*Grenze International Journal of Engineering and Technology, January Issue*

# AI Therapeutic Journaling: Privacy-First Multi- Stage System for Personalized Mental Health

Prof. Dr. Vijay Mane[1], Sanskruti Shinde[2], Sumit Raina[3] and Abhinav Tohare[4]
[1-4]Dept of Electronics and Telecommunications Vishwakarma Institute of Technology, Pune, India
Email: vijay.mane@vit.edu, sanskruti.shinde22@vit.edu, raina.sumit22@vit.edu, abhinav.tohare22@vit.edu

***Abstract*— Mental health journaling is a proven reflective tool used in Cognitive Behavioral Therapy and mindfulness practices, yet most digital solutions lack privacy, adaptability, and scalability. An AI therapeutic journaling platform with a multi-tier architecture that prioritizes privacy and integrates a responsive frontend, modular backend, and multi-stage AI pipeline is presented in this work.**

**The system's four layers-Immediate, Contextual, Pattern, and Long-Term Integration-deliver both real-time emotional feedback and deeper behavioral insights. Privacy is ensured through client-side encryption, federated learning with differential noise, and homomorphic search, keeping user data secure during computation and storage. The platform maintained sub-2-second latency, around 94% sentiment precision, and consistent personalization gains under high-load testing. The findings show a scalable, privacy-preserving AI system that improves therapeutic impact, engagement, and trust while providing a clinically relevant framework for flexible digital mental health journaling.**

***Keywords*— mental health technology, multi-stage AI analysis, privacy-preserving learning, multi-tier architecture, personalized therapeutic support.**

## I. INTRODUCTION

Journaling plays a vital role in therapies such as Cognitive Behavioral Therapy, Dialectical Behavior Therapy, and mindfulness practices, helping individuals build self- reflection and emotional awareness. Digital platforms have extended these benefits by allowing real-time emotional tracking and reflection. When integrated with artificial intelligence, journaling becomes adaptive-offering instant emotional feedback, contextual insights from past entries, and personalized coping strategies aligned with each user's psychological patterns.

Yet, current AI journaling tools remain limited. Personalization is often generic, privacy safeguards are weak due to server-side data storage, and performance issues like latency and scalability persist. Many systems also lack grounding in evidence-based therapeutic frameworks, reducing trust and clinical relevance.

This study proposes a privacy-first, multi-tier journaling architecture integrating frontend, backend, AI pipeline, and data management layers.

A multi-stage AI system-spanning Immediate, Contextual, Pattern, and Long-Term assessments delivers both real-time and longitudinal insights. Using client-side encryption, federated learning with noise injection, and homomorphic search, the platform ensures confidentiality while maintaining high responsiveness and adaptive personalization under real-world workloads.

*Grenze ID: 01.GIJET.12.1.37_1*

## II. LITERATURE REVIEW

Using a journaling interface that provides AI-generated prompts and summaries along with a clinician dashboard that displays patient data during consultations, MindfulDiary exemplifies the use of large language models (LLMs) in psychiatric support [1]. This arrangement translates unstructured patient tales into structured insights by bridging the gap between professional monitoring and personal reflection. Long-term assessments to verify long-term clinical benefits are absent from the system, nevertheless. Expanding on this concept, Journaling with Large Language Models presents a platform supported by LLM that combines interactive journaling with personal health records.[2]. Its three-panel interface-covering journaling, AI dialogue, and progress tracking-acts as a bridge between patients and healthcare providers. The authors propose patient-centered design principles that promote continuity of care, yet emphasize the need for comprehensive trials across varied populations.

PandaOmics extends AI use beyond mental health journaling by combining bioinformatics and multimodal omics analysis to identify biomarkers and therapeutic targets [3]. Validated through in vitro and in vivo studies, it forms part of a larger AI-driven drug discovery suite. However, its comparative performance against conventional discovery pipelines remains insufficiently studied, warranting broader validation across diverse disease domains. Personalization in mental health AI is further explored in the MindScape Study, which fuses behavioral sensing data with LLMs to deliver adaptive journaling experiences [4]. In an eight-week trial with 20 college students, the system showed improvements in positive affect (+7%) and reductions in negative affect (−11%), loneliness, anxiety, and depression. Although participants valued contextual prompts, the small sample and limited duration constrain generalizability, indicating a need for larger, more diverse studies. Similarly, DiaryMate employs LLMs for personal journaling through keyword-driven sentence generation and contextually relevant text suggestions [5].

