DEVELOPMENT OF A JOURNAL AI WITH APPLICATIONS OF ACCEPTANCE COMMI

By

Nathan Jeffrey Wang

A capstone project submitted for Graduation with University Honors

May 7, 2026

University Honors

University of California, Riverside

APPROVED

Dr. Weiwei Zhang

Department of Psychology

Dr. Begoña Echeverria, Howard H Hays, Jr. Chair

University Honors

ABSTRACT

This project outlines the development of a journaling application designed to enhance deep self reflection with LLMs. This application addresses the existing privacy and ethical risks prevalent in existing cloud-based journaling chatbots. The app is built as a native macOS application using Flutter for the frontend graphical user interface and Python for the backend engine. OLLAMA to securely host local large language models (LLMs) and integrated into the backend through Langchain. Local storage is managed through SQLite for data serialization and FAISS vector stores for Retrieval-Augmented Generation (RAG). The application guarantees absolute data privacy with zero external network dependencies. The system functions as a reflective mirror following the principles of Acceptance and Commitment Therapy (ACT). Core feature capabilities include value systems analysis, temporal cognitive tracking, and contradiction awareness. Through natural language processing, the AI extracts virtues and behavioral metrics from daily entries. This data is fed through the RAG pipeline to generate highly personalized prompts that highlight past versus present value-behavior inconsistencies. The goal of prompting is to enhance user introspection without generating clinical advice through the LLM. The future goal of this application is to serve as a therapeutic aid for clinicians to analyze the progress of patients.

ACKNOWLEDGMENT

I would like to express my gratitude to my faculty mentor, Dr. Weiwei Zhang, for his overall guidance as my project advisor. When I was starting the project, he provided invaluable research resources that helped me frame the direction. His expertise in cognitive processing of emotional experience improved the original implementation of the value extraction system. His contributions directly decreased the difficulties in development.

Furthermore, I would like to extend my appreciation to my peers for their continual engagement and collaborative exchange of ideas to enhance the application. Regardless of whether the discussion concerned visual aesthetics, edge cases, or models, these consistently elevated the quality of the features developed. Their feedback allowed the final application to account for general use cases that encapsulate the different ways people journal.

TABLE OF CONTENTS

Introduction

Journaling is an effective, low-cost method for managing mental health through daily mood and routine tracking. I, along with many others, enjoy journaling for the sake of self-reflection to create actions that improve my way of living. With the developments in LLM technology, a growing trend towards AI chatbot integration with journaling has spread. One of the earliest programs, ELIZA, demonstrated applications of Rogerian therapy to a chatbot. Even then, ELIZA was described as an "empathetic" agent, despite users' understanding the notion that ELIZA was a machine (Weizen baum, 1966). In modern day, LLMs outperform ELIZA in natural language processing and generation with vast database systems and training sets. It seems almost obvious that these LLMs could replace journaling and therapy. However, the fact that LLMs are machines should persist. I believe AI should only be used to analyze without guiding the user's actions. Issues with the current software ignore this ethical principle alongside privacy and data concerns. This project outlines the development of an LLM-based journaling application that utilizes analytic tools to assist users in self-reflection while addressing the aforementioned shortcomings. Designed for general use outside of formal therapy, the application ensures that all data processing is strictly localized, keeping user data safe from external networks.

Background

Journaling has been shown to be an effective preventative treatment that helps clients reflect and create coping strategies. In a study by Sohal et al. (2022), the group studied the efficacy of journaling in reducing symptoms for those with mental illnesses. The subjects had diagnoses including anxiety, depression, schizophrenia, PTSD, and more. The study found that journaling interventions generated the most impact in subjects with anxiety and depression, while not significantly changing subjects with more intense disease. The study also found interventions with

prior journaling technique training significantly improved outcomes (Sohal et al., 2022). I believe my project will have a positive impact on teaching users journaling techniques through prompts to ultimately reduce the learning curve.

