey psychological distress (Cork et al., 2019).

**3. Expand modality coverage to reach underserved populations**

AI mental health applications often default to textual interactions, excluding populations that prefer or rely on alternative communication modalities such as voice messages or video calls, including elderly individuals and people with disabilities (Bunyi et al., 2021). This restriction risks systematic exclusion of communities that communicate primarily through non-textual channels, compounding existing biases toward populations with high digital literacy and platform access. Importantly, our concern here is not that non-textual modalities carry richer diagnostic signal but rather that restricting data collection to typed text forecloses access to populations who sparsely produce it. Voice messages, for instance, can now be transcribed with high accuracy using contemporary automatic speech recognition systems and analyzed using standard NLP pipelines; accuracy is high for many high-resource languages, and rapidly improving multilingual models are extending this capability to a broader range of languages (Radford et al., 2023). This substantially broadens the demographic reach of mental health data collection without requiring new modeling infrastructure.

However, collecting and using sensitive data, via modalities like voice and video, which carry increased risks of identification and misuse, must prioritize robust privacy and security protections through an ethics-by-design approach (Brey and Dainow, 2024). Moreover, clear, meaningful informed consent procedures for those contributing their data should be in place that, among other things, transparently communicate the data privacy and security safeguards and risks, such as explicitly informing participants if and how their data can be withdrawn or erased after being incorporated into model training.

### 4. Engage communities in data collection and evaluation

Non-Western communities are often less likely to disclose mental health issues publicly due to stigma or cultural norms (Zhao et al., 2012), threatening the accuracy and applicability of language-based diagnostic models. To improve validity, researchers should proactively engage local communities and adopt culturally sensitive data collection practices such as anonymized structured clinical interviews, widely regarded as diagnostic gold standards (First, 1997). Emerging AI methods, such as in-context learning make carefully collected small-scale datasets particularly valuable.

When seeking ethics approval for cross-cultural mental health data collection, it is worth considering dimensions that standard IRB templates may not capture (Calia et al., 2023; Murray et al., 2016). First, seeking approval from a local ethics board in the target country or community, in addition to the home institution, can strengthen the legitimacy and cultural appropriateness of the research. Second, documenting whether participation carries social or legal risks specific to the target context, such as criminalization of mental health disclosures or employment consequences, can help protect participants in ways that generic risk assessments may overlook. Third, having consent procedures reviewed by local community members or practitioners can ensure linguistic accessibility and compatibility with collective or family-based decision-making norms. Finally, data agreements that specify community ownership and the right to withdraw or restrict use after collection can meaningfully strengthen data sovereignty. Reporting these considerations as part of model card or datasheet helps build transparency and community trust (Gerke, 2023).

## MODELING AND EVALUATION

### 5. Validate tools and check for outcome invariance

The development of culture-specific resources could improve the validity with which relationships between language and mental health are measured across cultures. The most widely used psychosocial dictionary, Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker, 2010), for analyzing linguistic patterns in mental health language is primarily designed for Standard American English. Similarly, current Large Language Models (LLMs) exhibit Anglocentric biases, leading to potential misinterpretations of expressed emotions (Havaldar et al., 2023). Concerted efforts to culturally adapt and validate existing Natural Language Processing (NLP) tools and develop new culturally tailored resources are essential.

Equally important is ensuring that outcome measures employed to evaluate AI mental health applications demonstrate measurement invariance across key demographic characteristics, including gender, race, ethnicity, and religion. Invariance should also be examined across language context: users interacting with interfaces in a non-native language may exhibit systematically different self-disclosure levels, response patterns, and help-seeking behaviors compared to native speakers (Dewaele et al., 2008), which can confound both model training and intervention evaluation. Assessing these effects and designing interfaces that mitigate them is a concrete step toward culturally grounded evaluation practice.

