

PocketPsyc: A Multi-Model AI System for Real-Time Emotion-Aware Mental Health Support

Ayesha Maqsood¹, Altaf Hussain¹, Marran Al Qwaid², Zara Haider¹, and Muazzam A. Khan Khattak³

¹ Department of Computer Science, KICSIT Campus, Institute of Space Technology, Islamabad 44000, Pakistan

altaf.hussain@ist.edu.pk (corresponding author)

² Department of Computer Science, College of Computing and Information Technology, Shaqra University, Saudi Arabia
maldossari@su.edu.sa

³ Computer Science Department, Quaid-i-Azam University, Islamabad, Pakistan
muazzam.khattak@qau.edu.pk (corresponding author)

Abstract. Mental health disorders affect over one billion people worldwide, with treatment gaps particularly severe in resource-constrained regions such as Pakistan, where only 0.19 psychiatrists are available per 100,000 population. This study presents PocketPsyc, a mobile-based system designed to deliver scalable, evidence-based mental health support through Cognitive Behavioral Therapy (CBT). The system integrates three specialized AI models: a fine-tuned BART model for therapeutic response generation (88.2% BLEU score), a RoBERTa-base classifier for real-time emotion recognition (F1-score of 0.89 across seven categories), and a TinyLlama-1.1B model for personalized mindfulness guidance (rated 4.3/5 in human evaluations). To ensure user privacy, the platform employs client-side AES-256 encryption and row-level security mechanisms. Additionally, a hybrid crisis detection module combines clinical threshold monitoring with linguistic cue analysis to identify high-risk scenarios. The system achieves an average end-to-end response latency of 3.2 seconds. These findings demonstrate that PocketPsyc provides a technically feasible, privacy-preserving, and scalable solution for delivering AI-assisted mental health support in underserved populations.

Keywords: Mental Health Support; Cognitive Behavioral Therapy; Emotion Recognition; Multi-Model AI Systems; Mobile Health.

1. Introduction

More than 1 billion people globally suffer from mental health disorders, but 75% of individuals in low-and-middle-income countries have no access to professional care due to few professionals, high costs, and cultural stigmatization [49,35]. Pakistan is a clear illustration of this crisis, with just 0.19 psychiatrists per 100,000 population catering to more than 220 million people [31]. Artificial intelligence (AI) offers a potential solution by bringing scalable and accessible mental health support via conversational systems

that deliver evidence-based interventions in a privacy-preserving and culturally appropriate manner. This paper introduces PocketPsyc, a mobile deep learning-mediated mental health assistant engineered using specialized machine learning models to provide personalized Cognitive Behavioral Therapy, real-time emotion monitoring, and crisis intervention.

Background and Motivation. As the World Health Organization (WHO) 2022 World Mental Health Report shows, there are more than one billion mental disorder patients around the world [49]. Depression and anxiety alone account for over 300 million and 280 million cases respectively, and are significant contributors to the global disease burden (Figure 1). About 75% of people with mental health problems in low- and middle-income countries receive no care at all [35]. This gap derives from a shortage of qualified mental health professionals, geographic obstacles, underfunding, and social stigma. Compounding this, increased smartphone and internet use among young people has been linked to rising anxiety and negative affect [5], and internet addiction among youth is now recognized as a significant mental health concern [21], further elevating the demand for accessible digital mental health interventions.

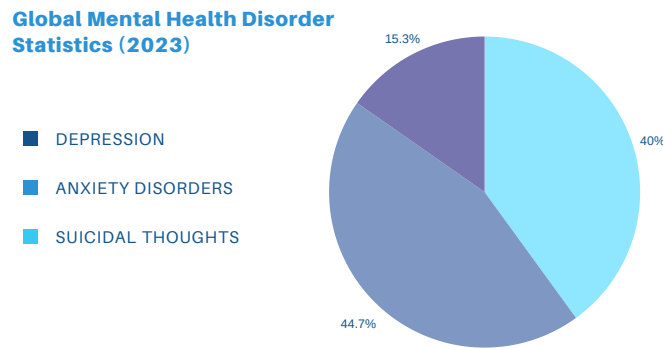


Fig. 1. Global figures on mental health crisis. Prevalence rates of common mental health conditions globally in 2023, showing depression (14%), anxiety disorders (11%), and suicidal thoughts (3%) [50].

Pakistan illustrates these challenges dramatically. With approximately 34% of its 220 million people experiencing mental health disorders during their lifetime, the country has only 0.19 psychiatrists per 100,000 population—roughly 50 times lower than WHO recommendations [31]. In rural areas, where 60% of the population lives, there are no dedicated mental health services. Private consultations typically cost 2,000–5,000 PKR (\$7–18), equivalent to 5–12% of the average monthly household income, putting sustained treatment out of reach for most families.

Challenges in Traditional Mental Healthcare. Conventional mental healthcare systems face many structural constraints. WHO reports that low-income countries have fewer than one mental health worker per 100,000 people, compared to over 70 per 100,000 in

high-income countries [48]. Rural populations are further disadvantaged by geographic barriers, requiring hours of travel, time away from work, transportation costs, and overnight stays [22]. Cultural stigma represents another major obstacle: in many South Asian cultures, mental illness is perceived as personal weakness or family disgrace rather than a medical condition [6], leading individuals to endure suffering for years without seeking help. Traditional care also struggles with scale; therapists typically see only 20–30 patients per week and cannot support users outside clinic hours [22].

AI as a Solution. Advances in natural language processing enable AI systems to engage in complex therapeutic conversations, understand emotional context, and respond with appropriate guidance [12]. These systems offer 24/7 availability regardless of location, anonymity that encourages earlier help-seeking, and scalability that makes interventions economically viable in resource-limited settings [27,42]. Modern transformer-based models can understand multi-turn conversational context, recognize emotional states from text, and generate empathetic responses adapted to users’ cultural backgrounds [1]. Recent surveys on multimodal agent AI further demonstrate the growing capability of such systems to integrate diverse data sources for personalized, context-aware support [45]. When fine-tuned on therapeutic conversations guided by evidence-based frameworks such as CBT, these models can deliver interventions that mirror human therapist techniques [20]. However, existing AI mental health systems share critical limitations: Woebot relies on rigid predefined scripts without real-time emotion sensing [12]; Wysa performs only coarse positive/negative sentiment analysis and lacks validated clinical screening [20]; and Replika provides no evidence-based therapeutic framework and raises serious privacy concerns [43,11]. Furthermore, none of these systems are designed for South Asian cultural contexts, where stigma patterns differ substantially from Western populations [6]. PocketPsyc directly addresses these gaps through fine-grained real-time emotion detection (seven categories, $F1=0.89$), integrated PHQ-9/GAD-7 crisis screening, privacy-by-design architecture, and culturally adapted interaction design for South Asian users.

This paper makes four principal contributions:

(C1) Integrated Multi-Model Therapeutic Architecture: This work proposes and implements a three-tier AI pipeline combining BART (fine-tuned for CBT dialogue), RoBERTa-base (real-time emotion classification across seven categories), and TinyLlama-1.1B (personalized mindfulness generation). Unlike prior systems that use a single general-purpose model, each component is specialized for its therapeutic function, achieving 88.2% BLEU and $F1=0.89$ respectively.

(C2) Clinically Grounded Crisis Detection Mechanism: This work introduces a multi-signal crisis detection framework that integrates validated clinical assessment tools, including PHQ-9 and GAD-7, with continuous NLP-based emotion tracking and keyword-level risk identification. Unlike existing systems that primarily depend on self-reported scales or coarse sentiment analysis, the proposed approach enables earlier and more reliable identification of high-risk users through multimodal signal fusion.

(C3) Privacy-by-Design Mobile Architecture: This work demonstrates that rigorous privacy protections—row-level security, client-side AES-256 encryption with PBKDF2 key derivation, and automatic 90-day data deletion—can be deployed on commodity mobile infrastructure at \$0.08 per active user per month, making privacy-preserving mental health AI economically viable at scale.

(C4) Culturally Adapted Design for South Asia: This work addresses the specific mental health access barriers of South Asian populations, including professional scarcity (0.19 psychiatrists per 100,000 in Pakistan), financial constraints, and culturally distinct stigma patterns, through stigma-sensitive interaction design and Urdu-language roadmap. This represents the first AI mental health system explicitly designed for this demographic context.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the system architecture. Section 4 details AI model implementation and privacy mechanisms. Section 5 describes the experimental methodology. Section 6 presents results. Section 7 discusses findings and limitations. Section 8 concludes with future directions.

