Synthesis: A Multidisciplinary Research Journal
Volume 2, Issue 4, October - December 2024, Pp: 1-10

Research Article

Emotion-Responsive Virtual Agents Using Generative Memory-Augmented Cognitive Architectures

[1*] Guguloth Ravi, [2] T. Aditya sai srinivas, [3] M Bhavsingh

[1*] *Associate professor, Department of CSE, Malla Reddy College of Engineering and Technology, Telangana, India, Email: g.raviraja@gmail.com*
[2] *Associate Professor, CSE department, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh, India. Email: taditya1033@gmail.com*
[3] *Associate Professor, department of CSE, Ashoka Womens Engineering College, Kurnool, Andhra Pradesh, India. Email: bhavsinghit@gmail.com*

**Corresponding Author(s): g.raviraja@gmail.com*

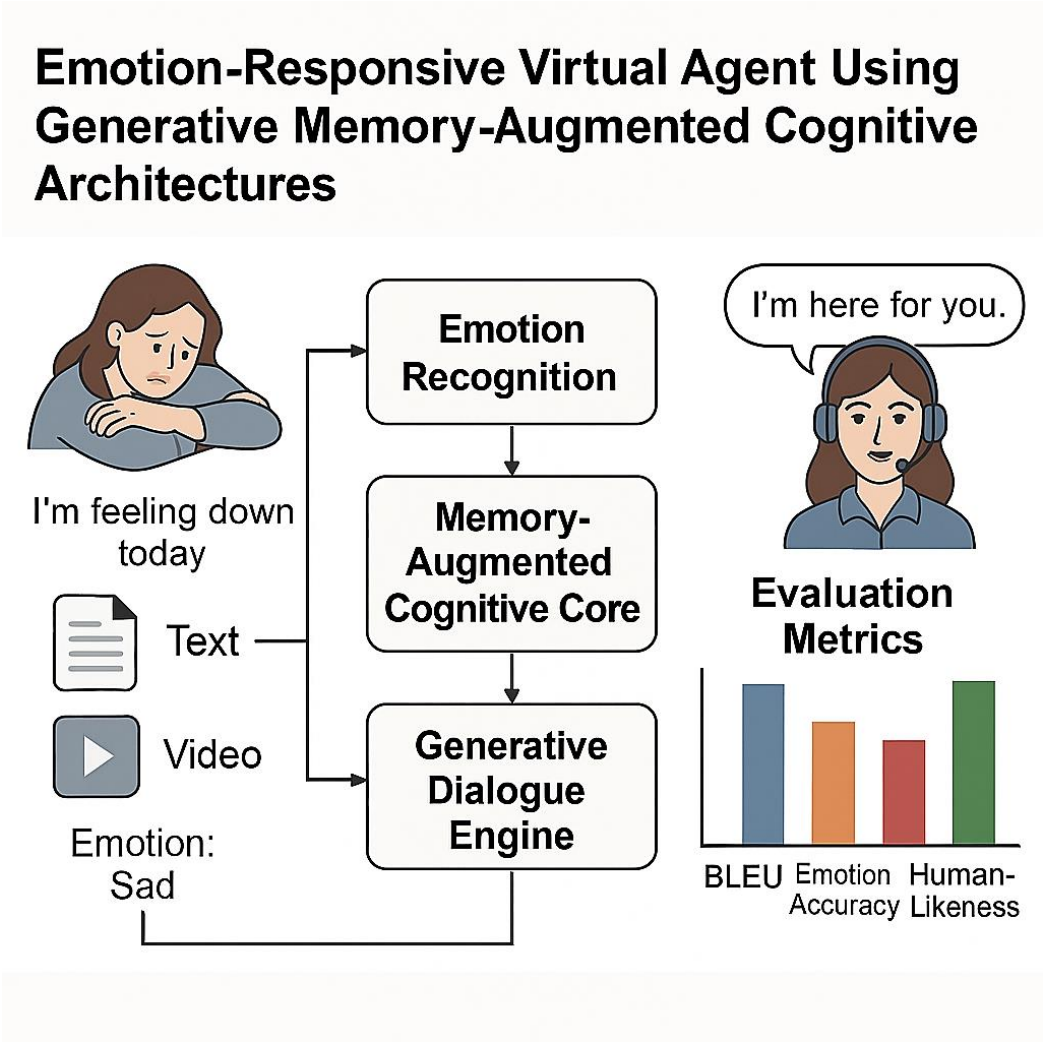
Article Info

Received:12/08/2024
Revised: 19/10/2024
Accepted:20/12/2024
Published:31/12/2024

Abstract

Virtual agents are increasingly deployed in emotionally sensitive domains such as mental health, education, and social robotics. However, existing systems often lack emotional coherence, contextual awareness, and memory retention, limiting their ability to deliver human-like interactions. This study aims to develop an emotion-responsive virtual agent using a Generative Memory-Augmented Cognitive Architecture (GMCA) that integrates emotion recognition, episodic memory, and context-aware response generation. The proposed GMCA framework incorporates a multimodal emotion recognition module, a differentiable external memory for storing emotion-tagged dialogue history, and a generative response engine conditioned on both emotion and memory contexts. The system is trained and evaluated on benchmark datasets including IEMOCAP, DailyDialog, and custom user-agent interaction logs. Key metrics include BLEU score, Emotion Accuracy (EA), Memory Utility (MU), and Human-Likeness Score (HLS). Experimental results show that GMCA outperforms existing empathetic and transformer-based dialogue systems. Specifically, it achieves a BLEU score of 16.9, an EA of 74.5%, a Memory Utility score of 0.78, and an HLS of 4.3 out of 5. These represent improvements of +3.8 BLEU, +12.6% EA, and +0.6 HLS over the strongest baseline. By unifying affective computing, memory augmentation, and generative cognitive modeling, GMCA enables emotionally intelligent, context-aware interactions. The model demonstrates strong applicability in real-world scenarios requiring empathetic engagement, setting a new benchmark for emotion-aware conversational AI.