| Recommendation | Priority | Minimum Compliance Criterion | Example Metric or Artifact |
|---|---|---|---|
| *Data Design* | | | |
| 1. Define the target population clearly | **Must Do** | Model card reports platform, language(s), country/region, age, gender distributions, and label source (self-report vs. clinical diagnosis vs. model prediction). | NIH-style demographic table in paper or supplementary materials. |
| 2. Identify cultural idioms of distress | **Good to Have** | Use a community-validated lexicon of culture-specific distress expressions for data collection and annotation. | Codebook of culture-specific distress expressions validated with local mental health practitioners. |
| 3. Expand modality coverage to reach underserved populations | **Must Do** | Data collection includes at least one non-textual modality (e.g., voice) with documented rationale for population reach; privacy safeguards for personally identifiable information explicitly documented. | IRB approval reported; demographic coverage of collected modalities documented; data anonymization and de-identification procedure described in supplementary material. |
| 4. Engage communities in data collection | **Good to Have** | Community involvement documented via structured participation log specifying who was consulted, at what stage, and how input changed data collection or annotation decisions. | Anonymized structured clinical interviews used for at least a subset of training data; participation log included in datasheet. |
| *Modeling and Evaluation* | | | |
| 5. Validate tools and check for outcome invariance | **Must Do** | Evaluation reports subgroup calibration and disparate impact across demographic groups alongside aggregate metrics. | Differential item functioning (DIF) analysis or confirmatory factor analysis across cultural groups; disparate impact table by subgroup. |
| 6. Disentangle cultural and platform effects | **Good to Have** | Sensitivity analyses conducted within platforms (linguistic features vs. outcomes) and across platforms (cross-platform generalization of mental health markers). | Performance reported per platform; cross-platform transfer experiment with at least two platforms (e.g., Reddit → Facebook). |
| 7. Measure and enhance therapeutic alliance | **Best Practice** | LLM-based alliance scoring applied to conversation logs; alliance scores tested for differential prediction of dropout or satisfaction across cultural subgroups. | Alliance rupture detection F1 reported by cultural group. |
| 8. Develop multicultural competency | **Best Practice** | Model evaluated on whether suggested rephrasing reduces culturally specific miscommunication markers relative to a baseline measured by conversation outcomes such as self-disclosure; annotator cultural background reported. | Culturally annotated clinical dialogue corpus; reduction in miscommunication flags; empathy ratings stratified by annotator cultural background . |
| *Deployment and Governance* | | | |
| 9. Adapt to structural and contextual factors | **Good to Have** | Accessibility or reach stratified by interface modality, input language proficiency, and device type; at least one ablation targets a low-resource deployment constraint. | Performance reported for voice vs. text; low-bandwidth variant evaluated on task completion and user satisfaction. |
| 10. Create resources beneficial to the target culture | **Best Practice** | At least one outcome measure preregistered in consultation with the target community; a dataset, lexicon, or model released publicly and remains accessible after the study concludes. | Community-defined outcome scale (e.g., family functioning) used alongside standard symptom measures; data or model weights deposited in an open repository. |

Table 1: Ten recommendations with priority level, minimum compliance criteria, and example metrics or artifacts. **Must Do**: absence constitutes a reportable methodological gap. **Good to Have**: strengthens validity and is expected at high-impact venues. **Best Practice**: aspirational; marks the current frontier of culturally grounded AI.

### 6. Disentangle cultural and platform-specific effects

Social media platform use is deeply influenced by cultural context (Burke et al., 2020), potentially confounding AI models that rely on social media language for mental health assessment. Without distinguishing between culture-specific platform usage patterns and genuine mental health indicators, models risk inaccurate assessments and biased conclusions. Therefore, rigorous sensitivity analyses are essential, both within platforms (examining how culturally distinct linguistic features relate to mental health conditions) and across different platforms (evaluating whether mental health markers identified on one platform, like Reddit, generalize effectively to another, like Facebook). Such careful testing ensures that AI mental health assessments accurately reflect mental health conditions, rather than platform-specific or culturally mediated communication norms.

### 7. Measure and enhance therapeutic alliance across cultures

Therapeutic alliance, the quality of trust and rapport between user and mental health support, is strongly influenced by cultural and demographic factors, such as preferences for emotional expression or interaction styles. Recent work has begun operationalizing alliance monitoring in LLM-based systems: Li et al. (2024) proposed an LLM-based approach to measuring therapeutic alliance in online counseling, while Chiu et al. (2024) introduced a framework for systematically comparing LLM therapy behaviors against high- and low-quality human sessions. These approaches provide concrete scaffolding for detecting alliance ruptures and cultural mismatches in real-time, though they have not yet been validated across cultural subgroups, a gap this recommendation directly targets. For instance, Chinese users may prefer emotionally expressive AI systems that foster personal connections, whereas European and American users favor more impersonal, controllable interactions (Ge et al., 2024), illustrating the cultural specificity that alliance models must account for. Monitoring private user conversations for alliance must carefully balance clinical responsibilities against user privacy and autonomy.

