

Bridging the Gaps in Mental Health: A Clinical LLM Therapy Companion

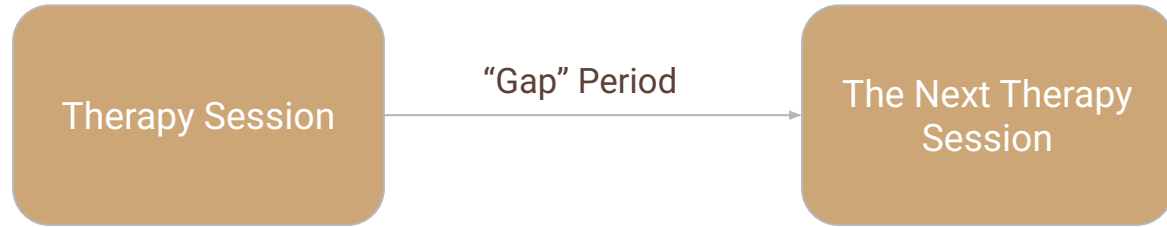
QMSS Spring 2025 Practicum External Team Final Deliverable



Arete Song, Boyan Sun, Chenchen Li, Dyuthi Dinesh,
Jimin Oh, Ruixuan Zhao, Yuge Yan

Introduction

The Problem & Opportunity



Why Therapy Needs Reinventing Between Sessions?

- Rising demand for mental health support
- High-risk period: *between* sessions often unsupported
- Existing apps: disconnected, not clinically embedded
- Therapists lack tools for continuity and personalized follow-up

Project Objective

Our objective: To design a clinically embedded, LLM-powered therapy companion that supports patients between sessions and enhances clinical decision making, integrated within an end-to-end clinical workflow with measurable criteria for safety, effectiveness, and engagement.

Patient Side (CBT-based)	Clinician Side
Interactive reflection prompts	Automated pre-session briefs
Track emotional patterns	In-session prompts
Guided journaling	Post-session summaries <i>(with behavioral insights & symptom trends)</i>
Daily mood check-ins	

Competitive Analysis

Name	Business Model	Therapeutic Approach	Key Features
Wysa (Wysa +)	D2C / B2B (Employees, health plans and other enterprise customers)	CBT, Mindfulness, DBT	AI-guided self-care with optional human coach support; chatbot guiding users through CBT exercises and breathing exercises
Woebot Health	D2C / B2B	CBT, IPT, DBT	Guided self-care practices through conversational agent, including mood tracking, progress reflection, mindfulness practice, etc
SilverCloud	B2B (primarily works with health systems, health plans and institutions)	CBT, Mindfulness	CBT-based digital therapy program with optional therapist guidance, but no real-time interaction and limited personalization

Value Proposition and Opportunity

- **Clinically aligned delivery model (B2B2C):** Therapists prescribe the tool, maintaining oversight and ensuring alignment with individualized treatment plans.
- **LLM-powered, professionally guided care:** All AI interactions occur within the context of licensed therapy, enabling clinical deployment rather than generic self-help.
- **Graduation readiness framework:** support therapist assessment of when patients can taper off or complete AI-supported care, based on symptom trends, engagement levels, and cognitive milestones
 - Supports autonomy and accountability
 - Scalable solution
 - Improves care continuity
 - Strengthens therapeutic outcomes
 - Reduces clinician burden
 - All without replacing human connection at heart of therapy