Like MindfulDiary, this design fosters natural engagement but lacks evidence on long-term outcomes. Future work should examine sustained impacts on mental health trajectories. Ethical concerns are addressed in Regulating AI in Mental Health, which critiques “responsible AI” frameworks as insufficient for relational care and advocates an ethics of care approach emphasizing empathy and human connection, though without empirical validation [6]. Growing public interest is reflected in Mental Health Applications of Generative AI in the United States, which forecasts a 114% rise in awareness by 2024 [7]. Yet, it remains unclear how this translates to actual use and sustained engagement. The Systematic Review and Meta- analysis of AI-based Conversational Agents found significant reductions in depression (Hedge’s $g = 0.64$) and distress (Hedge’s $g = 0.7$), especially in multimodal, generative systems integrated into mobile apps [8]. However, overall psychological well-being improvements were insignificant, and long-term safety remains uncertain. Disorder diagnosis, counseling, therapeutic support, professional training, decision-making, and optimization are the six primary areas that are outlined in Generative AI in Mental Health Care [9].

The study highlights the need for cautious human-AI cooperation to maintain clinical judgment by highlighting GAI's function as an additional tool rather than a substitute for physicians. In order to improve self-reflection and wellbeing, MindScape: Contextual AI Journaling combines time-series behavioral data with LLM-driven suggestions to further personalization [10]. Although it presents a new framework for behavioral intelligence, only proof-of-concept testing can validate it. ChatGPT in Mental Health Assessment emphasizes the drawbacks of general-purpose models by showing that while ChatGPT functions well in straightforward situations, it has trouble with more complicated evaluations, highlighting the need for human supervision[11]. Fairness is addressed in AI Response for Mental Health Support, which shows differences in empathy levels. LLMs showed less empathy for Black people than for other groups [12]. Together, these results highlight the significance of contextual knowledge, moral design, and fair performance in creating secure and efficient AI-assisted mental health systems. The paper emphasizes how response generation techniques have a significant impact on output quality and stresses the necessity of solid safety standards and bias mitigation in mental health AI. NLP and adaptive learning are used by the AI-Driven Journaling App for Emotional Well-Being to provide feedback and prompts that are culturally relevant [13]. Although sustained involvement is yet poorly studied, its experiments revealed encouraging emotion detection and cultural adaptability, showing potential to lessen the stigma associated with requesting help. AI Chatbots for Mental Health offers a scoping review of chatbot interventions, showing benefits in emotional well- being and behavioral change but persistent challenges in usability, retention, and clinical integration [14]. Collectively, these studies reveal a rapidly advancing yet still immature field, constrained by limited large-scale validation, weak personalization, and unresolved ethical concerns surrounding bias, privacy, and informed consent-barriers that must be addressed for equitable, trustworthy AI deployment in mental health contexts.

## III. SYSTEM ARCHITECTURE

### *A. High Level Overview*

The AI Therapeutic Journaling Platform uses a distributed, privacy-first multi-tier design connecting the interface, backend, AI pipeline, and storage. It supports modular scaling and secure updates. Core modules include journaling, orchestration APIs, multi-stage AI (NLP, RAG, pattern detection), and hybrid databases. Privacy is enforced through encryption and federated learning, with edge models for instant feedback and cloud models for deeper analysis- enabling secure, adaptive, and clinically reliable journaling.

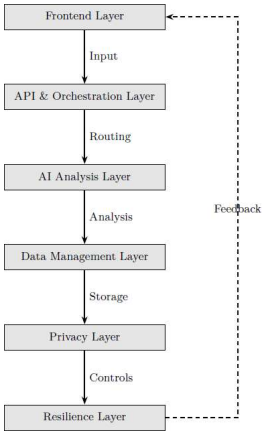


Fig. 1: System overview

### *B. Core Modules*

*User Onboarding Flow*

To maintain anonymity, users are given a randomly generated UUID and locally generated encryption keys (PBKDF2, 100k iterations). Servers only get non-sensitive verification data. The Policy Engine enforces client-side encryption with server-side rules for security and auditability, guided by therapy goals and privacy preferences.

*Daily Journaling Flow*

Local encryption is maintained for entries. While deeper phases (Immediate, Contextual, Pattern, and Long-Term) assess sentiment, context, and trends asynchronously, a lightweight model offers immediate emotion input. WebSocket-based differential sync preserves performance and privacy by sending only a small amount of data.

*AI Analysis Pipeline*

Text is transformed into emotional and semantic embeddings that are stored in a vector database. Cosine similarity, emotion, and temporal weighting are all combined in retrieval. High throughput and low latency are guaranteed by concurrent micro-workers and adaptive model selection (BERT for short text, LLMs for complicated inputs).

