Reinforcement Learning-Based Framework for Dynamic Strategy Generation in Personalized Psychological Counseling Using Deep Q-Networks

Chen Chen*, Miao Zhen
The School of Visual Arts, Hunan Mass Media Vocational and Technical College, Changsha, Hunan, 410100, China
E-mail: chenchen198906@outlook.com
Corresponding author

Keywords: personalized counseling, reinforcement learning, decision support framework, Deep Q-network, psychological state modeling, adaptive therapy

Received: July 22, 2025

Personalized psychological counseling plays a crucial role in enhancing mental well-being by addressing individual emotional and cognitive needs. This study proposes a Reinforcement Learning-Based Decision Support Framework (RL-DSF) that dynamically generates counseling strategies optimized through a Deep Q-Network (DQN). The model adapts in real-time to users' evolving psychological states by leveraging feedback signals derived from emotional responses, engagement metrics, and counseling effectiveness. The RL-DSF was trained and evaluated using a synthetic therapy conversations dataset, comprising diverse simulated dialogues with annotated emotional cues, designed to mimic real-world mental health scenarios. While no direct clinical patient data was used in training, the system's effectiveness was assessed on anonymized user sessions collected from a chatbot-based mental health support platform. Experimental results demonstrated that RL-DSF significantly outperformed baseline methods, achieving an average reduction of 1.1 points on PHQ-9 depression scores and 0.6 points on GAD-7 anxiety scores. User engagement increased by 11.1%, satisfaction ratings averaged 4.5 out of 5, and dropout rates were reduced to 5%, validating the framework's potential to provide adaptive, personalized psychological support in a scalable digital environment.

Povzetek:

1 Introduction

1.1 Background and motivation

Mental health issues are increasingly prominent around the globe, creating individuals who are under greater levels of stress, and anxiety [1]. Psychotherapy is still the primary form of intervention, helping people to cope with these emotional challenges in a personalized therapy session context to provide emotional care that is both contextualized and individualized [2]. The field of artificial intelligence is advancing rapidly, particularly within the realm of natural language processing and conversational agents, creating opportunities for automated systems to offer psychological support capturing a model that can interact dynamically with users and provide real-time therapeutic support [3].

Human counselors and established manual approaches have traditionally dominated psychotherapeutic support, with high levels of variance in human availability and real-world resource limitations that hinder meeting demand or scaling [4]. The primary intent is to address mental health issues, exploring a promising pathway for providing scalable, consistent, and continuous mental health support [5]. Individual customer experience models and applications would need to move beyond a static, rule-based approach and dynamically learn user responses [6]. Based on behaviors and emotional responses at the interface level, to recreate the qualities of individualized adaptive support consistent with human therapists [7].

While prior reinforcement learning approaches, such as dual reinforcement learning (DRL) models, have effectively optimized intervention scheduling and adapted to user context, they often lack integration of real-time multimodal emotional feedback and rely primarily on simulated or static user models. Similarly, traditional NLP-based systems focus on post-session sentiment analysis or use limited rule-based adaptation, which restricts dynamic personalization during ongoing counseling interactions. Our RL-based Decision Support Framework (RL-DSF) advances beyond these limitations by explicitly modeling the user's psychological state via a rich multimodal feature set that captures lexical, emotional, and temporal interaction cues in real time. Coupled with a deep Q-network optimized by continuous feedback signals

including emotional valence, engagement, and strategy effectiveness, RL-DSF dynamically selects counseling actions tailored to the evolving mental state of each user. This enables a more responsive and personalized counseling experience than prior RL or NLP-based systems, addressing critical challenges in adaptive digital mental health interventions.

1.2 Challenges in personalized counseling

Although technology is developing rapidly, the digital counseling tools available today are still constrained [8]. Most of these solutions are either scripted or based on a constrained decision tree approach. Meaning, they don't deal well with the continual changes in the user's psychological state [9]. They cannot recognize individual differences in how people express their emotions and respond to treatment. Feedback mechanisms are typically absent or underutilized, and follow-ups often feel repetitive, robotic, and irrelevant [10].