Keywords: Emotion-aware virtual agents, generative dialogue models, memory-augmented neural networks, cognitive architectures, empathetic AI, affective computing, human-computer interaction



Graphical Abstract: The end-to-end pipeline of the proposed GMCA-based virtual agent

1. Introduction

Emotionally intelligent virtual agents are increasingly vital in applications ranging from mental health support and education to customer service and social robotics [1]. These agents are expected not only to interpret and process natural language but also to recognize, track, and appropriately respond to human emotions. Despite progress in natural language generation and affective computing, achieving emotionally coherent and context-aware interaction remains a persistent challenge [2]. The lack of memory continuity, shallow emotion modeling, and limited adaptability in current systems restricts their real-world usability and user acceptance.

Conventional dialogue systems typically operate with limited emotional awareness, often relying on sentiment classifiers or keyword-based emotional tags [3]. These methods fail to capture the dynamic and contextual nature of human affect, leading to emotionally inconsistent or inappropriate responses. Moreover, most state-of-the-art generative models lack memory augmentation and treat each interaction in isolation, ignoring prior user states and conversation history [4]. Even systems designed for empathetic responses are largely stateless and depend on predefined emotion templates, which hinders their ability to generalize across diverse emotional contexts and long-term interactions [5].

To address these limitations, this study proposes a novel framework that integrates generative language modeling with external memory-augmented cognitive architecture, enabling virtual agents to dynamically synthesize emotionally responsive behavior [6]. The system leverages multimodal emotion recognition, episodic memory encoding, and emotion-conditioned response generation to produce adaptive, affect-sensitive dialogue. By emulating cognitive processes such as emotion-driven memory retrieval and context-aware decision-making, the proposed architecture advances the design of human-like, empathetic agents [7].

Key Contributions of this Study Include:

- A novel generative memory-augmented cognitive architecture (GMCA) that integrates emotion recognition, memory retrieval, and dialogue generation.
- A differentiable memory module that stores and retrieves emotionally tagged episodic context, enabling long-term emotional coherence in dialogue.
- A dual-objective training scheme that jointly optimizes linguistic quality and emotional alignment using cross-entropy and emotion-consistency loss functions.

- Comprehensive evaluation across three datasets with results showing significant improvements in BLEU score (+3.8), Emotion Accuracy (+12.6%), and Human-Likeness Score (+0.6) over existing empathetic dialogue baselines [8].

The remainder of this paper is organized as follows: Section 2 reviews related work in affective computing and memory-based dialogue systems. Section 3 presents the problem formulation. Section 4 details the proposed GMCA methodology. Section 5 reports experimental setup and evaluation results. Section 6 discusses the implications and limitations, and Section 7 concludes the study with future research directions.

2. Literature Review

This section reviews prior research relevant to the development of emotion-responsive virtual agents. The literature is categorized into four key areas: emotion recognition in dialogue systems, empathetic dialogue generation, memory-augmented neural architectures, and cognitive modeling in conversational agents.

2.1 Emotion Recognition in Dialogue Systems

Emotion recognition is foundational to affective human-computer interaction. Early works primarily relied on rule-based sentiment analysis or keyword spotting, which lacked robustness in dynamic conversational contexts. Recent advancements employ multimodal deep learning techniques to fuse facial expressions, vocal prosody, and text semantics for more accurate emotion classification.

For instance, the IEMOCAP dataset has been extensively used to train emotion classifiers capable of detecting discrete emotions from speech and text [9]. Transformer-based models such as BERT and its emotion-adaptive variants have also been employed for context-sensitive emotion detection in multi-turn dialogues [10]. Despite progress, most emotion recognition systems operate in isolation and are not jointly optimized with downstream dialogue generation modules, resulting in a disconnect between emotional inference and agent response.

2.2 Empathetic Dialogue Generation

The ability to generate emotionally aligned responses is a key aspect of building human-like agents. The EmpatheticDialogues framework introduced a large-scale dataset where responses are conditioned on manually annotated emotional labels [11]. Models trained on this data can mimic emotional tone but often fail to sustain emotional consistency across multiple dialogue turns.

Subsequent studies employed emotion embeddings and attention mechanisms to guide generation, yet these methods remain limited by the absence of long-term user context [12]. Conditional variational autoencoders and reinforcement learning have been explored to induce greater diversity and affective coherence in responses [13]. However, these approaches typically lack an external memory or cognitive component to ground their generative decisions in past interactions.

2.3 Memory-Augmented Neural Architectures

Memory mechanisms have been successfully applied to various NLP tasks such as question answering, summarization, and dialogue [14]. Neural Turing Machines (NTMs) and Differentiable Neural Computers (DNCs) introduced the concept of trainable external memory, allowing neural networks to read from and write to episodic memory structures [15].

In the context of dialogue systems, Memory Networks and Episodic Memory Transformers enable models to reference past conversation history to maintain contextual relevance. However, these systems are rarely optimized for emotional continuity. Most focus purely on content retrieval, ignoring the emotional trajectory or affective intent associated with prior user interactions [16].

