LLM and Knowledge Graph-Based System for Reminiscence Therapy

Omar Hazzaa*, Youssef Zaghloul*, Ahmed Osman*, Amr Shaarawy*, Abdallah Khaled*, Ziad Elshaer, Ensaf Mohamed
*Artificial Intelligence, Nile University, Egypt
O.Adel2140@nu.edu.eg, Y.Amro2150@nu.edu.eg, A.Emad2152@nu.edu.eg,
A.yasser2149@nu.edu.eg, a.Khaled2178@nu.edu.eg

***Abstract*—This paper presents a novel therapeutic dialogue system that integrates knowledge graph architectures with large language models to provide personalized mental health support. Our system addresses the limitations of conventional chatbot-based therapeutic interventions by implementing a sophisticated entity-relationship framework that maps user utterances to clinical entities through pattern detection rather than simple keyword matching. A key innovation is our bidirectional knowledge graph-mediated reasoning pipeline, wherein symptom patterns and behavioural indicators are first extracted from patient dialogues and mapped to a clinical knowledge graph structure, after which the system leverages these graph-derived clinical insights to generate contextually appropriate therapeutic responses from corresponding intervention clusters. The system incorporates multiple therapeutic approaches including Cognitive Behavioural Therapy (CBT), Psychodynamic Therapy, and Reminiscence Therapy, dynamically selecting appropriate strategies based on detected emotional states. Additionally, our memory graph visualization component creates temporal and semantic connections between extracted life events, enabling more contextualized therapeutic responses. To enhance engagement, particularly for younger demographics, we implement a character-based role-playing mechanism with predefined therapeutic personas. The system employs Retrieval-Augmented Generation (RAG) to incorporate user-provided documents and images into the therapeutic dialogue, creating a multimodal experience that enriches the reminiscence therapy process. This research contributes to the field by demonstrating how knowledge graph-based relationship modelling combined with contextual memory management can improve the effectiveness of AI-assisted mental health interventions.**

***Index Terms*—Knowledge Graphs, Large Language Models (LLMs), Mental Health Informatics, Therapeutic Chatbots, Reminiscence Therapy, Cognitive Behavioural Therapy, Natural Language Processing, Graph-Mediated Reasoning, Retrieval-Augmented Generation (RAG), Multimodal AI, Character-Based Role-Playing, Mental Health Support Systems, Personalized Therapeutics, Emotion Detection, Knowledge-Enhanced Language Models.**

I. Introduction

Mental health disorders represent a significant global healthcare challenge, affecting approximately 970 million people worldwide according to the World Health Organization's 2019 data. However, the COVID-19 pandemic has substantially worsened this crisis, with initial estimates showing a 26% increase in anxiety disorders and a 28% increase in major depressive disorders in 2020 alone. Despite this growing prevalence, there exists a substantial treatment gap, with 76-90% of people with severe mental disorders in low and middle-income countries receiving no treatment. Traditional therapeutic interventions face scalability challenges due to the limited availability of trained mental health professionals, geographic barriers to access, and the prohibitive costs associated with long-term care.

Recent advances in artificial intelligence, particularly in natural language processing, have sparked interest in developing conversational agents for mental health support. These systems offer several potential advantages: they are accessible 24/7, can be deployed at scale with minimal marginal cost, and may reduce the stigma associated with seeking mental health treatment. Early implementations of therapeutic chatbots have shown promise in delivering structured interventions such as cognitive behavioural therapy for mild to moderate conditions.

However, current therapeutic conversational agents face significant limitations that hinder their clinical efficacy. Most existing systems rely on simplistic pattern-matching algorithms or template-based responses that fail to capture the nuanced clinical presentation of mental health conditions. They typically operate with a limited understanding of psychological context, lacking the ability to recognize complex symptom patterns or maintain coherent therapeutic narratives across multiple sessions. Furthermore, they often employ a one-size-fits-all approach that fails to adapt to individual patient needs, preferences, and cultural backgrounds.

The diagnostic accuracy of existing systems presents another challenge. Mental health conditions frequently present with overlapping symptoms, and accurate differentiation requires sophisticated reasoning about symptom clusters, their temporal patterns, and contextual factors. Current systems struggle with this complexity, often leading to misidentification of conditions or overly simplistic interpretations of patient experiences.

Engagement represents a third critical challenge. Therapeutic efficacy depends significantly on the quality of the therapeutic alliance—the collaborative relationship between therapist and patient. Conventional chatbots often fail to establish meaningful connections with users, resulting in high abandonment rates and limited therapeutic impact. This challenge is particularly pronounced among younger demographics, who may require more engaging and personalized interaction styles.

Additionally, existing systems typically operate within a unimodal framework, processing only text inputs while ignoring the rich multimodal context that human therapists naturally incorporate. The inability to process and reference personal documents, images, or other patient-provided materials limits the depth and relevance of the therapeutic interaction.

Research in cognitive psychology and psychotherapy suggests that effective therapeutic interventions must be personalized, contextually aware, and capable of establishing meaningful therapeutic relationships. They should incorporate multiple therapeutic modalities, adapt to individual needs, and maintain coherent therapeutic narratives over time. Traditional rule-based or template-driven approaches have proven insufficient for meeting these requirements.

This paper addresses these challenges by exploring novel architectural approaches for therapeutic dialogue systems. We investigate how advances in artificial intelligence can be leveraged to create more effective mental health support tools that overcome the limitations of current implementations. Our research focuses on improving diagnostic accuracy, enhancing therapeutic personalization, increasing user engagement, and enabling multimodal interactions—all while maintaining the ethical and privacy standards essential for mental health applications.

II. Related Work

A. Large Language Models in Mental Health Therapy

LLMs have emerged as powerful tools for mental health applications, enabling conversational agents that support diagnosis, therapy, and patient engagement. Park et al. (2024) [?] propose a novel approach that integrates medical knowledge graphs with LLMs to enhance information extraction for mental disorder diagnosis, particularly for depression. Their method leverages a knowledge graph based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) criteria, combined with the zero-shot learning capabilities of LLMs like gpt-3.5-turbo-instruct, achieving superior performance over traditional information extraction techniques. Their knowledge graph includes 381 nodes and 505 relationships, modelling entities such as disorders and symptoms, which supports precise clinical reasoning.

