Multimedia Appendix I: Framework for AI Tool Assessment in Mental Health (FAITA-Mental Health)

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Subdimension** | **Maximum points** | **Scoring criteria** |
| **Credibility** | Proposed goal | 2 | 2:Specific, measurable, achievable mental health goals described**.**  1:General, non-specific mental health goals described.  0:No clear goals described. |
|  | Evidence-based content | 2 | 2:Exclusively uses evidence-based practices and principles, well-supported by current research in mental health interventions.  1:Employs a mix of evidence-based and non-evidence- based practices and principles; some efforts towards grounding in research are evident, but not consistently applied across all content or functionalities.  0: Lacks evidence-based practices and principles, with no clear grounding in current mental health research or established therapeutic methodologies. |
|  | Retention | 2 | 2:High retention or evidence of positive churn through users departing after meeting their goals, with over 70% of users actively engaged for a predefined period or successfully completing their journey towards set goals.  1:Moderate retention rates or some instances of positive churn, with 40-70% of users staying engaged for a predefined period or achieving their goals.  0:Low retention rates with fewer than 40% of users remaining engaged for a predefined period, and no clear evidence of positive churn. |
| **User experience** | Personalization and evolution | 2 | 2:Demonstrates strong personalization, adapts interactions in real time based on user input, and evolves through user feedback for continuous improvement.  1:Offers limited personalized interaction or adaptation based on user feedback.  0:Lacks significant personalization in its interactions and shows no clear signs of adaptation or improvement based on user feedback. |
|  | Interactivity quality | 2 | 2:Interactions are consistently natural, contextually appropriate, and supportive of mental health needs.  1:Interactions show some level of naturalness and appropriateness, but inconsistently.  0:The intervention frequently fails to provide natural or contextually appropriate interactions. |
|  | Feedback mechanism and support | 2 | 2:Provides clear, accessible channels for users to give feedback or seek support, with evidence of responsive support or acknowledgment.  1:Limited feedback mechanisms are available, with minimal support or acknowledgment.  0:No evidence of feedback channels or support mechanisms for users. |
| **User agency** | User autonomy, data protection, and privacy | 2 | 2: Employs advanced data protection measures (e.g., end-to-end encryption, secure data storage) and comprehensive user autonomy over personal health data, with explicit mechanisms including user consent, data sharing preferences, and the ability to access, amend, or delete personal information. Privacy policies, user consent, and other pertinent information are presented in clear, simple language.  1:Provides basic data protection and privacy controls, including some level of encryption and secure storage, with limited user autonomy over data consent, access, and management. Privacy policies, user consent, and other pertinent information may lack consistent clarity and simplicity.  0:Lacks information about data protection, user privacy, or autonomy mechanisms. Privacy policies, user consent, and other pertinent information are either absent or only available in complex legal jargon. |
|  | User empowerment | 2 | 2:Actively promotes user empowerment and minimizes potential dependency, offering tools and resources that encourage self-efficacy in managing mental health and offering opportunities for users to exercise informed  choice over intervention pathways aligned with their goals, preferences, and needs.  1:There are some elements of user empowerment (e.g., the user is presented with the choice to pursue one technique over another; the user is encouraged to  practice skills and techniques in real life beyond the context of the AI), but these may be limited or not fully developed.  0:Lacks efforts to empower users, missing opportunities to support active mental health management through choice or neglecting to encourage the application of techniques in users’ daily lives. |
| **Equity and inclusivity** | Cultural sensitivity and inclusivity | 2 | 2:Demonstrates explicit efforts to incorporate cultural diversity and inclusivity in interactions.  1:Some efforts toward cultural inclusivity are evident in interactions but are limited in scope or depth.  0:Little to no consideration for cultural diversity or inclusivity is apparent in interactions. |
|  | Bias and fairness | 2 | 2:Demonstrates proactive, thorough efforts to reduce bias and promote equitable support for all users, including the use of training data that is diverse  (representing different cultures, ethnicities, and other identity variables) and actively purged of biases.  1:Acknowledges bias and makes some efforts to reduce it but lacks comprehensive bias-mitigation methods.  0:Displays little to no effort to mitigate bias. |
| **Transparency** |  | 2 | 2: Includes clear, thorough details on the development team or creators, ownership, funding sources, business model, training and development approaches, and primary beneficiaries.  1: Shares some details on the development team or creators, ownership, funding sources, business model, training and development approaches, and beneficiaries, but lacks comprehensive details or full disclosure.  0: Lacks any information on the development team or creators, ownership, funding sources, business model, training and development approaches, and beneficiaries. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Safety and**  **crisis management** |  | 2 | 2:Demonstrates comprehensive safety protocols and crisis management features, including real-time crisis interventions, direct links to relevant crisis and emergency services, and strategies to optimize user  engagement with such resources.  1: Displays basic safety or crisis management features, such as the presence of crisis hotline numbers or links to emergency services, with minimal efforts to foster engagement with these features.  0:Lacks safety protocols or crisis management features. |