Journaling should be used to remember positive experiences. In a study correlating emotional condition and memory recall, researchers found that memories encoded with negative valence improve recalled details (Xie & Zhang, 2017). This notion implies our recollection of daily lives tends to hyper-fixate on negative details (Kensinger, 2009). Positive journaling is a common therapeutic technique to focus thoughts on positive aspects of the day. This technique improves self-regulation of emotion and thoughts (Smyth et al., 2018). For my journal application, the model will implement positive journaling prompts as a baseline for personalized prompts. Tracking the temporal frequency of positive and negative valence may allow the model to detect a quantitative change in mindset.

The application's therapeutic foundation is rooted in Acceptance & Commitment Therapy (ACT). ACT encourages mindful and continuous self-reflection by following the Hexaflex to boost psychological flexibility. The philosophy behind ACT argues that psychological flexibility allows balance between positive and negative thoughts and choosing appropriate actions to create change. A core component of the Hexaflex is "Values"—identifying what is truly important to the individual (Anusuya & Gayatridevi, 2025). This application focuses primarily on enforcing the Values section, analyzing how user entries align with or contradict their established value systems.

Design & Ethical Guardrails

Given the sensitive nature of journaling, strict ethical guardrails were implemented in the system's design. These guidelines were inspired by a blog post from Chen (2024), where she outlined her experience and issues with current journal AI technology. Initially, she was hesitant to write in an online journal due to the risks of data leakage. In the journal study above, the researchers found that subjects whose journals were regularly read by examiners struggled to show improvement for their conditions (Sohal et al., 2022). This idea warrants concern over using online services to write private details. After overcoming her initial concerns, Chen found that the emotional insight and advice generated from the software fell short as generalized statements trying to be personal. She also referenced the idea of "dataism", a trusting bias for numerical representation of concepts. Reducing emotions and life experiences to statistics might seem beneficial and unbiased; however, Chen (2024) warns the readers about making life choices out of an unfeeling machine (Chen, 2024). For my project, the following implementations address these issues.

- **No Therapeutic Advice:** The application is explicitly designed for pattern analysis, not clinical intervention. This mitigates liability concerns and prevents overgeneralized suggestions.
- **Self-Reflection Focus:** The LLM acts as a mirror, analyzing text patterns to bring awareness to contradictory values and behaviors. The system will encourage users to create actions for themselves after evaluating their behavioral patterns.
- **Absolute Privacy:** The system relies on entirely localized storage systems. LLM interactions are run via local models, and the program operates with zero network access requirements.

Due to the localization of the model, lower performance is expected. This ties into the application limits. Online LLMs utilize servers to run the models, which allows for more content safeguards. This application is not suitable for people with extreme mental illness or instability. The default model has limited protection, which can often be bypassed through prompt engineering. The program can scan for dangerous concepts by checking for specific words; however, this system cannot differentiate when terms signify harm versus emotional processing.

Development Process and System Architecture

The project was developed as a full-stack native macOS application with a graphical user interface (GUI), backend system processing, and multiple local databases. The LLM is stored locally as a list of weights. This allows the program to run offline without any network dependencies.

Technical Architecture

The frontend graphical user interface (GUI) was built using Flutter using the Dart framework. Flutter was chosen for aesthetic purposes along with modern customization options. The main program interacts with the backend by providing the user input as strings along with metadata. The backend engine processes the data using Python scripts and LLM model calls. The models are hosted by OLLAMA and integrated into Python using Lanchain. The local model used is qwen3:8b. The processed data is stored and managed in an SQLite database in JSON serialized formats. For Retrieval-Augmented Generation (RAG), the processed data is embedded by the nomic_embed_text model and stored in Facebook AI Similarity Search (FAISS) vector stores.

