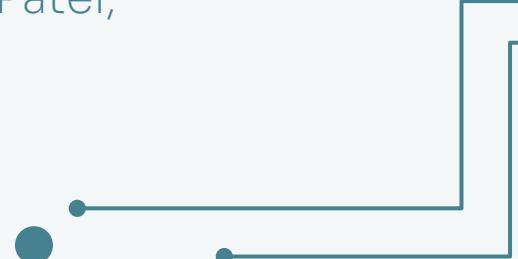




refrAlme

Intelligent CBT Journaling

Sineli Gamage, Hannah MacDonald, Shyam Patel,
Michael Garfagnoli, Bradley Igercich



1 in 5 U.S. adults live with a mental illness (**59.3M** in 2022)

Problem

AI-powered journaling tool rooted in **Cognitive Behavioral Therapy** (CBT)

Supports users to **reframe** unhealthy thoughts and **strengthen** positive ones

Solution

If 15% of adults are open to journaling, **~9M** potential users

Impact



Target Users

Reflective, Therapy-Adjacent Young Adults

About

Emotionally curious
Young Adults (18-35)

Motivation

- Build emotional awareness
- Find calm or perspective after a busy day
- Open to expressing feelings privately

Pain Points

- Lose consistency without emotional prompts
- Want a real reflective experience that feels friendly and personal



refraime-journal.com



Finish update

All Bookmarks



refrAlme

Your AI-powered wellness companion



Welcome to refrAlme

A safe space for reflection and growth, powered by compassionate AI

About refrAlme

refrAlme is an intelligent journaling app designed to help you **reframe** unhelpful thoughts and **strengthen** positive ones through AI-guided reflections rooted in Cognitive Behavioral Therapy (CBT).

Beta testers - Please let us know what you think before our feedback period closes on December 6th!

Begin your journey one word at a time

Sign in with Google

By continuing, you acknowledge that this app is for wellness support and is not a replacement for professional mental health care.

Data Privacy



Use

Personal entries are never used to train our AI models.



Encryption

Data is stored in an encrypted database, no access to third parties.



Deletion

Users have the ability to request deletion of all their entries at any time.



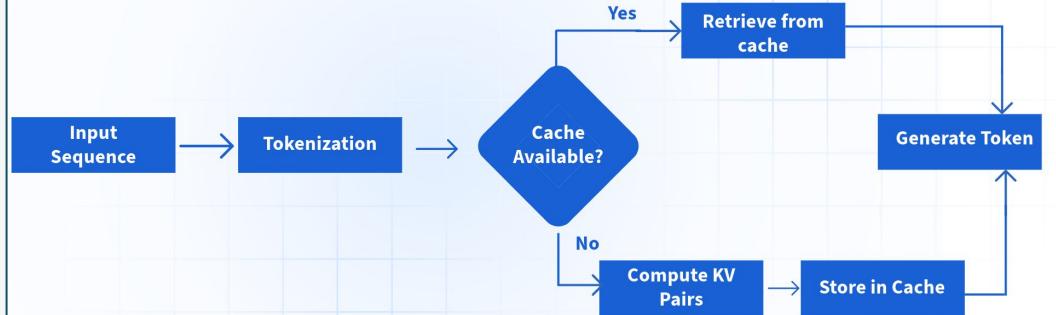
No External Sharing

Your entries never leave our system, they're processed within our secure environment and never sent to external APIs.