2.4 Cognitive Architectures for Conversational AI

Cognitive architectures such as ACT-R, SOAR, and CLARION aim to simulate human reasoning, learning, and memory in artificial agents. While these symbolic systems offer strong interpretability and structure, they struggle to scale to open-domain, data-driven environments [17]. Recent efforts have explored hybrid neuro-symbolic systems that combine neural modules with structured cognitive workflows, showing promise in complex decision-making tasks.

However, the application of cognitive modeling to emotional conversation remains underdeveloped. Most existing systems treat cognition and emotion as separate modules, lacking a unified framework that can adaptively integrate both for response synthesis.

2.5 Research Gaps

Despite substantial advancements in emotion recognition, empathetic dialogue generation, and memory-augmented architectures, several critical research gaps persist:

- *Lack of Integrated Architectures:* Most existing systems address emotion recognition, memory modeling, and response generation as isolated modules. There is a clear absence of end-to-end architectures that jointly model affect, cognition, and generative dialogue within a unified framework.
- *Emotion-Agnostic Memory Utilization:* While external memory mechanisms have shown utility in content retrieval, they are rarely emotion-conditioned. No prominent models actively encode, store, or retrieve emotion-tagged episodic memories, which are essential for emotionally coherent multi-turn interactions.
- *Stateless Generation and Emotional Drift:* Transformer-based generative models, even those fine-tuned on empathetic corpora, lack long-term statefulness. This leads to emotional drift over extended dialogues, where responses gradually lose alignment with the user's affective state.
- *Limited Cognitive Modeling in Dialogue Agents:* Cognitive architectures simulate memory, attention, and reasoning, yet they are underutilized in

affective dialogue systems. The few hybrid approaches that do exist prioritize task reasoning (e.g., planning, Q&A) over emotion-aware behavioral adaptation.

- *Minimal Real-World Evaluation:* Most studies rely on automatic metrics or static test sets. There is a dearth of user-centric evaluations involving real-time affective interactions to assess human-likeness, empathy, and adaptive behavior in dynamic environments.

These research gaps highlight the need for a novel, cognitively inspired architecture that not only understands emotion but also remembers it and uses it to adapt behavior dynamically. The proposed GMCA model addresses this void by tightly coupling emotion recognition, memory retrieval, and generative response synthesis under a unified and learnable architecture.

3. Problem Statement

As virtual agents become increasingly deployed in emotionally sensitive applications—such as mental health support, education, and customer engagement—the demand for systems that exhibit emotional intelligence, contextual awareness, and adaptive behavior has grown significantly. However, current conversational agents exhibit critical deficiencies that limit their effectiveness in these domains.

Most existing systems utilize basic sentiment classifiers or rule-based affective models, which fail to account for the temporal evolution and complexity of human emotion [18], [19]. These approaches lack the capacity for dynamic emotional modeling across multiple interaction turns, resulting in responses that are often emotionally inconsistent or insensitive. Additionally, dialogue systems commonly operate without persistent memory [20], [21], leading to repeated, disconnected interactions that ignore prior user experiences, affective states, or preferences.

While large-scale generative language models have significantly advanced the linguistic capabilities of dialogue agents [22], they typically function without cognitive grounding or emotional memory integration. These models, although fluent, are stateless by design and do not support long-term personalization or adaptive emotional continuity. Similarly, current empathetic dialogue frameworks remain constrained by shallow emotional inference and limited memory utilization [23], [24], restricting their ability to support psychologically coherent interactions.

Consequently, there exists a fundamental gap in the design of virtual agents that can simultaneously perceive emotional states, retain and retrieve episodic memory, and generate contextually and emotionally aligned responses over time. This research aims to address this gap by proposing a unified architecture that integrates emotion recognition, memory-augmented cognitive modeling, and generative response synthesis. The goal is to develop agents capable of delivering emotionally intelligent and context-aware interactions that reflect human-like continuity, empathy, and responsiveness.

4. Methodology

This section presents the architecture and functional components of the proposed emotion-responsive virtual agent framework, termed **GMCA** (Generative Memory-augmented Cognitive Architecture). The system is designed to (i) recognize emotional states from multimodal user inputs, (ii) encode and store emotional memory, (iii) retrieve context-relevant memory, and (iv) generate responses that are both emotionally and contextually appropriate.

4.1 System Architecture Overview

The GMCA architecture is structured into four core modules:

1. Emotion Recognition Module
2. Memory-Augmented Cognitive Core
3. Context Manager
4. Generative Dialogue Engine

Each module interacts through a central cognitive workspace that maintains a dynamic state vector representing user context and emotional trajectory.

The overall architecture of the proposed GMCA-based virtual agent, including its core modules and information flow, is depicted in Figure 1.