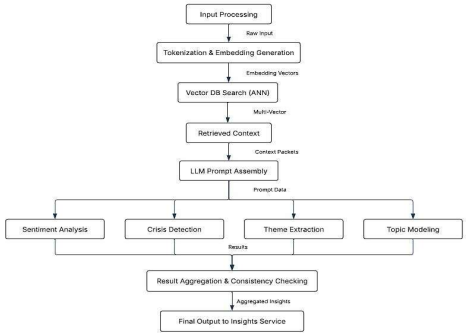


Fig. 2 : AI analysis pipeline

*Insights Service*
Combines AI results with clinical instruments (PHQ-9, GAD-7) and time-series models (ARIMA, change-point) to provide specific, context-aware CBT/DBT insights at the right times.

*Data Storage And Retrieval*
Uses PostgreSQL for metadata, vector stores for embeddings, time-series DBs for emotional logs, and Redis for caching. Searchable encryption and homomorphic operations enable secure querying.

*Privacy And Security Level*
Implements client-side encryption, federated learning with anonymized updates, and differential privacy. A Policy Engine maintains consent tracking and system-wide compliance.

*System Resilience & Fault Tolerance*
Ensures uptime through redundant microservices, predictive scaling, and geographic replication. Event-driven workflows and local model fallback maintain functionality during AI or network failures.

## IV. METHODOLOGY

### *A. Overview of methodology*

The methodology implements the multi-tier architecture from Section 6 into an end-to-end pipeline that transforms raw journaling input into personalized therapeutic insights. As shown in Fig. 2, the process includes secure client-side data capture, privacy-preserving preprocessing, multi-stage AI analysis (immediate, contextual, and longitudinal), resilient storage and retrieval, and adaptive feedback loops that refine personalization over time. Let $I_t$ represent the journal entry at time t, and let $I_t$ denote the set of insights produced for that entry. The pipeline may be abstractly defined as:

$$It = \mathcal{F}Insight(Jt, Hu, \mathrm{P}u, \Theta u)$$

where $Hu$ is the historical journal corpus for user $u$, P$u$ is the privacy policy vector, and $\Theta_u$ is the current personalization parameter set. This transformation $\mathcal{F}_{Insight}$ is not monolithic; it is an orchestrated sequence of functions described in Sections 7.2–7.9.

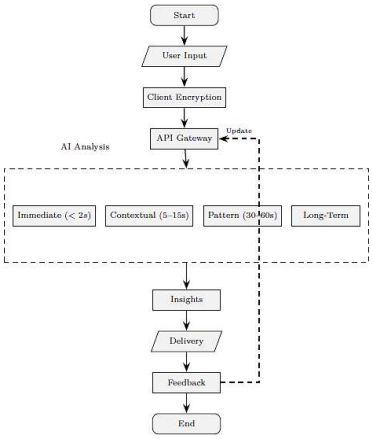


Fig. 3 : Overview of methodology

The macro-level implementation of the concept, showing interactions between system layers, is shown in Fig. 3. The flowchart emphasizes adherence to privacy regulations established during onboarding by labeling each arrow with the type of data shared, such as encrypted text, embeddings, or metadata. The workflow, data flow, and inter-layer dependencies of the technique are clearly visualized top- down in this picture.

### *B. Data Acquisition & Input Handling*

The At the journaling interface, where the client application immediately imposes encryption and policy tagging, the acquisition process starts. Each user is given a cryptographically secure UUID ( $ui$ ) upon registration in order to maintain stringent tenant isolation across the platform. A client-side key derivation function calculates the encryption key whenever a new entry ($Jt$)is created:

$$K_u = PBKDF2(p_u, s_u, 100000)$$

where $pu$is the user's password and $s_u$is a random salt generated per onboarding. This derived key never leaves the client device.

Perfect Forward Secrecy over TLS 1.3 is used for transmission to the backend. The transmission of just incremental variations ($\Delta J_t$) limits the attack surface and uses less bandwidth. Encrypted metadata P$u$ contains privacy preferences, so downstream services can enforce data handling policies without having to examine unencrypted material.

*C. Preprocessing & Client-Side Operations*

Before encryption, the journal entry undergoes preprocessing aimed at ensuring data safety and enhancing downstream analytical value. This includes text sanitization to remove harmful markup, tokenization, and local NLP-based inference for real-time feedback. Let the tokenized sequence be $T = (t_1, t_2, \dots, t_m)$ A lightweight sentiment model
$M_{local}$ computes:
$S_t, E_t = M_{local}(T)$
where S$S_t$ is a scalar sentiment polarity and $E_t$ is a vector of emotion probabilities. Crisis detection is performed using a boolean function:

$$C(J_t) = \begin{cases} 1 & if\ crisis\_keyword \cap J_t \neq \emptyset \\ 0 & otherwise \end{cases}$$

Detected crises trigger local alerts only, ensuring privacy. Encrypted and policy-tagged data is then queued for backend ingestion.

