A Framework for ECA-Based Psychotherapy

Santosh V. Patapati
Dept. of HCI
Cyrion Labs
santosh@cyrionlabs.org

Trisanth Srinivasan
Dept. of HCI
Cyrion Labs
trisanth@cyrionlabs.org

Himani Musku
School of Computer Science
Carnegie Mellon University
hmusku@andrew.cmu.edu

Amith Adiraju
Dept. of HCI
Cyrion Labs
aadiraju@cyrionlabs.org

***Abstract*—We present Tessa, an autonomous framework for structured bi-weekly psychotherapy intervention. Tessa alternates between screening sessions that use motivational interviewing and follow-up CBT sessions to help users identify, challenge, and reframe negative thoughts. Using text analysis with Large Language Models (LLMs), an Embodied Conversational Agent (ECA) adapts its dialogue and synchronizes non-verbal cues in real time. The framework incorporates a therapist-matching pipeline and clinical report generation to enable seamless hand-off to professional care. Taken together, Tessa provides accessible, stigma-free, and data-driven mental health support that improves personalized care and reduces barriers to therapy.**

***Index Terms*—Human-Computer Interaction, Psychotherapy**

I. Introduction

Many with mental health challenges are underserved due to limited access and social stigma. In response, digital therapy solutions are emerging as scalable alternatives. ECAs are of particular interest due to their ability to simulate human empathy. They mitigate stigma's impact by providing a non-judgmental, accessible interface. However, existing systems struggle to deliver longitudinal support, and past autonomous ECAs have not combined context-aware speech with coordinated non-verbal cues, limiting empathetic effectiveness.

We introduce Tessa, a therapeutic system that delivers structured bi-weekly psychotherapy. Tessa uses a dual-session cycle: an initial screening with Motivational Interviewing (MI) and a follow-up Cognitive Behavioral Therapy (CBT) session to identify, challenge, and reframe negative thoughts. Leveraging LLMs for real-time text analysis, Tessa adapts its dialogue and monitors 25 mental health dimensions to improve well-being. Its autonomous ECA delivers empathetic responses with synchronized gestures. Tessa also offers a streamlined pathway to professional care through therapist matching and clinical report generation. Our demonstration showcases Tessa's scalable approach to long-term mental health support.

II. Methodology

A. Bi-Weekly Interaction Structure

Tessa follows a bi-weekly cycle of Motivational Interviewing (MI) followed by a Cognitive Behavioral Therapy (CBT) session. MI and CBT complement eachother well [1].

Before the initial MI screening session, the user completes a 25-item questionnaire based on the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5), the clinical standard that outlines diagnostic criteria for mental disorders. The questionnaire combines the PHQ-8, GAD-7, and PDS [2] to cover a wide range of mood disorders and symptoms. Questionnaire responses are passed as context to Tessa during future sessions via its running state.

Fig. 1. Sample frame from a Tessa session. The ECA maintains a neutral listening position as the user speaks. Visual features are tracked in real-time.

In the first session (Screening), the ECA conducts a semi-structured interview using MI (Figure 2) to assess the user's mental state and boost motivation for change. To motivate users, the ECA must build rapport, and it does this by engaging in small talk and providing a transparent introduction.

Audio and video recordings, the session transcript, and prior questionnaire responses are input into a quad-modal transformer model [3] that predicts the user's well-being across 25 mental health dimensions, such as "Restless or Unsatisfying Sleep" and "Significant Weight Changes."

In the second bi-weekly session (Intervention), held a few days after screening, the ECA and user conduct a CBT session as a follow-up to MI. This session targets dimensions flagged during screening or by the ML model. CBT follows three steps: identify, challenge, and replace negative thoughts with constructive ones (Figure 3). The goal is for users to adopt more constructive thoughts in day-to-day life and thus face fewer mental health challenges. For this reason, we expect users' well-being across the 25 dimensions to improve over time. By session end, the user formulates a balanced perspective and actionable coping strategy. Each session thus builds on the last, creating a continuous therapeutic loop.

B. Multimodal Interaction Pipeline

Modeled after GenECA [4], each time the user presses the speaking button, an "Utterance Segment" is recorded. Audio and video signals are analyzed by lightweight analyzers to detect distress, while Automatic Speech Recognition (ASR) transcribes speech for text validation. Combined