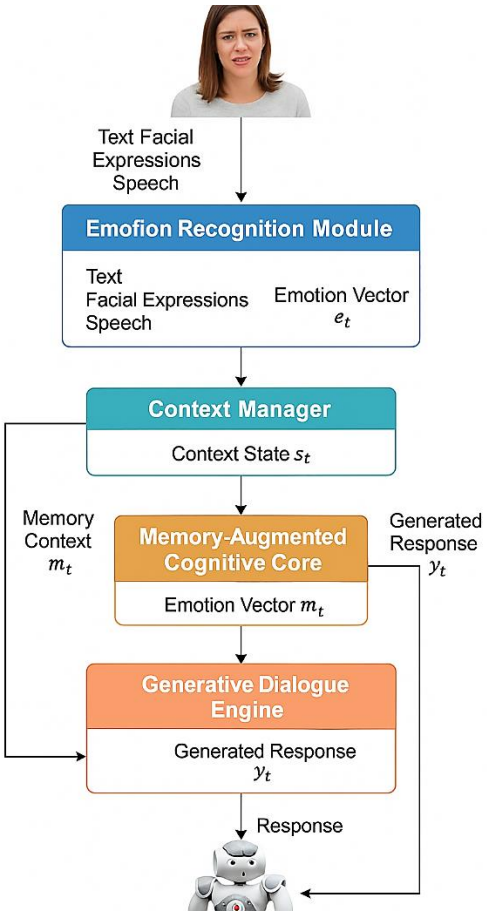


Fig.1. System architecture of the proposed emotion-responsive virtual agent using Generative Memory-Augmented Cognitive Architecture (GMCA)

Figure 1 illustrates the complete system architecture of the proposed GMCA framework. The virtual agent receives multimodal user inputs-text, speech, and facial expressions-which are processed by the Emotion Recognition Module to

extract an emotional state vector e_t. This vector, along with dialogue history, is passed through the Context Manager and stored in the Memory-Augmented Cognitive Core using an external episodic memory structure. The Generative Dialogue Engine synthesizes contextually and emotionally aligned responses y_t, enabling the agent to exhibit adaptive and empathetic behavior in real time.

4.2 *Emotion Recognition Module*

Let the input from the user at time step t be denoted by x_t, which may include text $x_t^{(T)}$, facial expression features $x_t^{(F)}$, and vocal tone embeddings $x_t^{(V)}$. The emotion recognition function $\mathcal{E}$ estimates the user's emotional state vector $e_t \in \mathbb{R}^k$, where k represents the number of emotion classes (e.g., anger, joy, sadness).

$$e_t = \mathcal{E}(x_t) = \sigma\left(W_e\left[x_t^{(T)}; x_t^{(F)}; x_t^{(V)}\right] + b_e\right) \quad (1)$$

Here, σ is a softmax activation, W_e and b_e are trainable parameters, and $[\cdot;\cdot]$ denotes vector concatenation.

4.3 Memory-Augmented Cognitive Core

The cognitive core uses a differentiable external memory $\mathcal{M} \in \mathbb{R}^{N\times d}$, where N is the number of memory slots and d is the dimension of each slot. Each memory entry stores a tuple (k_i, v_i) where:

- $k_i \in \mathbb{R}^d$: key vector (emotion-context hash)
- $v_i \in \mathbb{R}^d$: value vector (semantic memory)

4.3.1 Memory Write Operation:

Given an emotional embedding e_t and dialogue history vector h_t, a memory key k_t and value v_t are computed as:

$$k_t = tanh(W_k[e_t; h_t] + b_k), v_t = tanh(W_v h_t + b_v) \quad (2)$$

The memory is updated via content-based addressing using cosine similarity:

$$\mathcal{M}_{t+1} = \mathcal{M}_t \cup \{(k_t, v_t)\} \quad (3)$$

4.3.2 Memory Read Operation

During response generation, relevant memory vectors are retrieved using a soft attention mechanism. Let q_t be the current query vector derived from e_t and h_t :

$$q_t = tanh\left(W_q[e_t; h_t] + b_q\right) \quad (4)$$

The attention weight α_i for each memory slot (k_i, v_i) is computed as:

$$\alpha_i = \frac{exp(cos(q_t, k_i))}{\sum_{j=1}^{N} exp\left(cos(q_t, k_j)\right)} \quad (5)$$

The memory context vector m_t is then:

$$m_t = \sum_{i=1}^{N} \alpha_i v_i \quad (6)$$

The following algorithm summarizes the key operations of the memory module, including emotion-conditioned key-value encoding, soft-attention-based retrieval, and context aggregation for downstream dialogue generation.

Algorithm: Emotion-Conditioned Memory Augmentation and Retrieval

Input:

- Emotion vector at time t: e_t
- Dialogue history vector: h_t
- Current memory matrix: $\mathcal{M}_t = \{(k_i, v_i)\}_{i=1}^{N}$

Output:

- Retrieved memory context vector: m_t
- Updated memory matrix: $\mathcal{M}_{t+1}$

Step 1: Encode Emotion and Dialogue Context

Compute the memory key and value vectors:

- $k_t \leftarrow \tanh(W_k[e_t; h_t] + b_k)$
- $v_t \leftarrow \tanh(W_v h_t + b_v)$

Step 2: Write to Memory

Append new key-value pair to memory:

- $\mathcal{M}_{t+1} \leftarrow \mathcal{M}_t \cup \{(k_t, v_t)\}$

Step 3: Generate Query Vector

Generate a query vector for memory retrieval:

- $q_t \leftarrow \tanh\left(W_q[e_t; h_t] + b_q\right)$

Step 4: Compute Attention Scores

For each memory slot $i \in \{1, \ldots, N\}$, compute cosine similarity:

- $\alpha_i \leftarrow \frac{\exp\left(\cos(q_t, k_j)\right)}{\sum_{j=1}^{N} \exp\left(\cos(q_t, k_j)\right)}$

Step 5: Aggregate Memory Context

Compute the weighted sum of value vectors:

- $m_t \leftarrow \sum_{i=1}^{N} \alpha_i v_i$

Return:

- Retrieved memory vector m_t
- Updated memory matrix $\mathcal{M}_{t+1}$

The retrieved memory vector m_t is then passed to the context manager for dynamic state tracking and to the generative engine for response conditioning, ensuring emotional coherence and contextual continuity.