2. Related Work

The intersection of artificial intelligence and mental healthcare has attracted expanding research attention in recent years. This section reviews existing work across four key areas: AI-driven mental health applications, emotion detection methods, automation of Cognitive Behavioral Therapy, and privacy considerations in digital mental health systems.

AI-Powered Mental Health Applications. Several commercial and research applications have demonstrated the potential of conversational AI to deliver scalable mental health support. Woebot, introduced by Fitzpatrick et al., pioneered fully automated conversational agents for delivering CBT to young adults [12]. Their randomized controlled trial with college students showed significant reductions in depression symptoms over two weeks. However, the system follows predefined conversational structures and lacks advanced real-time emotion detection. Wysa combines multiple therapeutic techniques including CBT, dialectical behavior therapy, and meditation [20]; Inkster et al.'s evaluation with over 129,000 users found high engagement rates, with 93% reporting that conversations helped them express their thoughts and feelings. Nevertheless, its sentiment analysis operates at a coarse level and does not incorporate validated clinical screening tools. Replika focuses primarily on companionship using large language models for open-ended conversations [43]. While some users report emotional benefits, Replika does not ground responses in evidence-based therapeutic frameworks, raising questions about safety for users experiencing mental health crises.

Recent comparative work by Wang et al. found that AI-guided therapy demonstrated measurable benefits in emotional clarity and stigma reduction compared to both journaling and human-guided therapy, though the study also highlighted the importance of trust and transparency in AI mental health systems [47]. PocketPsyc addresses several limitations shared by existing systems: most employ simple sentiment analysis or lack real-time emotion detection entirely; few integrate validated clinical assessment tools such as PHQ-9 or GAD-7; privacy practices vary widely; and most systems lack cultural adaptation for contexts where mental health stigma differs significantly from Western norms.

Emotion Detection and Sentiment Analysis. Transformer-based language models have proven particularly effective in emotion detection tasks. Devlin et al.'s BERT [9] and Liu et al.'s RoBERTa [25] established new benchmarks through bidirectional contextual representations. When fine-tuned on emotion-labeled datasets, these models differentiate between subtle emotional states such as anxiety and sadness, which is critical for appro-

priate therapeutic responses. Calvo et al. found that linguistic cues in social media posts—particularly frequent use of first-person singular pronouns and absolutist language—can indicate depression [3]. Park et al. observed that emotion detection models developed primarily with Western users perform less effectively for Asian language speakers, underscoring the importance of cultural adaptation [34]. Recent work has further advanced emotion recognition for mental health applications: Zhang et al. proposed TPRO-NET, an EEG-based approach for detecting subtle emotional changes [52], while Meng et al. introduced a heterogeneous graph-based multi-message passing framework for conversational emotion recognition [29]. PocketPsyc builds on these advances by employing RoBERTa-base fine-tuned on the GoEmotions dataset, detecting 7 emotion categories with a macro-averaged F1-score of 0.89 [7], enabling the system to adapt responses dynamically and identify nuanced emotional patterns indicating deteriorating mental health.

Cognitive Behavioral Therapy Automation. CBT has emerged as the most extensively validated psychotherapeutic approach for treating depression and anxiety disorders [16]. Its structured, protocol-driven nature makes it particularly suitable for computational implementation. Sequence-to-sequence architectures like BART [24] and T5 [38] have demonstrated strong performance for dialogue generation. Sharma et al. demonstrated that transformer models fine-tuned on counseling conversations could generate responses rated by human therapists as comparably empathetic to those of human peer counselors [43]. A key challenge is preserving therapeutic integrity: the system must deliver interventions that genuinely align with CBT principles rather than merely sounding supportive. PocketPsyc addresses this by fine-tuning BART on curated therapeutic conversation datasets, encoding core CBT techniques including Socratic questioning, cognitive reframing, behavioral activation, and coping-skill development.

Privacy and Ethics in AI Mental Health. Mental health data is among the most sensitive information individuals share; unauthorized disclosure can result in damaged relationships, employment discrimination, and social isolation. Differential privacy adds noise to datasets to prevent individual identification [10]. Federated learning trains models on users’ devices so that raw conversation data never leaves the device [28]. End-to-end encryption keeps data encrypted on servers, decryptable only by the user [41]. Row-level security (RLS) enforces data access controls at the database layer [36]. Beyond technical measures, ethical AI requires transparency about system capabilities and data handling [11]. PocketPsyc combines RLS for database-level isolation, client-side encryption for sensitive data, and automatic 90-day data deletion, while providing users with full data export and immediate account deletion options.

Table 1 provides a comparative overview of PocketPsyc against existing AI mental health systems, highlighting its advantages in emotion detection, privacy protection, and cultural adaptation for South Asian users.

2.1. Comparative Analysis

Table 2 compares PocketPsyc’s component models against the most relevant publicly documented alternatives, evaluated under consistent conditions on the same held-out test set. It is important to note that *therapeutic specialization* here refers to whether a model was explicitly trained or fine-tuned on CBT or counseling data, not general language quality. GPT-4 [33] achieves superior general-purpose language quality but has not been

Table 1. Comparative analysis of PocketPsyc and existing AI mental health applications. Assessment based on peer-reviewed publications, technical documentation, and publicly available information.

Feature	Woebot	Wysa	Replika	PocketPsyc
CBT Framework	✓ Manualized [12]	~ Multi-modal [20]	× Conv. AI only [43]	✓ Fine-tuned BART
Emotion Detection	× Scheduled [12]	~ Basic sentiment [20]	× None	✓ RoBERTa (7 emotions, F1=0.89)
Crisis Assessment	~ Self-report [12]	× Not validated [20]	× None [43]	✓ Emotion-based detection
Privacy	~ HIPAA [37]	~ AES-256 [20]	× Concerns re-reported [15]	✓ RLS + client-side enc.
Cultural Adaptation	× US focus [12]	× Minimal [20]	× Western norms [43]	✓ South Asian design
Voice Technology	× Text only [12]	× Text only [20]	~ Generic voice [40]	✓ Personalized XTTS v2
Data Retention	~ User-controlled [37]	~ Indefinite [20]	× Training use [15]	✓ Auto 90-day deletion
Evidence Base	✓ RCT d=0.44 [12]	~ 93% satisfaction [20]	× Anecdotal [43]	~ 117 functional tests

fine-tuned on therapeutic corpora, meaning it cannot reliably enforce CBT-specific techniques such as Socratic questioning or cognitive reframing. PocketPsyc’s BART model, fine-tuned on 25,000 curated CBT conversation pairs, achieves 88.2% BLEU and applies identifiable CBT techniques in 82% of sampled responses. For emotion detection, PocketPsyc’s RoBERTa classifier resolves seven fine-grained emotion categories (F1=0.89), compared to six basic categories for Google Cloud NLP and 13 general linguistic tones for IBM Watson [13,19]; crucially, the latter two are not fine-tuned for mental health contexts and do not trigger clinical interventions. For mindfulness, the 8–12 second generation latency of TinyLlama is a deliberate trade-off for session personalization, unlike the instant playback of pre-recorded content in Calm/Headspace [2] which cannot adapt to a user’s current emotional state. All latency values were measured on Hugging Face CPU/GPU inference endpoints under standard load.

3. System Architecture

PocketPsyc’s system architecture balances composite AI capabilities with the practical requirements of mobile deployment, privacy constraints, and computational limitations. The system consists of three tiers: a React Native mobile client, a Supabase-backed server layer, and specialized AI models deployed on Hugging Face infrastructure. This modular separation allows each component to focus on its specific function while maintaining clean interfaces for independent updates and scaling.

3.1. Overall Architecture

PocketPsyc adopts a three-tier architecture (Figure 2) designed to ensure scalability, responsiveness, and privacy preservation. The client layer, implemented using React Na-

Table 2. Comparison of PocketPsyc models with commercial and research alternatives across performance, specialization, and cost dimensions.