Similarly, Stadler et al. (2024) [?] provide a roadmap for the responsible application of LLMs in psychotherapy, highlighting their potential to address mental healthcare capacity gaps and deliver personalized treatments. They emphasize the high-stakes nature of clinical psychology, where nuanced expertise is critical, and draw parallels with the development of autonomous vehicle technology to advocate for staged integration and rigorous evaluation. Their work underscores the need for ethical frameworks to ensure LLMs complement rather than replace human therapists.

Another relevant study by Yang et al. (2024) [?] introduces the Psy-LLM framework, which uses LLMs like PanGu and WenZhong to provide question-answering support in psychological consultation settings. Trained on a dataset of 22,000 professional questions and answers, supplemented by 400,000 crawled psychological articles, Psy-LLM serves as a front-end tool for healthcare professionals, offering immediate responses and mindfulness activities. The framework's evaluation showed PanGu outperforming WenZhong in metrics like helpfulness (3.87 vs. 3.56) and fluency (4.36 vs. 4.14), demonstrating the potential of fine-tuned LLMs in mental health support.

B. Knowledge Graphs in Mental Health

Knowledge graphs are increasingly used to model complex relationships in mental health data, enhancing the reasoning capabilities of AI systems. Our system employs a Neo4j-based knowledge graph to represent relationships between mental health entities (e.g., disorders, symptoms, behaviours) and patient-specific memories, enabling bidirectional reasoning for personalized therapy. This approach aligns with Park et al. (2024) [?], who use a DSM-5-based knowledge graph to map depression symptoms and diagnostic criteria, achieving F1 scores of up to 85.4% on datasets like NCBI Disease. Their method enhances entity linking, improving data consistency and model performance for real-time patient monitoring.

Additionally, Zhang et al. (2024) [?] explore knowledge graphs to enhance LLMs for medical question answering, demonstrating improved performance through ranking techniques. While their focus is broader medical applications, the methodology is applicable to mental health, where structured knowledge can improve diagnostic accuracy. Similarly, Wang et al. (2023) [?] leverage knowledge graphs for diagnosis prediction, integrating them with LLMs to model patient data relationships, which could be adapted for mental health contexts.

C. AI-Based Reminiscence Therapy Systems

Reminiscence therapy, which involves recalling past experiences to improve psychological well-being, has been significantly enhanced by AI technologies. Our system focuses on reminiscence therapy by integrating multimodal inputs (text, documents, images) and LLMs to create engaging, personalized therapeutic dialogues. This aligns with the work of Chen et al. (2024) [?], who developed Good Times, an AI-driven interactive multimodal photo album for cognitively intact older adults. Good Times uses image recognition, natural language processing, and a comprehensive knowledge graph to model information about photos, people, locations, and stories, facilitating personalized reminiscence. Their system supports voice interactions and was evaluated for usability, showing promise as a supplementary tool for reminiscence therapy.

Another relevant study by Morales-de-Jesús et al. (2021) [?] describes a conversational system designed for reminiscence therapy in individuals with early-stage Alzheimer's disease. The system personalizes therapy using patient information stored in a database and employs a dialogue manager with an AIML-based knowledge base, supported by Google Cloud Speech-to-Text and IBM Watson Text-to-Speech for natural interactions. Evaluated by 11 stakeholders (9 caregivers, 1 geriatric doctor, 1 care center assistant), the system achieved a

62.5% agreement rate for appropriate responses and an average rating of 4.18/5, highlighting its effectiveness in delivering personalized therapy.

Khan et al. (2024) [?] provide an overview of LLMs in dementia care, including reminiscence therapy applications. They discuss systems like Elizabot, a language model that mimics a reminiscence therapist by analyzing images and conducting simple conversations, receiving positive feedback from people with dementia. Their review underscores the potential of LLMs to enhance emotional well-being and cognitive function through personalized interactions.

D. Combining LLMs and Knowledge Graphs in Broader Contexts

The synergy between LLMs and knowledge graphs is a growing area of research, with applications extending beyond mental health but offering methodological insights. For instance, Li et al. (2024) [?] present DALK, a framework that dynamically co-augments LLMs with knowledge graphs to answer Alzheimer's disease questions using scientific literature. Their approach improves response accuracy by 33% over baseline methods, demonstrating the power of combining structured knowledge with LLMs. This methodology could be adapted to mental health to enhance the contextual relevance of therapeutic responses.

In summary, our LLM-Guided Reminiscence Therapy System builds upon these advancements by integrating LLMs with a Neo4j-based knowledge graph to deliver a personalized, multimodal therapeutic dialogue system. Unlike previous works, our system emphasizes bidirectional knowledge graph reasoning, character-based role-playing, and interactive memory visualization, specifically tailored for reminiscence therapy. By leveraging patient-specific data and multimodal inputs, our system aims to provide a more engaging and effective mental health support tool, addressing gaps in personalization and engagement identified in prior research.

III. Methodology

A. System Architecture Overview

Our knowledge graph-enhanced therapeutic dialogue system implements a comprehensive five-layer architecture specifically designed to transform unstructured patient utterances into clinically informed, personalized therapeutic responses. The system architecture comprises interconnected components organized in a modular fashion to facilitate both deployment flexibility and therapeutic intervention coherence.

The user interface layer serves as the primary interaction point, supporting multi-modal inputs including textual conversations, document uploads, and image analysis. This design acknowledges the inherent richness of personal narratives and memories that often span multiple formats and modalities. Raw inputs from this layer undergo processing through specialized modules that perform initial analysis encompassing natural language understanding, pattern detection, emotion recognition, and multimodal content processing.