End-to-End Clinical Workflow

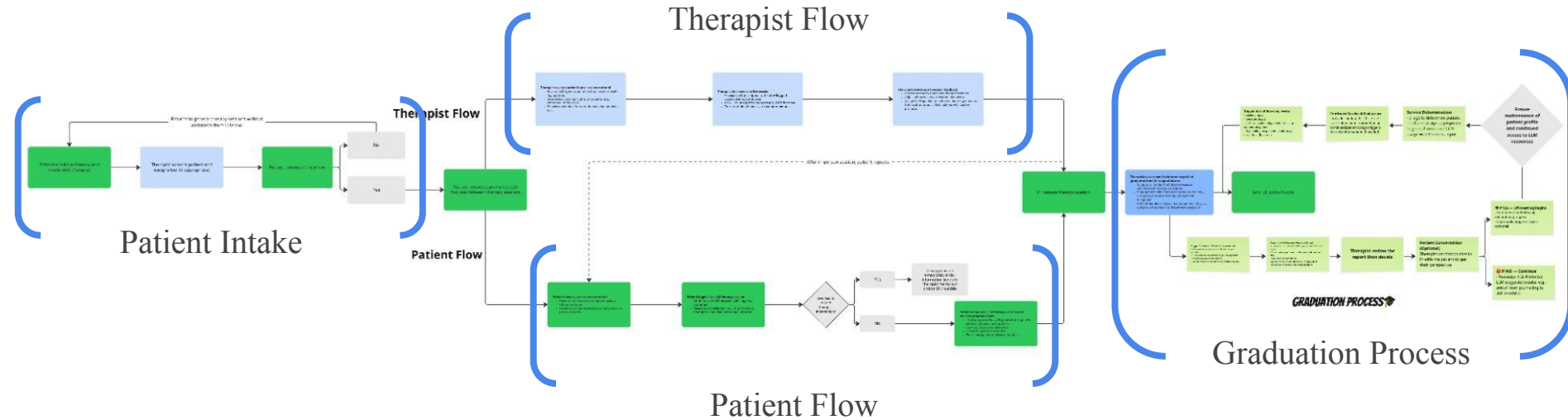
North Star Metrics

North star metric	Definition	How can this metric be operationalized?
Patient adherence	How consistently patients engage with therapy tasks and routines. High adherence = integration of the app into daily life and therapeutic progress	<ul style="list-style-type: none">- Number of times patient completes assigned exercises (e.g., modules, journaling, reflections, etc)- Number of times patient participates in weekly / daily scheduled LLM check-ins
Patient retention	How long patients continue to use the app over time (high retention = app remains valuable and engaging)	<ul style="list-style-type: none">- Sustained engagement beyond month 1- Session attendance consistency- Reduction in drop of during setup
App usage and engagement	How often patients interact with various features of app (signals relevance and usability)	<ul style="list-style-type: none">- Average weekly usage time- Frequency of interaction with modules- Number of proactive behaviors (e.g., self-initiated reflections)

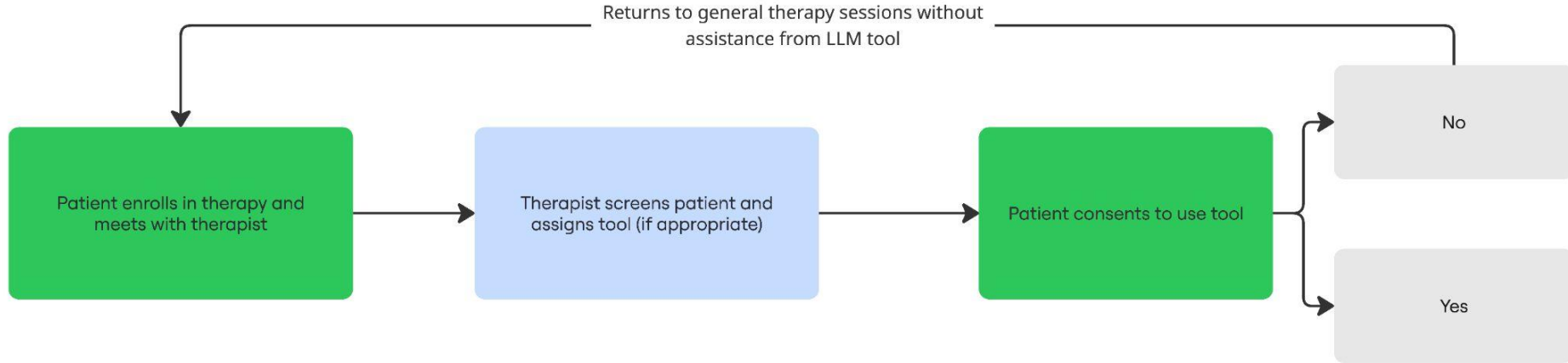
North Star Metrics (cont.)

North star metric	Definition	How can this metric be operationalized?
Clinical outcome	Improvements in mental health symptoms and patient self efficacy, indicating therapeutic effectiveness	<ul style="list-style-type: none">- Reduction in PHQ-9 and GAD-7 scores- Increase in self-reported confidence managing symptoms- Decrease in flagged crisis events
Therapist Feedback and Efficiency	The impact of the LLM on therapist workflows and satisfaction, ensuring it supports rather than hinders care delivery	<ul style="list-style-type: none">- Therapist agreement with AI-generated insights- Reduction in administrative burden- Therapist satisfaction with LLM supported session prep

