Facilitating Early Maladaptive Schema–Guided Polite and Empathetic Psychotherapeutic Support: An LLM-Driven MoE-RL-Based Dialogue System

Priyanshu Priya, Asif Ekbal
Department of Computer Science and Engineering, Indian Institute of Technology Patna, India
{priyanshu_2021cs26,asif}@iitp.ac.in

Abstract

In Psychotherapy, Early Maladaptive Schemas (EMS) are entrenched negative perceptions of self or others that perpetuate mental health challenges, contribute to treatment resistance and relapse, and obstruct therapeutic progress. Addressing EMS using appropriate psychotherapeutic support (PS) strategies helps resolve core emotional deficits, mitigate resistance, and improve client engagement. Moreover, adapting polite and empathetic communication based on clients' emotional states fosters trust, emotional safety, and a conducive therapeutic environment, which is critical for addressing EMS and achieving positive outcomes. Motivated by these insights, we introduce **MATE** - a novel E**M**S-guided polite and emp**A**thetic dialogue sys**T**em for psychoth**E**rapeutic support. **MATE** integrates a Large Language Model (LLM) with a Mixture of Experts-based Reinforcement Learning (MoE-RL) approach to overcome the limitations of traditional RL methods, such as large action spaces and generic responses. The LLM captures diverse semantic patterns from dialogue context. MoE-RL leverages dedicated psychotherapeutic, politeness, and empathy experts, along with a new reward function, comprising PS, politeness, empathy, contextual consistency, and diversity rewards to guide policy learning for effective response generation. Evaluations on the HOPE and PSYCON datasets demonstrate **MATE**'s efficacy in generating polite and empathetic psychotherapeutic responses based on clients' EMS and emotional cues while ensuring contextual consistency and diversity.

Introduction

In psychotherapy, Early Maladaptive Schemas (EMS) are enduring cognitive-emotional frameworks that originate from unmet core emotional needs in the past, shaping negative self-perceptions, interpersonal dysfunction, and maladaptive coping behaviors (Young, Klosko, and Weishaar 2006). These schemas act as barriers to mental health by perpetuating emotional distress, and reinforcing treatment resistance and relapse risk (Collins et al. 1990). EMS hinders psychotherapeutic progress by distorting experiences through a persistent negative lens (Young and Klosko 1994), making their targeted resolution critical to addressing core psychopathology and fostering sustainable recovery (Csigó, Münnich, and Molnar 2024).

Effective intervention necessitates psychotherapeutic support (PS) strategies tailored to clients' specific EMS (Wegener et al. 2013). For instance, individuals with an emotional deprivation schema may require nurturing support to counter ingrained feelings of unmet needs, while those with a failure schema benefit from competence-building interventions. Such EMS-guided PS strategies counteract maladaptive narratives, reduce resistance, and enhance engagement by addressing the unique emotional vulnerabilities associated with each schema.

Furthermore, individuals with EMS are often hypersensitive to perceived criticism or invalidation, which can evoke intense emotional reactions (Young, Klosko, and Weishaar 2006). Therefore, strategically incorporating polite and empathetic language attuned to clients' emotional state is essential to effectively address EMS and enhance psychotherapeutic outcomes. Politeness helps mitigate schema-driven defensiveness and regulate emotional reactivity by fostering a non-confrontational and non-judgmental environment (Budiarta et al. 2021). In parallel, empathetic communication promotes emotional validation and psychological safety, enabling clients to tackle distressing schema-related emotions and memories (Elliott et al. 2011).

Motivated by the need for courteous, affective, and schema-sensitive PS, we introduce **MATE**, a pioneering E**M**S-guided polite and emp**A**thetic dialogue sys**T**em for psychoth**E**rapeutic support that adapts PS strategies while ensuring politeness and empathy based on the clients' EMS and emotional cues. Such a system shows strong potential for AI-driven therapeutic support in high-demand settings like online mental health platforms and digital self-help tools, and contributes to the UN Sustainable Development Goals (SDG 3) by promoting mental well-being across all age groups (Heymann and Sprague 2023).

Recent dialogue systems employ reinforcement learning (RL) to model conversations as word-level generation over semantic representations. However, they face key challenges. First, the action space in RL-based dialogue generation is extremely large due to the innumerable possible language utterances, which makes policy learning unstable and leads to incoherent or suboptimal responses (Zhao, Xie, and Eskenazi 2019), leading to irrelevant responses, and fail to capture nuanced conversational goals, potentially misaligning both language quality (fluency, coherence, etc.) and goal accomplishment (task completion, user satisfaction, etc.) (Chow et al. 2022). Moreover, constructing the reward as a weighted sum of multiple objectives is suboptimal, as manually assigned weights may not accurately represent actual dialogue performance goals.

To address these challenges, we build **MATE** using a Large Language Model (LLM)-driven Mixture-of-Experts-based RL (MoE-RL) framework. This framework comprises (i) an LLM to capture diverse semantic cues from dialogue context for facilitating varied responses; (ii) three experts, *viz.* psychotherapeutic, politeness, and empathy to guide response generation; and (iii) an RL agent to dynamically select suitable experts via a learned expert identification policy. To generate consistent, diverse, polite, and empathetic responses with EMS-adaptive PS strategies, we introduce a composite reward function incorporating PS, politeness, empathy, contextual consistency, and diversity rewards.

To operationalize these components, we first design novel PS and empathy strategies grounded in the Motivational Interviewing framework (Miller and Rollnick 2012) and the two-dimensional empathy model (Davis et al. 1980), respectively. We then annotate two PS datasets, HOPE (Malhotra et al. 2022) and PSYCON (Mishra et al. 2023) with EMS labels (Young, Klosko, and Weishaar 2006), politeness strategies (Brown and Levinson 1987), emotions (Ekman 1992), and our proposed PS and empathy strategies through few-shot prompting of ChatGPT (OpenAI 2024) with human oversight, resulting in a comprehensive resource for analyzing psychological behaviors in therapeutic dialogue.

In summary, the key contributions are: (a) Investigate the impact of adapting PS strategies based on clients' EMS while incorporating politeness and empathy according to their emotional state in therapeutic dialogue. To the best of our knowledge, this work pioneers EMS-guided strategic modeling within psychotherapeutic interactions; (b) Propose psychologically-inspired PS and empathy strategies and annotate HOPE and PSYCON datasets with EMS, emotions, politeness, empathy, and PS strategies using LLM prompting and human-in-the-loop techniques; (c) Present **MATE**, LLM-driven MoE-RL framework to deliver EMS-guided, polite, and empathetic PS while ensuring contextual consistency and response diversity; (d) Devise novel experts and a reward function comprising PS, politeness, empathy, contextual consistency, and diversity rewards to generate contextually coherent, diverse, polite, and empathetic responses tailored to clients' EMS and emotions during psychotherapy[1].