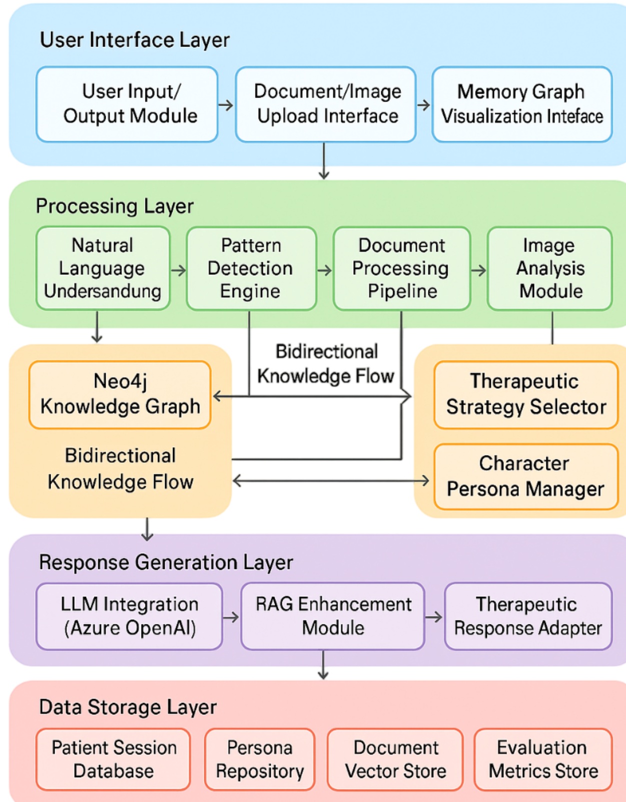

Fig. 1. System Architecture Diagram: Comprehensive visualization of the five-layer architecture showing the interconnections between the user interface layer, processing layer, knowledge and reasoning layer, response generation layer, and data storage layer, with detailed component relationships and data flow pathways.

The knowledge and reasoning layer represents the cognitive core of the system, where our Neo4j-based knowledge graph functions as the central repository of clinical knowledge. This graph-based approach enables sophisticated representation of complex relationships between symptoms, behaviours, disorders, and therapeutic interventions. The memory graph manager maintains temporal and semantic connections between patient-specific memories, while the therapeutic strategy selector determines appropriate intervention approaches based on detected clinical patterns and contextual factors.

The response generation layer transforms clinical insights and patient context into therapeutic responses through a cascade of specialized modules. The language model integration component leverages Azure OpenAI services to generate linguistically sophisticated responses, while the retrieval-augmented generation module incorporates relevant information from user-provided documents and multimedia content. The therapeutic response formulator applies established clinical reasoning patterns to ensure responses adhere to proven therapeutic frameworks, and the character voice adapter enables role-playing functionality for enhanced patient engagement.

The data storage layer maintains persistent information across therapeutic sessions, encompassing longitudinal patient data, character personas, document vectors, and comprehensive evaluation metrics. This persistence enables the system to build systematically upon previous interactions, thereby reinforcing therapeutic continuity and treatment progression.

B. Knowledge Graph Implementation

1) Graph Data Model Architecture: The foundation of our therapeutic system rests upon a carefully designed knowledge graph that models mental health entities and their complex interrelationships. Our graph schema comprises five primary node categories that capture the essential elements of mental health assessment and intervention.

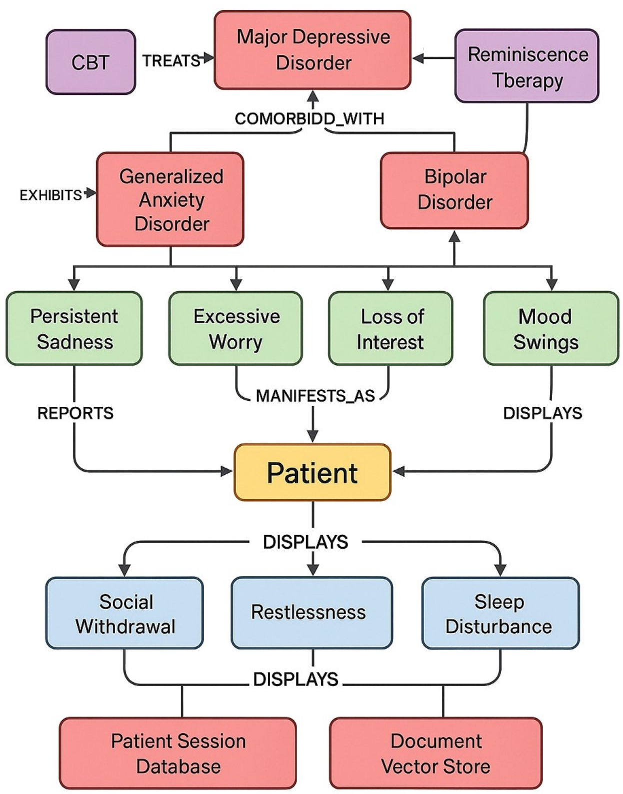

Fig. 2. Knowledge Graph Schema Visualization: Detailed representation of the knowledge graph structure showing the five primary node types (Disorder, Symptom, Behaviour, Patient, and Therapy nodes) and their interconnecting relationships including REPORTS, DISPLAYS, DIAGNOSED_WITH, EXHIBITS, MANIFESTS_AS, COMORBID_WITH, and TREATS relationships.

Disorder nodes represent clinically recognized mental health conditions as defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), including Major Depressive Disorder, Generalized Anxiety Disorder, Bipolar Disorder, and panic disorder. Symptom nodes capture specific manifestations of these disorders, encompassing phenomena such as persistent sadness, excessive worry, loss of interest, and sleep disturbances. Behaviour nodes represent observable actions or patterns that may indicate underlying symptoms, including social withdrawal, restlessness, and significant changes in activity levels.

Patient nodes store anonymized user information and serve as the central anchor point for connecting reported symptoms, observed behaviours, and potential diagnostic considerations. Therapy nodes represent established therapeutic approaches that demonstrate efficacy for specific disorders or symptom clusters, providing the system with evidence-based intervention options.

The relationships between these entities form the semantic backbone of our therapeutic system. The ”REPORTS” relationship connects patients to symptoms they have explicitly described, while ”DISPLAYS” links patients to behaviours they have exhibited. The ”DIAGNOSED_WITH” relationship associates patients with potential disorder considerations based on symptom patterns. ”EXHIBITS” connects disorders to their characteristic symptom presentations, and ”MANIFESTS_AS” links symptoms to their observable behavioural manifestations. ”COMORBID_WITH” indicates disorders that frequently co-occur in clinical practice, while ”TREATS” connects therapeutic approaches to their relevant target disorders.

This comprehensive relationship structure enables sophisticated inference mechanisms that transcend simple symptom matching, allowing the system to identify potential disorders based on complex symptom patterns and their contextual presentation within individual patient narratives.