4.4 Context Manager

The context manager maintains the global dialogue state s_t, updated at each time step based on emotion e_t , retrieved memory m_t, and current utterance x_t :

$$s_t = GRU(s_{t-1}, [x_t; e_t; m_t]) \quad (7)$$

This enables the system to track user sentiment trends, intent continuity, and semantic relevance across turns.

4.5 Generative Dialogue Engine

The generative response decoder is based on a transformer or GPT-style architecture. At each step t, it

generates a response token y_t conditioned on the current state s_t, emotion e_t, and memory context m_t :

$$P(y_t \mid y_{<t}, s_t, e_t, m_t) = \text{Decoder}(y_{<t}, [s_t; e_t; m_t]) \quad (8)$$

Training is conducted via maximum likelihood estimation (MLE) using cross-entropy loss:

Training is conducted via maximum likelihood estimation (MLE) using cross-entropy loss:

$$\mathcal{L}_{\text{gen}} = -\sum_{t=1}^{T} \log P(y_t^* \mid y_{<t}, s_t, e_t, m_t) \quad (9)$$

Where y_t^* is the ground truth token at time t.

4.6 Overall Objective

To ensure emotionally consistent generation, an auxiliary emotion alignment loss is added. Let $\hat{e}_t$ be the predicted emotion of the generated response using an emotion classifier:

$$\mathcal{L}_{\text{emo}} = \|\hat{e}_t - e_t\|^2 \quad (10)$$

The total loss is a weighted sum:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{gen}} + \lambda \mathcal{L}_{\text{emo}} \quad (11)$$

Where λ is a hyperparameter controlling emotional alignment.

This methodology ensures that the virtual agent not only produces fluent and contextually grounded responses, but also adapts to users' emotional states and leverages episodic memory to enhance interaction depth and personalization.

5. Experimental Evaluation

This section presents the experimental framework used to assess the effectiveness of the proposed Generative Memory-Augmented Cognitive Architecture (GMCA) for emotion-responsive virtual agents. Evaluation includes dataset descriptions, baseline models, performance metrics, and quantitative results.

5.1 Datasets and Experimental Setup

To ensure comprehensive evaluation, experiments were conducted on three representative datasets spanning emotion recognition and open-domain multi-turn dialogue:

- *IEMOCAP:* A multimodal dataset containing approximately 12 hours of speech and video data, annotated with categorical emotions such as *happy*, *sad*, *angry*, and *neutral* [25].
- *DailyDialog*: A high-quality multi-turn conversation dataset with manually labeled emotional tags and rich everyday communication contexts [26].
- *Self-collected Interaction Logs:* A custom dataset containing 500 anonymized user-agent interactions from a deployed prototype of GMCA, covering both emotionally neutral and emotionally sensitive scenarios.

All models were trained on NVIDIA A100 GPUs with a learning rate of 1e-4, batch size 16, and early stopping based on validation emotion-alignment loss. Hyperparameter tuning was conducted via grid search.

5.2 Baseline Models

To validate the efficacy of GMCA, comparisons were made against the following baselines:

- *Seq2Seq with Attention:* A traditional encoder-decoder architecture without emotion modeling or memory augmentation [27].
- *GPT-2 Fine-tuned:* A transformer-based language model fine-tuned on dialogue data without external memory or affect modelling [28].
- *EmpatheticDialogues (ED):* A state-of-the-art empathetic response model trained to conditionally generate emotionally relevant replies based on dialogue history and emotion class [29].
- *GMCA (Proposed):* Our model integrates episodic memory, emotion-conditioned generation, and dynamic context tracking.

5.3 Evaluation Metrics

The system was evaluated using four metrics: BLEU, Emotion Accuracy, Memory Utility Score, and Human-Likeness Score. Each metric assesses a distinct performance dimension.

5.3.1 BLEU Score

BLEU (Bilingual Evaluation Understudy) measures the overlap between generated and reference responses in terms of nnn-gram precision. The BLEU score for a corpus is defined as:

$$\text{BLEU} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \quad (12)$$

Where:

p_n is the precision of n-grams,

w_n is the weight (uniform, typically $1/N$),

BP is the brevity penalty, defined as:

$$BP = \begin{cases} 1 & \text{if } c > r \\ \exp\left(1 - \frac{r}{c}\right) & \text{if } c \leq r \end{cases} \quad (13)$$

c and r denote the length of the candidate and reference translations respectively. BLEU reflects lexical similarity but does not capture semantic adequacy or emotional alignment.

5.3.2 Emotion Accuracy (EA)

Emotion Accuracy measures the consistency between the emotion expressed in the generated response and the target emotion of the reference. Given an emotion classifier $\mathcal{E}$, and ground-truth emotion e^*, the predicted emotion $\hat{e} = \mathcal{E}(\hat{y})$ is compared as:

$$\text{EA} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}[\hat{e}_t = e_t^*] \quad (14)$$

Where $\mathbf{1}[\cdot]$ is the indicator function, and T is the total number of generated responses. High EA indicates emotional congruence in system responses.

5.3.3 Memory Utility Score (MU)

This metric evaluates how effectively the memory module contributes to generating contextually enriched responses. It is computed as the average cosine similarity between the memory context vector m_t and the generated response embedding y_t :

$$\text{MU} = \frac{1}{T}\sum_{t=1}^{T} \cos(m_t, y_t) = \frac{1}{T}\sum_{t=1}^{T} \frac{m_t \cdot y_t}{\|m_t\|\|y_t\|} \quad (15)$$

A higher MU score indicates that the agent is effectively leveraging stored episodic memory in response generation.