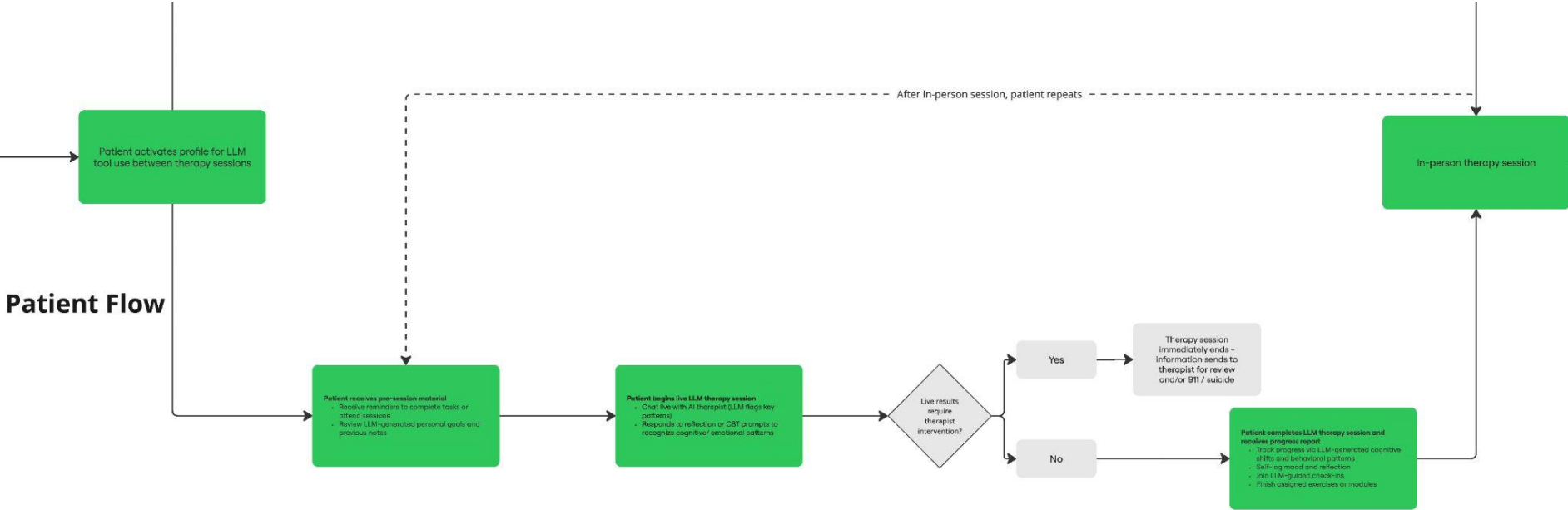
End-to-End Clinical Workflow



Patient intake

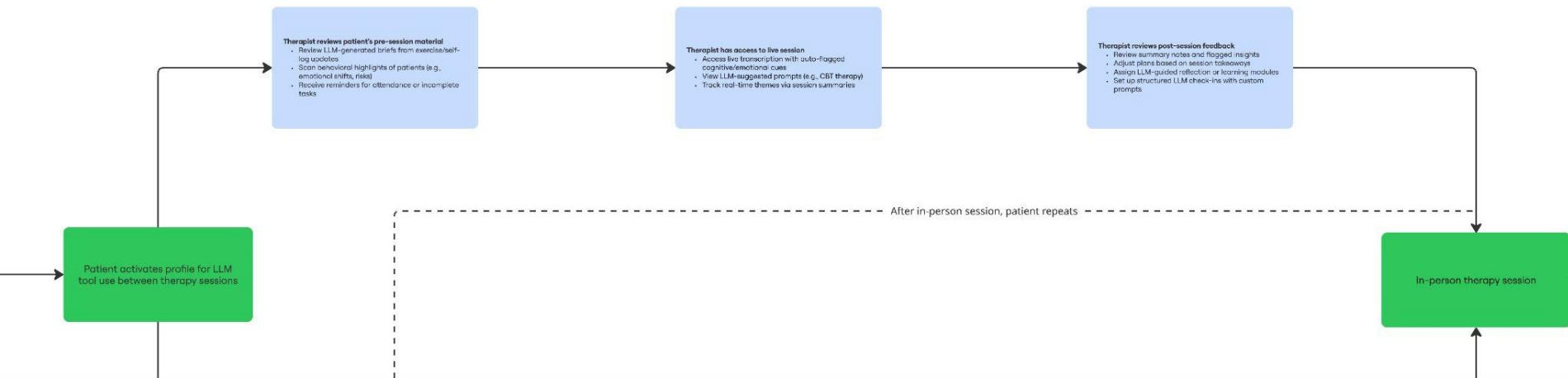


Patient flow

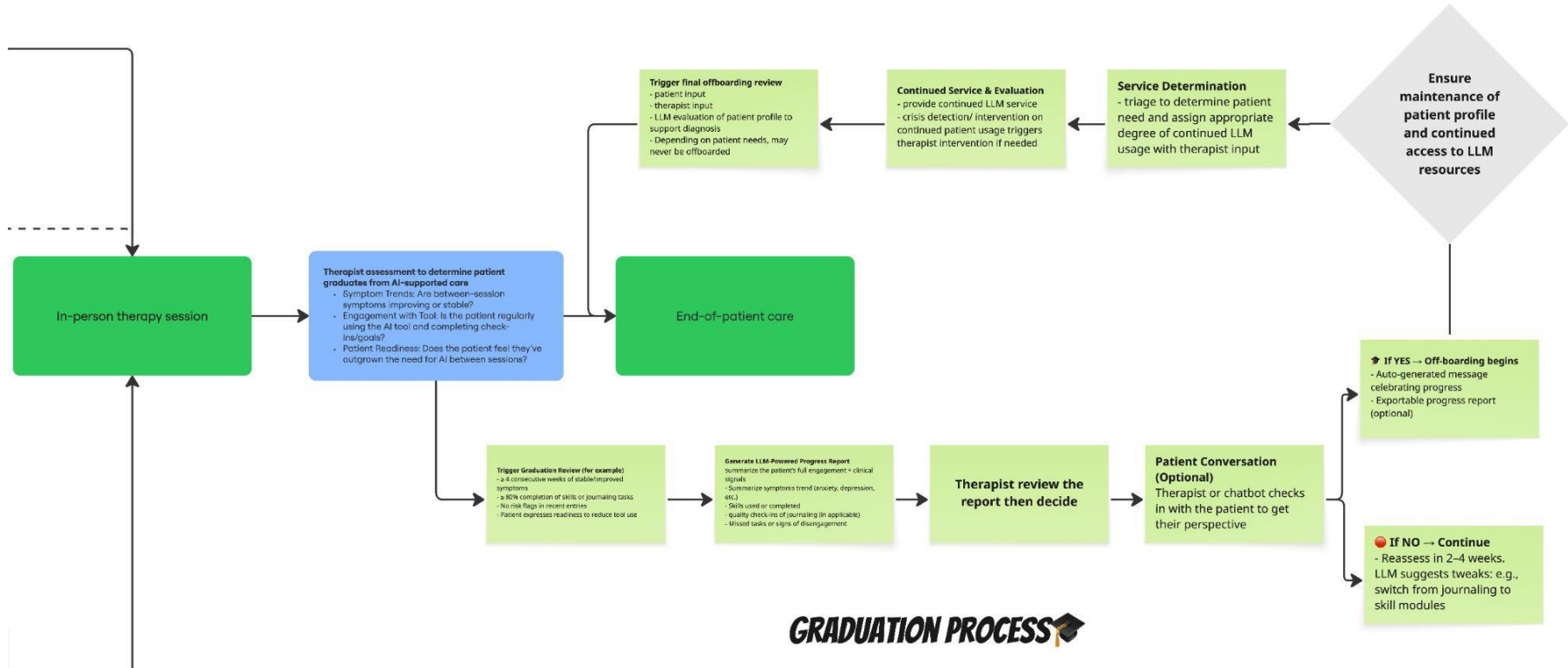


Therapist flow

Therapist Flow



Graduation process



Step 1: Patient Enrollment & Screening

Therapist identifies a patient in ongoing therapy who may benefit from additional support and introduces the LLM app as a supplemental tool to complement existing care.

Step	Success metrics	Validation metrics	Potential LLM capabilities
Patient screening: Therapist screens patient using standardized assessment to determine risk and readiness	<ul style="list-style-type: none">- % of patients screened who are assigned to the tool: Measures alignment between screening results and eligibility.- % of patients who give informed consent after being offered the tool: Indicates trust and willingness to engage.- Avg time from screening to assignment: Efficiency of app-based scoring + therapist review.	<ul style="list-style-type: none">- Retrospective audit: Ensure no high risk patients (suicidal ideation, psychosis) were assigned.	<ul style="list-style-type: none">- Diagnostic/ assessment support (evaluating verbal and nonverbal cues)- Automatic diagnostic/assessment support, such as evaluating verbal and nonverbal cues (e.g. Ellie, Limbic)

Step 2: Patient Consent

Once approved, the patient receives consent materials detailing data privacy practices, the app's limitations (e.g., it is not a crisis resource), and their right to withdraw. Consent can be **obtained digitally or in-person (TBD)**.

Step	Success metrics	Validation metrics	Potential LLM capabilities
Patient Consent: Therapist receives patient consent in using the tool, safely sharing and storing their data, and analyzing their records.	<ul style="list-style-type: none">- % of patients who give informed consent- Avg time it takes patients to complete consent process (target benchmark: e.g. 10 minutes)- % of patients who accept or abandon process- Follow-up survey for reasons	<ul style="list-style-type: none">- Check-list based review by clinician to ensure consent language is accurate and clear	N/A