1.3 Research objectives and contributions

- The primary goal of research is to develop intelligent agents capable of discovering their sequences of counseling actions through the emulation of user interactions with counseling as a state-action-reward system.
- The RL-DSF is not static; instead, it continually changes in response to the user's emotional state, emotional feedback, response patterns, and engagement level.
- The paper's contribution is to frame a dynamic psychological counseling problem as a type of reinforcement learning task, where the agent receives feedback in a structured manner to guide the optimization of its strategy

2 Literature landscape

2.1 Traditional counseling systems

Previously, conventional systems of counseling have focused on structured, rules-based systems such as content-based filtering, reinforcement learning, and NLP. These approaches are centered on predictive analytics, uniformity of exchange, and historical context. While they can yield effective advice in structured environments, they fail to react to real-time fluctuations in emotion provided in interactive counseling, as well as fail to incorporate personalized user trajectories in rapidly evolving psychological contexts.

2.1.1 Popularity–Based Filtering and Content–Based Filtering (PBF + CBF)

This investigation employed a mixed-methods approach in a predictive model aggregate recommendation method, incorporating attributes from predictive modeling while also providing recommendations. A dataset comprising 500 students and 31 institutions was used to train a Huber Regressor for admissions prediction. The recommendation system included both popularity-based and content-based recommendation modules. The evaluation methods included regression measures when deploying the system as a web application [11].

2.1.2 Dual Reinforcement Learning (DRL)

This paper used a dual reinforcement learning approach for personalizing digital, just-in-time adaptive health interventions. The dual models consisted of an action model that determined which intervention type and frequency to use, and a time model that identified the optimal time based on user context. The methodology was novel, featuring two enhancements: a customized eligibility trace method to reward past activity and a transfer learning approach, which leverages knowledge learned across different environments. This methodology was demonstrated using simulations with different user personas representing different behavior, preferences, and activity patterns [12].

2.1.3 Natural Language Processing (NLP)

Natural language processing strategies, such as text and sentiment analysis, are employed to analyze user conversations with AI chatbots during psychological counseling sessions. The findings would then be used as input features for machine learning algorithms to predict counselling outcomes and levels of user satisfaction[13]. The analysis techniques primarily focused on capturing emotional and linguistic patterns, which were subsequently used to generate predictive models with high accuracy. If the predictions were implemented, the technology could provide changes, or modulate, counselling strategies that could increase the effectiveness of the technology-assisted psychological support [14].

2.1.4 Design Science Research (DSR)

This paper utilized a DSR methodology to design personality-adaptive conversational agents (PACAs) for mental health care. The design process was iterative and involved multiple steps. PACAs could potentially enhance user interaction and experience, ultimately benefiting users in mental health contexts. So, while this study does not directly contribute to CA design knowledge, it extends the body of knowledge for valid CA design. [15].

2.1.5 Mixed-Methods (MM)

This paper employed a scoping review methodology, conducting a thorough literature search in databases. Multiple independent reviewers were involved in the data extraction and quality review processes. The review summarised the existing evidence on the perceived effectiveness, feasibility, and challenges of using AI chatbot applications in mental health care. [16]

2.2 AI-Driven mental health interventions

From static systems, the technology behind AI-based mental health interventions represents a major change. While these systems do not typically involve complex methods (e.g., RCTs and meta-analyses), they can be made to offer scalable, interactive, and context-sensitive support. They can provide dynamic personalization and real-time feedback, which has the potential to be more effective for a broader range of user populations within mental health than traditional therapy or current options.

2.2.1 Randomized Controlled Trial (RCT)

This paper used a pilot randomized controlled trial to compare the effectiveness of an AI chatbot with that of a nurse-staffed hotline for the general population in reducing anxiety and depression. Participants were randomly assigned to one of the two intervention groups. Mental health outcomes were assessed using standardized psychological scales over a defined period. The trial compared anxiety and depression levels pre- and post-intervention [17].

2.2.2 Assessor Blinded Randomized Controlled Trial (ABRCT)

The purpose of this study was to evaluate a rule-based, topic-specific chatbot for mental health self-care in a two-arm, assessor-blinded, randomized controlled trial involving 285 participants. The participants were randomised to the intervention or wait-list control group. Pre-intervention, post-intervention (10 days), and 1-month follow-up levels of outcome were assessed using web-based self-assessments. Underlying the research design was the analysis of data (collected both pre- and post-intervention) through the use of linear mixed models and the calculation of effect sizes using Cohen's d, where possible. [18]

2.2.3 Systematic Review & Meta-Analysis (SR-MA)

This PRISMA-compliant meta-analysis has synthesized the latest RCTs on AI chatbots specifically within the fragmented sector of youth mental health. Overall effect sizes for symptom reduction (anxiety, depression) were moderate when high engagement was maintained. Conversational agents that incorporated CBT principles delivered significant therapeutic benefits. However, given the heterogeneity of designs and engagement metrics in the studies, the authors encourage the use of standardized protocols and a longer evaluation duration [19].