Related Work

Psychotherapy has been extensively studied across diverse fields, including computational linguistics (Goldberg et al. 2020). Recently, human-agent psychotherapeutic dialogue systems have garnered significant attention as a promising approach for automated mental health support (Valizadeh and Parde 2022; Priya et al. 2023). Current advances in this domain have primarily focused on emotion and persona-aware modeling to create emotionally intelligent and personalized psychotherapeutic agents (Ghandeharioun et al. 2019; Mishra et al. 2023).

The advent of LLMs (Zhao et al. 2023) has further accelerated progress in developing automated psychotherapeutic support (Guo et al. 2024; Tu et al. 2025). Besides, Mixture-of-Experts (MoEs), which utilize n expert sub-networks, have become integral to enhancing the adaptability and performance of LLM-based systems (Gupta et al. 2022; Priya et al. 2025). More recently, the integration of politeness and empathy into psychotherapeutic dialogue systems (Saha et al. 2022; Mishra et al. 2023; Hossain et al. 2025) has emerged as a crucial research direction, given their proven effectiveness in enriching client experiences and facilitating positive therapeutic outcomes (Kim et al. 2018; Morris et al. 2018; Sharma et al. 2020). Politeness fosters respectful and considerate interactions, and empathy enables emotional attunement and support, thereby empowering agents to convey genuine concern and engagement akin to human therapists.

In clinical psychotherapy, EMS-guided interventions are central to addressing core emotional deficits and maladaptive behaviors for successful psychotherapeutic treatment (Wegener et al. 2013; Spicer et al. 2024). Despite significant advances in modeling clients' emotions and personas, the incorporation of EMS for guiding psychotherapeutic dialogue systems remains largely unexplored. To bridge this gap, we present a novel LLM-driven MoE-RL-based framework that combines the advantages of LLMs, MoEs, and RL to develop a robust EMS-guided polite and empathetic PS dialogue system. Following (Priya et al. 2025), we adopt a standard RL framework (Sutton and Barto 2018) and design novel experts and reward mechanisms tailored to our task. The proposed approach is scalable and adaptable in diverse therapeutic contexts and domains.

Dataset

The experiments are carried out on two PS dialogue datasets: HOPE (Malhotra et al. 2022) and PSYCON (Mishra et al. 2023). HOPE comprises 212 conversations ($\sim$12.8K utterances) extracted from counseling sessions involving one-on-one interactions between therapists and clients. These sessions, sourced from publicly available videos, span diverse demographic backgrounds and mental health topics. PSYCON comprises 1K therapist–client interactions ($\sim$25K utterances) addressing seven prevalent psychological conditions, *viz.* depression, anxiety, stress, bipolar disorder, disruptive and dissocial disorders, PTSD, and schizophrenia. These conversations are curated by a few-shot prompting of GPT-J (Wang and Komatsuzaki 2021), followed by manual oversight for quality control.

Dataset Annotation Scheme. We annotate the datasets along five key dimensions: *early maladaptive schemas* (EMS)*, emotions, politeness strategies, empathy strategies*, and *psychotherapeutic support* (PS) *strategies*. The clients' utterances are labeled with EMS and emotion tags, while therapists' utterances are annotated with politeness, empathy, and PS strategy labels.

Early Maladaptive Schemas and Emotions. In this work, we use 18 EMS labels derived from Schema Therapy

[1]Dataset, code, and appendix are available at https://github.com/priyanshu-profile/MATE; https://www.iitp.ac.in/~ai-nlp-ml/resources.html

Annotation Aspect	Labels
Early Maladaptive Schemas	Abandonment/Instability (AB), Mistrust/Abuse (MA), Emotional Deprivation (ED), Defectiveness/Shame (DS), Social Isolation/Alienation (SI), Dependence/Incompetence (DI), Vulnerability to Harm or Illness (VH), Enmeshment/Undeveloped Self (EM), Failure to Achieve (FA), Entitlement/Grandiosity (ET), Insufficient Self-Control/Self-Discipline (IS), Subjugation (SB), Self-Sacrifice (SS), Approval-Seeking/Recognition-Seeking (AS), Negativity/Pessimism (NP), Emotional Inhibition (EI), Unrelenting Standards/Hypercriticalness (US), and Punitiveness (PU)
Emotion	Anger, Fear, Joy, Sadness, Surprise, Disgust, Neutral
Politeness Strategy	Positive, Negative, Off-record
Empathy Strategy	Active Listening, Unconditional Positive Regard, Validation, Appropriate Self-Disclosure, Collaborative Approach, Managing Silence, Asking Empowering Questions, Holding Space for Emotions, Perspective Taking, and No Strategy
Psychotherapeutic Support Strategy	Open Questions, Reflections, Affirmations, Summarizing, Pros and Cons Chart, Learning Assessment, Elicit Change Talk, Readiness Rulers, Values and Strengths, Elicit Provide Elicit, and No Strategy

Table 1: Annotation aspects and their corresponding labels. Definitions and examples are given in 'Dataset Details' in appendix.

(Young, Klosko, and Weishaar 2006), and six emotions based on Ekman's emotions (Ekman 1992), extended to include an additional neutral category.

Politeness and Empathy Strategies. To facilitate respectful and emotionally attuned interactions during psychotherapy, it is essential for a PS dialogue system to incorporate both politeness and empathy strategies. The use of relevant politeness and empathy strategies helps mitigate potential threats to the clients' self-esteem while simultaneously conveying both emotional attunement and cognitive recognition of the individuals' circumstances. In this work, we adopt three core politeness strategies from Politeness Theory (Brown and Levinson 1987). However, due to the lack of well-defined empathy strategies for PS, we develop a taxonomy of 10 empathy strategies under the supervision of a domain specialist from a government-run institution, to capture emotional resonance and cognitive acknowledgment of clients' experiences. These strategies are informed by the two-dimensional empathy model (Davis et al. 1980) and a preliminary analysis of 60 randomly selected dialogues, each from the HOPE and PSYCON datasets. Three human annotators independently reviewed and annotated the sampled dialogues, resolved disagreements, and refined the strategies with the help of domain specialist. The κ (Fleiss 1971) score in the range of $0.48 < \kappa < 0.82$ for all categories indicates moderate to fair inter-annotator agreement (McHugh 2012).

Psychotherapeutic Support Strategies. Effective psychotherapy necessitates the structured development of PS strategies that directly address clients' underlying EMS. To this end, we propose a set of 11 psychotherapeutic support strategies grounded in the Motivational Interviewing model (Miller and Rollnick 2012), designed to foster motivation, uphold client autonomy, and strengthen a collaborative, client-centered therapeutic alliance. These strategies are designed in a way analogous to the formulation of empathy strategies. The κ scores ($0.58 < \kappa < 0.74$) across all categories, indicate substantial inter-annotator agreement.

Table 1 lists the EMS, emotion, politeness, empathy, and PS strategy labels.