*D. Multi-Stage AI Analysis Pipeline*

The AI Analysis Pipeline, shown in Fig. 4, is the central component of the methodology. Four steps make up this pipeline's processing of incoming entries: pattern analysis, contextual analysis, immediate analysis, and long-term integration.

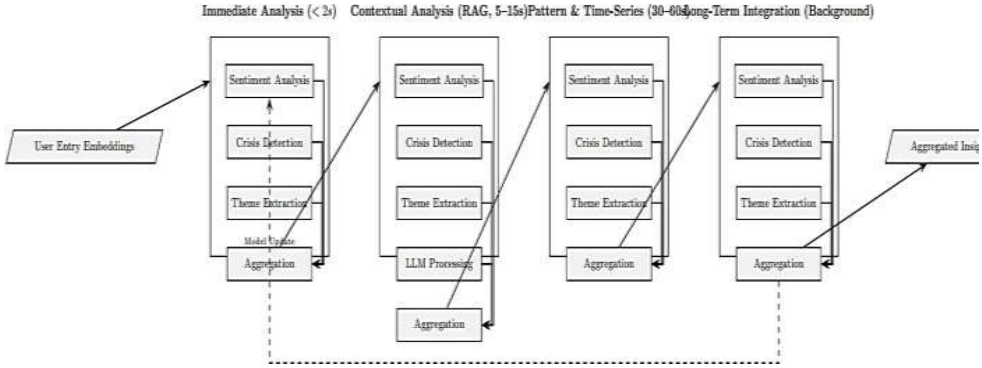


Fig. 4: AI pipeline process flow

Figure 4: Process flow of an AI pipeline The internal breakdown of the AI Analysis Layer into sequential stages and concurrent tasks is shown in Fig. 4. It demonstrates how lightweight models work in tandem with retrieval-augmented LLM inference, which is followed by consistency and aggregation checks. The Stage 2 contextual retrieval algorithm uses cosine similarity to mathematically embed $J_t$ into a vector $v_t$ and retrieve the k most comparable previous entries $\{v_1, \dots, v_k\}$ :

$$sim(v_t, v_i) = \frac{v_t \cdot v_i}{||v_t||\ ||v_i||}$$

The recovered set is chosen using an adaptive complexity function $\boldsymbol{\kappa}(M_{LL})$ . to produce an enriched prompt for the LLM $M_{LL}$ . The first algorithm, Multi-Stage AI Analysis, formalizes this procedure.

Algorithm 6.1. Multi-Stage AI Analysis
Input: encrypted entry E_enc, user ID U, policy vector P, retrieval size k
Output: encrypted insights IS_enc
1: decryptEntry(E_enc, U) → J_plain
2: sanitizeAndTokenize(J_plain) → Tokens
3: runLocalFeedback(Tokens) → sentiment, emotion, crisisFlag
4: encryptAndStore(J_plain, U, P)
5: generateEmbedding(J_plain) → v_curr
6: insertVectorDB(v_curr, U, P)
7: stage1Results ← runLightweightModels(J_plain)
8: if policyAllows(P) then
9: context ← secureRetrieveRAG(J_plain, k, P)
10: stage2Results ← runLLM(context, J_plain)
11: else stage2Results ← emptyContext()
12: stage3Results ← patternAnalysis(U, J_plain)
13: IS_enc ← aggregateAndEncrypt(stage1Results, stage2Results, stage3Results, U, P)
14: return IS_enc
15: end
Algorithm 1 provides a formal procedural definition of how an encrypted entry is decrypted client-side, processed through each stage of analysis, and re-encrypted for secure delivery of insights. It defines the control flow and data dependencies explicitly, enabling reproducibility and complexity analysis.

*E. Personalization & Adaptive Intelligence Mechanisms*

User models $\Theta u$ are continuously improved by the personalization subsystem using both explicit and implicit feedback.
While reading time, scroll depth, and re- engagement frequency are examples of implicit signals, user assessments of insights are examples of explicit feedback $Ft$. The following is the update rule for treatment modality weights $wu$:

$$w_u \leftarrow w_u + \eta(F_t - \bar{F}_t)$$

This adaptive update mechanism is captured in Algorithm 2
- Adaptive Personalization Update.
Algorithm 6.2. Adaptive Personalization Update Input: user ID U, feedback set F, learning rate η Output: updated user model UM_updated
1: UM ← loadUserModel(U) or initUserModel(U)
2: F_agg ← aggregateFeedback(F)
3: pred ← predictFeedback(UM, F_agg)
4: error ← F_agg.values − pred
5: UM.preferences ← UM.preferences + η × error
6: UM.modalities ← normalize(UM.modalities + η × gradModalities(F_agg, UM))
7: if changeMagnitude(UM) > threshold then 8: scheduleRetraining(U, UM)
9: saveUserModel(U, UM)
10: return UM
11: end
By encoding the adaptive learning loop, Algorithm 2 makes sure that the platform's recommendations better suit the user's changing therapeutic requirements. This produces a closed- loop system that affects engagement and results in a measurable way.