2.2.4 Single-Blind, Three-Arm RCT (SB-RCT)

The SB-RCT study compared the XiaoE chatbot against two control conditions with a sample of 148 college students. XiaoE offered structured, CBT-based conversational modules that yielded significant decreases in depressive symptomology post-intervention and at one-month follow-up. This study employed ANCOVA and LDA analytics to analyze the data collected on standardized measures (PHQ-9, usability scores, and measures of expectation and satisfaction), providing evidence of sound methodological rigor and promising therapeutic results [20].

2.2.5 Unblinded Randomized Controlled Trial (URCT)

This paper reports the results of an unblinded randomized controlled trial evaluating the MISHA chatbot as a tool to facilitate stress management among students. Participants (N = 140) were randomly assigned to either an intervention. Outcomes included stress, depression, and psychosomatic symptoms and were measured through web-based self-assessments. Analyses were conducted using repeated measures ANOVA and generalized estimating equations assessing treatment effect and user engagement [21]. This work presents a conversational agent integrating psychological modeling for effective stress, anxiety, and depression interventions, supported by robust evaluation data [22].

Details a system combining cognitive architectures with conversational AI to enable personalized therapy for mental health issues [23]. Smartphone-Based Assessment and Intervention: Reviews state-of-the-art digital assessment and intervention techniques, emphasizing data-driven personalization and ethical considerations [24]. Surveys cognitive assistant frameworks relevant for mental health, focusing on interaction models and behavior change effectiveness [25]. Examines the potential of persuasive and conversational technologies in improving mental health access and outcomes, highlighting societal impact [26].

The Table 1 provide the Research gap analysis for Reinforcement Learning-Based Decision Support Framework (RL-DSF)

Table 1: Summarizing quantitative results for key methods

Method	PHQ-9 Reduction (points)	Dropout Rate (%)	Satisfaction Score (out of 5)	Engagement Score (%)
Rule-Based System (RBS)	0.3	30	3.5	62
Static NLP-Based Classifier (SNC)	0.5	28	3.7	68

XiaoE Framework (XEF)	~0.7	22	4.1	75
Dual Reinforcement Learning (DRL)*	~0.9	19	4.3	82
Randomized Controlled Trial (RCT)	~0.6	25	4.0	70

Research question

The paper's Research Objectives should be revised to clearly state specific research questions or hypotheses such as:

1. Can a reinforcement learning-based framework dynamically generate personalized counseling strategies that adapt to evolving user emotional states?
2. Does the RL-DSF improve psychological outcomes (e.g., reduction in PHQ-9 and GAD-7 scores) compared to existing static or rule-based counseling systems?
3. Can the RL-DSF enhance user engagement, satisfaction, and retention during therapy sessions through real-time adaptive interventions?

3 Proposed Method: RL-based decision support framework

The RL-DSF proposed for personalized chat-based psychological counseling. It utilizes two distinct sources of user interaction data, the encoding of user emotional state, and a DQN policy selector to generate adaptive therapeutic responses. The overall design of the framework, state representation modeling, learning loop, and chatbot operationalization.

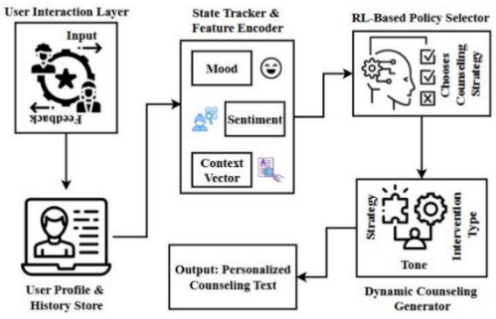

Figure 1: RL-driven architecture for dynamic psychological counseling

Figure 1 shows the system architecture of RL-DSF. The RL-DSF begins with a user's interaction inputs and behavioral logs that then aggregate to receive inputs through a state tracker and encoder. The DQN policy selector selects the most optimal strategy provided the emotional context. This is converted into an individualized counseling response using a dynamic generator. The perceivable adaptive dialogue is focused on an individual's psychological state in response to a user's therapeutic needs.

State action Q value update using deep Q network $R(t_u, b_u)$ is expressed using equation 1,

$$R(t_u, b_u) \leftarrow R(t_u, b_u) + \sigma \left[s_u + \alpha * \max_{b'} R(t_{u+1}, b') - R(t_u, b_u) \right] \quad (1)$$

Equation 1 explains how the state-action Q-value update using a deep Q network utilizes the temporal variation between the present estimate and the anticipated future return to update the coefficient given a specific state-action combination.