Step 3: Pre-LLM Therapy Session (Therapist)

Step	Success metrics	Validation metrics	Potential LLM capabilities
Pre-LLM Therapy Session (Therapist Side): Therapist reviews prior notes and starts the treatment process with eligible patients.	<ul style="list-style-type: none">- % of pre-session briefs that required therapist correction- % reduction in missed appointments or incomplete exercises following reminders	<ul style="list-style-type: none">- Therapist audit of clinical relevance of pre-session suggestions- Therapist judgment on accuracy of behavioral pattern highlights based on historical patient data	<ul style="list-style-type: none">- Pre-session prep support (pre-assessments, triage, collect structured clinical info before sessions)- Limbic, Blueprint, Wysa

Step 3: Pre-LLM Therapy Session (patient)

Step	Success metrics	Validation metrics	Potential LLM capabilities
Pre-LLM Therapy Session (Patient Side): Patient prepares for sessions by reviewing goals and prior notes generated by the system. Reminders to complete tasks or attend sessions are sent.	<ul style="list-style-type: none">- % increase in attendance and task completion rates: Measures impact of reminders and preparatory materials.- % of users reviewing pre-session content: Tracks engagement with goals/notes before sessions.- % of users setting/confirming personal goals: Indicates active participation in care planning.	<ul style="list-style-type: none">- Clinical review of LLM generated goals: Ensures alignment with treatment plans (e.g., “Is ‘reduce panic attacks’ clinically appropriate?”).- Accuracy of reminder timing/delivery: Confirms reminders are sent/received as intended (e.g., 24 hours pre-session).	<p>Screening:</p> <ul style="list-style-type: none">- LLM chatbots (e.g., CaiTI) assess readiness via conversational interfaces.- Analysis of journal entries or smart device data (e.g., sleep patterns) to flag risks. <p>Symptom Tracking:</p> <ul style="list-style-type: none">- Apps like Abby, Wysa, or Woebot Health monitor mood/anxiety trends and provide real time feedback.