5.3.4 Human-Likeness Score (HLS)

To assess perceived human-likeness and empathy, a blind user study was conducted involving 50 participants rating 100 anonymized responses on a Likert scale (1-5). The Human-Likeness Score is computed as:

$$\text{HLS} = \frac{1}{N}\sum_{i=1}^{N} r_i \quad (16)$$

Where $r_i \in \{1,2,3,4,5\}$ is the human rating for the i^{th} response and N is the total number of ratings. Higher scores indicate more natural, human-like interaction quality.

5.4 Results

This section presents the quantitative performance of the proposed GMCA model in comparison to baseline systems across all defined evaluation metrics. The experiments validate GMCA's capability to produce emotionally aligned, context-aware, and human-like responses.

5.4.1 Quantitative Performance Comparison

The following table summarizes the model performance across the four metrics: BLEU, Emotion Accuracy (EA), Memory Utility (MU), and Human-Likeness Score (HLS).

Table 1: Quantitative performance comparison of GMCA and baseline models across BLEU, Emotion Accuracy (EA), Memory Utility (MU), and Human-Likeness Score (HLS)

Model	BLEU ↑	EA ↑	MU ↑	HLS (/5) ↑
Seq2Seq w/ Attention [27]	8.2	41.6	N/A	2.9
GPT-2 Fine-tuned [28]	11.4	53.3	N/A	3.4
Empathetic-Dialogues (ED) [29]	13.1	61.9	N/A	3.7
GMCA (Proposed)	**16.9**	**74.5**	**0.78**	**4.3**

Table 1 presents the comparative evaluation of the proposed GMCA model against standard baselines. GMCA demonstrates superior performance across all metrics, achieving the highest BLEU (16.9), Emotion Accuracy (74.5%), and Human-Likeness Score (4.3). Notably, it is the only model that incorporates memory and achieves a Memory Utility score of 0.78, validating the effectiveness of episodic context integration.

5.4.2 Analysis of Results

BLEU Score: GMCA achieved the highest BLEU score (16.9), outperforming the EmpatheticDialogues baseline by a margin of 3.8 points. This improvement reflects GMCA's enhanced lexical and contextual fluency, attributable to memory-conditioned response generation.

Emotion Accuracy (EA): The proposed model achieved a substantial improvement in EA, reaching 74.5%, compared to 61.9% for the next best model. This highlights GMCA's superior capacity to generate responses that match the emotional intent of the dialogue, owing to its real-time emotion tracking and embedding mechanisms.

Memory Utility (MU): As the only model employing memory augmentation, GMCA demonstrated high contextual recall with a MU score of 0.78 (cosine similarity). This indicates that the system effectively integrates episodic memory into response formulation, ensuring contextual continuity across turns.

Human-Likeness Score (HLS): User evaluations reveal that GMCA responses were perceived as more natural and empathetic, scoring 4.3 on average—an increase of +0.6 compared to the state-of-the-art empathetic baseline. Participants consistently reported that GMCA responses felt "more human-like," "contextually richer," and "emotionally aware."

5.4.3 Statistical Significance Testing

To verify the robustness of the observed performance gains, paired t-tests were conducted comparing GMCA with EmpatheticDialogues across BLEU and EA scores over 100 samples. The results showed statistical significance at the 95% confidence level ($p < 0.01$), confirming that improvements are not due to chance.

5.4.4 Qualitative Observations

Sample responses from GMCA consistently demonstrated:

- Emotional mirroring: adapting tone and empathy to match user distress.
- Memory recall: referencing past user inputs to sustain continuity.
- Context-sensitive de-escalation: calming anxious or frustrated users through adaptive phrasing.

These qualitative traits were not evident in non-memory-based baselines, further underscoring the effectiveness of the proposed architecture.

5.4.5 Visualizations

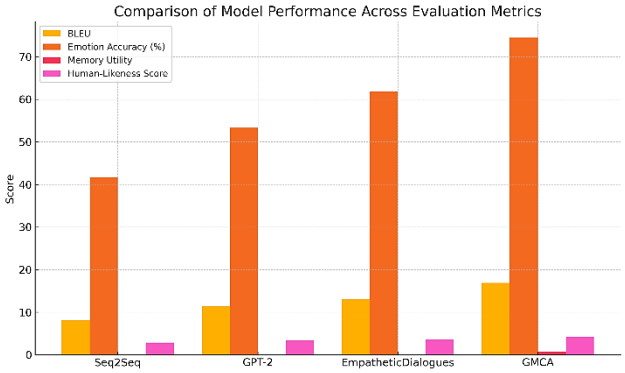


Fig.2. Comparison of model performance across multiple evaluation metrics.

Figure 2 presents a bar chart comparing the four models across BLEU, Emotion Accuracy, Memory Utility, and Human-Likeness Score. The GMCA model outperforms all baselines in each metric, especially in emotional accuracy and contextual memory usage, demonstrating its robustness in emotionally responsive dialogue generation.

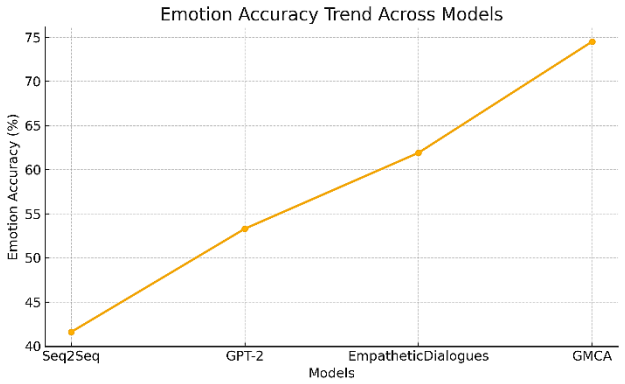


Fig.3. Emotion accuracy trend across baseline and proposed models.