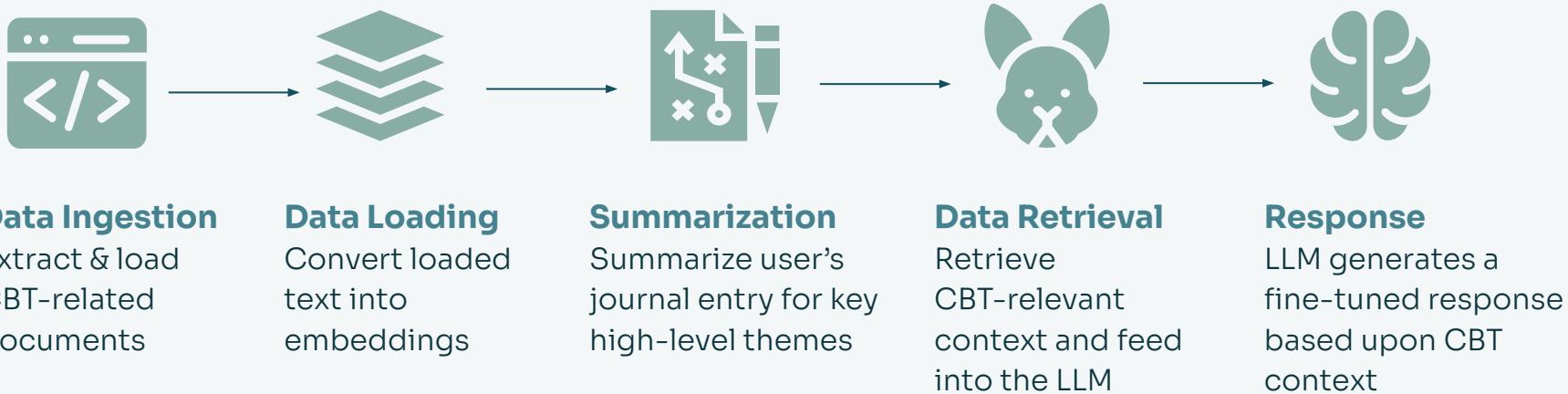
Model Configuration

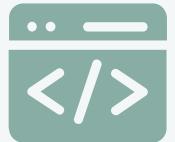
- Gemma2-9B-IT
- 4-bit AWQ
- vLLM
 - Continuous batching
 - KV Caching in blocks

Architecture Diagram of KV caching



RAG Architecture





RAG: Ingestion

- Pre-cleaned all CBT-related documents prior to loading in AWS.
- Focused on loading documents that contain instructional information about CBT as opposed to worksheets.

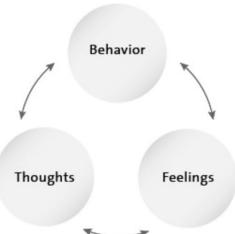
Full Website

In brief: Cognitive behavioral therapy (CBT)

Last Update: August 21, 2025; Next update: 2028.

Cognitive behavioral therapy (CBT) is one of the most common and best studied forms of psychotherapy. It is a combination of two therapeutic approaches, known as cognitive therapy and behavioral therapy.

The exact treatment approaches used will depend on the illness or problem to be treated. But the basic idea behind the therapy is always the same: What we think, how we behave, and how other people make us feel are all closely related – and they all affect our wellbeing.



Scraped Text

Cognitive behavioral therapy (CBT) is one of the most common and best studied forms of psychotherapy. It is a combination of two therapeutic approaches, known as cognitive therapy and behavioral therapy. The exact treatment approaches used will depend on...



RAG: Loading

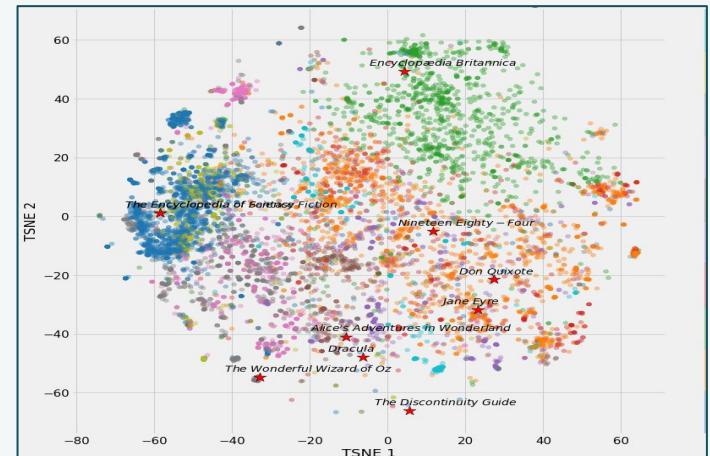
- Embeddings are created for loaded CBT documents within Amazon Bedrock
- Text is segmented into chunks of 384 tokens with a 50% overlap

Scraped Text

Cognitive behavioral therapy (CBT) is one of the most common and best studied forms of psychotherapy. It is a combination of two therapeutic approaches, known as cognitive therapy and behavioral therapy. The exact treatment approaches used will depend on...