Dataset Annotation Procedure. Motivated by recent progress in employing LLMs as annotation agents to reduce human annotation cost and effort (Gilardi, Alizadeh, and Kubli 2023), we prompt the ChatGPT (`GPT-4o-mini`) in a few-shot manner, accompanied by manual verification by human annotators for annotating the datasets with EMS, emotion, politeness, empathy, and PS strategy labels. The annotation proceeds through four steps:

1. **Initial Manual Annotation:** We randomly select 80 dialogues from each dataset and manually annotate them with EMS, emotion, politeness, empathy, and PS strategy labels following predefined guidelines under domain expert supervision. The resulting subsets are denoted as ANN-HOPE and ANN-PSYCON.
2. **Few-Shot Prompting for Strategy Prediction:** Using the manually annotated samples as exemplars, ChatGPT is prompted in a few-shot setup to predict EMS, emotion, politeness, empathy, and PS strategy labels for remaining dialogues. Each prediction is accompanied by a natural language explanation to improve transparency and interpretability. The labels and explanations for ANN-HOPE and ANN-PSYCON are generated using this approach. The prompt template is given in the 'Dataset Details' section of the appendix.
3. **Automated Verification of LLM Annotations:** To identify potentially incorrect LLM-generated labels, we develop 10 different RoBERTa-based (Liu et al. 2019) verifier models, one for each annotation aspect of the ANN-HOPE and ANN-PSYCON datasets. These verifiers serve as binary classifiers that determine whether an LLM-generated label $\hat{p}$ aligns with the ground-truth label p, based on the tuple $d = (u, p, \hat{p}, \hat{e})$, where u is the input utterance and $\hat{e}$ is the explanation. The verifier is modeled as $\mathbb{P}_\theta(w|u, \hat{p}, \hat{e})$, where $w \in \{0, 1\}$ and $w = 1$ indicates correct LLM annotation ($\hat{p} = p$)[2]. The utterances with incorrect predictions, as flagged by the verifiers, are selected for further review.
4. **Manual Re-annotation of Discrepant Samples:** The utterances flagged as potentially misclassified by the verifier models are re-annotated by annotators to construct high-quality gold-standard datasets. We obtain a substantial inter-annotator agreement with Fleiss' Kappa (κ) scores of $\langle 0.76, 0.85, 0.83, 0.81, 0.78\rangle$ and $\langle 0.79, 0.82, 0.84, 0.80, 0.77\rangle$ in Step 1, and $\langle 0.81, 0.87, 0.86, 0.83, 0.80\rangle$ and $\langle 0.83, 0.89, 0.88, 0.85, 0.82\rangle$ in Step 4 for the HOPE

[2]We achieve accuracies of 73.5%, 71.2%, 87.7%, 85.4%, and 89.9% on ANN-HOPE, and 88.3%, 80.1%, 78.6%, 83.7%, and 81.9% on ANN-PSYCON for EMS, emotion, politeness, empathy, and PS strategy classifiers, respectively.

and PSYCON datasets, respectively for EMS, emotions, politeness, empathy, and PS strategy labels, respectively.

Annotators Details. Among the three annotators involved in the dataset annotation, two hold Ph.D. degrees in Linguistics, and the third holds a Master's degree in Computer Science focused on Natural Language Processing. All annotators are proficient in English, experienced in annotation tasks, and well-versed in emotions, politeness, and empathy. All the annotators are briefed about the EMS labels thoroughly and foundational psychotherapeutic principles with the help of the domain expert. These annotators are compensated in accordance with the institute's guidelines.

Methodology

Problem Formulation

We define a psychotherapeutic support dialogue as $D = \{(c_v, t_v)\}_{v=1}^{N}$, where c and t represent turns from client and therapist, respectively. For a given a dialogue context $\mathcal{H} = \{c_1, t_1, \ldots, c_{n-1}, t_{n-1}\}$ (an alternating sequence of $(n-1)$ turns between client and therapist), and the target client turn c_n, the objective is to generate EMS-guided polite and empathetic psychotherapeutic support response $t_n(=r)$.

MATE - MoE-RL-based Dialogue System

The overall architecture of the proposed **MATE** framework is illustrated in Figure 1. It begins by encoding the dialogue context into a sequential representation of states via a dialogue context encoder, which serves as the semantic encoding policy. To capture therapeutic nuances, a mixture-of-experts module, specifically designed to model psychotherapeutic support, politeness, and empathy strategies, interprets the semantics of each state and outputs possible action candidates under the expert identification policy. The most relevant experts are then selectively activated to update the conversational state. Based on this updated state, the framework generates the therapist's response, which is subsequently evaluated and optimized through a well-designed set of rewards. These rewards assess the degree to which the response adheres to the desired psychotherapeutic support, politeness, and empathy strategies, while simultaneously promoting contextual consistency and response diversity, thereby ensuring both psychotherapeutic relevance and conversational naturalness.

Dialogue Context Encoder. We utilize the Phi-3-mini-128k-instruct model (Abdin et al. 2024) as the dialogue context encoder, denoted by $ENC(\cdot)$. The input sequence I constructed by prepending a special token [CLS] to $\mathcal{H}$ and c_n, that is, I = [CLS] $\oplus\ \mathcal{H} \oplus c_n$ is encoded by $ENC(\cdot)$ to yield hidden representations $HR_I = \{hr_{\text{[CLS]}}, hr_{c_1}, hr_{t_1}, \ldots, hr_{c_{n-1}}, hr_{t_{n-1}}, hr_{c_n}\}$. The representation of the [CLS] token ($hr_{\text{[CLS]}}$) serves as a contextual representation of the entire input sequence.

Multi-task Learning of Mixture-of-Experts. To model the nuanced behavioral dynamics in psychotherapeutic conversations, we introduce three dedicated experts, namely a *Psychotherapeutic Expert* (PE), a *Politeness Expert* (PoE), and an *Empathy Expert* (EE). Each expert is trained to predict the respective behavioral strategy in the therapist's future turn, guided by the dialogue history and the client's cognitive or affective cues. In particular, PE models the transition of therapists' psychotherapeutic support strategies (PSS) guided by the client's evolving Early Maladaptive Schemas (EMS) to enable schema-adaptive therapeutic support. PoE captures emotional shifts in clients' responses and adapts the politeness strategy (PoS) to maintain socially appropriate and emotionally attuned therapist behavior. EE learns empathy strategy (ES) transitions as a function of clients' emotional dynamics to support empathetic attunement in dialogue.