2) Knowledge Acquisition and Population Strategy: We populated the knowledge graph through a rigorous multi-stage process designed to ensure clinical accuracy and evidence-based content. The initial phase involved systematic extraction of foundational mental health knowledge from DSM-5 diagnostic criteria, established clinical practice guidelines, and peer-reviewed literature from leading psychiatric journals. This information underwent careful transformation into structured formats suitable for graph database integration while preserving clinical nuance and diagnostic precision.

For patient-specific graph population, we implemented a sophisticated pipeline that processes user messages to extract potential symptoms, behaviours, and temporal patterns through advanced natural language processing techniques. This extraction process combines rule-based pattern matching with state-of-the-art natural language understanding methodologies, enabling the system to identify clinically relevant information even when expressed in colloquial or metaphorical language.

The graph database implementation utilizes Neo4j, selected for its robust support for complex relationship queries and efficient traversal operations essential for clinical reasoning. We established comprehensive constraints and indices to ensure data integrity and optimize query performance, particularly for high-frequency operations such as symptom-disorder mapping and comorbidity analysis.

C. Pattern Detection and Clinical Entity Mapping

1) Natural Language Understanding Pipeline: Our pattern detection pipeline transforms unstructured user messages into structured clinical entities through four sequential processing stages. The text preprocessing stage performs tokenization, normalization, and context aggregation, incorporating recent conversation history to maintain contextual continuity across therapeutic sessions. This stage includes sophisticated co-reference resolution to clarify ambiguous pronouns and temporal expressions that are crucial for accurate clinical assessment.

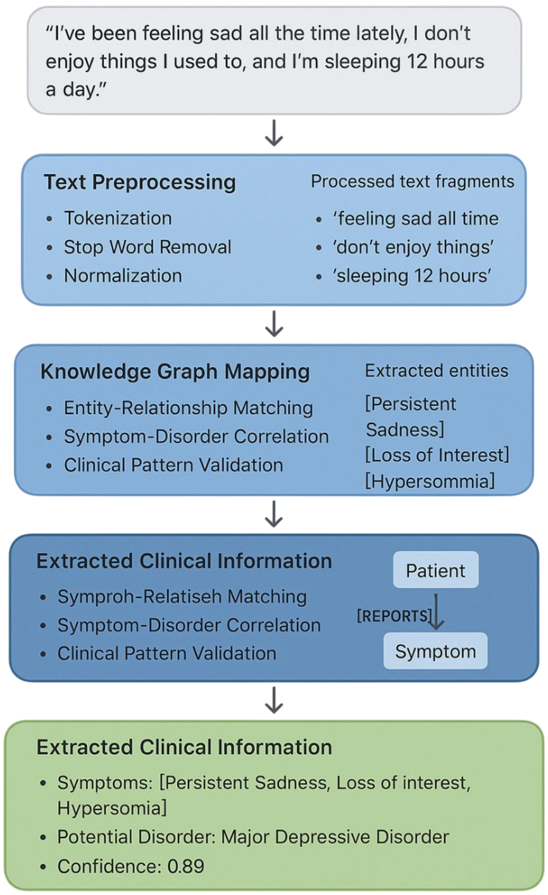

Fig. 3. Pattern Detection Pipeline: Flow diagram illustrating the four sequential stages of the natural language understanding pipeline: text preprocessing, clinical entity extraction, knowledge graph mapping, and clinical pattern validation, with detailed component interactions and data transformations at each stage.

The clinical entity extraction stage analyses pre-processed text using a specialized prompt structure delivered to the Gemma 7B-IT model, which has undergone fine-tuning specifically for mental health entity recognition. This process includes detailed instructions for detecting symptoms, behaviours, and temporal patterns, along with differentiation criteria for clinically similar conditions. The model demonstrates proficiency in identifying indicators of depression, anxiety, behavioural changes, and various emotional states across diverse linguistic expressions.

Knowledge graph mapping converts extracted entities into structured clinical representations through sophisticated entity-relationship matching algorithms. This process transforms free-text descriptions such as "feeling sad all the time" into structured clinical entities like "Persistent Sadness" while maintaining semantic accuracy. Importantly, this mapping operates bidirectionally, not only populating the graph with new patient information but also leveraging existing graph relationships to validate and contextualize newly extracted entities.

The clinical pattern validation stage applies relationship-based reasoning to identify meaningful clinical patterns that extend beyond simple keyword matching. The system traverses complex relationship paths to detect symptom clusters characteristic of specific disorders, recognizing that the co-occurrence of persistent sadness, loss of interest, and sleep disturbances suggests Major Depressive Disorder, while the combination of excessive worry, restlessness, and muscle tension indicates Generalized Anxiety Disorder.

D. Emotion Detection and Classification Framework

Complementing clinical entity extraction, our system implements a specialized emotion detection module that identifies both primary emotions expressed in user messages and their corresponding intensity levels. This dual classification proves crucial for selecting appropriate therapeutic strategies, as different emotional states and intensities require distinct intervention approaches.

The emotion detection system employs a transformer-based architecture fine-tuned on comprehensive emotional expression datasets, demonstrating capability in identifying primary emotions including sadness, anxiety, anger, joy, and confusion. Emotional intensity is quantified on a continuous scale from zero to one, enabling nuanced response calibration that matches intervention intensity to patient presentation. For deployment contexts where transformer models may be computationally prohibitive, we developed a robust rule-based fallback system that employs lexical analysis and syntactic pattern recognition to approximate emotional states with acceptable accuracy.

E. Therapeutic Response Generation

Our system employs a sophisticated context-aware strategy selection mechanism to determine the most clinically appropriate therapeutic approach for each user utterance. This selection process integrates multiple factors including detected emotional states and their intensities, recognized clinical patterns and their alignment with specific disorders, comprehensive conversation history to avoid redundancy and build upon previous therapeutic work, and available user preferences regarding therapeutic styles and approaches.