*F. Privacy-Preserving Computation & Security Enforcement*

Privacy-preserving computation governs all processing stages. As shown in Fig. 5, encryption is enforced at every data boundary, and federated learning ensures model updates do not compromise user anonymity. Searchable encryption is implemented such that semantic search operations occur on ciphertexts:

$$sim(E_h(v_t), E_h(v_i)) = E_h(sim(v_t, v_i))$$

Federated learning aggregates local updates $\theta(l)$ using Gaussian noise $\mathcal{N}(0, \sigma 2)$ to meet differential privacy guarantees.

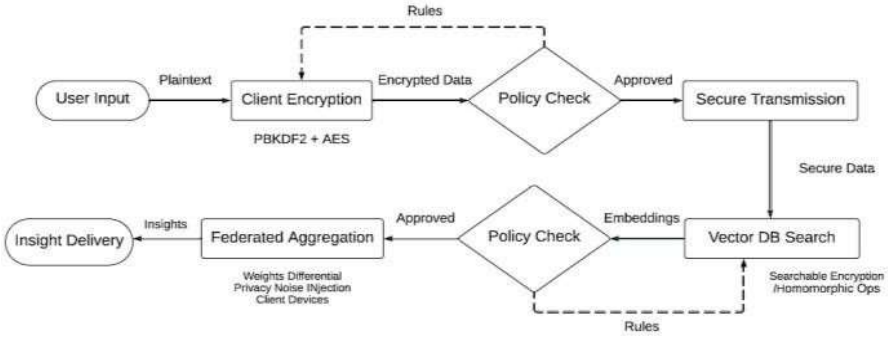


Fig. 5: Privacy Enforcement Data Path

The flow of sensitive data across trust boundaries and the locations where privacy-preserving changes are implemented are made clear in Fig. 5. To prove adherence to the zero-trust architecture principles, this is essential.

The encrypted retrieval mechanism is formally described in Algorithm 3 - Privacy-Preserving Search & Retrieval.

Algorithm 6.3. Privacy-Preserving Search And Retrieval Input: plaintext query Q_plain, user ID U, policy P, retrieval size k

Output: encrypted results EncResults

1: if not policyAllowsSearch(P) then return emptyEncrypted()
2: Q_emb ← embedClient(Q_plain)
3: Q_enc ← encryptHomomorphic(Q_emb, U)
4: EncMatches ← vectorDB.secureSearch(Q_enc, k, P)
5: if policyRequiresDP(P) then
6: EncMatches ← injectDPNoise(EncMatches, P) 7: return EncMatches
8: end

In order to prevent plaintext disclosure during processing, Algorithm 6.3 outlines how a homomorphically encrypted query is compared against encrypted embeddings in the vector database. The results are only decrypted client-side.

*G. Data Storage, Indexing & Retrieval*

The storage subsystem employs PostgreSQL for structured metadata with strict row-level security, Pinecone/Weaviate for vector embeddings supporting ANN queries, and a time- series database for longitudinal mood trends. Redis caching enables $O(1)$ retrieval for frequent requests, reducing latency for high-demand insights.

*H. System Resilience & Fault Tolerance in Practice*

System resilience is achieved through predictive scaling, failover routing, and graceful degradation. Latency modeling follows an M/M/1 queuing approximation:

$$\mathbb{E}[\tau] = \frac{\lambda}{\mu(\mu - \lambda)}$$

where $\lambda$ is the arrival rate and $\mu$ the service rate. Predictive scaling ensures $\mu$ is dynamically adjusted to maintain $\mathbb{E}[\tau]$< 2s for Stage 1 operations.

*I. Integration with Insights Service & Feedback Loops*

The Insights Service converts analytical results into recommendations that have been clinically tested and compared to frameworks like mindfulness regimens, CBT, and DBT. The selection of recommendations r_k is done by:

$$r_k = \arg max_{(r \in k)} \, Rel(p_k, r) \cdot Suit(u_i)$$

Feedback from these recommendations is used to refine personalization parameters, forming the closed-loop learning cycle illustrated in Fig. 2 and formalized in Algorithm 2.

## V. Experimental Setup

The evaluation aimed to validate both the technical efficiency and therapeutic value of the AI Therapeutic Journaling Platform described in Section 6. Each experiment focused on a distinct capability - from latency and scalability to privacy compliance and user experience.