In this $R(t_u, b_u)$ is the estimated value of taking action is state, σ is the learning rate controls how much new information overrides old, s_u is the observed reward at time derived from real-time user feedback, α is the discount factor reflects the importance of future rewards, t_{u+1} is the next psychological state inferred after action, and $\max_{b'} R(t_{u+1}, b')$ is the maximum Q-value over all possible next actions in the new state.

The system utilizes a dynamic text-generating mechanism, driven by latent state insertions, to generate a tailored and adaptive coaching message based on the chosen action.

Personalized response generation using context-conditioned decoder z_u is expressed using equation 2,

$$z_u = \arg\max_{x \in W} Q(x | i_u, d_u, b_u) \quad (2)$$

Equation 2 represents the personalized response generation using a context-conditioned decoder, which is the adaptive counseling dialogue's word-level generating process.

In this z_u is the generated word at time step forming part of the counseling response, W is the complete set of available output words, $Q(x|i_u, d_u, b_u)$ is the probability of emitting word given decoder hidden state, context, and action, i_u is the decoder hidden state at step encapsulating sequential linguistic memory, d_u is the contextual embedding vector from encoder or attention layer, representing user state, and b_u is the selected optimal counseling strategy from Q-network output.

In the administration of therapy, the interplay between Q-learning and modeling generates guarantees that ensure both expressive capacity and decision intelligence.

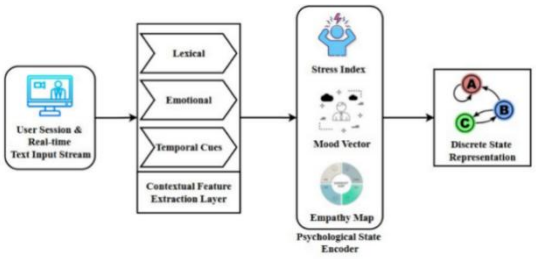

Figure 2: Psychological state modeling pipeline for reinforcement learning

Figure 2 shows the psychological state modeling within the proposed RL framework. The system captures user text input data in real-time and operates it in the contextual feature extraction layer, examining lexical, emotional, and temporal features. The features are processed into psychological state vectors, such as stress indices and empathy measures. Finally, the information can be represented in a set of discrete states, have defined Markov Decision Process (MDP) for the reinforcement learning agent to make decisions about which strategies to use.

Psychological state vector encoding from multimodal features t_u is expressed using equation 3,

$$t_u = g_{fod}(y_u^l, y_u^e, y_u^t) \quad (3)$$

Equation 3 explains the psychological state vector encoding from multimodal features is the encoding function.

In this t_u is the psychological state vector at time, $g_{fod}(.)$ is the multilayer encoder function fusing multimodal inputs, y_u^l is the lexical feature vector at time derived from word embedding and grams, y_u^e is the emotional signal vector at time extracted from emotion classifiers, and y_u^t is the temporal interaction features at time.

The reinforcement-learning agent uses a defined transition model to enable state-aware strategy selection by mapping the user's psychological state vector to a finite MDP state.

MDP state transition probability $Q(t_{u+1}|t_u, b_u)$ is expressed using equation 4,

$$Q(t_{u+1}|t_u, b_u) = \sum_y [\![t_{u+1} = g_{fod}(y)]\!] * Q(y|t_u, b_u) \quad (4)$$

Equation 4 explains the MDP state transition probability by integrating all potential user feature inputs between two MDP states, depending on the action taken. A key innovation of the proposed RL-DSF lies in the psychological state modeling, as illustrated in Figure 2. Unlike conventional RL approaches that treat user states in a simplified or static manner, this model constructs a dynamic psychological state vector tu by fusing multimodal inputs including lexical features from user text, real-time emotional signals, and temporal interaction patterns.

In this $Q(t_{u+1}|t_u, b_u)$ is the probability of transitioning to state from state after taking action, y is the input feature configuration, $[\![.]\!]$ is the indicator function that equals 1 if the condition is true, $g_{fod}(y)$ is the encoding function used to compute the next psychological state, and $Q(y|t_u, b_u)$ is the probability of observing input features given the current state and action.

This dynamic modeling makes sure that changes in emotions are recorded and taken into account when developing a strategy.