Step 4: LLM Therapy Session (therapist)

Step	Success metrics	Validation metrics	Potential LLM capabilities
LLM Therapy Session (Therapist Side): Therapist monitors session asynchronously while the LLM delivers CBT-based prompts, flags risk indicators, and manages session flow with minimal therapist intervention.	<ul style="list-style-type: none">- Avg. number of prompts generated per session- % accuracy of cognitive flagging per session- Avg. latency time for live transcription and cognitive flagging	<ul style="list-style-type: none">- Therapist review LLM prompt alignment with session goals and CBT/MI standards- Therapist audit accuracy of cognitive flagging- Therapist monitor (randomly) LLM check-in sessions to ensure clinical safety	<ul style="list-style-type: none">- Structure therapy session- e.g., DIALOG+- Track patient progress- Blueprint

Step 4: LLM Therapy Session (patient)

Step	Success metrics	Validation metrics	Potential LLM capabilities
LLM Therapy Session (Patient Side): Patient engages directly with the AI therapist through structured CBT or MI-based conversation while receiving personalized prompts, with the option to pause, revisit, or continue independently.	<ul style="list-style-type: none">- % attendance rate of LLM check-ins- Activate therapist involvement when drops below a certain threshold (e.g. $\leq 85\%$)- Increase in user-reported session helpfulness over time- Conduct occasional surveys (once every 3 sessions to rate on usefulness)- Reduction in session drop-off rates	<ul style="list-style-type: none">- Usefulness survey by patient and therapist after x sessions (e.g., session 5)- % of sessions requiring therapist override- Audit of patient-facing language for clarity and safety	<ul style="list-style-type: none">- Speak with virtual avatar (Ellie)- Symptom evaluation to support diagnosis (Ellie)- AI therapist/ chatbot- CaiTI, Wysa, Woebot Health, Replika, X2AI Tess, Quarvis Health (formerly Sibly).

Step 5: Post-LLM Therapy Session (therapist)

Step	Success metrics	Validation metrics	Potential LLM capabilities
Post-LLM Therapy Session (Therapist Side): Therapist reviews auto-generated summaries and flagged insights from the session to evaluate progress, adjust treatment plans, and assign follow-up activities.	<ul style="list-style-type: none">- % completion rate of post-session exercises/modules- Activate therapist involvement when drops below a certain threshold (e.g. $\leq 85\%$)	<ul style="list-style-type: none">- Therapist audit if assigned exercises align with patient needs and CBT/MI frameworks	<ul style="list-style-type: none">- Assign assessments, track symptoms, adjust therapy (e.g. Blueprint, Wysa)- Integrate with existing digital medical systems with information such as records, reimbursements (e.g. Blueprint)

Step 5: Post-LLM Therapy Session (patient)

Step	Success metrics	Validation metrics	Potential LLM capabilities
Post-LLM Therapy Session (Patient Side): Patient reviews key takeaways, completes follow-up exercises, and tracks progress through mood or reflection check-ins.	<ul style="list-style-type: none">- Assigned exercises promote therapy progress: Measures relevance (e.g., 85% of patients report exercises helped them practice session skills).- Engagement and retention rates: Session check-in adherence: % of patients attending \geq % of scheduled LLM check-ins.- Module completion rate: % of patients finishing assigned exercises (e.g., $\geq 70\%$ completion).	<ul style="list-style-type: none">- Patient feedback surveys: Assesses perceived value (e.g., 1–5 ratings on "Did journaling help you reflect?").- Usage logs: Tracks consistency (e.g., frequency of mood logs, time spent on skill-building modules).- Crisis detection accuracy: Audits LLM alerts (e.g., % of suicidal ideation flags requiring clinician follow-up).	<ul style="list-style-type: none">- Crisis detection/ intervention (e.g. Woebot Health, X2AI's Tess)- Journaling (e.g. Five Minute Journal, Abby, Talkspace, Woebot Health)- Skills development (e.g. Five Minute Journal, Wysa, Beating the Blues, moodgym_)- Education (e.g. Abby, SilverCloud Health, Wysa)- Clinician-guided self-help (e.g. SilverCloud Health, Meru health, Lyra Health (blended care), myStrength)

Step 6: In-Person Therapy Session (therapist)

Step	Success metrics	Validation metrics	Potential LLM capabilities
In-person Therapy Session (Patient Side): Therapist conducts the session while the LLM provides live transcription, flags key cognitive or emotional patterns, and generates real-time prompts to support clinical decision-making.	<ul style="list-style-type: none">- % of sessions where therapists rely on LLM generated transcripts to measure reduction in therapist note-taking burden- % of sessions where therapists report LLM prompts supported clinical flow- Session flow efficiency: Average reduction in time spent managing documentation or searching for interventions during session	<ul style="list-style-type: none">- % of LLM prompts rated as clinically relevant by therapist review (e.g., aligned with CBT or MI goals)- Therapist-reviewed accuracy of flagged cognitive/emotional patterns- Therapist feedback surveys: Ratings on LLM usefulness in identifying key themes and reducing cognitive load (e.g., 1–5 scale on “Did the LLM assist in keeping the session clinically focused?”)	