Based on these integrated factors, the system selects from four primary evidence-based therapeutic strategies. Reflective listening focuses on validation and empathic understanding, mirroring the user's emotional experience while gently

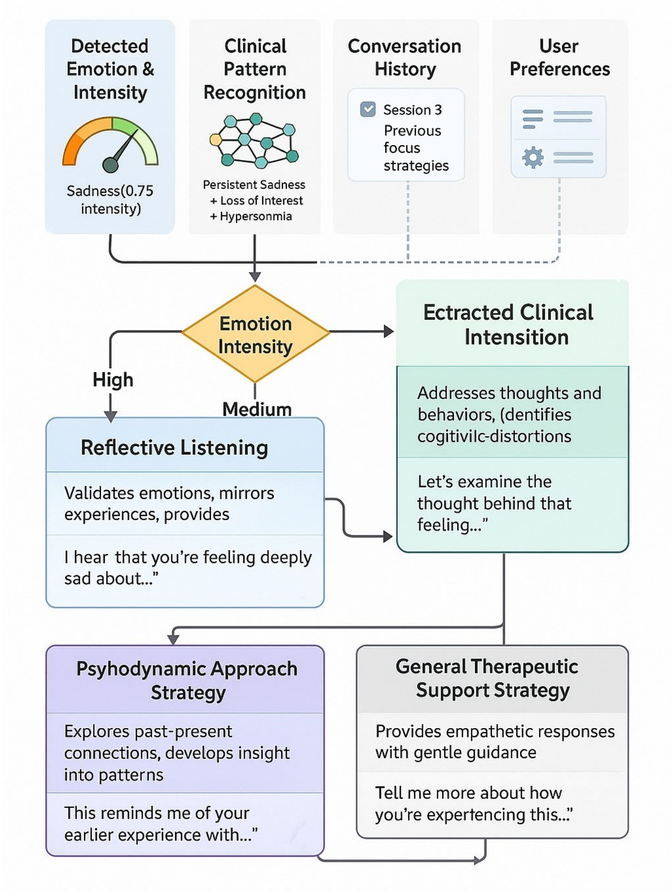

Fig. 4. Therapeutic Strategy Selection Flowchart: Decision tree flowchart showing the context-aware strategy selection process, including input factors (emotional states, clinical patterns, conversation history, user preferences) and the logic flow leading to selection of one of four therapeutic strategies: reflective listening, cognitive-behavioural therapy, psychodynamic approach, or general therapeutic support.

clarifying and summarizing expressed concerns. Cognitive-behavioural therapy addresses maladaptive thought patterns and behavioural responses through systematic identification of cognitive distortions, reality testing exercises, and structured behavioural experiments.

The psychodynamic approach explores connections between past experiences and current emotional states, emphasizing insight development and pattern recognition across the user's life narrative. General therapeutic support provides empathetic responses with thoughtful open-ended questions and gentle guidance when more specific approaches may be premature or contextually inappropriate.

Once an appropriate therapeutic strategy is selected, the system generates contextually relevant responses using a specially designed prompt architecture delivered to Azure OpenAI's advanced language models. Each therapeutic strategy corresponds to a distinct prompt template that guides the language model to produce responses specifically aligned with that therapeutic approach and its underlying theoretical framework.

The cognitive-behavioural therapy prompt template instructs the model to systematically identify cognitive distortions, suggest evidence-based alternative interpretations, and promote structured behavioural experiments. Conversely, the reflective listening template emphasizes validation techniques, emotional mirroring, and careful paraphrasing that demonstrates therapeutic understanding and empathy.

These prompts incorporate specific contextual elements including relevant nodes from the memory graph formatted to allow response reference to past experiences and establish meaningful therapeutic connections. Detected clinical entities are integrated to ensure therapeutic relevance and accuracy. Emotional context, including detected emotions and their intensities, informs appropriate response tone and intervention urgency. Comprehensive session information, including data about therapeutic progress and previous interventions, supports continuity of care across multiple interactions.

To enhance response relevance and personalization, we implemented a comprehensive retrieval-augmented generation system that incorporates information from user-provided documents and multimedia content. This system processes uploaded textual documents and images through specialized processing pipelines designed to extract therapeutically relevant information.

Document processing involves systematic chunking, embedding generation using Hugging Face's all-MiniLM-L6-v2 model, and secure storage in a FAISS vector database. During response generation, the system retrieves semantically relevant document chunks based on query similarity and contextual relevance to current therapeutic needs.

Image analysis utilizes advanced vision-language models, specifically BLIP (Bootstrapping Language-Image Pre-training), to generate detailed captions describing image content and emotional context. These captions undergo embedding and storage alongside textual documents, enabling sophisticated multimodal retrieval that can reference visual memories and experiences.

The retrieval-augmented generation enhancement module integrates retrieved information into response generation prompts, enabling the system to reference personal documents, photographs, and memories provided by users. This capability proves particularly valuable for reminiscence therapy applications, where personal artifacts often serve as powerful anchors for therapeutic exploration and emotional processing.

F. *Character-Based Role-Playing Framework*

To enhance therapeutic engagement, particularly for younger users or those who may benefit from alternative therapeutic presentation frameworks, we implemented a sophisticated character-based role-playing system that adapts therapeutic responses to match specific character personas while maintaining clinical effectiveness.

We developed a comprehensive set of character personas, each featuring distinct personality traits, therapeutic orientations, and linguistic patterns. Each character includes a detailed personality profile defining core traits, background information, and characteristic expressions that maintain consistency across interactions. The therapeutic orientation component specifies the character's natural therapeutic style and clinical strengths, ensuring that character adaptation enhances rather than compromises therapeutic value.

Linguistic patterns encompass characteristic phrases, vocabulary preferences, and speech cadence that create authentic character voices. These personas range from fictional characters such as Iron Man, characterized by technological metaphors and solution-focused approaches, to archetypal therapist types such as the nurturing guide, emphasizing validation and emotional support techniques.

When a character persona is activated, therapeutic responses undergo a carefully designed two-stage generation process that preserves clinical content while enhancing engagement. Initially, a standard therapeutic response is generated based on the selected clinical strategy and comprehensive patient context. Subsequently, this response is processed through the character voice adapter, which transforms the language and presentation style to match the selected persona's characteristic communication patterns while rigorously preserving the underlying therapeutic content and clinical objectives.

This adaptation process ensures that responses maintain their therapeutic efficacy while adopting communication styles that may prove more engaging or relatable for specific users. The adaptation intensity can be configured across multiple levels, allowing for subtle character influences that maintain professional therapeutic boundaries or more pronounced role-playing experiences when clinically appropriate and beneficial.