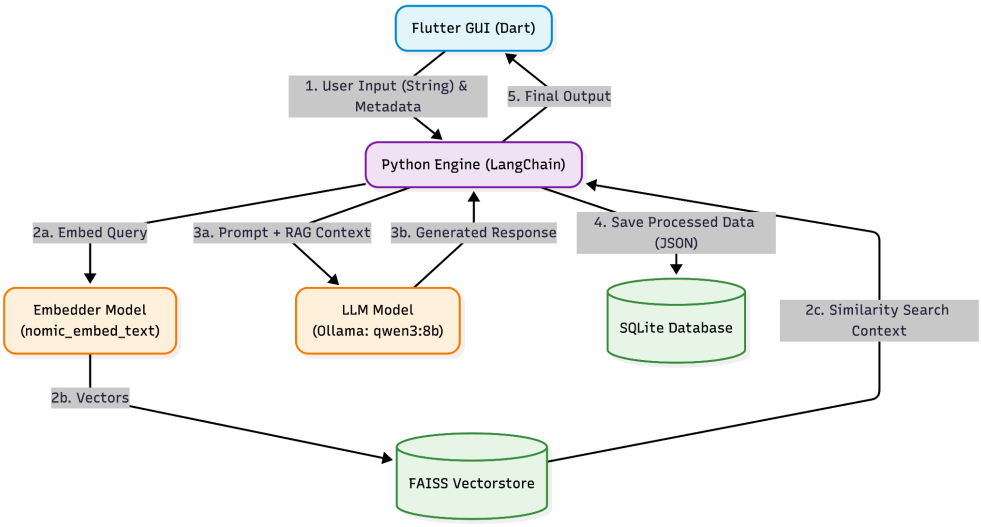


Figure 1: Comprehensive system interactions diagrams. Full-stack architecture is distinguished by color: blue-frontend, purple-backend, orange-models, green-database. Data is represented in gray boxes.

Prompt Engineering

Prompt engineering is a technique to enhance the precision and accuracy of model queries. Prior to model input, queries are formatted to encapsulate the overall role. Modifications such as text structuring, input-to-output examples, or feeding the model's output back for more context are used to ensure proper output formatting. Prompt engineering allows models to be accurate while maintaining versatility (Hsieh et al., 2025). Fine-tuning and distillation improve performance in specific tasks while losing external capabilities. In this project, prompt engineering is used for the following tasks.

- **Directing Journal Intention Processing:** The model screens the raw input to guess the user's intentions behind journaling. Text modalities include routine, deep reflection, venting, or general information.
- **Feature Extraction:** Model acts as a psychoanalyst. Sentiment analysis and other feature metrics are analyzed from the text (see Section 5).

- **Data Structuring:** Model acts as a format converter. After feature extraction, the model bundles data and formats it into JSONs for vectorstore and SQLite storage.
- **Prompt Generation:** Model acts as a curious mediator between the current self and past self. The model uses current and past features to create a question prompt that maximizes new information gained from the user.

Retrieval-Augmented Generation (RAG)

RAG improves model accuracy by searching for relevant context from a list of documents. In my project, RAG is utilized to query previous journal entries for matching value systems or situations and add to the current prompt generation as context. The model can create prompts that relate the current journal entries to similarities and contradictions from past journals. The RAG pipeline is shown visually in Figure 1. High-level RAG workflow is described below.

1. **Raw Ingestion:** Entries are recorded as raw text via the Flutter interface.
2. **Metadata Processing:** Feature extraction is performed on raw text to generate metadata (see Section 5). Metadata is appended to raw text.
3. **Text Chunking:** Text and metadata are converted to structured JSON objects.
4. Vector Encoding: JSON objects are encoded into high-dimensional vector spaces that index source text.
5. **Context Retrieval:** User queries trigger similarity searches in the FAISS vectorstore.
6. **Generation:** The LLM utilizes different retrieved features from the historical context to synthesize personalized prompts.

This workflow allows for more relevant and precise prompt generation that allows the LLM to compare current psychological profiles with historical ones.

Model Comparison & Selection

Several open-source models were evaluated for running locally on macOS. While Llama 3.1:1b provided excellent input latency, it lacked task precision for general prompt-engineered tasks. For instance, JSON outputs failed to be consistent or lacked elements. Llama 3.1:3b was balanced but struggled with rigid data formatting and complex dynamic prompting. Ultimately, qwen 3:8b was selected for its performance on sentiment analysis and output formatting. Model latency was moderate, but can be further improved via distillation.

Features of the Application

The application features a comprehensive suite of cognitive and journaling tools designed to deepen self-reflection through targeted data analysis.