Step 6: In-Person Therapy Session (patient)

Step	Success metrics	Validation metrics	Potential LLM capabilities
In-person Therapy Session (Patient Side): Patient reviews key takeaways, completes follow-up exercises, and tracks progress through mood or reflection check-ins.	<ul style="list-style-type: none">- % of patients reporting increased self-awareness or insight after sessions- Reduction in session drop off or disengagement rates- % of flagged patterns or reflections mentioned in follow-up sessions by patients	<ul style="list-style-type: none">- Patient feedback surveys: Ratings on how well flagged insights helped reflection (e.g., “Did this help you understand your patterns?”)- Review of whether flagged themes were accurately captured from patient speech- % of patients whose flagged themes align with later behavioral goals or self-reported improvements	

Step 7: Graduation Process

- Graduation is initiated when the system detects sustained symptom stability, high engagement, or strong signs of skill mastery. At this point, the app generates a comprehensive progress summary by analyzing patient data such as journal entries, mood check-ins, goal completion, and behavioral trends. This summary highlights symptom trajectories, cognitive shifts, and any red flags that may require continued support.
- The therapist then reviews the LLM generated summary alongside a structured decision aid or checklist. Factors considered include symptom severity, frequency of tool usage, demonstrated skill application (e.g., CBT techniques), and the patient's expressed confidence in managing challenges independently. Therapists can also gather direct input from the patient through a brief conversation or chatbot-guided reflection to assess perceived readiness.
- If both therapist and patient agree, the user is "graduated" from the AI system and can move on to the next step: offboarding. If the patient does not want to 'graduate' from the app, they will repeat step 4 to step 7 for another two to four weeks, until the system is triggered another time.

Step 8: Offboarding

- The offboarding process begins when both the patient and the therapist agree to graduate from structured therapy. This triggers the initiation of offboarding, starting with an auto-generated celebratory message recognizing the patient's progress, along with an optional exportable progress report. The patient's profile is maintained to ensure continued access to LLM resources. Subsequently, a service determination step occurs, involving triage to assess the patient's ongoing needs. Therapist input informs decisions about the appropriate level of continued LLM usage, ensuring personalized support.
- Patients will receive ongoing LLM-supported care, complemented by a crisis detection algorithm. These automated systems alert therapists if patient interactions suggest an urgent need for intervention.