Each expert shares a unified processing approach, while differing in the type of target label it predicts and the conditioning client information it uses (e.g., EMS or emotions). Given the hidden representation HR_I obtained from the context encoder, each expert E $\in \{\text{PE}, \text{PoE}, \text{EE}\}$ first applies a task-specific transformation using a dedicated MLP:

$$
\begin{aligned}
HR_I^{\text{E}} &= \text{MLP}^{\text{E}}(HR_I) \\
&= \{hr_{\text{[CLS]}}^{\text{E}}, hr_{c_1}^{\text{E}}, hr_{t_1}^{\text{E}}, \ldots, hr_{c_n}^{\text{E}}\}
\end{aligned} \tag{1}
$$

Each expert models a sequence composed of alternating client cues (either EMS or emotions) and corresponding therapist strategy labels (either PSS, PoS, or ES). These sequences are defined as:

$$S^{\text{PE}} = \{\text{ems}_1, \text{pss}_1, \ldots, \text{ems}_{n-1}, \text{pss}_{n-1}, \text{ems}_n\} \tag{2}$$

$$S^{\text{PoE}} = \{\text{emo}_1, \text{pos}_1, \ldots, \text{emo}_{n-1}, \text{pos}_{n-1}, \text{emo}_n\} \tag{3}$$

$$S^{\text{EE}} = \{\text{emo}_1, \text{es}_1, \ldots, \text{emo}_{n-1}, \text{es}_{n-1}, \text{emo}_n\} \tag{4}$$

Here, ems_i and emo_i represent the EMS and emotion of the client at the i^{th} client turn, respectively, and pss_i, pos_i, and es_i denote the PSS, PoS, and ES adopted by the therapist at the i^{th} turn, respectively. We randomly initialize trainable embeddings $h_{x_j} \in \mathbb{R}^d$ for each sequence element $x_j \in \{\text{ems}_j, \text{pss}_j, \text{emo}_j, \text{pos}_j, \text{es}_j\}_{j=1}^{n}$, resulting in respective embedding sequences:

$$\mathbf{H}^{\text{PE}} = \{h_{\text{ems}_1},\ h_{\text{pss}_1},\ \ldots,\ h_{\text{ems}_n}\} \tag{5}$$

$$\mathbf{H}^{\text{PoE}} = \{h_{\text{emo}_1},\ h_{\text{pos}_1},\ \ldots,\ h_{\text{emo}_n}\} \tag{6}$$

$$\mathbf{H}^{\text{EE}} = \{h_{\text{emo}_1},\ h_{\text{es}_1},\ \ldots,\ h_{\text{emo}_n}\} \tag{7}$$

Next, we concatenate the contextual embedding of the [CLS] token with the respective sequence embedding to form the combined expert-specific context representation:

$$h_{\text{CTX-E}} = hr_{\text{[CLS]}}^{\text{E}} \oplus \mathbf{H}^{\text{E}}, \quad \text{for E} \in \{\text{PE}, \text{PoE}, \text{EE}\} \tag{8}$$

These concatenated representations are passed through a mean pooling layer to compute fixed-size context-aware embeddings:

$$\bar{h}_{\text{CTX-E}} = \text{MeanPool}(h_{\text{CTX-E}}) \tag{9}$$

Following Tu et al. (2022), we model the temporal dynamics in each sequence using a dedicated GRU per expert. The GRU takes the pooled context representation and previous hidden state as input, producing the final expert-specific contextual state for the n^{th} turn:

$$\mathbf{S}_n^{\text{E}} = \text{GRU}^{\text{E}}(\mathbf{S}_{n-1}^{\text{E}}, \bar{h}_{\text{CTX-E}}) \tag{10}$$

We then project each GRU output through a linear transformation followed by softmax activation to predict the probability distribution over the respective strategy label:

$$\mathbb{P}^{\text{E}} = \text{softmax}(W^{\text{E}}\mathbf{S}_n^{\text{E}} + b^{\text{E}}) \tag{11}$$

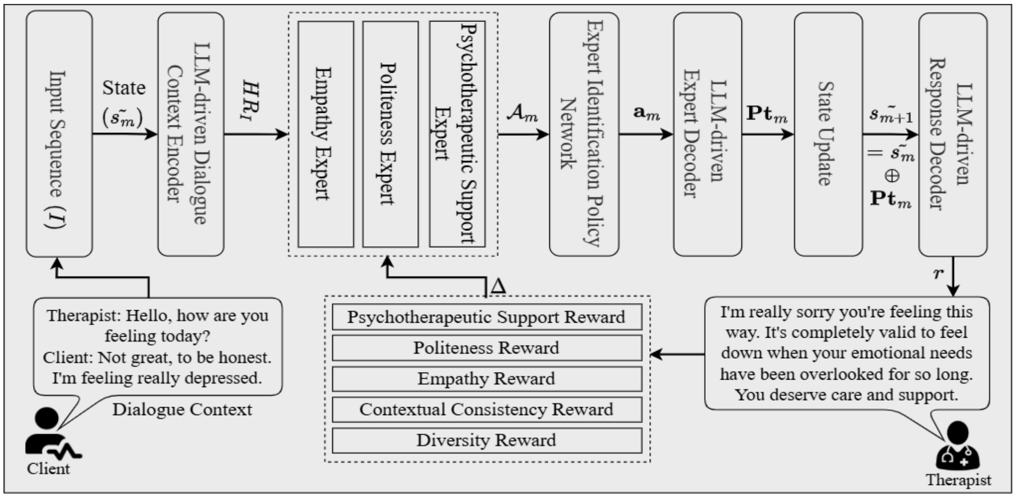

Figure 1: Architecture of the proposed dialogue system - **MATE**.

Here, where W^{E} and b^{E} ($\text{E} \in \{\text{PE}, \text{PoE}, \text{EE}\}$) are learnable parameters. The experts are trained to predict the next strategy label for the therapist's turn using the cross-entropy (CE) loss:

$$\mathcal{L}^{\text{E}} = \frac{1}{N} \sum_{p=1}^{N} \text{CE}(\hat{\mathbf{y}}_p^{\text{lbl}}, \mathbf{y}_p^{\text{lbl}}) \tag{12}$$

Here, $\text{lbl} \in \{\text{PSS}, \text{PoS}, \text{ES}\}$. $\hat{\mathbf{y}}_p^{\text{lbl}}$ is predicted probability of respective strategy label (from Eq. 11), and $\mathbf{y}_p^{\text{lbl}}$ is respective ground-truth strategy label.

Multi-Task Learning Objective. We enforce consistency among the experts while preserving their individual specializations by aligning their average pooled representation z_I^{avg} with the overall contextual embedding $hr_{[\text{CLS}]}$ using a Mean Squared Error (MSE) loss:

$$\mathcal{L}^{\text{MSE}} = \lambda \cdot \frac{1}{D_z} \sum_{k=1}^{D_z} (hr_{[\text{CLS}]}[k] - z_I^{\text{avg}}[k])^2 \tag{13}$$

Here, λ is a weighting coefficient and D_z is dimension of $hr_{[\text{CLS}]}$. Finally, we train the entire multi-task mixture-of-experts (MoE) framework using a joint loss:

$$\mathcal{L}^{\text{MOE}} = \mathcal{L}^{\text{PE}} + \mathcal{L}^{\text{PoE}} + \mathcal{L}^{\text{EE}} + \mathcal{L}^{\text{MSE}} \tag{14}$$

Mixture-of-Experts-based Reinforcement Learning (MoE-RL). We leverage the standard reinforcement learning paradigm (Sutton and Barto 2018) to build **MATE**.