Value Systems Analysis

Using natural language processing techniques, journal entries are evaluated for strength of reflected virtues, arousal levels, locus of control, and relevant situational context (like relationships or work). These values are then structured into JSON format (see Figure 3b). After embedding these features as a single unit, journal entries with similar or contrasting values are easily representable in vector space. The RAG pipeline uses these embeddings to find relevant entries and generate personalized prompts. The accumulation of these value systems builds a cognitive profile. Future prompts may notice deviations from previously established value systems and ask the user to elaborate.

Cognitive Database

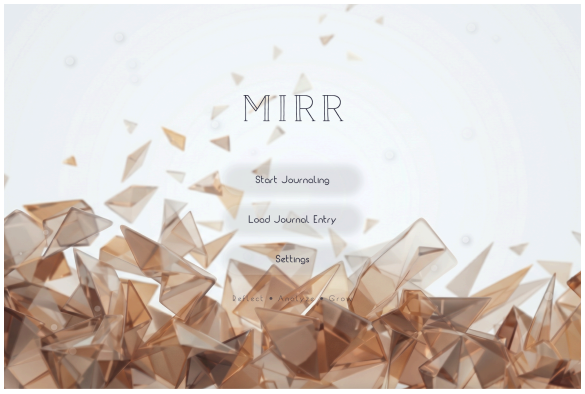
Expanding on the cognitive profile, temporal information allows the system to detect if the user is creating actionable change. By tracking changes in the intensity and presence of personal values over time, the system can learn to correlate growth to situations or actions. This feature is the

most experimental due to the difficulty in capturing and explaining change through value analysis alone.

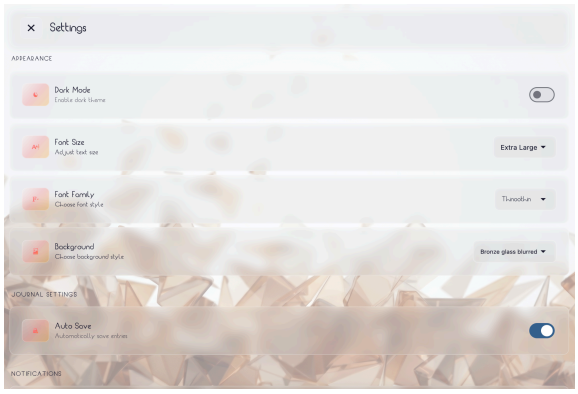
Contradiction Awareness

Contradiction awareness is activated during the RAG context retrieval phase. If retrieved documents contain context contradicting current values, the system will compare the situations to identify the source of conflict. For instance, if a past stated goal or value conflicts with present behaviors, the system will create a prompt to address the discrepancy. This technique tracks behavioral inconsistencies and shifting priorities that may contribute to forgetfulness. The LLM then surfaces this contradiction to the user without offering prescriptive advice, querying the user about their thoughts or proposal of action. It acts as a mirror, generating prompts to engage self-reflection of values and behaviors.

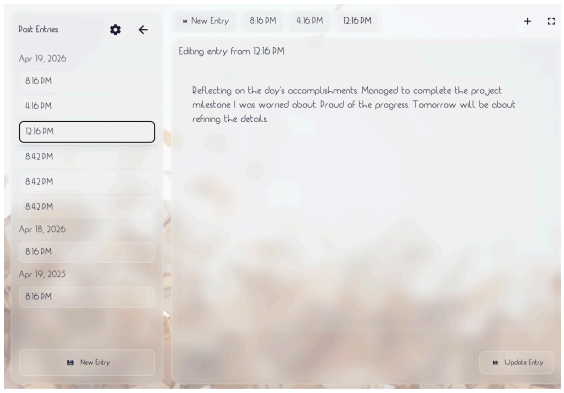
Prototype Showcase



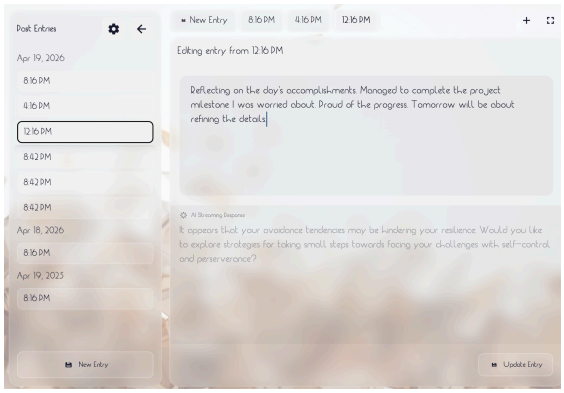
(a) Main screen options with full background and animations.