Embeddings





RAG: Summarization

- Journal entries are summarized by an LLM to capture key high-level themes and insights, providing a more accurate context and enabling CBT correlations to be drawn from meaningful patterns rather than unnecessary details

Journal Entry

*"I just felt like **he does not miss me** **the way that I miss him** and that he really does not care about me the way that I care about him. **I know that these things aren't true, but our conversation left me feeling frustrated and sad.**"*



Summary

*The user expresses feelings of sadness and frustration **stemming from perceived lack of reciprocation** in their romantic relationship. ... **suggesting potential for cognitive distortions and catastrophizing.***



RAG: Retrieval

- Finds the chunks of text that contain the most relevant CBT information to the users journal entry
- This context is fed into the prompt to provide CBT-relevant information for the LLM to use in its output

Journal Entry

*"I just felt like **he does not miss me the way that I miss him** and that he really does not care about me the way that I care about him. **I know that these things aren't true, but our conversation left me feeling frustrated and sad.**"*

Retrieved Context

A woman who has a **core belief that she is "unlovable"** may also believe that if she works hard to please her partner, he may love her anyway. However, if her partner breaks up with her, her core belief of being "unlovable" **may be reactivated, and increase her vulnerability to depression or anxiety.**

RAG Model Response Example



Baseline (No RAG & Limited Prompt Eng)

It sounds like working many tasks was challenging at times. Have you noticed any specific thoughts during those moments when it felt particularly frustrating? Reframing these thoughts could make things feel less overwhelming. For example, instead of thinking "I can't handle this," maybe try telling yourself, "This situation feels difficult right now, but I have the skills to manage." How does that sound as a possible way to reframe your thoughts next time you face challenges at work?

Overly clinical;
giving direct
CBT advice

Tells the user
what to do

Feels artificial or scripted. Longer.

RAG Model Results

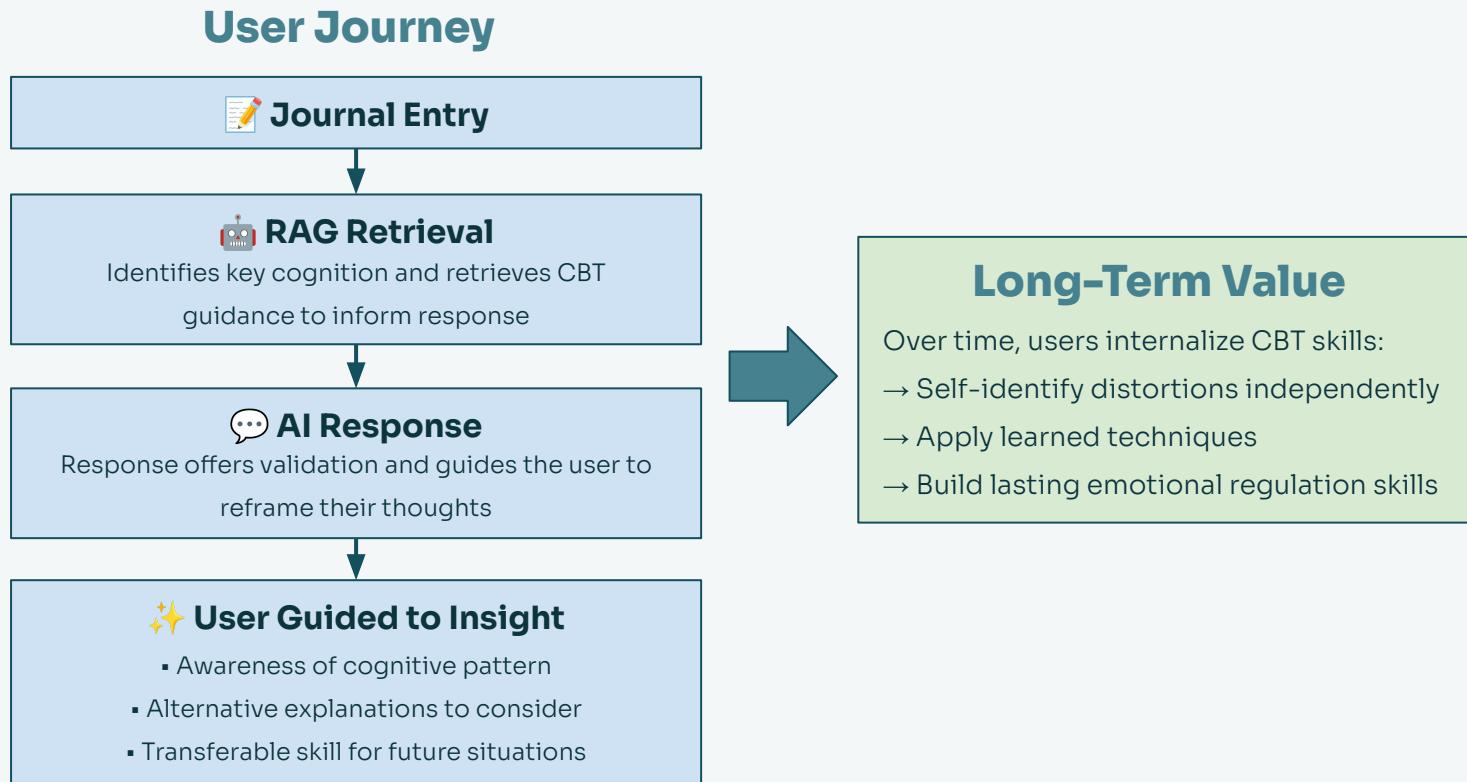
Sounds like quite a busy day! Frustration is normal when things get tough, especially when we're juggling multiple tasks. Can you think of any situations where your thoughts might have been exaggerated or all-or-nothing during those difficult moments?