State Representation. The initial state is initialized with dialogue context as: $\tilde{s}_1 = \mathcal{H} \in \mathcal{S}$ (for brevity, we remove the subscript m from context $\mathcal{H}$). At each time step m, the prompt token sequence $\mathbf{Pt}_m$ generated by the policy-identified expert (i.e., action) updates the current state. Specifically, the observed state at step m is represented as $\tilde{s}_m = \{\mathcal{H}, \mathbf{Pt}_1, \ldots, \mathbf{Pt}_{m-1}\} \in \mathcal{S}$, which is encoded via the context encoder to acquire $HR_{\mathcal{S},m}$ and $hr_{\mathcal{S},m}$. To maintain temporal continuity, we concatenate the sequence of prior state representations: $\tilde{s}_m = hr_{\mathcal{S},1} \oplus hr_{\mathcal{S},2} \oplus \ldots \oplus h_{\mathcal{S},m}$ to get the current state embedding. For $m < M$ (M: maximum iterations), zero-padding is applied to $\tilde{s}_m$ to ensure consistent dimensionality across state embeddings. Notably, the sequence length constraints of the underlying Phi-3 model are taken into consideration.

Action. At each step m, the action space $\mathcal{A}_m$ comprises multi-task experts, conditioned on the observed state $\tilde{s}_m$. The agent learns to determine an expert $\mathbf{a}_m \in \mathcal{A}_t$ as the selected action at each step. Once the expert is selected, a Phi-3-based dialogue context decoder generates expert prompt token sequence $\mathbf{Pt}_m$ for $\mathbf{a}_m$.

Policy. In addition to leveraging a dialogue context encoder as a semantic encoding policy encoder, we construct a dedicated policy network for expert identification, optimized via the `REINFORCE` algorithm with baseline (Sutton and Barto 2018), which consists of an actor model and a critic model. The actor learns a policy $\Pi_\theta(\mathbf{a}_m \mid \tilde{s}_m, \mathcal{A}_m)$ for choosing the optimal expert action $\mathbf{a}_m$ from $\mathcal{A}_m$, conditioned on the current state $\tilde{s}_m$. The critic network estimates the value $V_\psi(\tilde{s}_m)$ for state $\tilde{s}_m$ to establish a baseline for `REINFORCE`. We compute a shared latent feature $\mathbf{h}_m$ of $\tilde{s}_m$ using a two-layer transformation with ELU activation and dropout:

$$\mathbf{h}_m = \xi\left(\left(\xi\left(\tilde{s}_m \mathbf{W}^{(1)}\right)\right) \mathbf{W}^{(2)}\right) \tag{15}$$

where $\xi(\cdot)$ denotes the ELU activation followed by dropout. The actor's policy over expert actions is computed as:

$$\Pi_\theta(\mathbf{a}_m \mid \tilde{s}_m, \mathcal{A}_m) = \text{softmax}\left(\mathcal{A}_m \otimes \mathbf{h}_m \mathbf{W}_\theta\right) \tag{16}$$

and the critic value estimate is:

$$V_\psi(\tilde{s}_m) = \mathbf{h}_m \mathbf{W}_\psi \tag{17}$$

Here, $\otimes$ represents the Hadamard product, and $\mathcal{A}_m$ is optionally modulated using a binary mask vector for action pruning. In our setting, we use an all-ones vector for $\mathcal{A}_m$ since the number of experts is limited. $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \mathbf{W}_\theta, \mathbf{W}_\psi$ are trainable parameters.

Rewards. For effective policy learning, we reinforce the action taken at each step by evaluating how well the generated response $\hat{r}$ at updated state $\tilde{s}_{m+1}$ facilitates polite and empathetic psychotherapeutic support based on the client's early maladaptive schema and emotional state, while ensuring contextual consistency and diversity. To this end, we define two categories of rewards, *viz.* Goal-oriented and Language-oriented rewards.

(1) ***Goal-oriented rewards*** guide the therapist toward the responses exhibiting pertinent psychotherapeutic support, politeness, and empathy strategies while taking into account the client's schema and emotion. It includes:

- *Psychotherapeutic Support Reward* defined as $\mu_{\text{psy}} = f_{\text{psy}}(r) - \delta_{\text{psy}} \cdot f_{\text{psy}}(\hat{r})$ aims to ensure that the generated response aligns with the intended psychotherapeutic support strategy. Here, $f_{\text{psy}}(\cdot)$ denotes the schema-adaptive support strategy classifier[3] that outputs the support strategy probability as support strategy score.
- *Politeness Reward* defined as $\mu_{\text{pol}} = f_{\text{pol}}(r) - \delta_{\text{pol}} \cdot f_{\text{pol}}(\hat{r})$ reinforces the generation of responses aligned with predefined politeness strategy to ensure respectful and socially appropriate interaction. Here, $f_{\text{pol}}(\cdot)$ is emotion-adaptive politeness strategy classifier[3] that provides the politeness strategy probability as politeness score.
- *Empathy Reward* defined as $\mu_{\text{emp}} = f_{\text{emp}}(r) - \delta_{\text{emp}} \cdot f_{\text{emp}}(\hat{r})$ encourages that the generated response adhere to the established empathy strategy to reflect emotionally intelligent interactions. Here, $f_{\text{emp}}(\cdot)$ is the emotion-adative empathy strategy classifier[3] that provides the empathy strategy probability as empathy score.

(2) ***Language-oriented rewards*** encourage dialogue quality in terms of coherence and diversity. It includes:

- *Contextual Consistency Reward* formulated as $\mu_{cc} = 1/2[\text{MS}(\mathcal{H}, \hat{r}) + \text{MS}(r, \hat{r})]$ ensures that the generated response is coherent with the dialogue context. Here, MS denotes the MoverScore (Zhao et al. 2019) that provides the semantic similarity score between two sequences as a consistency score.
- *Diversity Reward* defined as $\mu_{\text{d}} = EAD(\hat{r})$ encourages that generated responses are diverse and non-repetitive, enabling engaging interactions during psychotherapy. Here, EAD denotes the Expectation Adjusted Difference (Liu et al. 2022) that provides the diversity score.

Cumulative Reward is formulated as:

$$\Delta = \frac{1}{1 + \exp\left(-\rho \cdot \sum_{p \in \{\text{psy},\text{pol},\text{emp},\text{cc},\text{d}\}} \upsilon_p \cdot \mu_p\right)} \tag{18}$$

In the rewards, $\delta_{\text{psy}}, \delta_{\text{pol}}, \delta_{\text{emp}} \in [1, 2]$ acts as penalization factor. υ_p are weights assigned to each of the rewards, ρ is the scaling parameter that controls how sharply the cumulative reward responds to variations in the weighted sum of rewards. A higher ρ amplifies the effect of individual rewards, while a lower ρ smoothens the fluctuations in rewards, enabling stable training and balanced reward optimization.