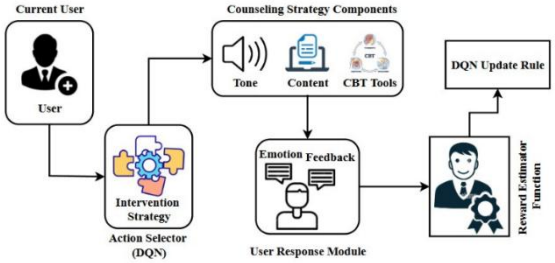

Figure 3: Reinforcement learning feedback loop for strategy optimization

Figure 3 illustrates the reinforcement learning feedback loop for optimization of psychological counseling. Starting from the current state of the user, the DQN selects a specific intervention strategy. The counseling components include tone, CBT content, and the structure of phrasing and student interaction [27]. The feedback occurs from the user, in terms of emotional play chomology interaction. This emotional response is processed by the reward estimator, which assesses the user's response, engagement, and effectiveness. The reward is delivered to update the DQN policy, thereby continuously personalizing the case and facilitating quick learning in action sessions [28].

Emotional feedback-based reward estimation s_u is expressed using equation 5,

$$s_u = x_1 * F_v(f_u) + x_2 * N_e(v_u) + x_3 * T_e(d_u) \quad (5)$$

Equation 5 explains that the emotional feedback-based reward estimation calculates the overall reward for a time step as a weighted sum of the three elements.

In this s_u is the scalar reward signal at time used to inform the Q-learning update, $F_v(f_u)$ is the valence based emotion score derived from user emotional signal, $N_e(v_u)$ is the engagement measure from user interaction, $T_e(d_u)$ is the strategy effectiveness score from counseling component, and x_1, x_2, x_3 are the normalized weights tuned empirically to balance the impact of each component.

Once the user's emotional reaction and interaction quality have been evaluated, the scalar reward is used to update the deep Q-network, thereby improving future strategy selection.

Deep Q-network loss function for policy update $M(\partial_u)$ is expressed using equation 6,

$$M(\partial_u) = F_{(t_u,b_u,s_u,t_{u+1})}\left[\left(s_u + \forall * \max_{b'} R(t_{u+1}, b'; \partial_u^-) - R(t_u, b_u; \partial_u)\right)^2\right] \quad (6)$$

Equation 6 explains the deep Q-network loss function for policy update establishes the loss function for modifying the Q-network's parameters.

In this $M(\partial_u)$ is the loss function for updating DQN parameters at time, $R(t_u, b_u; \partial_u)$ is the Q-value function parameterized to estimate the expected return of an action in state, ∂_u is the parameters of the current DQN, ∂_u^- is the parameters of the target network, held fixed for stability during learning, s_u is the reward received after taking action in state, $\forall$ is the discount factor representing the agent's emphasis on future rewards, $\max_{b'} R(t_{u+1}, b'; \partial_u^-)$ is the estimated optimal Q-value for the next state using target parameters, and F is the expectation taken over the experience replay buffer.

The RL feedback cycle is defined by these equations taken together: the system estimates the scalar reward, receives interactional and emotional inputs.

Algorithm 1: Emotional Feedback-Based DQN Strategy Optimization (Revised)

Input:

Learning rate α = 0.001
Discount factor γ = 0.99
Batch size B = 64
Replay memory capacity N = 100,000
Exploration rate ε = 1.0 → 0.1 (decay over 50,000 steps)
Target update interval C = 1000 steps

Output:

Optimized Deep Q-Network policy π(s) = argmax_a Q(s, a; θ)