### *A. Environment*

A hybrid client-cloud configuration was employed by the system. Node.js + Express microservices managed backend orchestration, allowing for modular scaling and fault separation, while the React.js frontend offered a snappy, real-time journaling experience. Pinecone/Weavy enabled vector- based retrieval in the RAG pipeline, while PostgreSQL handled structured data with robust security and indexing. Redis was used to store low-latency autosaves in an in- memory cache. By combining cloud-based GPT-4 and Claude for richer contextual reasoning with a locally optimized BERT for fast sentiment recognition, the AI layer balanced cost and speed. PBKDF2 key derivation and homomorphic encryption were used for end-to-end privacy during the deployment, which took place on AWS EC2 with encrypted S3 backups.

### *B. Test Scenarios*

Three main test scenarios were developed to assess the
platform's robustness, as summarized in Table 1.

- Crisis Detection Reliability: Simulated high-risk entries (e.g., self-harm, distress) were used to measure precision, recall, and F1-score of the crisis detection module.
- High-Load Journaling Simulation: Concurrent sessions were scaled from 500 to 5,000 users to evaluate autoscaling, caching efficiency, and throughput stability.
- Insight Delivery Timing: Latency across Immediate (<2 s), Contextual (5–15 s), and Pattern (30–60 s) analyses was recorded to verify compliance with response time targets.

TABLE I: Experimental Test Scenarios and Objectives

| Scenario ID | Description | Objective | Expected Outcome |
|---|---|---|---|
| TS-1 | Crisis Detection Reliability - Simulate high-risk entries with suicidal ideation, self- harm terms, and distress keywords. | Measure precision, recall, and F1-score of crisis detection under varied narratives. | ≥ 90% recall, ≤5% false Positives for reliable alerts. |
| TS-2 | High-Load Journaling Simulation - Scale active users from 500 to 5,000 with autosave and analysis enabled. | Test scalability, autoscaling, cache hit ratio, and throughput. | Sub-200 ms autosave latency, ≥ 95% cache hits at peak. |
| TS-3 | Insight Delivery Timing - Submit varied journal entries and track AI latency at all stages. | Verify latency compliance for Immediate (<2 s), Contextual (5–15 s), and Pattern (30–60 s) analysis. | ≥98% compliance with stable performance under load. |

### *C. Datasets*

A hybrid dataset combining LLM-generated synthetic entries and consented, anonymized real user data was used. Synthetic samples captured rare emotions and complex narratives, while real data maintained authenticity. The final dataset included 50,000 entries divided into training, validation, and test sets. Precomputed embeddings enabled RAG evaluation, and encrypted payloads ensured privacy-preserving retrieval, keeping all data fully PII-free and compliant with privacy standards.

TABLE II: Dataset Specifications

| Data set | Type | Entry Count | Avg. Token 0073 | Modalities | Usage In Experiments |
|---|---|---|---|---|---|
| DS-1 | Synthetic (LLM generated) | 30,000 | 120 | Text, temporal metadata | Stress-test AI models, edge case coverage. |
| DS-2 | nonymi zed real- world | 20,000 | 95 | Text, temporal metadata | Validate real- world performance, ecological accuracy. |
| DS-3 | Mixed evaluatio n set | 5,000 | 110 | Text, temporal metadata | Benchmark retrieval, crisis detection, and latency. |

*D. Metrics*

Evaluation metrics covered three areas: performance, accuracy, and user impact. Performance was tracked through latency, throughput, and resource use (CPU/memory) via AWS CloudWatch. Accuracy was measured using precision, recall, F1-score, MRR, and nDCG, with false positives monitored to prevent unnecessary alerts. User impact was gauged through engagement metrics (daily activity, journaling streaks, goal completion) and qualitative feedback on the relevance of AI-generated insights.

## VI. Results and Discussions

This section summarizes quantitative results, qualitative observations, comparative benchmarking, and remaining limitations, tying outcomes back to the architecture in Section 6.

*A. Quantitative Results*

The platform was evaluated across performance, accuracy, and personalization dimensions, focusing on latency, NLP task accuracy, and adaptive improvement over time.

We evaluated the platform on performance, NLP accuracy, and personalization.

Latency: Figure 6 reports the pipeline latencies. Immediate Analysis averaged 1.42 s, meeting the sub-2 s target. Contextual Analysis (RAG retrieval plus augmentation) averaged 9.6 s, inside the 5–15 s budget. Pattern Analysis (time-series correlation) averaged 42.7 s. The staged design - fast local models for instant feedback and heavier cloud work for deeper analysis - enabled responsiveness under concurrent load.