Policy Optimization and Training. We adopt M-step iterative framework where the primary objective of the agent is to maximize the expected cumulative reward:

$$\mathbb{J}_\Theta = \mathbb{E}_\Pi \left[\sum_{m=1}^{M} \beta^m \Delta_{m+1} \right] \tag{19}$$

with Θ and β denote trainable parameters and discount term, respectively. The agent is optimized with loss $\mathcal{L}^{\text{AGT}}$, and its policy gradient is given by:

$$\nabla_\Theta \mathbb{J}_\Theta = \mathbb{E}_\Pi \left[\nabla_\Theta \log \Pi_\theta(\mathbf{a}_m \mid \tilde{s}_m, \mathcal{A}_m)(\hat{A} - V_\psi(\tilde{s}_m)) \right] \tag{20}$$

where $\hat{A}$ denotes discounted cumulative reward from first state to last one. At the final step, we use the hidden state representation $HR_{\mathcal{S},M+1}$ of the terminal state $\tilde{s}_{M+1}$ to generate the response. The decoder is trained using the loss:

$$\mathcal{L}^{\text{DEC}} = -\sum_{w=1}^{W} \log \mathbb{P}(r_w \mid HR_{\mathcal{S},M+1}, r_{<w}) \tag{21}$$

Warm Start. We initialize MATE using Phi-3-mini-128k-instruct model. The model is fine-tuned on initial states using a warm start process, where optimization is guided by:

$$\mathcal{L}^{\text{WS}} = \mathcal{L}^{\text{MOE}} + \mathcal{L}^{\text{DEC}} \tag{22}$$

Joint Training. The MATE is finally optimized by minimizing the joint loss:

$$\mathcal{L}^{\text{JOINT}} = \mathcal{L}^{\text{AGT}} + \mathcal{L}^{\text{DEC}} + \frac{1}{M+1} \sum_{m=1}^{M+1} \mathcal{L}^{\text{MOE},m} \tag{23}$$

Experiments

We compare MATE with 6 baselines: ARDM (Wu et al. 2021), PersRFI (Shi et al. 2021), GPT-Critic (Jang, Lee, and Kim 2022), INA (Ahmad et al. 2023), ProCoT (ChatGPT) (Deng et al. 2023), e-Therapist (Mishra et al. 2023), and Phi-3-FT (Touvron et al. 2023). For automatic evaluation, we adopt Perplexity (PPL) (Brown et al. 1992), BLEU (B-3) (Papineni et al. 2002), BERTScore-F1 (BS-F1) (Zhang et al. 2019), Distinct-3 (D-3) (Li et al. 2015), and Response Length (LEN) to evaluate responses' language quality. To evaluate goal accomplishment, we design goal-oriented metrics following prior works (Sharma et al. 2021; Mishra et al. 2023), including Schema-Support Strategy Consistency (SSSC), Emotion-Politeness Strategy Consistency (EPSC), Emotion-Empathy Strategy Consistency (EESC), and Engagingness (ENG) metrics. For human evaluation, we assess Fluency (FLU), Contextual Coherence (COH), and Naturalness (NAT) to measure language quality of responses, and employ SSSC, EPSC, and EESC to evaluate goal accomplishment from a human perspective. Due to space constraints, we provide 'Implementation Details', 'Baselines Details', and 'Evaluation Metrics Details' in 'Experiment Details' section of the appendix.

Results and Analysis

Automatic Evaluation. Table 2 presents automatic evaluation results comparing MATE with several baselines. MATE achieves lowest PPL scores across both datasets, which indicates its improved fluency and language modeling capabilities. Further, it obtains superior dialogue quality, as reflected by its lexical (B-3) and semantic richness (BS-F1)

[3] Due to space constraints, classifier details are given in 'Experiment Details' section of the appendix.

Models	PPL ↓	B-3 ↑	D-3 ↑	BS-F1↑	SSSC ↑	EPSC ↑	EESC ↑	ENG ↑	LEN ↑	PPL ↓	B-3 ↑	D-3 ↑	BS-F1↑	SSSC ↑	EPSC ↑	EESC ↑	ENG ↑	LEN ↑
	HOPE									PSYCON								
ARDM	19.87	3.68	29.92	0.633	0.479	0.646	0.493	0.415	13.36	3.76	3.66	35.88	0.762	0.358	0.393	0.418	0.269	19.15
PersRFI	18.74	4.06	30.41	0.641	0.486	0.654	0.505	0.422	14.80	3.95	3.93	37.21	0.783	0.381	0.409	0.443	0.286	19.71
GPT-Critic	17.51	4.87	32.23	0.654	0.495	0.668	0.526	0.438	16.93	4.74	4.73	38.56	0.808	0.407	0.426	0.476	0.312	21.51
e-Therapist	15.86	5.63	33.74	0.668	0.522	0.684	0.549	0.462	18.70	2.52	5.68	32.12	0.892	0.556	0.571	0.673	0.479	23.89
ProCoT (ChatGPT)	19.21	3.51	29.75	0.629	0.471	0.639	0.465	0.435	17.55	9.84	3.46	34.78	0.812	0.504	0.533	0.608	0.442	27.81
Phi-3-FT	15.12	7.03	34.89	0.688	0.671	0.735	0.653	0.481	19.62	2.41	6.87	43.33	0.861	0.671	0.696	0.701	0.506	29.96
MATE	**13.47**	**8.95**	**38.12**	**0.772**	**0.726**	**0.796**	**0.683**	**0.527**	**23.41**	**1.09**	**8.31**	**46.07**	**0.901**	**0.819**	**0.843**	**0.853**	**0.698**	**32.85**

Table 2: Automatic evaluation results. Results are statistically significant at 5% significance level based on t-test (Welch 1947).

Models	SSSC	EPSC	EESC	FLU	COH	NAT
			HOPE			
ARDM	2.32	2.28	2.13	2.94	2.87	2.76
PersRFI	2.89	2.74	2.71	3.21	3.00	3.09
GPT-Critic	3.14	3.07	2.93	3.43	3.22	3.38
e-Therapist	2.51	2.48	2.33	3.05	2.97	2.88
ProCoT (ChatGPT)	3.45	3.42	3.30	3.68	3.54	3.51
Phi-3-FT	3.82	3.64	3.51	3.91	3.87	3.80
MATE	**4.35**	**4.32**	**4.19**	**4.62**	**4.58**	**4.41**
			PSYCON			
ARDM	2.48	2.61	2.37	2.85	2.73	2.52
PersRFI	2.91	2.85	2.68	3.10	3.04	3.01
GPT-Critic	3.18	3.04	3.01	3.41	3.36	3.35
e-Therapist	2.66	2.70	2.55	2.98	2.90	2.84
ProCoT (ChatGPT)	3.39	3.36	3.28	3.62	3.59	3.44
Phi-3-FT	3.84	3.72	3.55	3.86	3.83	3.75
MATE	**4.38**	**4.34**	**4.28**	**4.59**	**4.61**	**4.45**