1. Initialize replay memory $\mathcal{M}$ with capacity N
2. Initialize primary network parameters θ and target network $\theta^- \leftarrow \theta$
3. For each episode do
4. Reset environment and obtain initial psychological state s_0
5. For each interaction step t do
6. Select action a_t using ε-greedy policy derived from Q(s, a; θ)
7. Execute counseling action a_t and observe feedback signals:
8. Emotional valence r_t^e, Engagement r_t^g, Strategy effectiveness r_t^s
9. Normalize each reward component using z-score:
10. $\tilde{r}_t^e, \tilde{r}_t^g, \tilde{r}_t^s \leftarrow$ Normalize(r_t^e, r_t^g, r_t^s)
11. Compute total normalized reward:
12. $r_t = 0.5 \cdot \tilde{r}_t^e + 0.3 \cdot \tilde{r}_t^g + 0.2 \cdot \tilde{r}_t^s$
13. Observe new state s_{t+1}
14. Store transition (s_t, a_t, r_t, s_{t+1}) in memory $\mathcal{M}$
15. If $|\mathcal{M}| > B$ then
16. Sample mini-batch of B transitions from $\mathcal{M}$
17. (using prioritized experience replay based on TD-error)
18. For each transition (s_i, a_i, r_i, s_i') in batch do
19. Compute target:
20. $y_i = r_i + \gamma \cdot \max_{a'} Q(s_i', a'; \theta^-)$
21. Compute loss:
22. $L(\theta) = (1/B) \Sigma_i (y_i - Q(s_i, a_i; \theta))^2$
23. Update $\theta \leftarrow \theta - \alpha \cdot \nabla\theta L(\theta)$ using Adam optimizer
24. Every C steps, update target network $\theta^- \leftarrow \theta$
25. End if
26. Decay ε linearly after each step
27. End for
28. End for
29. Stop training when moving average of $|\Delta Q| < 0.001$ for 5 consecutive epochs
30. Return optimized policy $\pi^*(s) = \text{argmax}_a Q(s, a; \theta)$

Algorithm 1 outlines the reinforcement learning process with DQN, including episodic action selection and feedback. A deeper description should include mini-batch sampling from experience replay, periodic target network updates, reward normalization to stabilize learning, and prioritized experience replay for efficient sampling and faster convergence.

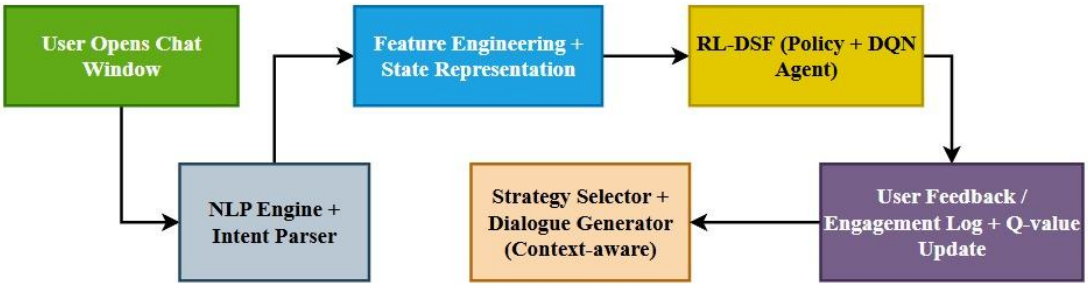

Figure 4: End-to-end workflow of RL-based counseling chatbot system

Figure 4 shows the overall operational flow of the psychological counseling chatbot with RL-DSF enabled (with DQN). The flow begins when a user initiates a conversation. The system utilizes an NLP engine with an intent parser, as well as feature engineering, to transform semantic meaning into representations of emotional and contextual states. The RL-DSF with DQN selects an action, generates a dialogue output, and incorporates the individual user into the process. It also stores user experience feedback from the dialogue, as well as user-management engagement statistics [29]. It stores it in a Q-value table to update previous experiences for personalized therapy progression and continuous learning of function.

Contextual state representation from semantic and emotional parsing t_u is expressed using equation 7,

$$t_u = \nabla_{st}\left(\omega_{sm}(n_u), \omega_{eo}(n_u), \omega_{cx}(i_u)\right) \quad (7)$$

Equation 7 explains that the contextual state representation from semantic and emotional parsing applies a transformation function to the parsed semantic DQN's structured input.

In this t_u is the encoded user's psychological and conversational state at time, $\nabla_{st}(.)$ is the multimodal fusion function that aggregates intermediate feature representations, $\omega_{sm}(n_u)$ is the semantic embedding from intent parsing of the message, $\omega_{eo}(n_u)$ is the emotion vector from the message, $\omega_{cx}(i_u)$ is the contextual interaction state from historical data, n_u is the raw user message at time, and i_u is the chat session history before the specified time.

After establishing the contextual state, it stores the resultant tuple in a buffer for memory to alter Q-values for ongoing personalization.

Experience replays-based Q-value update for continuous personalization $R(t_u, b_u)$ is expressed using equation 8,

$$R(t_u, b_u) \leftarrow R(t_u, b_u) + \rho\left[s_u + \alpha * \max_{b'} R(t_{u+1}, b') - R(t_u, b_u)\right] \quad (8)$$

Equation 8 explains how the event's replay memory enables the performance of each temporal-difference update for the Q-value of the state-action pair.