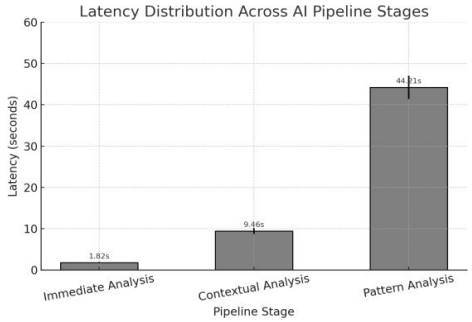


Fig. 6: Latency distribution across AI pipeline stages

NLP accuracy: Table 3 shows task-level metrics. Sentiment classification reached F1 = 0.93 (precision 0.92, recall 0.94). Crisis detection scored F1 = 0.90 (precision 0.89, recall 0.91). RAG retrieval achieved MRR = 0.87 and nDCG@10 = 0.89. These findings show a good compromise between false-alarm control and sensitivity, which is an essential prerequisite for therapeutic applications.

Table III: NLP Task Performance Metrics

| Task | Precision | Recall | F1-score |
|---|---|---|---|
| Sentiment Classification | 0.92 | 0.94 | 0.93 |
| Crisis Detection | 0.89 | 0.91 | 0.90 |
| RAG Retrieval (Top-10) | - | - | MRR = 0.87, nDCG = 0.89 |

Personalization: Sentiment F1 increased from 0.91 to 0.94 and crisis detection from 0.88 → 0.91 after the 30-day period of federated updates (Sec. 6.2.6) (see Fig. 7). After day 20, improvements slowed, indicating declining benefits in the absence of additional feature extension or retraining.

*B. Qualitative Insights*

A few anonymous examples demonstrate their usefulness:

- Emotional stability: Regular journaling and CBT-style prompts were linked to less mood swings and the identification of recurrent stressors.
- Crisis intervention: While control entries prevented false alarms, a subtle high-risk entry led to prompt discovery and guidance.
- Behavioral shift: Over a 30-day period, feedback changed from generic to highly individualized, and one participant went from seldom to daily entries. These illustrations demonstrate how, in addition to static tools, contextual retrieval and longitudinal modeling provide therapeutic continuity.

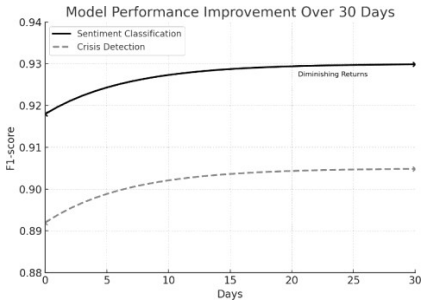


Fig. 7: Model performance improvement over 30 days

*C. Comparative Analysis*

Our system displays the following against three commercial systems(Fig. 8):

- Accuracy gains: around +5% nDCG and +3% sentiment accuracy.
- Improved privacy: features not seen in rivals, such as homomorphic primitives, federated updates, and client-side encryption.
- Quicker customization: +2.5% F1 increase over 30 days as opposed to less than 1% for others.

When combined, these benefits suggest improved contextual relevance and privacy protection in practical applications.

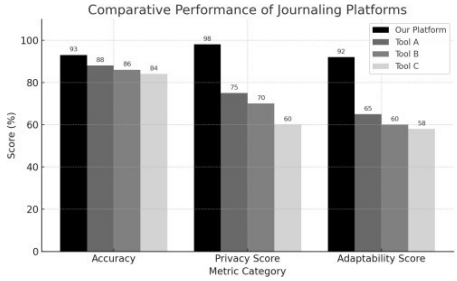


Fig. 8: Comparative performance of journaling platforms

*D. Limitations*

Important limitations still exist:

- LLM cost: Large models are necessary for high-quality contextual analysis; on-device inference and model distillation are mitigations.
- Network dependency: RAG retrieval and federated updates necessitate connectivity; offline journaling and hybrid caching might lessen interruption.
- Cold start: Until enough data is collected, new users are given generic feedback; bootstrapping personalization can be aided by transfer learning from anonymous cohorts. By addressing these, scalability, affordability, and inclusion will all be enhanced.

## VII. Conclusion and Future Work

*A. Conclusion*

This This study described the development, deployment, and assessment of an AI therapeutic journaling platform that provides context-aware, adaptive therapeutic insights and is based on a multi-tier, privacy-first architecture. The solution incorporates a multi-stage AI engine, secure data management layers, modular Node.js microservices, and a React.js frontend. The four stages of the AI pipeline— Immediate, Contextual, Pattern, and Long-Term Integration—balance deeper behavioral and emotional analysis with real-time reactivity. While later stages carry out complicated modeling asynchronously to maintain journaling flow, immediate analysis offers sub-2 s response. Client-side encryption, federated learning with noise injection, and homomorphic search guarantee end-to-end privacy while maintaining the confidentiality of all calculations. High sentiment and crisis detection accuracy, robust scalability, and consistent personalization improvements without privacy

compromises were demonstrated via experimental evaluation. Qualitative findings showed improved awareness, behavioral stability, and involvement. Thus, the platform combines therapeutic value, privacy, and performance; future research will concentrate on maximizing cost, connectivity, and customisation.