Table 3: Human evaluation results. Results are statistically significant at 5% significance level based on Welch's t-test (Welch 1947). All metrics are rated on a scale of 1 to 5.

combined with its diverse (D-3) and longer (LEN) response generation ability. In particular, on HOPE, it obtains a notable improvement of 27.3%, 9.3%, 12.2%, and 19.3% in B-3, D-3, BS-F1, and LEN, respectively, compared to second-best baseline Phi-3-FT. On PSYCON, it achieves a significant gain of 21.0%, 6.3%, 4.6%, and 9.6% in B-3, D-3, BS-F1, and LEN, respectively. Further, **MATE** achieves highest SSSC, EPSC, EESC, and ENG scores on both datasets, which substantiates the design of novel experts and reward function in guiding the model to generate polite and empathetic PS responses aligned with the clients' EMS and emotions. Specifically, it achieves 0.726, 0.796, 0.683, and 0.527 scores for SSSC, EPSC, EESC, and ENG, respectively, on HOPE with a notable increase of 8.2%, 8.3%, 4.6%, and 9.6% compared to Phi-3-FT. Similarly, on PSYCON, it obtains an improvement of 22.1%, 21.1%, 21.7%, and 37.9% in SSSC, EPSC, EESC, and ENG, respectively.

Phi-3-FT often produces polite and empathetic responses during psychotherapy, but these tend to lack diversity and coherence, frequently resorting to repetitive, generic statements such as "*I understand what you're going through, I am here to help you*". In contrast, e-Therapist's outputs frequently fail to align with the client's EMS and emotional cues, which can be justified given that it lacks dedicated experts and rewards guiding it to generate responses aligned with the client's schema and emotional states. ProCoT (ChatGPT) demonstrates strong natural language generation abilities, maintaining contextual consistency and diversity, but it struggles to deliver psychotherapeutic responses that are polite, empathetic and aligned with clients' emotions and EMS. For instance, when a client says "*My family thinks I have a drinking problem.*", it replies simply with "*No issues. Tell me when it started?*", which lacks warmth and courteousness, essential for positive psychotherapeutic outcomes. Overall, these comprehensive evaluation results clearly demonstrate that **MATE** consistently outperforms all baselines across all metrics on both datasets, underscoring the effectiveness of its advanced experts and rewards in facilitating high-quality EMS-guided polite and empathetic psychotherapeutic support dialogues.

Human Evaluation. Table 3 presents human evaluation results for **MATE** and baselines. It is evident that **MATE** obtains better scores of 4.35, 4.32, 4.19, 4.62, 4.58, and 4.41 for SSSC, EPSC, EESC, FLU, COH, and NAT, respectively, with an improvement of +0.53, +0.68, +0.68, +0.71, +0.71, and +0.61 points for these metrics, compared to second best baseline, Phi-3-FT on HOPE. A similar performance improvement is observed in the PSYCON dataset. The superior SSSC, EPSC, and EESC scores indicate the effectiveness of psychotherapeutic support, politeness, and empathy experts and rewards in facilitating EMS-guided polite and empathetic responses during psychotherapy that foster respectful and amicable support, eventually contributing to positive psychotherapeutic outcomes. Also, the high FLU, COH, and NAT scores reflect the pivotal role of contextual consistency and diversity rewards in generating contextually coherent and engaging responses that enhance the naturalness of overall psychotherapeutic interactions. Owing to space limitations, 'Human Evaluation Process' is presented in the 'Experiment Details' section of the appendix.

Additional Analysis. Due to space constraints, we include more analyses - (1) Ablation of Experts and Rewards, (2) Ablation of Multi-task Learning of MoE, (3) Ablation of Expert Policy and Training Paradigm, (4) Effect of Iteration Steps (5) Out-of-Domain Evaluation, and (6) Case Study under 'Additional Analysis' section of the appendix.

Conclusion

In this work, we presented **MATE**, an EMS-guided, polite, and empathetic dialogue system for psychotherapeutic support. To build this system, we annotated clients' utterances in HOPE and PSYCON datasets with EMS and emotions, and therapists' utterances with politeness strategies, along with novel psychotherapeutic support and empathy strategies, using ChatGPT in a few-shot setup. **MATE** integrates an LLM with MoE-RL approach, incorporating new experts and rewards for psychotherapeutic support, politeness, and empathy. Evaluations show that **MATE** generates polite and empathetic responses attuned to clients' EMS and emotions, thereby facilitating effective psychotherapeutic support.

Acknowledgments

Priyanshu Priya gratefully acknowledges the financial support provided by the Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowship, administered by the Department of Science and Technology, Ministry of Science and Technology, Government of India.

References

Abdin, M.; Aneja, J.; Awadalla, H.; Awadallah, A.; Awan, A. A.; Bach, N.; Bahree, A.; Bakhtiari, A.; Bao, J.; Behl, H.; et al. 2024. Phi-3 technical report: A highly capable language model locally on your phone, 2024. *URL https://arxiv.org/abs/2404.14219*.

Ahmad, Z.; Saurabh, S.; Menon, V.; Ekbal, A.; Ramnani, R.; and Maitra, A. 2023. INA: An Integrative Approach for Enhancing Negotiation Strategies with Reward-Based Dialogue Agent. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2536–2549.

Brown, P.; and Levinson, S. C. 1987. *Politeness: Some universals in language usage*. 4. Cambridge university press.

Brown, P. F.; Della Pietra, S. A.; Della Pietra, V. J.; Lai, J. C.; and Mercer, R. L. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics*, 18(1): 31–40.

Budiarta, T.; Nurkamto, J.; Sumarlam, S.; and Purnanto, D. 2021. Expressive speech acts of politeness in the counselling process. *Langkawi: Journal of The Association for Arabic and English*, 7(2): 212–226.

Chow, Y.; Tulepbergenov, A.; Nachum, O.; Ryu, M.; Ghavamzadeh, M.; and Boutilier, C. 2022. A mixture-of-expert approach to rl-based dialogue management. *arXiv preprint arXiv:2206.00059*.

Collins, J. F.; Elkin, I.; Sotsky, S. M.; and Docherty, J. P. 1990. Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, 147(6): 711–718.

Csigó, K.; Münnich, Á.; and Molnar, J. 2024. The importance of examining early maladaptive schemas in the diagnosis and treatment of obsessive-compulsive disorder. *Frontiers in Psychiatry*, 15: 1360127.

Davis, M. H.; et al. 1980. A multidimensional approach to individual differences in empathy.

Deng, Y.; Liao, L.; Chen, L.; Wang, H.; Lei, W.; and Chua, T.-S. 2023. Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 10602–10621.

Ekman, P. 1992. Facial expressions of emotion: New findings, new questions.