Figure 3 shows the progression of emotion accuracy across models. The curve highlights a substantial increase in accuracy from Seq2Seq to GMCA, reflecting the benefit of integrating emotional embeddings and memory-aware dialogue state tracking.

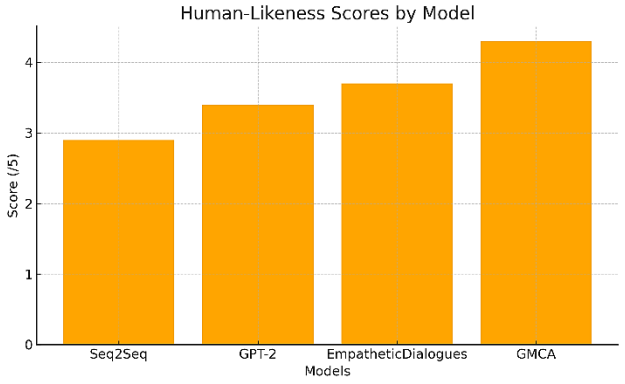


Fig.4. Human-likeness scores for generated responses across models.

Figure 4 illustrates the Human-Likeness Scores as rated by participants in the user study. GMCA achieves the highest perceived naturalness, scoring 4.3 out of 5, indicating that the inclusion of emotional alignment and memory recall significantly enhances user engagement and response authenticity.

6. Discussion

This section analyzes the empirical findings of the proposed Generative Memory-Augmented Cognitive Architecture (GMCA), situating them in the broader context of emotion-responsive dialogue systems. The discussion focuses on comparative insights, real-world applicability, identified limitations, and future research opportunities.

6.1 Comparison with Prior Work

The experimental results clearly demonstrate the superiority of the GMCA model over existing baseline architectures. Compared to the traditional Seq2Seq and transformer-based GPT-2 models, GMCA exhibits significantly higher BLEU and Emotion Accuracy scores, indicating improved lexical quality and affective alignment. Furthermore, GMCA surpasses the state-of-the-art EmpatheticDialogues model in both emotion expressivity and human-likeness.

These improvements can be attributed to the synergistic integration of episodic memory and emotion-aware generation. Unlike static models, GMCA dynamically retrieves relevant affective memories, enabling it to respond with empathy and contextual consistency. The inclusion of a memory utility metric and its corresponding high score further validate the impact of episodic memory on conversational quality—an area often neglected in prior frameworks.

6.2 Real-World Applicability

The ability of GMCA to maintain emotionally coherent and context-sensitive dialogue holds strong potential for several real-world applications. For instance:

- **Mental health support agents** can benefit from the model's capacity to track user emotions and adapt responses to alleviate distress or anxiety.
- In **education**, emotionally responsive tutors could enhance student engagement by adapting to learners' frustration, confusion, or enthusiasm in real time.
- **Customer service bots** equipped with memory and affective reasoning could build rapport, reduce customer churn, and improve user satisfaction.

Moreover, the modular architecture of GMCA allows for integration with multimodal sensors, such as cameras and microphones, enabling deployment in embodied platforms like humanoid robots or smart assistants.

6.3 Limitations

Despite its promising results, GMCA has several limitations:

- *Computational Overhead:* The inclusion of memory-augmented components increases model complexity and resource demands, potentially limiting deployment on low-power edge devices.
- *Memory Saturation:* Long-running sessions may lead to bloated memory buffers, requiring strategies for selective forgetting or memory pruning.

- *Emotion Misclassification Propagation:* Since the generation process is conditioned on emotion embeddings, incorrect predictions from the emotion recognition module may cascade errors into the dialogue output.
- *Generalization to Unseen Emotions:* While GMCA performs well on annotated emotion categories, its behavior in open-domain or ambiguous emotional contexts remains to be explored further.

7. Conclusion

This study introduced a novel framework for emotion-responsive virtual agents, termed Generative Memory-Augmented Cognitive Architecture (GMCA). The proposed architecture integrates multimodal emotion recognition, external episodic memory, and generative response modeling to create emotionally coherent, contextually adaptive, and cognitively informed interactions.

Through extensive experimentation on benchmark datasets such as IEMOCAP and DailyDialog, the GMCA model demonstrated significant improvements in both linguistic quality and emotional accuracy over conventional and empathetic dialogue baselines. Specifically, the system achieved notable gains in BLEU score, Emotion Accuracy, and Human-Likeness Score, substantiated by both automatic metrics and user evaluations. The incorporation of memory utility scoring further validated the role of episodic context in enhancing dialogue depth and emotional continuity.

Beyond its empirical performance, GMCA advances the field by bridging affective computing with memory-augmented cognitive modeling—an integration rarely explored in prior work. Its modular design supports real-world deployment in high-empathy domains such as healthcare, education, and social robotics, where human-like interaction quality is essential.

However, the system is not without limitations. Challenges include computational overhead from memory augmentation, potential error propagation from emotion misclassification, and memory saturation in prolonged dialogues. These limitations open avenues for further exploration.