### 8. Develop multicultural competency

LLMs offer significant potential to advance multicultural competence in mental health care by adapting to the cultural values and communication preferences underlying patient behaviors and treatment responses (Liu et al., 2026). Given patients' strong preferences for therapists sharing similar cultural or ethnic backgrounds (Cabral and Smith, 2011), LLMs could augment therapist training by generating culturally tailored therapeutic scenarios, identifying culturally embedded biases in clinical language, and suggesting more culturally sensitive communication alternatives. For instance, AI-generated guidance could assist therapists in adapting interaction styles to avoid culturally specific misinterpretations, such as perceptions of rudeness arising from direct communication with traditional East Asian clients.

## DEPLOYMENT AND GOVERNANCE

### 9. Adapt to structural and contextual factors

Effective AI mental health applications require careful consideration of structural barriers faced by real individuals, such as shared device access, limited internet connectivity, varying literacy, and language proficiency. Tailoring technologies to these contextual realities can significantly enhance user satisfaction and perceived usefulness.

Establishing a centralized, transparent registry for mental health applications could be a promising way to strengthen accountability and build public trust. Such a registry could disclose valuable information, including the demonstration of safety and effectiveness (e.g., clinical trial outcomes), robust privacy and security protections, and regulatory compliance (e.g., compliance with medical device regulations) (Gerke et al., 2020).

### 10. Create resources beneficial to the target culture

AI mental health research and applications must explicitly prioritize the generation of knowledge that benefits the cultures and communities they aim to serve. Researchers should contextualize mental health symptoms within specific cultural and social realities, thereby avoiding inadvertent imposition of external, often Western, frameworks, which can distort or misrepresent culturally embedded behaviors. For instance, conventional outcome measures, such as individual symptom reduction, may inadequately reflect culturally significant goals in collectivist societies, where communal harmony and family well-being are paramount. When designing AI-driven mental health applications, developers should measure outcomes consistent with the beliefs and values of the target patient populations.

Furthermore, developers must rigorously adhere to applicable legal frameworks, including data privacy laws and medical device and/or AI regulations. Even when regulatory classification as wellness tools potentially bypasses stricter medical device scrutiny or a country has no comprehensive privacy law, an ethics-by-design approach remains essential, and developers retain discretion to adopt standards beyond minimum legal requirements (Gerke, 2022). Sharing anonymized data with proper safeguards (e.g., a ban on reidentification) and trained models will help compare the generalizability of mental health markers across cultures and translate research insights into technology.

## Conclusion

Cultural grounding should be treated as a core requirement in the design and deployment of AI systems for mental health. The principle of *culture by design* demands deeper interdisciplinary engagement, including ethical data governance, overcoming data silos, culturally adapted informed consent processes, robust regulatory compliance, fostering patient trust, and seamless integration into existing local mental health infrastructures. The framework presented here is intended as a practical foundation for future interdisciplinary dialogue and research on the sociotechnical complexities essential for developing and deploying equitable, culturally grounded AI mental health applications worldwide.

## Limitations

Our recommendations synthesize prior work across NLP, HCI, clinical psychology, psychiatry, and global mental health; however, we do not empirically quantify their relative impact on downstream model performance, safety, or equity. Our treatment of regulation, ethics, and governance is intentionally broad and non-jurisdiction-specific. Legal requirements governing privacy, informed consent, data sharing, and medical device classification vary substantially across countries; nothing in this discussion should be construed as legal guidance. Finally, several recommendations involve inherent trade-offs. Collecting richer multimodal or culturally specific data may improve representational validity while simultaneously increasing privacy risks, annotation complexity, and barriers to data sharing. Likewise, tailoring systems to a particular community may improve local utility at the cost of cross-population comparability or scalability. We regard these tensions not as arguments against culturally grounded design, but as central challenges for future interdisciplinary research.

## Acknowledgments

S.C.G. was partly supported by NIH-NIMHD (R01MD018340), the Penn Global Research Engagement Fund, and the Leonard Davis Institute of Health Economics. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. S.G.'s work was supported by a Novo Nordisk Foundation Grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL Programme — grant number NNF23SA0087056).

## References

Carlos Aguirre, Keith Harrigian, and Mark Dredze. 2021. Gender and racial fairness in depression research using social media. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2932–2949. Association for Computational Linguistics.

Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

Philip Brey and Brandt Dainow. 2024. Ethics by design for artificial intelligence. *AI and Ethics*, 4(4):1265–1277.