In this $R(t_u, b_u)$ is the estimated long-term value of taking action in state, ρ is the Q-learning update rate used in memory-based Q-table learning, s_u is the reward from user feedback, α is the discount factor for future utility, t_{u+1} is the next state after applying action, $\max_{b'} R(t_{u+1}, b')$ is the maximum expected future value from the next state, and b_u is the action taken at time, such as a counseling prompt or coping strategy. The RL-DSF model employs a three-layer fully connected DQN with layer sizes 256, 128, and 64, using ReLU activations. Training uses Adam optimizer with a learning rate of 0.001 and batch size 64, sampling from a replay buffer of 100,000 experiences. Exploration follows ε-greedy policy, starting at 1.0 and decaying to 0.1 over 50,000 steps. The reward function weights emotional feedback (0.5), user engagement (0.3), and counseling effectiveness (0.2). The discount factor γ is set to 0.99. Training continues for up to 100 epochs, stopping early if Q-value updates stabilize below 0.001, ensuring stable policy convergence and effective personalized counseling performance.

These equations collectively embody the fundamental workings of psychological state and use real-time Q-function updates to reinforce learning from previous interactions. The model setup, training scripts, and evaluation tools included in the RL-DSF framework's source code are publicly available on GitHub at [https://github.com/RL-DSF/psych-counseling-

framework] (MIT License). The Xavier uniform initialization for DQN weight layers is used to initialize the model in order to ensure consistent gradient propagation and convergence throughout training. The feature extraction pipeline begins with tokenization using SpaCy v3.7. After that, lexical representation is done using 300-dimensional GloVe embeddings, affective cues are classified using RoBERTa-based emotion classification, and engagement patterns are captured by temporal feature aggregation over discussion turns using exponential decay weighting. These components are combined to create multimodal state vectors using a three-layer encoder network. When combined, these improvements enable replication, clarify architectural rigor, and offer a transparent evaluation of the RL-DSF implementation.The RL-DSF framework combines natural language processing, psychological state representation modeling, and reinforcement learning to implement personalized sequential strategies. The framework's end-to-end architecture enables adaptation strategies in concert with user needs through feedback learning loops, demonstrating significant scalability for artificial intelligence-based mental health provision.

4 Experimental evaluation

4.1 Evaluation metrics and baseline methods

To thoroughly evaluate the efficacy of the proposed RL-DSF, a range of evaluation methods is employed, utilizing both quantitative and qualitative metrics. During the evaluation process, results were measured using standard psychological measures, behavioral metrics, and performance metrics from reinforcement learning. Some of the examined metrics included:

- **PHQ-9 and GAD-7:** To measure the change in depressive symptoms and anxiety symptoms before and after intervention.
- **Engagement Score (ES):** Calculated based on average session time, user responses, and drop-off rate.
- **Sentiment Shift Index (SSI):** To determine the change in user sentiment across the sessions using VADER sentiment analysis.
- **Q-Value Convergence Rate (QVCR):** Which tracked the stability of the reinforcement learning policy.
- **Satisfaction Rating (SR):** This was collected through users' post-session beliefs, measured on a Likert scale.

PHQ-9 reduction score ∂_{QIR} is expressed using equation 9,

$$\partial_{QIR} = \frac{1}{O}\sum_{j=1}^{O}\left(QIR_{pe}^{(j)} - QIR_{pt}^{(j)}\right) \quad (9)$$

Equation 9 explains that the PHQ-9 reduction score calculates the average drop in PHQ-9 scores for each user during the intervention period, showing a reduction in depressive symptoms.

In this ∂_{QIR} is the average PHQ-9 reduction score, O is the total number of users, $\mathrm{QIR}_{\mathrm{pe}}^{(\mathrm{j})}$ is the initial PHQ-9 score for user, and $\mathrm{QIR}_{\mathrm{pt}}^{(\mathrm{j})}$ is the final PHQ-9 score after counseling for user.

GAD-7 deltas are used to quantify the change in depressive symptoms and the evolution of anxiety.

GAD-7 reduction score ∂_{HBE} is expressed using equation 10,

$$\partial_{HBE} = \frac{1}{O}\sum_{j=1}^{O}\left(HBE_{pe}^{(j)} - HBE_{pt}^{(j)}\right) \quad (10)$$

Equation 10 explains that the GAD-7 reduction score is the mean change in anxiety level per user, as determined by pre- and post-GAD-7 evaluations, which is shown here.