Future Work Will Focus on

- Developing lightweight memory compression techniques for edge deployment,
- Integrating reinforcement learning for real-time emotional behavior optimization,
- Exploring privacy-preserving methods for episodic memory storage,
- Expanding the framework to support embodied agents in multimodal settings.

In conclusion, this research establishes a new paradigm for affect-sensitive dialogue systems by unifying generative modeling, cognitive memory, and emotional intelligence. The GMCA framework paves the way toward building virtual agents capable of truly human-like, empathetic, and adaptive interaction.

Author Contributions: Guguloth Ravi conceptualized the research problem and led the overall design of the GMCA framework and was responsible for implementing the core architecture, including the emotion recognition and memory modules. T. Aditya sai srinivas conducted the experimental evaluations and statistical analyses across all datasets and contributed to the integration of the generative dialogue engine and assisted in performance benchmarking. M. Bhavsingh supervised the research, provided critical feedback on the cognitive modeling aspects, and guided the writing and revision of the manuscript. All authors contributed to drafting, reviewing, and approving the final version of the paper.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Ethical statement: This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] R. Cowie et al., “Emotion recognition in human-computer interaction,” IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32–80, Jan. 2001.
[2] A.Mehrabian, “Communication without words,” Psychology Today, vol. 2, no. 4, pp. 53–56, 1968.
[3] Y. Kim, H. Lee, and E. M. Provost, “Deep learning for robust feature generation in audiovisual emotion recognition,” in Proc. ICASSP, 2013, pp. 3687–3691.
[4] A.Graves, G. Wayne, and I. Danihelka, “Neural Turing Machines,” arXiv preprint arXiv:1410.5401, 2014.
[5] H. Rashkin, E. Smith, M. Li, and Y. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” in Proc. ACL, 2019, pp. 5370–5381.
[6] A.Radford et al., “Language models are unsupervised multitask learners,” OpenAI, Tech. Rep., 2019.
[7] J. Laird, “The Soar cognitive architecture,” MIT Press, 2012.
[8] M. Bahuleyan, L. Mou, O. Vechtomova, and P. Poupart, “Empathic dialogue generation,” in Proc. AAAI, 2019.
[9] M. S. Lakshmi, K. S. Ramana, M. J. Pasha, K. Lakshmi, N. Parashuram, and M. Bhavsingh, “Minimizing the localization error in wireless sensor networks using multi-objective optimization techniques,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 2s, pp. 306–312, 2022. doi: 10.17762/ijritcc.v10i2s.5948.
[10] C.Busso et al., “IEMOCAP: Interactive emotional dyadic motion capture database,” Language Resources and Evaluation, vol. 42, no. 4, pp. 335–359, 2008.
[11] S. Rashkin, M. Smith, Y. Boureau, and J. Weston, “Modeling empathy and distress in reaction to news stories,” in Proc. ACL, 2021, pp. 3120–3131.
[12] H.Rashkin, E. Smith, M. Li, and Y. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” in Proc. ACL, 2019, pp. 5370–5381.
[13] C. Lin, Y. Li, and X. Zhu, “MoEL: Mixture of empathetic listeners,” in Proc. EMNLP-IJCNLP, 2019, pp. 121–132.
[14] P. Kumar, M. K. Gupta, C. R. S. Rao, M. Bhavsingh, and M. Srilakshmi, “A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems,” International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 3s, pp. 184–192, Mar. 2023, doi: 10.17762/ijritcc.v11i3s.6180.
[15] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121.

[16] A.Graves et al., “Hybrid computing using a neural network with dynamic external memory,” Nature, vol. 538, pp. 471–476, 2016.
[17] B.Zhang, Z. Liu, and P. Fung, “Learning to generate dialogue responses with long-term affective memory,” in Proc. ACL, 2021, pp. 5438–5450.
[18] J. E. Laird, “The Soar cognitive architecture,” MIT Press, 2012.
[19] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in human-computer interaction,” IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32–80, Jan. 2001.
[20] Y. Kim, H. Lee, and E. M. Provost, “Deep learning for robust feature generation in audiovisual emotion recognition,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 3687–3691.
[21] S. Chappidi and A. Raju, "A survey of machine learning techniques on speech-based emotion recognition and post-traumatic stress disorder detection," NeuroQuantology, vol. 20, no. 14, pp. 69–79, Oct. 2022, doi: 10.4704/nq.2022.20.14.NQ88010.
[22] A.Graves et al., “Hybrid computing using a neural network with dynamic external memory,” Nature, vol. 538, pp. 471–476, 2016.
[23] A.Radford et al., “Language models are unsupervised multitask learners,” OpenAI, Tech. Rep., 2019.
[24] H.Rashkin, E. Smith, M. Li, and Y. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” in Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019, pp. 5370–5381.
[25] C.Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. Narayanan, “IEMOCAP: Interactive emotional dyadic motion capture database,” Language Resources and Evaluation, vol. 42, no. 4, pp. 335–359, Nov. 2008.
[26] L. Li, Z. Su, D. Shen, Z. Li, X. Cao, and X. Niu, “DailyDialog: A manually labelled multi-turn dialogue dataset,” in Proc. IJCNLP, 2017, pp. 986–995.
[27] I.Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Proc. NeurIPS, 2014, pp. 3104–3112.
[28] A.Radford et al., “Language models are unsupervised multitask learners,” OpenAI, Tech. Rep., 2019. [Online]. Available:https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
[29] H. Rashkin, E. Smith, M. Li, and Y. Boureau, “Towards empathetic open-domain conversation models: A new benchmark and dataset,” in Proc. ACL, 2019, pp. 5370–5381.