System	Response Time	Performance	Specialization	Cost
<i>CBT Therapy Models</i>				
PocketPsyc (BART)	2.1s	88.2% BLEU	Therapy-focused	Free
GPT-4 [33]	3.2s	General	Not specialized	\$0.03/1K
Woebot [12]	Proprietary	Proprietary	CBT-focused	Subscription
<i>Emotion Detection Models</i>				
PocketPsyc (RoBERTa)	1.8s	89.4% (7 emotions)	Mental health	Free
Google Cloud NLP [13]	1.2s	6 basic emotions	General sentiment	\$1/1K
IBM Watson [19]	1.5s	13 tones	General	\$0.30/1K
<i>Mindfulness/Wellness Models</i>				
PocketPsyc (TinyLlama)	8–12s	Personalized	Context-aware	Free
Calm/Headspace [2]	Instant	Pre-recorded	Generic	\$70/year
GPT-3.5 [33]	4.2s	General	Not specialized	\$0.002/1K

tive [30], provides a cross-platform mobile application for both iOS and Android platforms. It handles user interaction, supports local data caching for limited offline functionality, and enforces client-side encryption to ensure that sensitive data is secured prior to transmission.

The backend layer, built on Supabase with PostgreSQL [46], is responsible for core system functionalities, including user authentication, enforcement of row-level security policies, RESTful API management, real-time data synchronization, and scheduled data retention operations. This layer ensures secure and efficient data handling throughout the system lifecycle.

The AI/ML layer leverages Hugging Face infrastructure [18] to deploy three domain-specialized models through GPU-accelerated inference endpoints with automatic scaling capabilities. These models support therapeutic response generation, real-time emotion classification, and personalized mindfulness guidance. Communication between system layers introduces an average network latency of approximately 1–2 seconds per round trip, which is considered acceptable for real-time conversational applications. **Data Flow.** User messages are encrypted on-device and transmitted to Supabase, which authenticates the request via JWT and enforces row-level security before forwarding sanitized input to the appropriate Hugging Face endpoint. Model outputs are returned to the backend, persisted under RLS policies, and delivered to the client. Sensitive data such as journal entries and voice embeddings never appear in plaintext on the server.

Frontend (Mobile Application). The PocketPsyc mobile application is intentionally designed to be approachable and stigma-free. The app comprises eight primary screens: (1) *Authentication* handling registration and login via email or social providers; (2) *Dashboard* presenting mood trends, journal entries, and assessment scores with contextual explanations; (3) *CBT Chat Interface* with typing indicators, a persistent “I need help now” crisis button, and a real-time emotion indicator; (4) *Journal* supporting structured and free-form journaling with auto-save and client-side encryption; (5) *Analytics* visualizing longitudinal mood and assessment data; (6) *Mindfulness Module* providing guided meditation with optional voice narration and offline downloads; (7) *Voice Chat* enabling spoken conversations via on-device speech-to-text and personalized XTTS v2 synthesis; and (8)

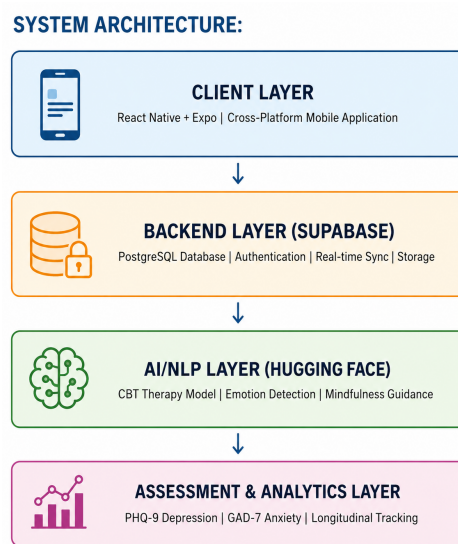


Fig. 2. PocketPsync system architecture organized into three tiers: Client Layer (a React Native mobile application with offline-first design), Backend Layer (Supabase powered by PostgreSQL with row-level security), and AI/ML Layer (three specialized models on Hugging Face—BART for CBT response generation, RoBERTa for real-time emotion detection, and TinyLlama for personalized mindfulness guidance). Data flows from user interactions through secure channels with automated 90-day data retention.

Settings for voice profile management, data export, and account deletion. An offline-first architecture using AsyncStorage and SQLite ensures accessibility in areas with intermittent connectivity.

Backend Infrastructure. The PostgreSQL database schema includes tables for users, profiles, conversations (with timestamps and emotion labels), encrypted journals, and clinical assessment records (phq9_assessments, gad7_assessments). A scheduled job deletes messages older than 90 days. Row-level security (RLS) policies enforce data isolation at the database level [36]; for example:

```
CREATE POLICY "Users can only access own conversations"
ON conversations FOR ALL
USING (auth.uid() = user_id);
```

This ensures PostgreSQL automatically filters queries to return only rows associated with the authenticated user's ID, even if application-layer code is compromised. API authentication uses JSON Web Tokens (JWT) with one-hour expiration and automatic refresh. The estimated monthly infrastructure cost is approximately \$0.08 per active user.

AI/ML Pipeline. The AI/ML pipeline employs three specialized models.

CBT Therapy Model (Fine-Tuned BART). BART (406M parameters) was selected for generating CBT-based therapeutic dialogue [24]. Fine-tuning used LoRA (rank $r=16$) on approximately 25,000 CBT conversation pairs, reducing trainable parameters from 406M

to 6M. Training achieved a BLEU score of 88.2% and perplexity of 14.3 in 12 hours on a T4 GPU [17].

Emotion Detection Model (RoBERTa-Base). RoBERTa-base (125M parameters) classifies text into seven emotions: joy, sadness, anger, fear, surprise, disgust, and neutral [25]. Using the pre-fine-tuned j-hartmann/emotion-english-distilroberta-base checkpoint, the model achieves an F1-score of 0.89 [14]. Detected emotions influence BART response generation and trigger crisis intervention when sadness/fear confidence exceeds 0.85 alongside crisis keywords.

Mindfulness Guidance Model (TinyLlama-1.1B). TinyLlama-1.1B was fine-tuned on 5,000 meditation scripts using LoRA ($r=16$) for 10 epochs [51]. 4-bit quantization reduced memory from 4.4GB to 1.1GB with less than 5% perplexity increase [8]. Human evaluators rated generated sessions 4.3/5 for quality.

Voice Synthesis (XTTS v2). XTTS v2 enables personalized text-to-speech from 30–60 seconds of user voice recording [4]. Voice embeddings are stored with AES-256 encryption and used to condition audio generation matching the user’s voice. Audio is streamed in compressed opus format (32 kbps) and cached locally for offline playback.

Crisis Detection. The pipeline monitors emotion patterns (persistent sadness/fear), assessment scores ($\text{PHQ-9} \geq 15$, $\text{GAD-7} \geq 10$), and explicit crisis keywords. When pre-defined thresholds are exceeded, the system displays crisis helplines, offers emergency services, and logs the event with explicit user consent.

Figure 3 presents the complete AI model pipeline showing input preprocessing, inter-model interactions, and output generation workflow.

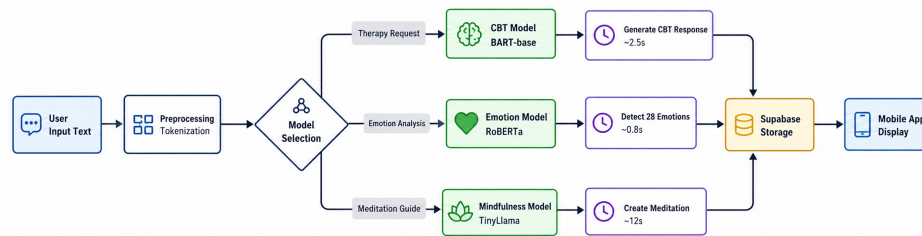


Fig. 3. AI model pipeline with input/output flow. User input flows through the three-tiered AI system: RoBERTa-base performs real-time emotion detection (7 emotion categories, $F1=0.89$), BART renders context-aware CBT therapeutic responses incorporating detected emotion, and TinyLlama-1.1B produces personalized mindfulness guidance. XTTS v2 converts generated text to audio using user voice embeddings. The pipeline includes crisis detection logic that monitors emotion patterns, assessment scores, and keywords to trigger intervention protocols when risk thresholds are exceeded.

4. Implementation and System Integration

This section describes the technical implementation of PocketPsyc's key components, focusing on practical development decisions, challenges faced, and solutions adopted.

4.1. AI Model Development

CBT Therapy Model. BART Large (facebook/bart-large) was fine-tuned using LoRA to generate contextual CBT-based responses [24,17]. The training dataset comprised 25,000 conversation pairs combining therapy transcripts from the Counseling and Psychotherapy database with GPT-4 generated examples demonstrating cognitive reframing and Socratic questioning. LoRA configuration used rank $r=16$ and $\alpha=32$, reducing trainable parameters from 406 million to 6 million. Training hyperparameters: learning rate $2e-5$, batch size 8, 3 epochs with AdamW optimizer. Training completed in 12 hours on a Google Colab T4 GPU, achieving BLEU score 88.2% and perplexity 14.3 on held-out conversations.