Implicitly applies CBT through gentle questioning and normalization

Encourages self-reflection with open-ended questions

Warm and conversational. Feels more natural and human. Shorter.

From Model Output to User Action



Evaluation Methodology: Cognitive Therapy Rating Scale



LLM-As-A-Judge

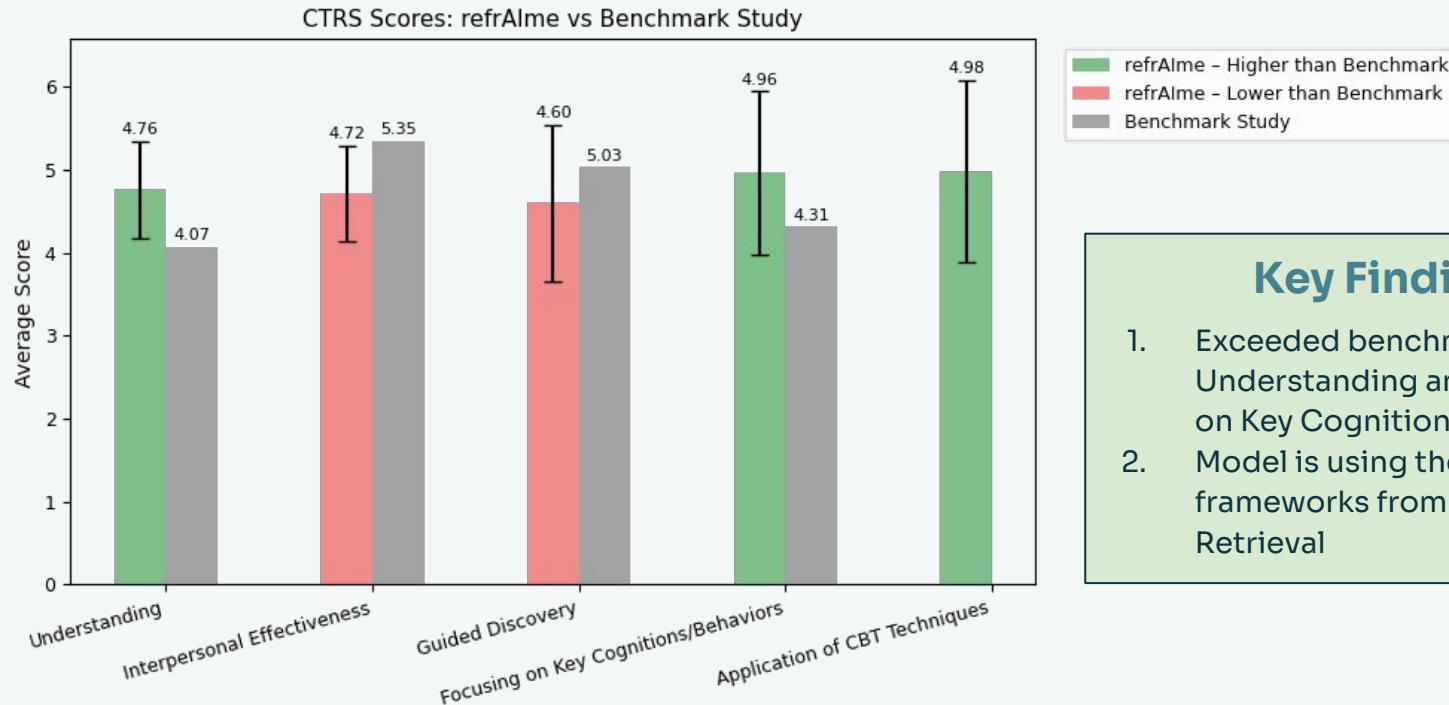
LLM scores 50 User/Journal Assistant dialogues from 1-6 using the Cognitive Therapy Rating Scale on:

- Understanding
- Interpersonal Effectiveness
- Guided Discovery
- Focusing on Key Cognitions
- Application of CBT Methods

Why This Metric:

1. Widely used in CBT therapy assessment
2. Measures therapeutic competence across multiple key dimensions
3. Allows comparison to other LLM-generated benchmarks from similar work

Evaluation Results: Cognitive Therapy Rating Scale



Key Findings

1. Exceeded benchmark study in Understanding and Focusing on Key Cognitions metrics
2. Model is using the CBT frameworks from the RAG Retrieval

Real-World Evaluation Results

Users	
26	
Entries	
54	
Feedback	
29	

How emotionally supportive and helpful was this session?

1 = Not at all 5 = Extremely

3.8

How well did this session help you reflect on your thoughts or feelings?

1 = Not at all 5 = Very Well

4.0

What could have made this session more helpful or relevant?

"Replies were consistent and polite"
"Feels like it would've just continued to ask questions forever"

Avg. Number of AI Exchanges per Session

2.9 turns

Avg. Entry Length

302 chars

Key Learnings:

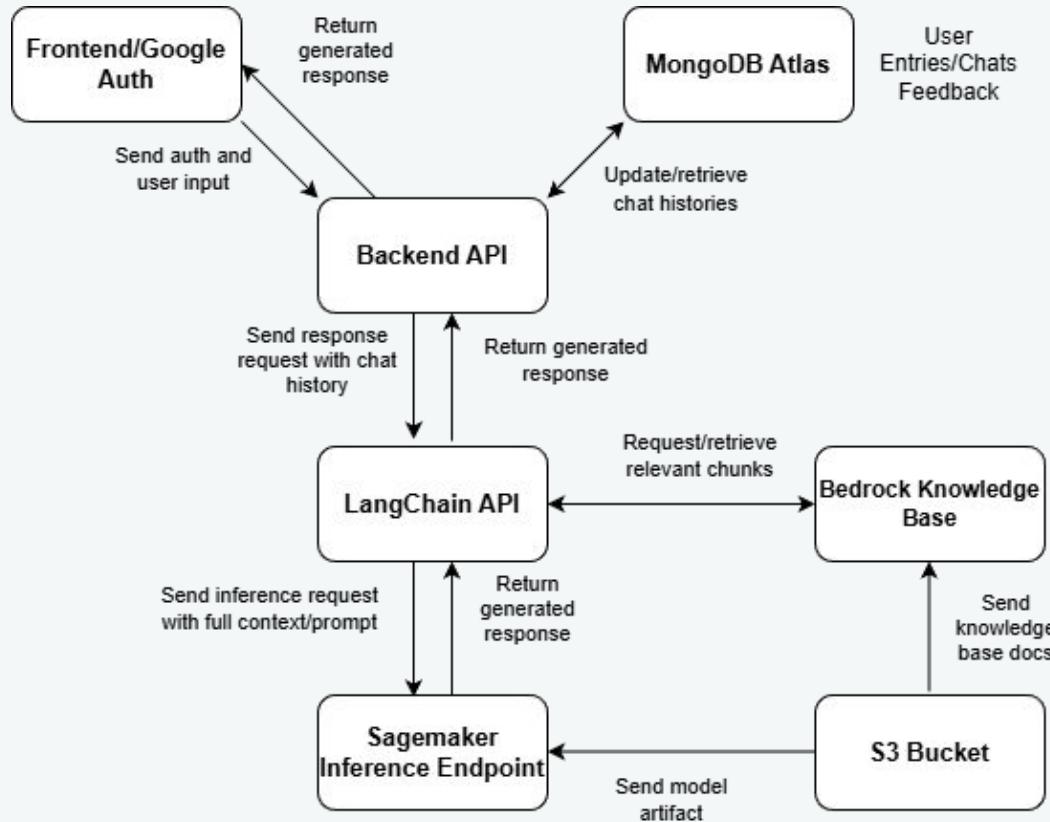
- Users find the app easy to use, AI asks good questions
- A gap between user expectations and app functionality