(b) Settings screen for user configuration and journal entry loading.



(c) Journal screen displaying past entries on left and open tabs on top.



(d) Journal screen after display model response after data ingest pipeline.

Figure 2: Application screens showcasing flushed out GUI and interface options.

System Testing and Validation

To ensure application functionality and reliability, various testing methods were performed on the GUI, the backend scripts, and the database systems.

Frontend and GUI Testing

Testing the GUI mainly involved running the compiled program and verifying the implementation of all visual features. The settings menu was tested to ensure user configurations successfully updated backend parameters and modified the existing program appropriately. To improve usability, keyboard shortcuts were implemented and successfully tested. For tab navigation: Ctrl+W closes tabs, Ctrl+T opens new tabs, and Esc exits journal full-screen mode.

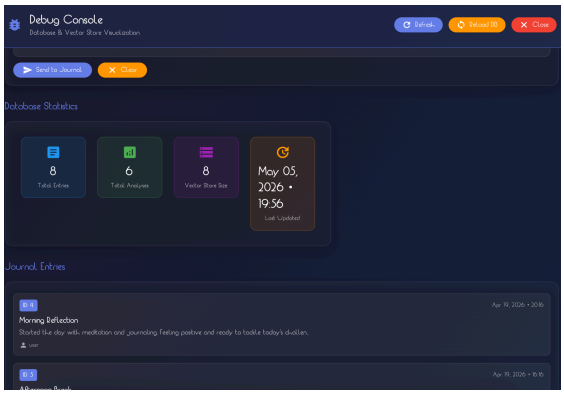
Some user interface (UI) edge cases were also tested for functionality. Closing all active tabs returns the user to the main menu. Additionally, changes to a journal tab will persist when moving to a different tab without losing unsaved changes. The data will only be lost if the tab is closed and the user chooses to exit without saving.

Backend and Database Testing

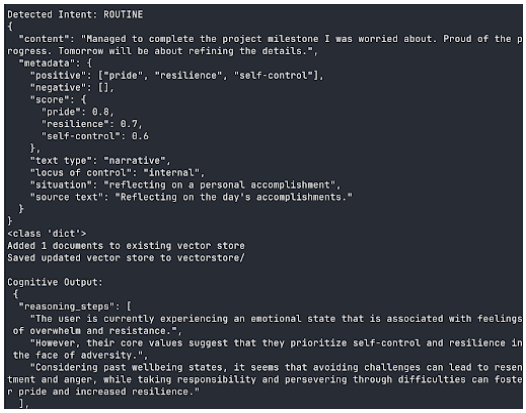
Backend testing focused on the data pipeline (see Figure 1), ensuring the system properly ingests and deletes journal entries. Multiple test entries were created and saved to the databases. Feature extraction was performed on these entries to verify that the model interactions were valid. The output of the models was verified to contain the proper metadata and JSON formatting.

To aid in backend testing and validation, a dedicated debug screen was implemented (see Figure 3a). The screen allows developers to visualize the contents of the databases directly without SQLite calls. This screen was used to confirm the existence of past journals in the database and the JSON structuring after feature extraction. The debug screen was also used to launch the application to visualize the database when the database was empty. This ensured null states were handled without fatal program failure.

Testing Demonstration



(a) Debug screen displaying tools and database information



(b) Example journal entry ingestion and JSON formatting

Figure 3: Testing and Data Validation Tools

Future Direction

While the current build works as a prototype, future development will focus on expanding its capabilities beyond value analysis and preparing the software for wider distribution. Some suggestions for future implementations are listed below.