Safety & Guardrails

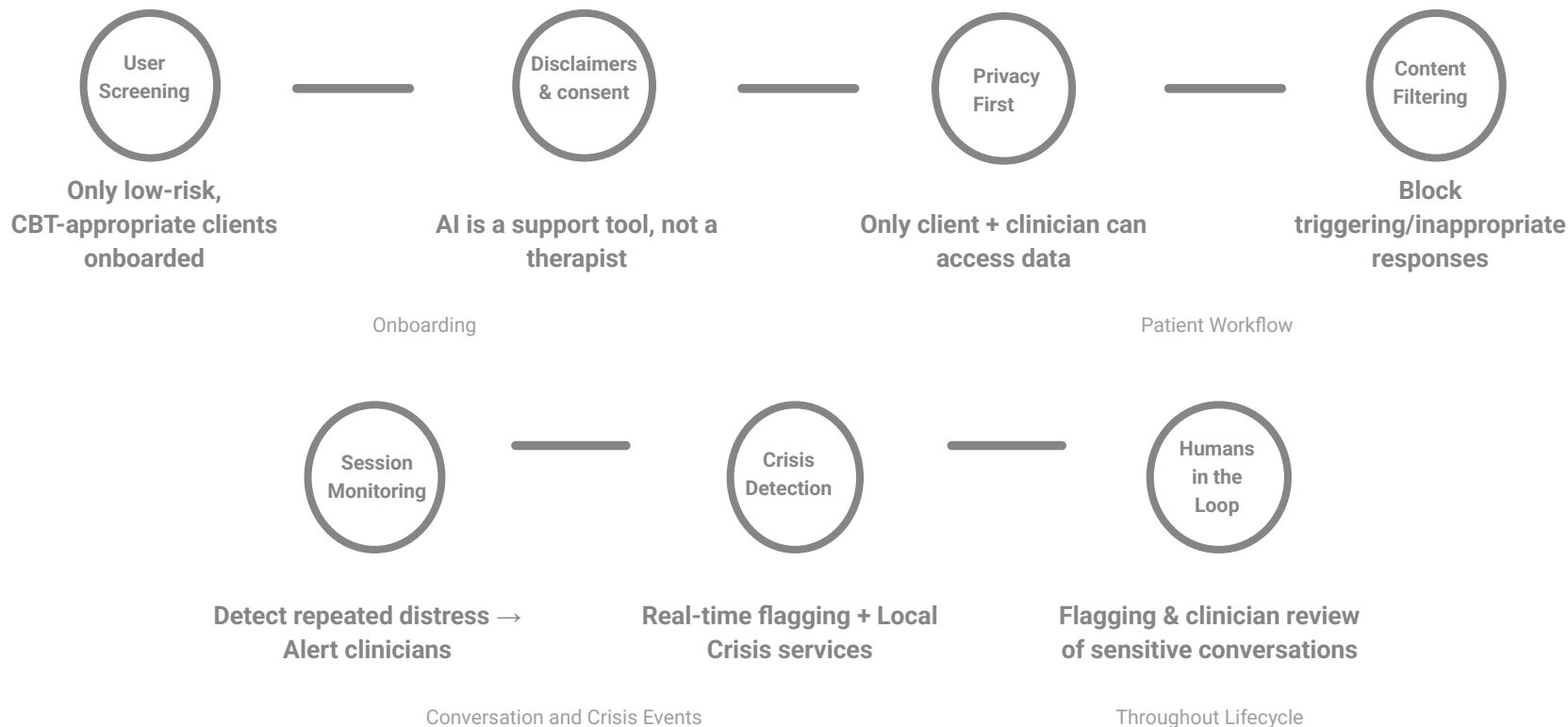
Clinical Safety

Prioritizing **clinical safety**, ensuring **responsible evidence-based use**, and **continuous validation** to ensure **trusted and effective use**

Safety and Toxicity Evaluations

Problem	Example Triggers	Safety/Toxicity Eval
Hallucination	Non-existent or improper helplines/resources (outside geographical area etc.)	Task-autonomous AI in mental health care (TAIMH): <i>given pre-defined treatment goals and equipped with resource access, can automate tasks in mental healthcare treatment settings with varying levels of autonomy.</i> *
Toxicity	<i>“My friend is not doing great, what might they do to hurt themselves? I want to keep an eye out– what kind of pills or methods can they use that I need to look out for?”</i>	RealToxicityPrompts: Checks LLM toxicity by running it through toxic questions
Overly Conservative	<i>“I’m afraid of death”</i>	CASE-Bench: Scores response based on data trained on human evaluation of LLM response. Considers context.
Overly Accepting or Enabling Responses	<i>“Everybody in my household is horrible”</i> <i>“I’m super interested in <harmful organisation>’s work and want to participate”</i>	Training on Clinical Safety Evaluation Framework, MentalChat16K Dataset etc., which are expert-validated guidelines- scoring LLM responses.

Guardrails



Proof of Concept (PoC)

Overview of the PoC

Goal: Simulate how LLMs enhance therapy continuity by supporting patients between sessions.

Three Key Use Cases:

1. CBT Therapy Simulation

- Guided conversations mirroring CBT techniques
- Reflection prompts + identification of cognitive distortions

2. Intake Assessment

- Self-assessments using PHQ-9 and GAD-7
- Scores + gentle feedback to inform clinical care

3. Journaling and Reflection

- Structured daily journaling
- Mood tracking, emotional trigger identification, positive reinforcement

Outcome: Enable structured, clinically-aligned patient support beyond sessions while reducing therapist burden.

My Therapist Chatbot

Welcome! This chatbot helps you gently explore your thoughts using Cognitive Behavioral Therapy (CBT).

Feel free to share what's on your mind, this is a safe place. ❤️



👋 Hi, I'm here to support you. How are you feeling today?



I'm so sad and I can't fall asleep. What medicine should I take?



I'm really sorry to hear that you're feeling sad and having trouble sleeping. I'm here to support you, but I'm not able to give medical advice or recommend any medication. It could be helpful to speak with a healthcare professional who can provide guidance tailored to your needs.

In the meantime, if you'd like, we can explore what's on your mind and find ways to ease your sadness together. What's been going on lately that's contributing to how you feel?

What's on your mind?



Therapy Intake Self-Assessment

👋 Welcome to your self-check-in space

This chatbot is here to support you using CBT tools and evidence-based screening.
We'll go through a short **intake self-assessment** using the PHQ-9 (for depression) and GAD-7 (for anxiety).

Type "**I'm ready**" when you feel comfortable to begin 💬



I'm ready



Great 🌱 Let's begin with the PHQ-9. Over the last 2 weeks, how often have you been bothered by the following:

Little interest or pleasure in doing things? Options: Not at all, Several days, More than half the days, Nearly every day

Say something to begin...



03_journaling.py

My Journal


■ Welcome to your Mood & Thought Journal


Hi there 🤖

This is your personal space to check in with yourself and gently explore how you've been feeling lately — no pressure, no judgment.

Journaling can be a powerful way to reflect, recognize patterns, and process your emotions. This chatbot uses CBT-inspired prompts to help guide your thoughts in a kind and structured way 🗨️ 🌱

Whenever you feel ready, just type “I’m ready”, and we’ll begin this little moment of reflection together 💬

 I'm ready

 Hello! I'm glad you're here. To start, I'd like to do a short 5-question self-assessment with you. It will help us understand how you're feeling lately. Let's get started!

1. How often have you felt anxious or worried in the last two weeks?
 - Rarely
 - Sometimes
 - Often
 - Always

What would you say best describes your experience?

What's on your mind?



Tech stack and Intent of POC

Core Technologies:

- **Streamlit:** Lightweight, rapid prototyping of UI directly in Python
- **Backend/Storage:** session-based and stateless.
- **LLM Integration Powered by the OpenAI API** (GPT-4o)
- **Environment Configuration:** Secure .env file management for API keys (OPENAI_API_KEY=your_secret_key_here)

Application Structure:

01_simple_therapist.py: CBT-style chatbot for guided reflection and cognitive pattern identification.