Jessica Bunyi, Kathryn E. Ringland, and Stephen M. Schueller. 2021. Accessibility and digital mental health: Considerations for more accessible and equitable mental health apps. *Frontiers in Digital Health*, 3:742196.

Moira Burke, Justin Cheng, and Benjamin de Gant. 2020. Social comparison and facebook: Feedback, positivity, and opportunities for comparison. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA. Association for Computing Machinery.

R. R. Cabral and Timothy B. Smith. 2011. Racial/ethnic matching of clients and therapists in mental health services: A meta-analytic review of preferences, perceptions, and outcomes. *Journal of Counseling Psychology*, 58(4):537–554.

Clara Calia, Amit Chakrabarti, Emmanuel Sarabwe, and Anna Chiumento. 2023. Maximising impactful, locally relevant global mental health research conducted in low and middle income country settings: Ethical considerations. *Wellcome open research*, 7:240.

Ann Cavoukian. 2009. Privacy by design: The 7 foundational principles. Technical report, Information and Privacy Commissioner of Ontario, Toronto, Canada.

Yu Ying Chiu, Ashish Sharma, Inna Wanyin Lin, and Tim Althoff. 2024. A computational framework for behavioral assessment of llm therapists. *arXiv preprint arXiv:2401.00820*.

Young Min Cho, Sunny Rai, Lyle Ungar, João Sedoc, and Sharath Guntuku. 2023. An integrative survey on mental health conversational agents to bridge computer science and medical perspectives. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11346–11369, Singapore. Association for Computational Linguistics.

Catrin Cork, Bonnie N. Kaiser, and Ross G. White. 2019. The integration of idioms of distress into mental health assessments and interventions: A systematic review. *Global Mental Health*, 6:e7.

Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D Hoffman, and 1 others. 2022. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, 23(226):1–61.

Jean-Marc Dewaele, Konstantinos V Petrides, and Adrian Furnham. 2008. Effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language learning*, 58(4):911–960.

Michael B First. 1997. Structured clinical interview for dsm-iv axis i disorders. *Biometrics research department*.

Xiaoqi Ge, Chenyue Xu, Donna Misaki, Hazel R. Markus, and Jeanne L. Tsai. 2024. How culture shapes what people want from AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–15, Honolulu, HI, USA. ACM.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92.

Sara Gerke. 2022. Digital home health during the COVID-19 pandemic: Challenges to safety, liability, and informed consent, and the way to move forward. In Carmel Shachar, Christopher Robertson, I. Glenn Cohen, Timo Minssen, and W. Nicholson Price II, editors, *The Future of Medical Device Regulation: Innovation and Protection*, pages 141–160. Cambridge University Press, Cambridge.

Sara Gerke. 2023. "Nutrition Facts Labels" for Artificial Intelligence/Machine Learning-based medical devices-the urgent need for labeling standards. *Geo. Wash. L. Rev.*, 91:79.

Sara Gerke. 2024. A comprehensive labeling framework for Artificial Intelligence (AI)/Machine Learning (ML)-based medical devices: From AI facts labels to a front-of-package AI labeling System—Lessons learned from food labeling. *Emory LJ*, 74:1297.

Sara Gerke, Ariel D Stern, and Timo Minssen. 2020. Germany's digital health reforms in the covid-19 era: lessons and opportunities for other countries. *NPJ digital medicine*, 3(1):94.

Roxana Girju and Marina Girju. 2022. Design considerations for an nlp-driven empathy and emotion interface for clinician training via telemedicine. In *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 21–27.

Global Burden of Disease Collaborative Network. 2021. Global Burden of Disease Study 2021 (GBD 2021) Incidence, Prevalence, and Years Lived with Disability 1990–2015. `https://ghdx.healthdata.org/record/ihme-data/gbd-2015-incidence-prevalence-and-ylds-1990-2015`. Accessed: 2024.

Gordon C. Nagayama Hall, A. Y. Ibaraki, Ellen R. Huang, C. Nathan Marti, and Eric Stice. 2016. A meta-analysis of cultural adaptations of psychological interventions. *Behaviour Therapy*, 47(6):993–1014.

Keith Harrigian, Carlos Aguirre, and Mark Dredze. 2021. On the state of social media data for mental health research. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 15–24.

Shreya Havaldar, Bhumika Singhal, Sunny Rai, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual language models are not multicultural: A case study in emotion. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 202–214, Toronto, Canada. Association for Computational Linguistics.

Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Bonnie N Kaiser and Lesley Jo Weaver. 2019. Culture-bound syndromes, idioms of distress, and cultural concepts of distress: New directions for an old concept in psychological anthropology. *Transcultural psychiatry*, 56(4):589–598.

Simon Kemp. 2023. Digital 2023 April Global Statshot Report. `https://datareportal.com/reports/digital-2023-april-global-statshot`. Twitter advertising audience share data at UN subregion level. Accessed: April 2023.

Heather A Kissel and Bruce H Friedman. 2023. Participant diversity in psychophysiology. *Psychophysiology*, 60(11):e14369.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, and 1 others. 2021. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5637–5664. PMLR.

Roberto Lewis-Fernández, Neil Krishan Aggarwal, Sofie Bäärnhielm, Hans Rohlof, Laurence J Kirmayer, Mitchell G Weiss, Sushrut Jadhav, Ladson Hinton, Renato D Alarcón, Dinesh Bhugra, and 1 others. 2014. Culture and psychiatric evaluation: operationalizing cultural formulation for dsm-5. *Psychiatry: Interpersonal and biological processes*, 77(2):130–154.

Anqi Li, Yu Lu, Nirui Song, Shuai Zhang, Lizhi Ma, and Zhenzhong Lan. 2024. Understanding the therapeutic relationship between counselors and clients in online text-based counseling using llms. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1280–1303.

Anqi Li, Lizhi Ma, Yaling Mei, Hongliang He, Shuai Zhang, Huachuan Qiu, and Zhenzhong Lan. 2023. Understanding client reactions in online mental health counseling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10358–10376.

Chen Cecilia Liu, Hiba Arnaout, Nils Kovačić, Dana Atzil-Slonim, and Iryna Gurevych. 2026. Tailored emotional llm-supporter: Enhancing cultural sensitivity. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 535–574.

Anthony J Marsella and Ann Marie Yamada. 2000. Culture and mental health: An introduction and overview of foundations, concepts, and issues. In *Handbook of multicultural mental health*, pages 3–24. Elsevier.

Anthony J Marsella and Ann Marie Yamada. 2010. Culture and psychopathology: Foundations, issues, directions. *Journal of Pacific Rim Psychology*, 4(2):103–115.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 220–229, New York, NY, USA. Association for Computing Machinery.

SM Murray, N Kass, T Mendelson, and J Bass. 2016. The ethics of mental health survey research in low- and middle-income countries. *Global Mental Health*, 3:e12.

NapoleonCat. 2024. Facebook users by country—global statistics. `https://stats.napoleoncat.com`. Data sourced from Meta Advertising Tools via NapoleonCat. Licensed under Creative Commons Attribution 4.0 International. Accessed: July 2024.

Yarin Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek. 2019. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, volume 32.

Sachin R Pendse, Kate Niederhoffer, and Amit Sharma. 2019. Cross-cultural differences in the use of online mental health support forums. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–29.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning*.

Sunny Rai, Khushi Shelat, Devansh Jain, Ashwin Kishen, Young Min Cho, Maitreyi Redkar, Samindara Hardikar-Sawant, Lyle Ungar, and Sharath Chandra Guntuku. 2025a. Cross-cultural differences in mental health expressions on social media. In *Proceedings of the 3rd Workshop on Cross-Cultural Considerations in NLP (C3NLP 2025)*, pages 132–142.

Sunny Rai, Elizabeth C Stade, Salvatore Giorgi, Ashley Francisco, Lyle H Ungar, Brenda Curtis, and Sharath C Guntuku. 2024. Key language markers of depression on social media depend on race. *Proceedings of the National Academy of Sciences*, 121:e2319837121.

Sunny Rai, Khushang Zaveri, Shreya Havaldar, Soumna Nema, Lyle Ungar, and Sharath Chandra Guntuku. 2025b. Social norms in cinema: A cross-cultural analysis of shame, pride and prejudice. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 11396–11415.

Semrush. 2023. Reddit.com traffic analytics and market share—country level. `https://www.semrush.com/website/reddit.com/overview/`. Global traffic share by country for Reddit.com, March 2023. Accessed: March 2023.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Harry C Triandis. 1996. The psychological measurement of cultural syndromes. *American psychologist*, 51(4):407.

World Health Organization. 2024. How AI could help improve access to mental health treatment. Accessed via World Economic Forum.

Chao Zhao, Pamela Hinds, and Ge Gao. 2012. How and to whom people share: The role of culture in self-disclosure in online communities. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pages 67–76, New York, NY, USA. Association for Computing Machinery.