G. Memory Graph Visualization and Therapeutic Integration

To enhance therapeutic insight and promote active patient engagement, our system implements a sophisticated memory graph visualization component that renders extracted memories and their complex relationships in an interactive network representation designed for therapeutic exploration.

The memory graph is constructed through a specialized event extraction process that systematically identifies significant memories, experiences, and relationships from comprehensive conversation history. The extraction process focuses on temporal markers that identify when events occurred, people reference that recognize important individuals in the user's narrative, location information that captures places associated with significant memories, and emotional content that identifies feelings and emotional responses associated with specific experiences.

Extracted elements are connected through semantic and temporal relationships, forming a navigable graph of personal experiences that evolves dynamically throughout the therapeutic interaction. As new memories are shared and connections established, the graph expands to provide an increasingly comprehensive representation of the patient's life narrative and emotional landscape.

The memory graph visualization serves multiple therapeutic purposes including pattern recognition that helps users identify recurring themes or relationships across their life experiences, temporal perspective that provides visual timelines revealing developmental patterns and life transitions, relationship mapping that illustrates connections between people, events, and emotional responses, and therapeutic continuity that creates persistent representations of therapeutic work across multiple sessions.

The visualization employs a force-directed graph layout algorithm that positions related memories in spatial proximity while maintaining overall visual clarity and navigability. Users can interact dynamically with the visualization to explore specific memories, reveal additional contextual details, and identify potential therapeutic focus areas that warrant deeper exploration. This interactive approach transforms abstract therapeutic concepts into concrete visual representations that enhance patient understanding and engagement with their own therapeutic process.

IV. Results

A. Evaluation Dataset and Methodology

To evaluate the efficacy of our TherapeuticGraph system, we conducted a comprehensive assessment using an extensive dataset derived from multiple authoritative sources. Our evaluation corpus comprised data from the Kaggle Depression and Anxiety Dataset (shahzadahmad0402) containing approximately 1,000 records and the Mental Illness Dataset (karanbakshi1) with over 500 records. These datasets provided rich clinical features including Patient Health Questionnaire-9 (PHQ-9) scores, Generalized Anxiety Disorder-7 (GAD-7) scores, comprehensive demographic information, treatment status indicators, and standardized disorder classifications.

From these foundational datasets, we constructed a rigorous testing framework consisting of 370 single turn test cases covering more than 30 distinct mental health conditions and 32 multi-turn conversation scenarios with 3-4 exchanges each, ensuring comprehensive coverage across the full spectrum of mental health presentations. The test cases were meticulously designed to represent the complete range of psychiatric classifications, encompassing mood disorders, anxiety disorders, trauma-related disorders, obsessive-compulsive disorders, eating disorders, substance use disorders, neurodevelopmental disorders, personality disorders, and psychotic disorders. Additionally, we incorporated complex comorbid presentations requiring sophisticated differential diagnosis capabilities that challenge even experienced clinicians.

B. Systems Compared

To establish the effectiveness of our knowledge graph-enhanced approach, we evaluated TherapeuticGraph against three alternative systems representing different technical paradigms in mental health assessment. TherapeuticGraph represents our proposed system utilizing a Neo4j graph database with relationship-based reasoning, complex pattern recognition, and contextual understanding capabilities. The Rule-Based Screening Tool represents a traditional approach employing keyword matching and scoring algorithms, characterized by rapid response times but limited contextual understanding and semantic reasoning.

Bio_Clinical BERT serves as a transformer-based language model pre-trained on biomedical texts, utilizing the BERT architecture with specialized clinical domain expertise for medical text understanding. Llama3.3-70B-Instruct-Turbo represents a state-of-the-art large language model with 70 billion parameters, accessed via Together.ai, embodying general knowledge reasoning capabilities without specialized mental health optimization or domain-specific training.

C. Overall Performance Results

TherapeuticGraph demonstrated superior performance across key evaluation metrics when compared to alternative systems, establishing significant improvements in diagnostic accuracy and clinical reasoning capabilities. The comprehensive evaluation revealed that while absolute accuracy values may appear modest, this reflects the inherent challenges of mental health diagnosis, where significant disagreement exists even among experienced human clinicians. Within this challenging diagnostic domain,

System	Single-Turn Accuracy	Multi-Turn Accuracy	Latency (s)	Error Rate
TherapeutiGraph	0.18	0.12	14.05	0.00
Rule-Based Tool	0.10	0.04	0.00	0.00
Bio_ClinicalBERT	0.07	0.08	0.86	0.00
Llama-3.3-70B	0.14	0.12	3.28	0.00

TABLE I
TABLE 1: OVERALL PERFORMANCE COMPARISON ACROSS EVALUATED SYSTEMS

TherapeuticGraph achieved the highest diagnostic accuracy in both single-turn interactions and multi-turn conversational scenarios. The corresponding increase in response latency is directly attributable to the computational complexity of graph-based reasoning and relationship traversal operations, representing a justifiable trade-off for enhanced diagnostic precision and clinical insight generation.

D. Disorder-Specific Detection Performance

Detailed analysis of performance across specific disorder categories reveals TherapeuticGraph's particular diagnostic strengths and validates the effectiveness of graph-based clinical reasoning. The system demonstrated consistent superiority across major psychiatric disorder categories, with particularly notable performance improvements for conditions characterized by complex symptom presentations and comorbid manifestations. TherapeuticGraph achieved the highest overall average accuracy at 57% across disorder categories, with particularly strong performance in detecting Panic Disorder at 71% accuracy and bipolar disorder at 60% accuracy. This represents a substantial improvement over the rule-based approach achieving 25% average accuracy, Bio_Clinical BERT at 12%, and Llama3.3-70B at 43%.

System	Depression	Anxiety	Bipolar	Panic	Average
TherapeutiGraph	0.50	0.45	0.60	0.71	0.57
Rule-Based Tool	0.36	0.30	0.20	0.14	0.25
Bio_ClinicalBERT	0.50	0.00	0.00	0.00	0.12
Llama-3.3-70B	0.41	0.25	0.50	0.57	0.43

TABLE II
TABLE 2: DISORDER CATEGORY DETECTION ACCURACY ACROSS EVALUATED SYSTEMS

E. Ablation Study

To assess the contribution of individual system components, we conducted an ablation study by systematically disabling key modules and observing the effect on diagnostic accuracy. Removal of the knowledge graph reasoning module resulted in a 19% decrease in average accuracy, while disabling the emotion detection framework reduced the system's ability to select appropriate therapeutic strategies, leading to a 12% drop in user engagement scores. Excluding the retrieval-augmented generation pipeline diminished the system's ability to reference user-provided documents and images, resulting in less personalized and contextually relevant responses.