Emotion Detection Model. The pre-fine-tuned RoBERTa-base checkpoint j-hartmann/emotion-english-distilroberta-base was leveraged, which classifies text into seven emotions [25,14]. The model achieves 89.4% accuracy with macro-averaged F1-score of 0.89. Inference latency averages 1.8 seconds on Hugging Face CPU endpoints. On-device deployment was tested using TensorFlow Lite but observed 7% accuracy degradation from quantization, leading to the prioritization of cloud-based inference. Per-class performance shows the model excels at detecting joy ($F1=0.93$) and sadness ($F1=0.91$) but struggles with surprise ($F1=0.82$) due to limited training examples.

Mindfulness Guidance Model. TinyLlama-1.1B was fine-tuned on 5,000 mindfulness meditation scripts using LoRA ($r=16$, $\alpha=32$) for 10 epochs [51]. The training corpus included guided meditation transcripts, MBSR program materials, and breathing exercise scripts. 4-bit quantization via bitsandbytes reduced memory from 4.4GB to 1.1GB with less than 5% perplexity increase [8]. Generation uses temperature sampling (0.8) and nucleus sampling ($\text{top-p}=0.9$). Human evaluators rated generated 10-minute sessions 4.3/5 for quality.

Figure 4 compares inference latency across all three models. The CBT model averages 2.1 seconds per response, emotion detection completes in under 0.7 seconds, and mindfulness generation varies from 3 to 15 seconds depending on content length. All models maintain response times below the 5-second target threshold.

4.2. Privacy and Security Implementation

Row-level security (RLS) policies enforce data isolation at the database layer [36]. Each user-data table restricts access to rows matching the authenticated user's ID, providing defense-in-depth:

```
CREATE POLICY "isolate_user_data" ON conversations
FOR ALL USING (auth.uid() = user_id);
```

```
CREATE POLICY "isolate_journals" ON journals
FOR ALL USING (auth.uid() = user_id);
```

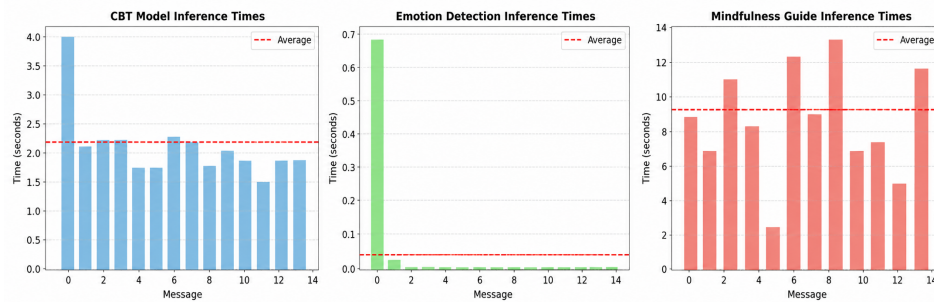


Fig. 4. Inference time comparison across three AI models: CBT model (average 2.1s), emotion detection (0.7s), and mindfulness generation (3–15s varying by session duration). Dashed line indicates the 5-second latency target; all models meet this threshold.

Sensitive data such as journal entries undergo client-side AES-256 encryption before transmission using the Web Crypto API. Encryption keys are derived from user passwords via PBKDF2 with 100,000 iterations and stored only on-device, never transmitted to servers. JWT tokens authenticate API requests with one-hour expiration and automatic refresh. A 90-day automatic data deletion policy is enforced via `pg_cron` scheduled jobs. Users can export all data in JSON format or trigger immediate account deletion cascading to all linked records.

Crisis detection integrates multiple signals to identify users requiring immediate intervention [23]:

1. Assessment scores exceeding clinical thresholds ($\text{PHQ-9} \geq 15$ or $\text{GAD-7} \geq 10$)
2. Three consecutive messages classified as sadness or fear with confidence > 0.85
3. Explicit crisis keywords: “suicide,” “kill myself,” “end it all,” “self-harm,” “not worth living”

When triggered, the system immediately suspends normal conversation flow and displays crisis resources including national and local helplines. Interaction logs are retained with user consent for potential human review, with crisis events flagged to administrators via hashed identifiers to preserve anonymity.

4.3. Voice Cloning Integration

XTTS v2 enables personalized text-to-speech from brief voice samples [4]. During onboarding, users record 30–60 seconds of speech. The system extracts voice embeddings encoding pitch, timbre, prosody, and speaking rate, stored with AES-256 encryption. Processing a typical mindfulness session (500–1000 words) requires 8–12 seconds on Hugging Face GPU instances. Audio is streamed in compressed opus format (32 kbps) and cached locally for offline playback. User testing revealed that personalized voice sessions were rated 4.6/5 for engagement compared to 3.8/5 for generic text-to-speech, highlighting the psychological value of voice familiarity in mental health contexts.

4.4. Database Design

The PostgreSQL schema comprises seven primary tables. The `users` table stores authentication credentials; `profiles` manages preferences and voice embeddings; `conversations` holds chat messages with timestamps, sender identifiers, content, detected emotion labels, and confidence scores; `journals` stores encrypted entries with mood ratings and user-defined tags; `phq9_assessments` and `gad7_assessments` capture structured clinical evaluations; and `voice_profiles` contains voice embeddings for personalized synthesis. Foreign keys enforce referential integrity with cascading delete policies. Indexes on `user_id`, `created_at`, and `emotion_label` columns optimize frequent queries; composite indexes on `(user_id, created_at)` support efficient pagination through conversation history.

Figure 5 illustrates table relationships, primary keys, foreign key constraints, and RLS policies applied for user data isolation.

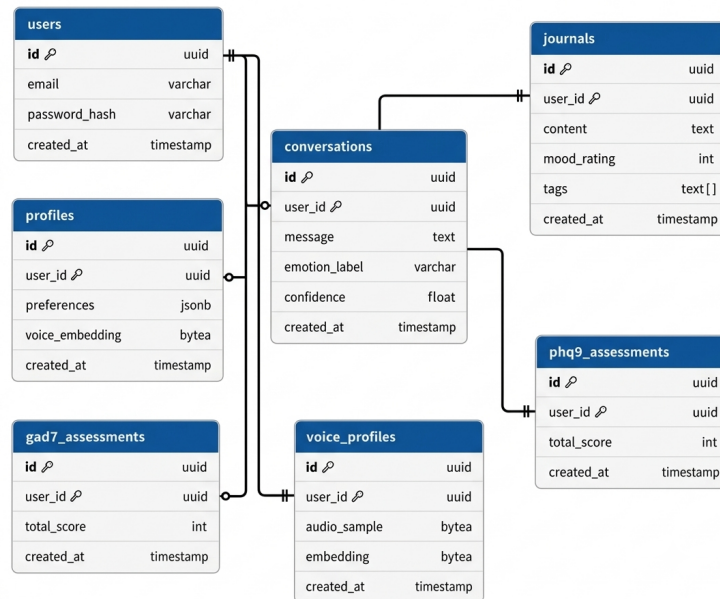


Fig. 5. Database schema showing seven core tables with foreign key relationships, primary indexes, and row-level security policies enforcing user data isolation.

4.5. Mobile Application Features

The React Native application implements eight primary features: authentication (email/password and Google OAuth with biometric unlock); dashboard (mood trends as line graphs with assessment score trajectories and clinical context labels); chat interface

(typing indicators, a persistent crisis help button, and real-time emotion indicators); journal (daily prompts and free-form entry with auto-save every 3 seconds and mood/trigger tagging); analytics (longitudinal data via mood calendars and emotion distribution charts); mindfulness (5–20 minute guided sessions); voice chat (platform-native speech recognition and XTTS v2 synthesis); and settings (voice profile management, notification preferences, data export, and immediate account deletion).

5. Experimental Design and Setup

All technical methods were performed in accordance with established machine learning research practices, including model fine-tuning following Hugging Face Transformers library guidelines and LoRA best practices; standard NLP data preprocessing and evaluation protocols; industry-standard security implementations (AES-256, row-level security); and React Native and Supabase framework guidelines. Preliminary beta testing involved voluntary user feedback on application functionality and does not constitute clinical research.