- **Model Optimization:** At the moment, prompt generation is the most complex task and time consuming task. The current qwen3:8b model performs well on generation while maintaining moderate latency. Model distillation could compress the LLM size improving performance and space. The general tasks performed by prompt engineering can be replaced with smaller models (see Section 4.2).
- **Expanded Hexaflex Framework:** The current iteration focuses heavily on the "Values" pillar of the ACT Hexaflex. Future updates aim to integrate broader ACT concepts, particularly Cognitive Defusion and Self-as-Context. Cognitive defusion might ask the user how they might reframe negative thoughts in an accepting manner. Self-as-Context features could track the user's self-perception to help distinguish the user's identity from transient thoughts and emotions.
- **Self-Coaching Tools:** The system could identify long-term psychological conflict-resolution patterns to show the user how they resolved a similar issue in the past. The prompts would encourage the creation of personally proven coping mechanisms rather than seeking external guidance.
- **Safeguard Implementation and Harm Prevention:** Before the application can be released, safety testing and safeguard implementations must be completed. Local models lack cloud based guardrails of commercial APIs. Safety classifiers must be developed and rigorously tested to detect mental instability or signs of self-harm. The system must be

programmed to recognize these edge cases to stop the reflection loop and provide emergency resources.

- **Open Source Release:** After the implemented safety measures, the final objective is to open-source the project. This will involve polishing the Flutter frontend for seamless cross platform compatibility and building a website with installation protocols. The installation pipeline will automatically configure OLLAMA and the required models and databases.

Conclusion

This capstone project successfully demonstrates the development of a fully localized, AI-augmented journaling application. Utilizing the ACT framework with local RAG architecture, this software offers an alternative to cloud-based chatbots. The application successfully extracts virtues to personalize journal prompts.

The system presents several technical and ethical limitations. The AI relies exclusively on self-reported information. Thus, it cannot create a fully accurate representation of the individual's personality, nor can definite change be measured. Furthermore, the technical constraints of running models locally limit the system's reasoning depth. It remains imperative that this tool functions as a reflective mirror rather than a diagnostic tool.

Despite these limitations, the underlying technology holds potential for application to clinical practices. The software excels in data analysis of emotional and behavioral trends. Clinicians could use this tool to track the improvement of treatment plans. For therapy, therapists could track patient progress between sessions or uncover hidden behavioral trends that could serve as talking points.

Moving forward, future iterations of this project will concentrate on optimizing the language models through distillation and expanding the capabilities to encompass broader ACT

tools. Implementing rigorous safety safeguards is a priority before open-sourcing the project. By recognizing the constraints of current LLMs, this project establishes a foundation for technological integration into mental health management.

References

Anusuya, S. P., & Gayatridevi, S. (2025). Acceptance and commitment therapy and psychological well-being: A narrative review. *Cureus*, 17(1), e77705. https://doi.org/10.7759/cureus.77705

Chen, A. (2024, June 24). What I learned from sharing my private self with an AI journal: Psyche ideas. *Psyche*. https://psyche.co/ideas/what-i-learned-from-sharing-my-private-self-with-an-ai-journal

Hsieh, M.-Y., et al. (2025). Impact of prompt engineering on the performance of ChatGPT variants across different question types in medical student examinations: Cross-sectional study. *JMIR Medical Education*, 11, e78320. https://doi.org/10.2196/78320

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, 1(2), 99–113. https://doi.org/10.1177/1754073908100432

Smyth, J. M., et al. (2018). Online positive affect journaling in the improvement of mental distress and well-being in general medical patients with elevated anxiety symptoms: A preliminary randomized controlled trial. *JMIR Mental Health*, 5(4), e11290. https://doi.org/10.2196/11290

Sohal, M., et al. (2022). Efficacy of journaling in the management of mental illness: A systematic review and meta-analysis. *Family Medicine and Community Health*, 10(1), e001154. https://doi.org/10.1136/fmch-2021-001154

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM,* 9(1), 36–45. https: //doi.org/10.1145/365153.365168

Xie, W., & Zhang, W. (2017). Negative emotion enhances mnemonic precision and subjective 17

feelings of remembering in visual long-term memory. *Cognition*, 166, 73–83. https://doi.org/10.1016/j.cognition.2017.05.025