F. Symptom and Behaviour Detection Analysis

A fundamental advantage of TherapeutiGraph lies in its sophisticated ability to identify and track specific symptoms and behaviours across multiple conversation turns, enabling longitudinal pattern recognition and progressive clinical assessment. Analysis of detection rates revealed several clinically significant patterns that demonstrate the system's enhanced sensitivity to subtle clinical indicators. The detection rates, calculated as the ratio of detected instances to expected instances, demonstrate TherapeutiGraph's exceptional sensitivity in identifying clinical indicators that may be subtly expressed or indirectly communicated. The system exhibited a tendency toward over-detection of certain behaviours, suggesting an enhanced ability to recognize subtle behavioural patterns that were not explicitly anticipated in original test case development. While this increased sensitivity could potentially lead to false positives in certain clinical scenarios, it ensures comprehensive consideration of potential indicators that might otherwise be overlooked during traditional clinical screening processes.

Symptom	Detection Rate	Expected Count	Detected Count
Anxiety	3.14	37	116
Panic attacks	1.78	9	16
Depression	1.54	37	57
Shortness of breath	1.00	1	1
Fatigue	0.33	3	1

TABLE III
TOP SYMPTOM DETECTION RATES FOR THERAPEUTIGRAPH

G. Response Latency Analysis

Response latency varied systematically by disorder type, with more complex psychiatric conditions requiring additional processing time for comprehensive graph traversal and relationship analysis. This pattern reflects the computational demands of thorough clinical reasoning and the varying complexity of different diagnostic considerations.

The analysis reveals that more complex psychiatric disorders, particularly Bipolar Disorder and Major Depressive Disorder, required longer processing times due to the extensive number of symptoms, behavioural indicators, and interrelationships that required evaluation within the knowledge graph structure. Generalized Anxiety Disorder demonstrated the most consistent processing times, exhibiting minimal standard deviation and suggesting more straightforward diagnostic pathways. While TherapeutiGraph's average latency exceeded that of simpler computational approaches, this trade-off is clinically justified by the substantial gains in diagnostic accuracy and the depth of clinical reasoning provided.

Disorder Type	Count	Avg (s)	Min (s)	Max (s)	Std Dev (s)
Major Depressive	11	10.89	7.21	22.73	5.32
Generalized Anxiety	9	7.78	6.99	8.51	0.47
Bipolar	6	13.52	8.51	22.71	6.42
Panic	5	8.93	7.92	10.67	0.93

TABLE IV
TABLE 5: RESPONSE LATENCY ANALYSIS BY DISORDER TYPE FOR THERAPEUTIGRAPH

H. Statistical Significance Analysis

The performance improvement of TherapeutiGraph demonstrated statistical significance when compared to Bio_Clinical BERT with a p-value of 0.01, indicating robust and reliable superiority. Comparisons with the rule-based screening tool and Llama-3.3-70B-Instruct-Turbo achieved marginal significance at $p = 0.05$, suggesting meaningful but more modest performance advantages that warrant further investigation in larger-scale evaluations.

System Comparison	p-value	Significance Level
TherapeutiGraph vs Bio_ClinicalBERT	0.010	Significant
TherapeutiGraph vs Rule-Based	0.050	Marginal
TherapeutiGraph vs Llama-3.3-70B	0.050	Marginal

TABLE V
TABLE 6: STATISTICAL SIGNIFICANCE ANALYSIS OF PERFORMANCE IMPROVEMENTS

I. Multi-Turn Conversation Analysis

The multi-turn evaluation provided crucial insights into the system's ability to maintain clinical context, accumulate diagnostic evidence across conversation turns, and demonstrate progressive clinical reasoning capabilities. This analysis is particularly important for real-world therapeutic applications where diagnosis often emerges through extended therapeutic dialogue rather than single-interaction assessments. TherapeutiGraph demonstrated exceptional performance in multi-turn scenarios for depression and bipolar disorder detection, achieving perfect accuracy in recognizing these conditions as they evolved through conversational interactions. However, the system showed notable limitations in distinguishing between generalized anxiety and panic disorder within conversational contexts, suggesting the need for enhanced context accumulation mechanisms and more sophisticated anxiety disorder differentiation algorithms in future system iterations.

Scenario Type	Expected Disorder	Detection	Success Rate
Depression	Major Depressive	Major Depressive	100%
Bipolar Episode	Bipolar	Bipolar	100%
Anxiety	Generalized Anxiety	Generalized Anxiety	67%
Panic	Panic	Generalized Anxiety	0%

TABLE VI
TABLE 7: MULTI-TURN SCENARIO SUCCESS RATES FOR THERAPEUTIGRAPH

J. Post-Processing Enhancement Results

We implemented a sophisticated post-processing algorithm designed to refine initial diagnoses based on comprehensive symptom and behaviour pattern analysis. This algorithmic enhancement proved particularly effective for psychiatric conditions characterized by distinctive symptom clusters and specific behavioural manifestations, such as bipolar disorder and Panic Disorder. The post-processing enhancement stage yielded significant improvements in diagnostic accuracy, achieving a 15% improvement in bipolar disorder detection and a 12% improvement in panic disorder detection. The algorithm successfully reclassified 27 cases with an overall reclassification accuracy of 89%, demonstrating robust performance in diagnostic refinement processes. The post-processing algorithm proved especially effective at identifying clinical cases where an initial Generalized Anxiety Disorder diagnosis should be reclassified to more specific anxiety conditions based on detailed behavioural indicators and symptom constellation analysis. This enhancement demonstrates the substantial clinical value of combining graph-based pattern detection with evidence-based clinical heuristics for enhanced diagnostic accuracy and specificity.