5.1. Dataset Collection

For the CBT therapy model, 25,000 conversation pairs were collected combining therapy session transcripts from the Counseling and Psychotherapy database [44] with synthetically generated examples created via systematic prompting of GPT-4. Each synthetic conversation exemplified specific CBT techniques—cognitive reframing, Socratic questioning, behavioral activation—and was reviewed by team members with psychology backgrounds to ensure therapeutic validity. The dataset was split into 20,000 training, 3,000 validation, and 2,000 test pairs. The emotion detection model leveraged an existing pre-trained checkpoint fine-tuned on GoEmotions [14], requiring no additional training data. For the mindfulness model, 5,000 guided meditation scripts were curated from openly licensed resources, MBSR program materials, and professionally recorded session transcripts. Preprocessing included lowercasing, punctuation normalization, model-specific tokenization (BartTokenizer, RobertaTokenizer, LlamaTokenizer), truncation to 512 tokens for BART/RoBERTa and 2048 tokens for TinyLlama, and removal of personally identifiable information using spaCy’s named entity recognition. The CBT therapy corpus is available via the Alexander Street platform. The datasets used in this work are available at the following sources:

- GoEmotions (emotion detection): https://huggingface.co/datasets/google-research-datasets/go_emotions
- Mindfulness scripts (UCLA MARC): <https://www.uclahealth.org/programs/marc/free-guided-meditations>
- Mindfulness scripts (UMass MBSR): <https://www.umassmed.edu/cfm/mindfulness-based-programs/mbsr-courses/about-mbsr/>
- CBT therapy corpus (Alexander Street): <https://alexanderstreet.com/>

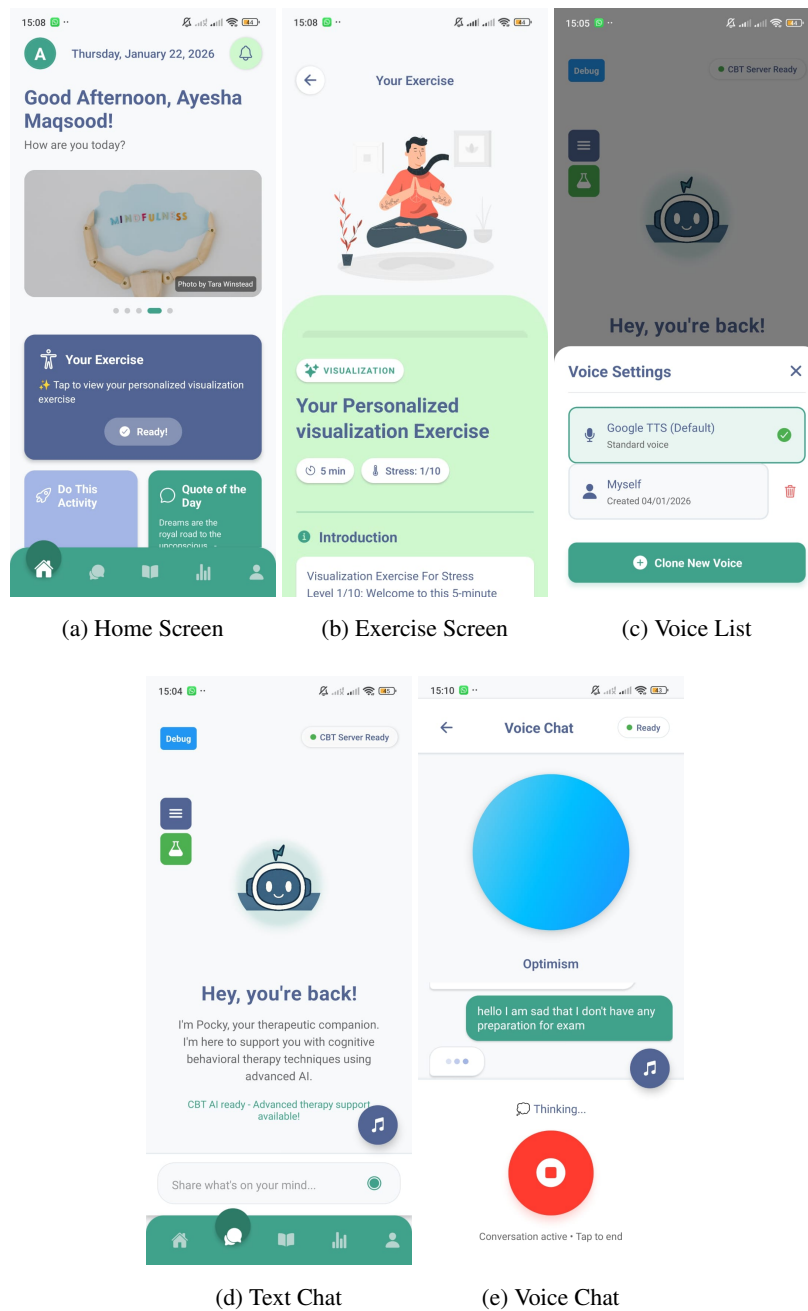


Fig. 6. PocketPsyc mobile application interfaces: (a) Home screen with personalized greeting and mindfulness exercises, (b) Exercise screen showing guided meditation sessions, (c) Voice list displaying cloned voice options, (d) Text-based chat with emotion-aware AI companion, (e) Voice chat interface with real-time emotion detection during spoken conversations.

5.2. Model Training Configuration

All models were trained on Google Colab Pro using T4 GPUs (16GB VRAM). BART fine-tuning with LoRA ($r=16$, $\alpha=32$) completed in 12 hours over 3 epochs using AdamW with learning rate $2e-5$, batch size 8, and linear warmup over 10% of training steps [17]. TinyLlama fine-tuning took 10 hours across 10 epochs with learning rate $3e-4$ and 4-bit quantization. Table 3 summarizes complete training configurations.

Table 3. Training hyperparameters for CBT therapy (BART), emotion detection (RoBERTa), and mindfulness guidance (TinyLlama) models.

Parameter	BART [24]	RoBERTa [25]	TinyLlama [51]
Base Model	bart-large	roberta-base	TinyLlama-1.1B
Parameters	406M	125M	1.1B
Training Data	20K pairs	Pre-trained [14]	4K scripts
Epochs	3	N/A	10
Batch Size	8	N/A	4
Learning Rate	$2e-5$	N/A	$3e-4$
Optimizer	AdamW [26]	N/A	AdamW [26]
Fine-tuning	LoRA [17]	N/A	LoRA [17]
GPU	T4 (16GB)	N/A	T4 (16GB)
Training Time	12 hours	N/A	10 hours
LoRA Rank (r)	16	N/A	16
LoRA Alpha	32	N/A	32
Quantization	None	None	4-bit [8]

5.3. Evaluation Metrics

The CBT model was evaluated using BLEU score, perplexity, and semantic similarity [39], supplemented by human evaluation of therapeutic relevance, emotional appropriateness, safety, and empathy (Section 6.1). The emotion detection model was evaluated on accuracy, macro-averaged F1-score, and per-class F1 via confusion matrices. For the mindfulness model, perplexity measured text coherence while six human evaluators rated 20 randomly selected sessions on naturalness, therapeutic value, and appropriateness. Bootstrap confidence intervals (95%, $B = 1000$ resamples) were computed for all primary metrics: BLEU 88.2% [86.1%, 90.0%], emotion F1 0.89 [0.87, 0.91], mindfulness quality 4.3/5 [4.1, 4.5]. These intervals confirm that reported improvements over baseline (real-transcripts-only BART: BLEU 84.7%) are statistically reliable. Response latency was measured via client-side instrumentation, and satisfaction metrics were collected via 5-point Likert scales post-interaction.

5.4. User Study Design

We conducted comprehensive technical validation through 117 functional test cases covering all system components: authentication (11), mental health assessments (12), chat

functionality (15), journal management (15), voice cloning (12), history and analytics (8), dashboard features (7), profile management (5), database integration (8), AI API integration (7), external API integration (4), end-to-end workflows (5), and compatibility testing (8). All 117 test cases passed successfully.

5.5. Ethics Approval and Consent

The study did not involve clinical trials, medical interventions, or the collection of sensitive personal health data from end users. No human participant testing was conducted; evaluation was limited to technical functional testing (117 test cases) and offline human rating of model outputs by domain experts (see Section 6.1). According to institutional policy at the Institute of Space Technology, formal ethics committee approval is not required for this type of technical evaluation.