Key Technical Takeaways

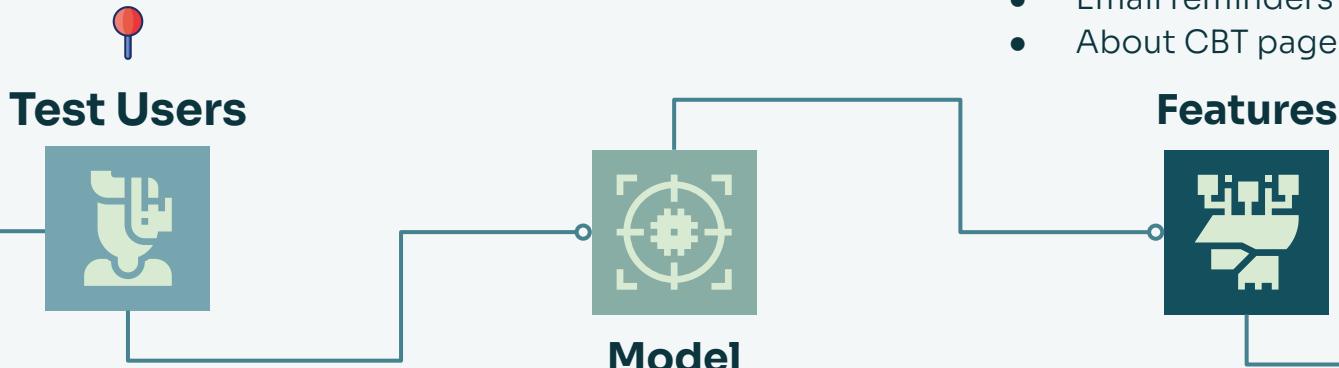
- **Rigid Input Schema**
 - Helps avoid hallucinations; Properly organize info in large prompts
- **RAG Document Selection**
 - Perspective matters, not just general subject matter
- **Model Matters!**

Model	Hallucination Frequency	Takeaways
Gemma 9B 	Low	High quality responses ; Reliable performance
Qwen 14B	Low (Colab), Medium (Langchain API)	Best quality responses; Inconsistent performance with Langchain AI schema
Mistral 7B	Medium	Decent quality responses; Frequent hallucinations
Llama 8B	High	Decent quality responses; Many hallucinations

Deployment



Roadmap



- Improve ability of users to ask AI questions
- Provide relevant worksheets & resources
- Allow users to highlight parts of their journal entry for AI to analyze



Our Mission



Our Mission:

Empower users to take small, daily steps toward better mental health through AI-guided journaling rooted in Cognitive Behavioral Therapy.



Goal:

Build a tool that feels less like an app and more like a supportive friend that helps users think, feel, and grow.

Thank You!



Contributions

Sineli: Backend

Hannah: Frontend

Brad: RAG Pipeline

Michael: Prompt Engineering, Evaluation

Shyam: LLM, Deployment