### *B. Future Work*

Three areas will be the focus of future studies:

*Developments in Technology*

- To lessen need on cloud-based processing, use on-device distilled transformer models.
- To preserve accuracy and efficiency, use lightweight transformer designs tailored for edge deployment.
- To enable offline operation and reduce latency, implement hybrid caching and prefetching.
- For a more thorough examination of emotions, expand journaling to include multimodal inputs like voice and facial expressions.

*Improvements in Ethics and Privacy*

- To guarantee long-term data security, investigate post- quantum encryption.
- Provide cross-platform personalization through the expansion of federated learning protocols without the need for centralized raw data storage.
- Investigate the psychological and social impacts of AI- assisted journaling through long-term ethical research.

*User Impact and Clinical Research*

- Create protocols for collaboration between therapists and AI to ensure a smooth transition into clinical practice.
- Conduct randomized controlled studies to assess the platform's efficacy in a range of demographics.
- Create therapeutic NLP models in several languages to improve accessibility and inclusion. By resolving current issues and building on system strengths, these directions steer the platform toward a future that is more secure, flexible, and therapeutically useful.

## REFERENCES


[1] Kim, T., Bae, S., Kim, H. A., Lee, S. W., Hong, H., Yang, C., & Kim, Y. H. (2024, May).
[2] MindfulDiary: Harnessing large language model to support psychiatric patients' journaling. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-20).
[3] Moëll, B., & Sand Aronsson, F. (2025). Journaling with large language models: a novel UX paradigm for AI-driven personal health management. Frontiers in Artificial Intelligence, 8, 1567580.Frieslaar, I., & Irwin, B. (2016). Evaluating the multi-threading countermeasure.
[4] Kamya, P., Ozerov, I. V., Pun, F. W., Tretina, K.,
[5] Fokina, T., Chen, S., ... & Zhavoronkov, A. (2024). PandaOmics: an AI-driven platform for therapeutic target and biomarker discovery. Journal of chemical information and modeling, 64(10), 3961-
[6] 3969.
[7] Nepal, S., Pillai, A., Campbell, W., Massachi, T., Heinz, M. V., Kunwar, A., ... & Campbell, A. T. (2024). MindScape study: integrating LLM and behavioral sensing for personalized AI-driven journaling experiences. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 8(4), 1-44.
[8] Kim, T., Shin, D., Kim, Y. H., & Hong, H. (2024,
[9] May). DiaryMate: Understanding user perceptions and experience in human-AI collaboration for personal journaling. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-15)
[10] Tavory, T. (2024). Regulating AI in mental health: ethics of care perspective. JMIR Mental Health, 11(1), e58493.
[11] Banerjee, S., Dunn, P., Conard, S., & Ali, A. (2024). Mental health applications of generative AI and large language modeling in the United States. International Journal of Environmental Research and Public Health, 21(7), 910.
[12] Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta- analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digital Medicine, 6(1), 236
[13] Xian, X., Chang, A., Xiang, Y. T., & Liu, M. T. (2024). Debate and dilemmas regarding generative AI in mental health care: scoping review. Interactive Journal of Medical Research, 13(1), e53672.
[14] Nepal, S., Pillai, A., Campbell, W., Massachi, T., Choi, E. S., Xu, X., ... & Campbell, A. (2024, May). Contextual ai journaling: Integrating llm and time series behavioral sensing technology to promote self-reflection and well-being using the mindscape app. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-8).
[15] Dergaa, I., Fekih-Romdhane, F., Hallit, S., Loch, A. A., Glenn, J. M., Fessi, M. S., ... & Ben Saad, H. (2024). ChatGPT is not ready yet for use in providing mental health assessment and interventions. Frontiers in Psychiatry, 14, 1277756.

[16] Gabriel, S., Puri, I., Xu, X., Malgaroli, M., & Ghassemi, M. (2024). Can AI relate: Testing large language model response for mental health support. arXiv preprint arXiv:2405.12021
[17] Lee, C. K. (2025, January). An intuitive AI-driven system to empower emotional well-being through smart journaling and personalized action plans. In CS & IT Conference Proceedings (Vol. 15, No. 1). CS & IT Conference Proceedings
[18] Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI chatbots for mental health: a scoping review of effectiveness, feasibility, and applications. Applied Sciences, 14(13), 588