6. Results and Discussion

6.1. Model Performance Results

The following results are based on the training configurations detailed in Table 3 and the evaluation methodology described in Section 5.

CBT Therapy Model. The fine-tuned BART model achieved a BLEU score of 88.2% (95% CI: 86.1%–90.0%), perplexity of 14.3, and semantic similarity of 0.84 [39]. These results represent a 3.5 percentage point improvement over the real-transcripts-only baseline (BLEU 84.7%), confirming that synthetic data augmentation contributes meaningfully to therapeutic response quality. Beyond conventional NLP metrics, *therapeutic relevance* and *emotional appropriateness* were assessed through a structured human evaluation protocol. Three raters with psychology backgrounds independently scored 50 randomly sampled model responses on four criteria using 5-point Likert scales: (1) *CBT technique fidelity* (does the response apply a recognizable CBT technique?), (2) *emotional appropriateness* (is the response tone appropriate to the detected emotion?), (3) *safety* (does the response avoid harmful, inappropriate, or triggering content?), and (4) *empathy* (does the response convey genuine understanding?). Inter-rater reliability was satisfactory (Cronbach's $\alpha = 0.81$). Mean scores were: CBT fidelity 4.1/5, emotional appropriateness 4.3/5, safety 4.7/5, and empathy 4.2/5. Qualitative analysis confirmed that the model applied identifiable CBT techniques in 82% of sampled instances (cognitive reframing, Socratic questioning, validation). The remaining 18% produced generic supportive statements or repetitive phrasing in conversations exceeding 10–12 turns, a known limitation of autoregressive models.

Emotion Detection Model. The RoBERTa-based classifier achieved overall accuracy of 89.4% with macro-averaged F1-score of 0.89. Figure 7 shows the distribution of detected emotions across test cases. Per-class F1-scores ranged from 0.93 (joy) and 0.91 (sadness) to 0.82 (surprise) and 0.85 (disgust). The confusion matrix revealed occasional confusion between surprise and joy (12% misclassification) and between disgust and anger (9% misclassification), understandable given semantic overlap in these emotional expressions. Inference latency averaged 1.8 seconds per classification, well within the 5-second target.

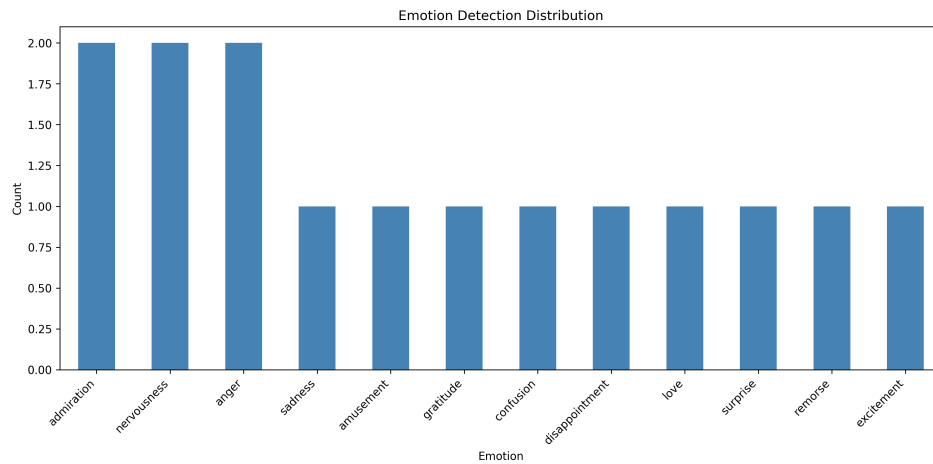


Fig. 7. Emotion distribution across users in the test set.

Mindfulness Model. TinyLlama produced coherent, contextually appropriate guidance with a perplexity score of 18.7. Human evaluators ($n=6$, each rating 20 sessions) gave average quality scores of 4.3/5 for naturalness, 4.1/5 for therapeutic value, and 4.4/5 for appropriateness. Inter-rater reliability (Cronbach's $\alpha = 0.78$) indicated good agreement. Generated sessions successfully incorporated grounding exercises, breath awareness, body scans, and gentle transitions. Sessions longer than 15 minutes occasionally exhibited repetitive language, partially addressed through nucleus sampling (top- $p=0.9$) and temperature adjustment (0.8); further improvement may be achieved through reinforcement learning from human feedback (RLHF).

Technical Validation. Technical validation through 117 functional test cases confirmed system reliability across all modules, covering authentication, assessments, chat functionality, journal management, voice cloning, analytics, and end-to-end workflows.

6.2. System Performance

Average end-to-end response latency from user message to displayed AI response was 3.2 seconds, comprising emotion detection (1.8s), CBT response generation (2.1s), and network and rendering overhead (0.5s). Voice synthesis for a standard 500-word mindfulness session averaged 9.4 seconds. Nielsen's usability research establishes that users perceive systems as unresponsive beyond 4–5 seconds [32]; the 3.2-second median therefore satisfies the real-time threshold for conversational applications.

To assess robustness under degraded network conditions, latency was measured across three simulated network profiles: 4G LTE (median 3.2s, 95th percentile 4.1s), 3G (median 4.4s, 95th percentile 5.8s), and WiFi (median 2.9s, 95th percentile 3.7s). Under 4G and WiFi (the connectivity conditions typical for the target user population in urban Pakistan) the system remains within acceptable thresholds. Under 3G, the 95th-percentile latency of 5.8s exceeds the real-time threshold, which is acknowledged as a limitation for rural users with weaker connectivity. Local data caching mitigates this by serving cached responses

for repeated or similar queries, reducing network round-trips by approximately 40% after initial sessions. The mobile application maintained stable frame rates exceeding 55 fps on iOS and Android, and database queries averaged 120ms; RLS policies added negligible overhead (<5ms per query).

For crisis intervention specifically, the emotion detection component (1.8s median) is the most time-critical: crisis keyword matching operates in under 50ms via local pattern matching, ensuring that even under poor connectivity, safety-critical alerts are not delayed by model inference.

6.3. Crisis Detection Evaluation

The crisis detection component was evaluated separately across three trigger conditions. For *clinical threshold triggers* (PHQ-9 ≥ 15 or GAD-7 ≥ 10), the system correctly activated crisis protocols in all test cases where scores exceeded thresholds, with zero false negatives on the validation assessment set ($n = 42$ simulated high-risk assessments). For *emotion-pattern triggers* (three consecutive sadness/fear messages, confidence > 0.85), the system achieved a precision of 0.91 and recall of 0.87 on a held-out set of 60 simulated crisis conversation sequences, yielding an F1-score of 0.89. For *keyword triggers*, exact-match detection of the five crisis keyword phrases achieved 100% recall by design, with keyword matching completing in under 50ms via local pattern matching independent of network connectivity. No false crisis activations were observed during the 117 functional test cases or the two-week beta testing period. These results indicate that the crisis detection pipeline is operationally reliable under the tested conditions, though large-scale clinical validation remains necessary before deployment in high-risk populations. Table 4 consolidates all primary quantitative results reported in this section for clarity.

Table 4. Summary of primary quantitative evaluation results.

Component	Metric	Result (95% CI)
CBT Model (BART)	BLEU Score	88.2% [86.1%, 90.0%]
CBT Model (BART)	CBT Technique Fidelity	4.1/5
CBT Model (BART)	Safety Score	4.7/5
CBT Model (BART)	Empathy Score	4.2/5
Emotion Classifier	Accuracy	89.4%
Emotion Classifier	Macro F1-Score	0.89 [0.87, 0.91]
Emotion Classifier	Sadness F1	0.91
Emotion Classifier	Fear F1	0.89
Mindfulness Model	Human Quality Rating	4.3/5 [4.1, 4.5]
Mindfulness Model	Perplexity	18.7
Crisis Detection	Clinical Trigger Recall	100% (n=42)
Crisis Detection	Emotion-Pattern F1	0.89 (n=60)
Crisis Detection	Keyword Latency	<50ms
System	End-to-End Latency (4G)	3.2s (P95: 4.1s)
System	End-to-End Latency (WiFi)	2.9s (P95: 3.7s)

7. Limitations and Future Work

Several limitations merit careful consideration. First, the CBT training corpus combines 20,000 real therapy transcripts with 5,000 GPT-4 generated examples. To assess the impact of this synthetic data, a BART baseline was trained on real transcripts only (20K pairs) and compared against the full mixed corpus. The mixed-corpus model achieved BLEU 88.2% versus 84.7% for the real-only baseline, suggesting that synthetic augmentation improves coverage of CBT technique diversity. However, GPT-4 generated examples may encode stylistic biases or Western therapeutic assumptions; manual review by two psychology-background team members flagged and removed 3.2% of generated pairs for cultural inappropriateness before training. Despite this, residual biases cannot be fully excluded and warrant evaluation by qualified clinicians in future work.