Elliott, R.; Bohart, A. C.; Watson, J. C.; and Greenberg, L. S. 2011. Empathy. *Psychotherapy*, 48(1): 43.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5): 378.

Ghandeharioun, A.; McDuff, D.; Czerwinski, M.; and Rowan, K. 2019. EMMA: an emotion-aware wellbeing chatbot. In *2019 8th international conference on affective computing and intelligent interaction (ACII)*, 1–7. IEEE.

Gilardi, F.; Alizadeh, M.; and Kubli, M. 2023. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30): e2305016120.

Goldberg, S. B.; Flemotomos, N.; Martinez, V. R.; Tanana, M. J.; Kuo, P. B.; Pace, B. T.; Villatte, J. L.; Georgiou, P. G.; Van Epps, J.; Imel, Z. E.; et al. 2020. Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *Journal of counseling psychology*, 67(4): 438.

Guo, Z.; Lai, A.; Thygesen, J. H.; Farrington, J.; Keen, T.; and Li, K. 2024. Large language model for mental health: A systematic review. *arXiv preprint arXiv:2403.15401*.

Gupta, S.; Mukherjee, S.; Subudhi, K.; Gonzalez, E.; Jose, D.; Awadallah, A. H.; and Gao, J. 2022. Sparsely activated mixture-of-experts are robust multi-task learners. *arXiv preprint arXiv:2204.07689*.

Heymann, J.; and Sprague, A. 2023. Meeting the UN Sustainable Development Goals for mental health: why greater prioritization and adequately tracking progress are critical. *World Psychiatry*, 22(2): 325.

Hossain, A.; Priya, P.; Tripathi, A. M.; Verma, P.; and Ekbal, A. 2025. MENDER: Multi-hop Commonsense and Domain-specific CoT Reasoning for Knowledge-grounded Empathetic Counseling of Crime Victims. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)*, 501–516.

Jang, Y.; Lee, J.; and Kim, K.-E. 2022. GPT-critic: Offline reinforcement learning for end-to-end task-oriented dialogue systems. In *10th International Conference on Learning Representations, ICLR 2022*. International Conference on Learning Representations, ICLR.

Kim, J.; Kim, Y.; Kim, B.; Yun, S.; Kim, M.; and Lee, J. 2018. Can a Machine Tend to Teenagers' Emotional Needs? A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6.

Li, J.; Galley, M.; Brockett, C.; Gao, J.; and Dolan, B. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

Liu, S.; Sabour, S.; Zheng, Y.; Ke, P.; Zhu, X.; and Huang, M. 2022. Rethinking and refining the distinct metric. *arXiv preprint arXiv:2202.13587*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Malhotra, G.; Waheed, A.; Srivastava, A.; Akhtar, M. S.; and Chakraborty, T. 2022. Speaker and time-aware joint contextual learning for dialogue-act classification in counselling conversations. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, 735–745.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

Miller, W. R.; and Rollnick, S. 2012. *Motivational interviewing: Helping people change*. Guilford press.

Mishra, K.; Priya, P.; Burja, M.; and Ekbal, A. 2023. e-THERAPIST: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 13952–13967.

Morris, R. R.; Kouddous, K.; Kshirsagar, R.; and Schueller, S. M. 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of medical Internet research*, 20(6): e10148.

OpenAI. 2024. ChatGPT. https://chatgpt.com/.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

Priya, P.; Chigrupaatii, R.; Firdaus, M.; and Ekbal, A. 2025. GENTEEL-NEGOTIATOR: LLM-Enhanced Mixture-of-Expert-Based Reinforcement Learning Approach for Polite Negotiation Dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 25010–25018.

Priya, P.; Mishra, K.; Totala, P.; and Ekbal, A. 2023. PARTNER: A Persuasive Mental Health and Legal Counselling Dialogue System for Women and Children Crime Victims. In Elkind, E., ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 6183–6191. International Joint Conferences on Artificial Intelligence Organization. AI for Good.

Saha, T.; Reddy, S.; Das, A.; Saha, S.; and Bhattacharyya, P. 2022. A Shoulder to Cry on: Towards A Motivational Virtual Assistant for Assuaging Mental Agony. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2436–2449.

Sharma, A.; Lin, I. W.; Miner, A. S.; Atkins, D. C.; and Althoff, T. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the Web Conference 2021*, 194–205.

Sharma, A.; Miner, A.; Atkins, D.; and Althoff, T. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 5263–5276.

Shi, W.; Li, Y.; Sahay, S.; and Yu, Z. 2021. Refine and Imitate: Reducing Repetition and Inconsistency in Persuasion Dialogues via Reinforcement Learning and Human Demonstration. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 3478–3492.

Spicer, L.; DeCicco, E.; Clarke, A.; Ambrosius, R.; and Yalcin, O. 2024. Understanding early maladaptive schemas in autistic and ADHD individuals: exploring the impact, changing the narrative, and schema therapy considerations. *Frontiers in Psychology*, 15: 1436053.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement learning: An introduction*.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Tu, Q.; Gao, S.; Li, Y.; Cui, J.; Wang, B.; and Yan, R. 2022. Conversational recommendation via hierarchical information modeling. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2201–2205.

Tu, S.; Powers, A.; Doogan, S.; and Choi, J. D. 2025. TRUST: An LLM-Based Dialogue System for Trauma Understanding and Structured Assessments. *arXiv preprint arXiv:2504.21851*.

Valizadeh, M.; and Parde, N. 2022. The AI doctor is in: A survey of task-oriented dialogue systems for healthcare applications. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 6638–6660.

Wang, B.; and Komatsuzaki, A. 2021. GPT-J-6B: A 6 billion parameter autoregressive language model.

Wegener, I.; Alfter, S.; Geiser, F.; Liedtke, R.; and Conrad, R. 2013. Schema change without schema therapy: The role of early maladaptive schemata for a successful treatment of major depression. *Psychiatry*, 76(1): 1–17.

Welch, B. L. 1947. The generalization of 'STUDENT'S'problem when several different population varlances are involved. *Biometrika*, 34(1-2): 28–35.

Wu, Q.; Zhang, Y.; Li, Y.; and Yu, Z. 2021. Alternating Recurrent Dialog Model with Large-scale Pre-trained Language Models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1292–1301.

Young, J. E.; and Klosko, J. S. 1994. *Reinventing your LIFE: The breakthough program to end negative behavior... and feel great again*. Penguin.

Young, J. E.; Klosko, J. S.; and Weishaar, M. E. 2006. *Schema therapy: A practitioner's guide*. guilford press.

Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K. Q.; and Artzi, Y. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Zhao, T.; Xie, K.; and Eskenazi, M. 2019. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. *arXiv preprint arXiv:1902.08858*.

Zhao, W.; Peyrard, M.; Liu, F.; Gao, Y.; Meyer, C. M.; and Eger, S. 2019. MoverScore: Text generation evaluating with contextualized embeddings and earth mover distance. *arXiv preprint arXiv:1909.02622*.

Zhao, W. X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2).