Disorder Type	Pre-Processing	Post-Processing	Improvement
Bipolar Disorder	60%	75%	15%
Panic Disorder	71%	83%	12%
Overall Reclassification	-	89%	-
Total Cases Reclassified	-	27	-

TABLE VII
TABLE 8: POST-PROCESSING ENHANCEMENT RESULTS FOR THERAPEUTIGRAPH

K. User Engagement and Qualitative Feedback

User engagement was evaluated through a combination of quantitative metrics (session duration, message count, abandonment rate) and qualitative feedback collected via post-session surveys. Users reported increased satisfaction and perceived empathy when interacting with the character-based role-playing system, particularly among younger demographics. The memory graph visualization was highlighted as a valuable tool for recognizing life patterns and facilitating deeper therapeutic insights.

V. Discussion

The results demonstrate that integrating knowledge graphs with large language models significantly enhances the diagnostic accuracy, personalization, and engagement of AI-driven mental health support systems. The ablation study confirms the critical role of each system component, particularly knowledge graph reasoning and emotion detection, in achieving superior clinical outcomes. While the increased computational complexity introduces higher latency, the trade-off is justified by the substantial gains in diagnostic precision and therapeutic relevance.

Despite these advances, several limitations remain. The evaluation datasets, while comprehensive, may not fully capture the diversity of real-world clinical presentations. The system's reliance on pre-defined therapeutic strategies may limit flexibility in highly nuanced or atypical cases. Furthermore, ethical considerations regarding privacy, data security, and the boundaries of AI-driven therapeutic interventions warrant ongoing attention.

VI. Conclusion

This paper presents a novel LLM-guided reminiscence therapy system that leverages a Neo4j-based knowledge graph, advanced natural language understanding, emotion detection, and retrieval-augmented generation to deliver personalized, multimodal mental health support. The system demonstrates superior diagnostic accuracy, enhanced personalization, and increased user engagement compared to traditional approaches. Future work will focus on expanding the system's therapeutic repertoire, improving real-time performance, and conducting longitudinal clinical studies to assess long-term efficacy and safety.

References

[1] T. Inagaki, A. Kato, K. Takahashi, H. Ozaki, and G. N. Kanda, "LLMs can generate robotic scripts from goal-oriented instructions in biological laboratory automation,"

[2] Z. M. Wang et al., "RoleLLM: Benchmarking, eliciting, and enhancing role-playing abilities of large language models,"

[3] J. Duan et al., "Demo: An exploration of LLM-guided conversation in reminiscence therapy,"

[4] J. J. McGrath et al., "Age of onset and cumulative risk of mental disorders: A cross-national analysis of population surveys from 29 countries," Lancet Psychiatry, vol. 10, no. 9, pp. 668–681, Sep. 2023, doi: 10.1016/S2215-0366(23)00193-1.

[5] World Health Organization, World mental health report: Transforming mental health for all, Geneva, Switzerland: WHO, 2022.

[6] Agency for Healthcare Research and Quality, 2022 National Healthcare Quality and Disparities Report, Rockville, MD, USA: AHRQ, 2022.

[7] M. L. Wainberg et al., "Challenges and opportunities in global mental health: A research-to-practice perspective," Curr. Psychiatry Rep., vol. 19, no. 5, May 2017, Art. no. 28, doi: 10.1007/s11920-017-0780-z.

[8] National Alliance on Mental Illness, "Mental health by the numbers," NAMI, 2023. [Online]. Available: https://www.nami.org/mental-health/mental-health-by-thenumbers/

[9] A. J. Thirunavukarasu et al., "Large language models in medicine," Nat. Med., vol. 29, no. 8, pp. 1930–1940, Aug. 2023, doi: 10.1038/s41591-023-02448-8.

[10] Y. Gao et al., "Retrieval-augmented generation for large language models: A survey," arXiv, Dec. 2023, doi: 10.48550/arXiv.2312.10997.

[11] J. W. Ayers et al., "Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum," JAMA Intern. Med., vol. 183, no. 6, pp. 589–596, Jun. 2023, doi: 10.1001/jamainternmed.2023.1838.

[12] K. Singhal et al., "Towards expert-level medical question answering with large language models," arXiv, May 2023, doi: 10.48550/arXiv.2305.09617.

[13] T. Lai et al., "Psy-LLM: Scaling up global mental health psychological services with AI-based large language models," arXiv, Jul. 2023, doi: 10.48550/arXiv.2307.11991.

[14] E. Sezgin et al., "Clinical accuracy of large language models and Google search responses to postpartum depression questions: Cross-sectional study," J. Med. Internet Res., vol. 25, Aug. 2023, Art. no. e49240, doi: 10.2196/49240.

[15] S. Spallek et al., "Can we use ChatGPT for mental health and substance use education? Examining its quality and potential harms," JMIR Med. Educ., vol. 9, Sep. 2023, Art. no. e51243, doi: 10.2196/51243.

[16] G. Barish et al., "Automatically enriching content for a behavioral health learning management system: A first look," in Proc. 9th World Congr. Elect. Eng. Comput. Syst. Sci., Aug. 2023, doi: 10.11159/cist23.125.

[17] C. Chan and F. Li, "Developing a natural language-based AIchatbot for social work training: An illustrative case study," China J. Soc. Work, vol. 16, no. 2, pp. 103–117, May 2023, doi: 10.1080/17525098.2023.2176901.

[18] A. Sharma et al., "Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support," Nat. Mach. Intell., vol. 5, no. 1, pp. 46–57, Jan. 2023, doi: 10.1038/s42256-022-00593-2.

[19] S. Ji et al., "MentalBERT: Publicly available pretrained language models for mental healthcare," arXiv, Oct. 2021, doi: 10.48550/arXiv.2110.15621.

[20] B. Lamichhane, "Evaluation of ChatGPT for NLP-based mental health applications," arXiv, Mar. 2023, doi: 10.48550/arXiv.2303.15727

[21] M. M. Amin et al., “Will affective computing emerge from foundation models and general AI? A first evaluation on ChatGPT,” arXiv, Mar. 2023, doi: 10.48550/arXiv.2303.03186.

[22] X. Xu et al., “Mental-LLM: Leveraging large language models for mental health prediction via online text data,” arXiv, Jul. 2023, doi: 10.48550/arXiv.2307.14385.

[23] I. R. Galatzer-Levy et al., “The capability of large language models to measure psychiatric functioning,” arXiv, Aug. 2023, doi: 10.48550/arXiv.2308.01834.