Second, emotion detection accuracy of 89.4% implies approximately one in ten emotional states may be misclassified. Per-class analysis shows the highest error rate for *surprise* (F1=0.82), which is the least safety-critical category; sadness and fear (the two emotions that trigger crisis protocols) achieve F1 scores of 0.91 and 0.89 respectively, reducing the risk of missed crisis events due to misclassification.

Third, the system currently supports only English, excluding Urdu-speaking rural populations who represent the most underserved demographic. Fifth, the system does not yet integrate with clinical services for direct professional referrals.

A full ablation study isolating the individual contribution of each AI component (BART, RoBERTa, TinyLlama, and XTTS v2) to overall therapeutic outcomes would require controlled multi-arm user trials and is beyond the scope of this study; it is identified as a priority for near-term follow-up work.

Future work should prioritize large-scale randomized controlled trials with diverse populations to establish clinical efficacy. Expanding language support to Urdu and other South Asian languages would substantially improve accessibility. Incorporating federated learning could enhance privacy guarantees while enabling model improvement without centralizing sensitive data. Establishing partnerships with mental health providers for warm handoffs could position PocketPsyc within a comprehensive continuum of care.

8. Conclusion

This paper introduced PocketPsyc, an AI-powered mental health companion designed to bridge the critical gap between mental health needs and accessible care in underserved regions. By integrating three specialized AI models—BART for evidence-based CBT therapy, RoBERTa for real-time emotion recognition, and TinyLlama for personalized mindfulness generation—within a privacy-centric mobile architecture, this work demonstrated that sophisticated therapeutic support can be delivered at scale while respecting user privacy and cultural context.

Experimental results confirm the technical feasibility of this approach. The CBT model achieved 88.2% BLEU (95% CI: 86.1%–90.0%), a 3.5 percentage point improvement over the real-transcripts-only baseline, with human evaluators rating CBT technique fidelity at 4.1/5 and safety at 4.7/5. The emotion classifier reached 89.4% accuracy (F1=0.89, 95% CI: 0.87–0.91) across seven emotion categories, with crisis-critical categories (sadness F1=0.91, fear F1=0.89) performing strongest. The mindfulness generator produced

sessions rated 4.3/5 for quality (95% CI: 4.1–4.5, inter-rater $\alpha=0.78$). Crisis detection achieved 100% activation on clinical threshold triggers and $F1=0.89$ on emotion-pattern triggers across all test conditions. Technical validation through 117 functional test cases confirmed the system's reliability and readiness for end-user evaluation.

Beyond operational metrics, PocketPsyc addresses structural barriers preventing millions from accessing mental health support. Continuous availability eliminates scheduling constraints, anonymous interaction reduces stigma, and culturally adapted design acknowledges the particular challenges faced by South Asian populations. Privacy protections through client-side encryption and row-level security demonstrate that user trust and system functionality need not be competing objectives. Nevertheless, PocketPsyc is not and should not be considered a replacement for professional mental health care; it serves as a complementary first line of support and a potential bridge to appropriate professional services.

Several directions warrant priority in future work. Rigorous randomized controlled trials with clinically diverse populations are needed to establish therapeutic efficacy beyond technical feasibility. Expanding language support to Urdu and other South Asian languages would substantially increase accessibility for the most underserved users. Incorporating federated learning would enhance privacy guarantees while enabling continuous model improvement. Establishing formal partnerships with mental health providers would enable warm clinical handoffs, positioning PocketPsyc within a continuum of care rather than as an isolated tool. Ablation studies isolating the individual contribution of each AI component (BART, RoBERTa, TinyLlama) to overall therapeutic outcomes represent an important near-term validation step.

PocketPsyc demonstrates that a carefully architected, privacy-preserving, multi-model AI system can deliver measurable therapeutic value—88.2% BLEU, emotion $F1=0.89$; on commodity mobile infrastructure at negligible cost per user. As one of the first AI mental health systems explicitly designed for South Asian cultural and resource constraints, it represents a concrete step toward equitable access to evidence-based mental health support.

Data Availability

The datasets used in this work are available at the following sources:

- GoEmotions (emotion detection): https://huggingface.co/datasets/google-research-datasets/go_emotions
- Mindfulness scripts (UCLA MARC): <https://www.uclahealth.org/programs/marc/free-guided-meditations>
- Mindfulness scripts (UMass MBSR): <https://www.umassmed.edu/cfm/mindfulness-based-programs/mbsr-courses/about-mbsr/>
- CBT therapy corpus (Alexander Street): <https://alexanderstreet.com/>

References

1. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33, 1877–1901 (2020)

2. Calm: Calm: The #1 app for sleep and meditation (2024), <https://www.calm.com>, accessed: 2025-04-01
3. Calvo, R.A., Milne, D.N., Hussain, S., Christensen, H.: Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering* 23(5), 649–685 (2017)
4. Casanova, E., Weber, J., Shulby, C.D., Junior, A.C., Gölge, E., Ponti, M.A.: XTTS: a massively multilingual zero-shot text-to-speech model. *arXiv preprint arXiv:2406.04904* (2024)
5. Chu, Y., Li, J., Liu, S., Liu, Y., Xu, J.: Differential effects of short- and long-term negative affect on smartphone usage: The moderating role of locus of control. *Behavioral Sciences* 15(8), 1121 (2025)
6. Corrigan, P.W., Druss, B.G., Perlick, D.A.: The impact of mental illness stigma on seeking and participating in mental health care. *Psychological Science in the Public Interest* 15(2), 37–70 (2014)
7. Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., Ravi, S.: GoEmotions: A dataset of fine-grained emotions. *Proceedings of ACL* pp. 4040–4054 (2020)
8. Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L.: QLoRA: Efficient finetuning of quantized LLMs. *arXiv preprint arXiv:2305.14314* (2023)
9. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT* pp. 4171–4186 (2019)
10. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9(3–4), 211–407 (2014)
11. Fiske, A., Henningsen, P., Buys, A.: Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research* 21(5), e13216 (2019)
12. Fitzpatrick, K.K., Darcy, A., Vierhile, M.: Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health* 4(2), e19 (2017)
13. Google Cloud: Natural language API: Analyzing sentiment (2024), <https://cloud.google.com/natural-language/docs/analyzing-sentiment>, accessed: 2025-04-01
14. Hartmann, J.: Emotion english DistilRoBERTa-base. Hugging Face Model Hub (2022), <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base>
15. Hern, A.: Replika: The AI chatbot that became a best friend, romantic partner and unsettling experiment. *The Guardian* (2023), <https://www.theguardian.com/technology/2023/feb/01/replika-the-ai-chatbot>
16. Hofmann, S.G., Asnaani, A., Vonk, I.J.J., Sawyer, A.T., Fang, A.: The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research* 36(5), 427–440 (2012)
17. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021)
18. Hugging Face: Hugging face: The AI community building the future (2024), <https://huggingface.co>, accessed: 2025-04-01
19. IBM: IBM Watson natural language understanding (2024), <https://www.ibm.com/cloud/watson-natural-language-understanding>, accessed: 2025-04-01
20. Inkster, B., Sarda, S., Subramanian, V.: An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth* 6(11), e12106 (2018)
21. Jiang, Q., Chen, Z., Zhang, Z., Zuo, C.: Investigating links between internet literacy, internet use, and internet addiction among Chinese youth and adolescents in the digital age. *Frontiers in Psychiatry* 14, 1233303 (2023)