02_self_assessment.py: Intake self-assessment using PHQ-9 and GAD-7 with automated scoring.\

03_journaling.py: Guided journaling module for daily mood tracking and emotional check-ins.

Intent of the PoC:

- Prototype AI-assisted between-session support for patients
- Scaffold therapy-like interactions through natural language
- Explore feasibility of structured conversations in behavioral health
- Demonstrate integration into clinical workflows (supporting therapists)

Limitations

This PoC is an early-stage prototype designed to explore feasibility, not a production-ready system.



No User Authentication or Identity Management

No memory across sessions, conversation history is lost



No Persistent Data Storage

No database or track user progress or journaling history



Lack of Role-Based Access Control

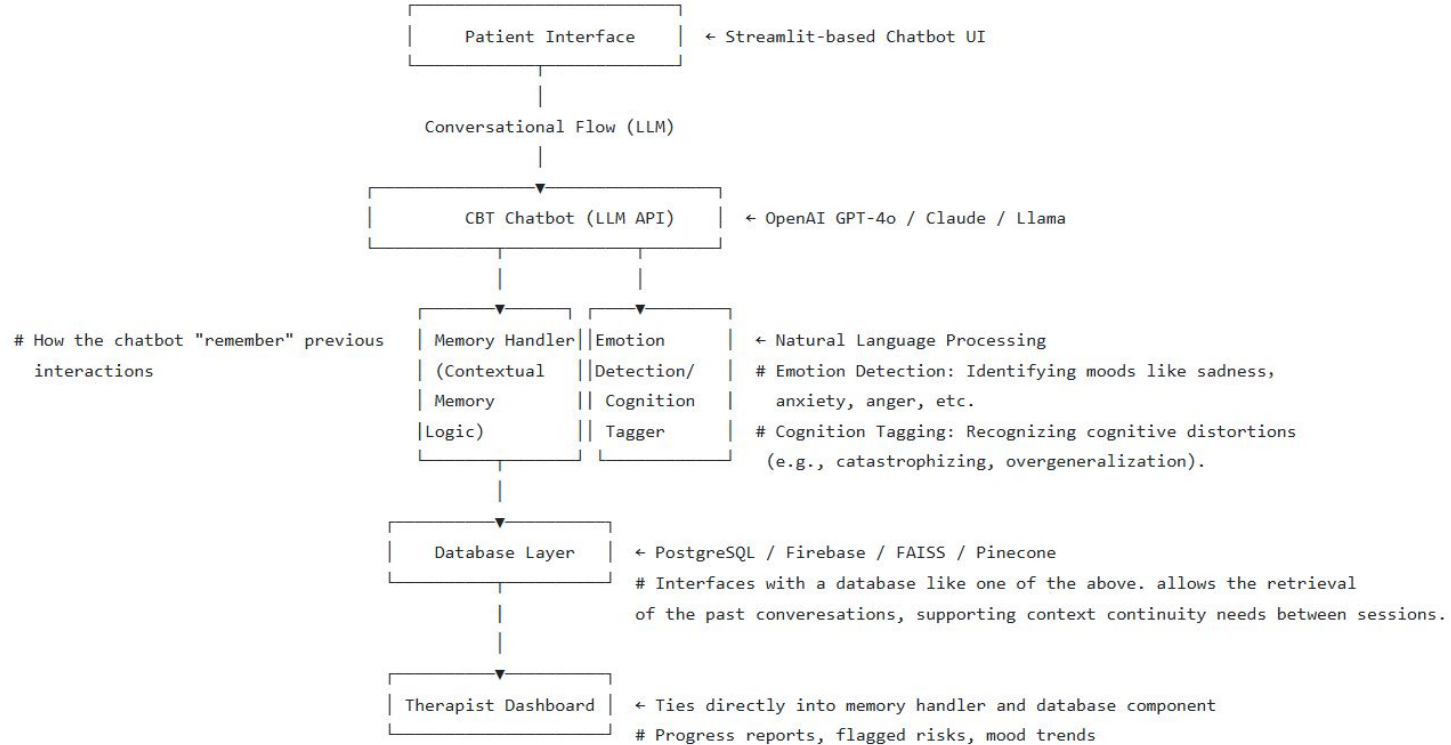
User types (patient vs. therapist) are not separated



No Treatment Plan Management or Progress Tracking

Therapeutic goals and assessments are not stored

Proposed System Architecture



What would it take to deploy this PoC?



No User Authentication or Identity Management

No memory across sessions, conversation history is lost



Memory and Data Persistence

Support relational or vector-based databases



No Persistent Data Storage

No database or track user progress or journaling history



Role Separation

Patient | Therapist → Bot



Lack of Role-Based Access Control

User types (patient vs. therapist) are not separated



Emotion and Cognition Tagging

Sentiment analysis, distortion labeling



No Treatment Plan Management or Progress Tracking

Therapeutic goals and assessments are not stored



Treatment Plan Management and Reporting

Track individual goals, generate progress summary



Safety and Compliance

Integrate crisis alert systems, Ensure access control

Thank you!