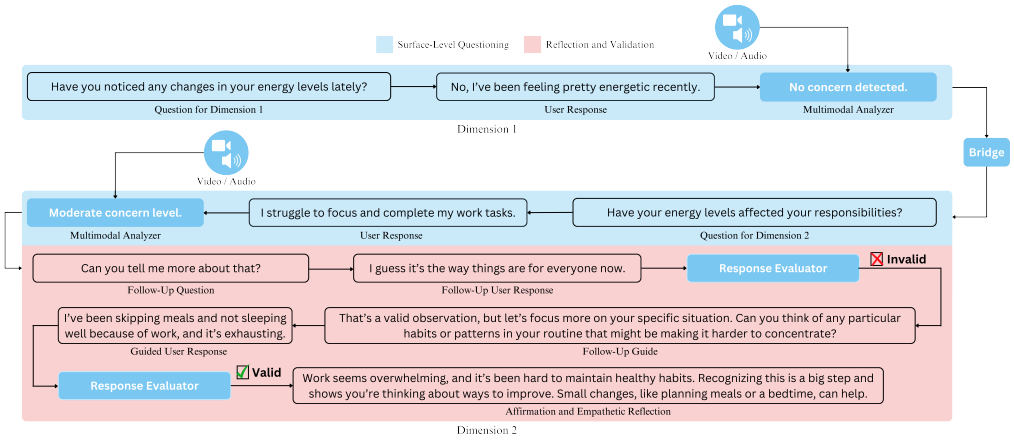

Fig. 2. Diagram of Tessa's MI process during screening. Tessa cycles through 25 mental health dimensions by asking surface-level questions. Lightweight classifiers analyze audio and video to detect stress, while an LLM performs chain-of-thought analysis to identify concerns. When a concern is identified, Tessa enters reflection-validation, asking follow-up questions for deeper insight. If follow-up responses are unconstructive, a guide intervenes. Conversational bridges transition between dimensions for coherent dialogue throughout the session.

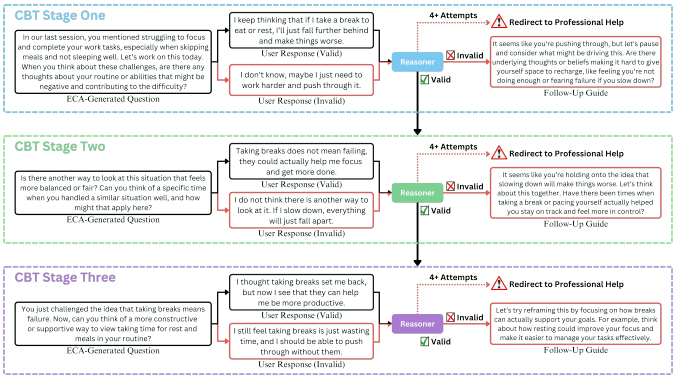

Fig. 3. Diagram of CBT dialogue pipeline. Tessa cycles through dimensions flagged in MI screening and generates personalized questions using previous session context. User responses are analyzed for validity, and if invalid, a follow-up guide is offered (red arrows indicate invalid response path). After four unsuccessful prompts, the user is redirected to professional support.

multimodal and text analysis generates an ECA response using an LLM (MentalLLaMA-7B-chat [5]), which directs synchronized hand gestures and voice modulation. Real-time audio segments trigger appropriate backchannel behaviors. The resulting lip-sync animations, audio, and nonverbal cues are rendered in a Unity 3D environment. To our knowledge, this is the first ECA framework for psychotherapy capable of delivering real-time empathetic speech and well-timed gestures without human control.

C. Therapist Matching

Tessa is not meant to operate in isolation, as the client-clinician relationship is crucial. It bridges users to traditional therapy with automated therapist-matching and clinical report generation. After using Tessa, users can opt to be matched with a practitioner. The system uses a database of 15,000 Texas licensed counselors, updated weekly via a web scraper. A matching algorithm integrates geographic proximity, insurance compatibility, age specialty alignment, and text-based specialty matching using BERT to yield composite scores (0–1.0) for each therapist. Users then choose one of their top matches.

D. Clinical Reporting & Hand-Off

When handing off to a therapist, Tessa generates a clinical report summarizing the user's interaction history and mental health indicators. This intake-style report covers three domains: Presenting Problems, Behavioral Health History, and Strengths and Risks, plus quantitative trends across 25 dimensions. This streamlined report provides the therapist with a comprehensive view of the user's state for faster handoff.

III. Conclusion & Demonstration

Tessa is a scalable, autonomous psychotherapy system that delivers structured digital care and facilitates professional support. Its accessible, stigma-free design integrates AI-powered dialogue with traditional treatment, making it applicable in diverse settings such as clinics, schools, and workplaces. Our ongoing work focuses on developing a mobile application for upcoming clinical trials to evaluate the framework.

For the demo, we present a simplified version of the Tessa ECA (Figure 1). Attendees interact at a high level, sharing general feelings (e.g., happy, frustrated) without revealing personal details that may make them uncomfortable. Visualizations will display the ECA's real-time thought process, tracked multimodal signals, distress markers, and more.

References

[1] C. L. Randall and D. W. McNeil, "Motivational interviewing as an adjunct to cognitive behavior therapy for anxiety disorders: A critical review of the literature," *Cognitive and Behavioral Practice*, vol. 24, Aug 2017.

[2] E. Brinkmann et al., "Psychometric evaluation of a screening question for persistent depressive disorder," *BMC Psychiatry*, vol. 19, Apr 2019.

[3] Santosh V. Patapati, "Integrating large language models into a tri-modal architecture for automated depression classification on the daic-woz," arXiv preprint arXiv:2407.19340, 2024.

[4] Santosh Patapati and Trisanth Srinivasan, "Geneca: A general-purpose framework for real-time adaptive multimodal embodied conversational agents," arXiv preprint, 2025.

[5] K. Yang et al., "Mentallama: Interpretable mental health analysis on social media with large language models," in *WWW '24*, May 2024.