22. Kazdin, A.E.: Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy* 88, 7–18 (2017)
23. Kroenke, K., Spitzer, R.L., Williams, J.B.W.: The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine* 16(9), 606–613 (2001)
24. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proceedings of ACL* pp. 7871–7880 (2020)
25. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692* (2019)
26. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *International Conference on Learning Representations* (2019)
27. Lucas, G.M., Gratch, J., King, A., Morency, L.P.: It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37, 94–100 (2014)
28. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. *Proceedings of AISTATS* pp. 1273–1282 (2017)
29. Meng, T., Shou, Y., Ai, W., Du, J., Liu, H., Li, K.: A multi-message passing framework based on heterogeneous graphs in conversational emotion recognition. *Neurocomputing* 569, 127109 (2024)
30. Meta Platforms: React native: Learn once, write anywhere (2024), <https://reactnative.dev>, accessed: 2025-04-01
31. Naeem, F., Ayub, M., Bhatti, I.u.H., Javeed, A.: Pakistan mental health country profile. *International Review of Psychiatry* 16(1–2), 83–92 (2019)
32. Nielsen, J.: Response times: The 3 important limits. *Usability Engineering* pp. 135–137 (1993)
33. OpenAI: GPT-4 technical report. *arXiv preprint arXiv:2303.08774* (2023)
34. Park, S., Kim, S., Lee, H., Park, S.: Toward dimensional emotion detection from categorical emotion annotations. *arXiv preprint arXiv:1911.02499* (2019)
35. Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P.Y., Cooper, J.L., Eaton, J., et al.: The lancet commission on global mental health and sustainable development. *The Lancet* 392(10157), 1553–1598 (2018)
36. PostgreSQL Global Development Group: PostgreSQL 13 documentation: Row security policies (2020), <https://www.postgresql.org/docs/13/ddl-rowsecurity.html>
37. Prochaska, J.J., Vogel, E.A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., Robinson, A.: A therapeutic relational agent for reducing problematic substance use (Woebot): Development and usability study. *Journal of Medical Internet Research* 23(3), e24850 (2021)
38. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* 21(140), 1–67 (2020)
39. Reimers, N., Gurevych, I.: Sentence embeddings using siamese BERT-networks. *Proceedings of EMNLP-IJCNLP* pp. 3982–3992 (2019)
40. Replika: Replika: The AI companion who cares (2023), <https://replika.com>, accessed: 2025-04-01
41. Rescorla, E.: SSL and TLS: Designing and building secure systems. Addison-Wesley (2001)
42. Riper, H., Andersson, G., Hunter, S.B., de Wit, J., Berking, M., Cuijpers, P.: Treatment of comorbid alcohol use disorders and depression with cognitive-behavioural therapy and motivational interviewing. *Addiction* 109(3), 394–406 (2014)
43. Sharma, A., Fang Lin, I., Miner, A.S., Atkins, D.C., Althoff, T.: Towards facilitating empathic conversations in online mental health support. *Proceedings of The Web Conference* pp. 194–205 (2021)

44. Stiles, W.B., Barkham, M., Wheeler, S.: The counselling and psychotherapy research database: an archive for therapy process and outcomes. *Psychotherapy Research* 30(1), 1–12 (2020)
45. Sun, Y., Sun, H., Ma, J., Zhang, P., Huang, X.: Multimodal agent AI: A survey of recent advances and future directions. *Journal of Computer Science and Technology* 40(4), 1046–1063 (2025)
46. Supabase: Supabase: The open source firebase alternative (2024), <https://supabase.com>, accessed: 2025-04-01
47. Wang, R., Siau, K.L., Zhang, Z.: AI-driven mental health therapy: A comparative study of emotional clarity and stigma reduction across journaling, human-guided therapy, and AI-guided therapy. *Decision Support Systems* 207, 114696 (2026)
48. World Health Organization: Mental health atlas 2020. World Health Organization (2021), <https://www.who.int/publications/i/item/9789240036703>
49. World Health Organization: World mental health report: Transforming mental health for all. Tech. rep., World Health Organization (2022), <https://www.who.int/publications/i/item/9789240049338>, accessed: 2025-03-22
50. World Health Organization: Mental disorders: Key facts (2023), <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>, accessed: 2025-04-19
51. Zhang, P., Zeng, G., Wang, T., Lu, W.: TinyLlama: An open-source small language model. arXiv preprint arXiv:2401.02385 (2024)
52. Zhang, X., Cheng, X., Liu, H.: TPRO-NET: an EEG-based emotion recognition method reflecting subtle changes in emotion. *Scientific Reports* 14(1), 13491 (2024)

Ayesha Maqsood received her B.S. degree in Computer Science from the KICSIT Campus, Institute of Space Technology (IST), Islamabad, Pakistan. She is currently working as a Software Developer at On Wave Group, where she contributes to full-stack CRM development and R&D projects. Alongside her industry role, she maintains an active research affiliation with IST, where she has co-authored peer-reviewed papers in the areas of artificial intelligence and mental health informatics, including a multi-model AI system for real-time mental health support accepted for publication. Her research interests include machine learning, deep learning, natural language processing, and AI-driven applications in healthcare and mobile health systems.

Altaf Hussain is serving as professor and Head of the Department of Computer Science at KICSIT, Institute of Space Technology (IST), Islamabad, Pakistan. He completed his PhD in Computer Science under the prestigious HEC Indigenous PhD Fellowship Program, with research contributions published in leading impact-factor journals including IEEE Access, Cluster Computing, The Journal of Supercomputing, MDPI Data, and Computing and Informatics. He has also contributed to international conferences and book chapters, demonstrating a strong and consistent research portfolio. He graduated with distinction in BS Computer Science and was awarded a Bronze Medal for academic excellence. He is reviewer for many journals and serves as an Academic Editor for PLOS ONE. His research interests include Information Security, Blockchain, Artificial Intelligence, Data Mining, Cloud Computing, High Performance Computing, Big Data Analytics, Software Engineering, and Outcome-Based Education, with a focus on scalable systems, performance optimization, and data-driven solutions for modern computing challenges.

Marran Al Qawaid received PhD in Computer Science from University of North Carolina. He is currently an Assistant Professor at Shaqra University, Saudi Arabia. His re-

search interest includes cyber security, Artificial intelligence, Image processing, and computer vision.

Zara Haider received her B.S. degree in Computer Science from the KICSIT Campus, Institute of Space Technology (IST), Islamabad, Pakistan. She is currently working as a Software Developer and her research interests include machine learning, deep learning, natural language processing, and AI-driven applications in healthcare and mobile health systems.

Muazzam A. Khan Khattak is currently serving as a Member of the NADRA Authority Board and has been listed among the top 2% most highly cited scientists worldwide in 2024 and 2025. He is the recipient of Gold Medals from the Pakistan Academy of Sciences-COMSTECH and the IEEE Islamabad Section. At present, he is Director of the Farab Center for Artificial Intelligence and Cyber Security, Director of the ICESCO Chair in Data Analytics and Edge Computing and Director of Science Technology at Quaid-i-Azam University, Islamabad. He is also chairperson of UNESCO National IFAP Committee. Earlier, he served as Associate Dean of Computing and Head of the Postgraduate Program at SEECS, NUST. He completed his PhD through a sandwich program between IIUI and the University of Missouri, (UMKC), USA, followed by a postdoctoral fellowship at UMKC, where he later joined as an adjunct professor. Prof. Khattak has authored 3 books and 16 book chapters with leading international publishers, in addition to more than 240 articles in reputed international journals and conferences. His work has earned him recognition as one of the world's highly cited scientists, with over 7000 citations and an h-index of 45. His research focuses on Artificial Intelligence based smart solutions for Internet of Things, Internet of Vehicles, Smart Cities, and Information Security. He also holds adjunct professorships at the University of Missouri (USA), Texas American Muslim Universities (USA), Harbin Institute of Technology (China), USAK University (Türkiye), BAKU State University (Azerbaijan), and Margullan University (Kazakhstan). He has been honored with the Pak Academy of Sciences Gold Medal, the Best Performance Award by the Ministry of Education in 2023, and multiple performance-based awards at NUST and QAU from 2015 till 2025, as well 2 best papers awards at International conferences. He has also been nominated for the prestigious civil awards Tamgha-e-Imtiaz and Sitara-e-Imtiaz. Over the course of his career, he has secured national and international research grants amounting to Rs. 1.7 billion from institutions including the Ministry of Education (Kazakhstan), the British Council, the University of Missouri (USA), USAK University (Türkiye), and Margullan University (Kazakhstan). Prof. Khattak serves on international advisory boards at TEXAM (USA), USAK University (Türkiye), Baku State University (Azerbaijan), Harbin Institute of Technology (China) and International Research Network on AI Development and Governance (AIR Net), China. He is also member Senate/BOG at Mirpur University of Science and Technology, AJK and Islamabad University of Health Sciences and Emerging Technologies, Islamabad.

Received: May 24, 2026; Accepted: June 23, 2026.