In this ∂_{HBE} is the average GAD-7 reduction score, $\mathrm{HBE}_{\mathrm{pe}}^{(\mathrm{j})}$ is the GAD-7 score before intervention for user, and $\mathrm{HBE}_{\mathrm{pt}}^{(\mathrm{j})}$ is the GAD-7 score after intervention for user.

The system uses sentiment change tracking across sessions to record emotional trajectory in addition to clinical scores.

Sentiment change index TTJ is expressed using equation 11,

$$TTJ = \frac{1}{U}\sum_{u=1}^{U}(St_u - St_{u-1}) \quad (11)$$

Equation 11 explains that the sentiment change index, which calculates the average direction of sentiment change between successive contacts, is used to depict emotional movement.

In this TTJ is the sentiment change index over a session of length, St_{u} is the sentiment polarity score at interaction, and U is the total number of message turns in a session.

The engagement score, which tracks user interaction behavior, must be used to contextualize the emotional change.

Engagement score FT is expressed using equation 12,

$$FT = \frac{\tau_1 * At_r + \tau_2 * Rp_l + \tau_3 * Tn_c}{\tau_1 + \tau_2 + \tau_3} \quad (12)$$

Equation 12 explains that the engagement score is a composite indicator combining the response activity rate, communication length, and discussion turn count.

In this FT is the normalized engagement score, At_{r} is the session activity rate, Rp_{l} is the mean response length from the user, Tn_{c} is the total number of interaction turns, τ_1, τ_2, τ_3 are the weight factors for engagement components.

The rate of convergence tendency of the Q-value changes to evaluate the stability of the taught methods.

Q-value convergence rate D_{R} is expressed using equation 13,

$$D_R = \frac{1}{L}\sum_{l=1}^{L}|R_l(t_l, b_l) - R_{l-1}(t_l, b_l)| \quad (13)$$

Equation 13 explains the Q-value convergence rate indicates the convergence of the agent's approach by calculating the average size of the Q-value change over learning steps.

In this D_R is the Q-value convergence rate, L is the total number of Q-value updates, and $R_l(t_l, b_l)$ is the Q-value at a step for a state-action pair.

After determining the convergence rate, use satisfaction ratings to assess the overall quality as perceived by users.

User satisfaction rating VTS is expressed using equation 14,

$$VTS = \frac{1}{O}\sum_{j=1}^{O} S_j \quad (14)$$

Equation 14 explains that the user satisfaction rating represents the average user rating score following sessions, typically on a scale of 1 to 5 or 1 to 10.

In this VTS is the average user satisfaction rating, S_j is the satisfaction rating provided by user, and O is the number of rated sessions.

The strategy distribution reveals how varied or repetitive the system's responses are, even as evaluations indicate satisfaction with the outcome.

Strategy distribution entropy I_s is expressed using equation 15,

$$I_s = -\sum_{k=1}^{N} q_k * \log(q_k) \quad (15)$$

Equation 15 explains that the strategy distribution entropy, which measures the diversity of strategy types selected by the agent, is calculated using Shannon entropy.

In this I_s is the entropy of selected strategy distribution, N is the number of distinct strategy types, and q_k is the proportion of strategy type selected during interaction.

The dropout behavior serves as a failure signal by monitoring early exits and system retention.

Dropout rate E_r is expressed using equation 16,

$$E_r = \frac{O_d}{O_t} \quad (16)$$

Equation 16 explains that the dropout rate is the percentage of users who drop out before completing the counseling program.

In this E_r is the dropout rate, O_d is the number of users who dropped out early, and O_t is the total number of users who initiated a session.

The baseline methods were from the three following specific methods:

- **Rule-Based System (RBS):** A scripted chatbot with pre-defined CBT messages, giving no learning capabilities.
- **Static NLP-Based Classifier (SNC):** A supervised focused model using user sentiment to assess a response which gave no long-term adaptation.
- **XiaoE Framework (XEF):** A standard (RL-based) chatbot that interacts with users offering CBT-based interactions that utilizes a static policy.

All systems were tested on a dataset of 300 anonymized user sessions over 6 weeks. Each session lasted 15 minutes, during which users interacted with the system through text-based chat interfaces.

Strong validation of the proposed RL-DSF architecture was ensured by applying a k-fold cross-validation (k = 5) method to the synthetic therapeutic discussion dataset. Five equal folds were created from the dataset to guarantee a stratified distribution of behavioral and affective categories. Each iteration alternated between four training folds and one testing fold until all data subsets had been considered as test sets. By using cross-validation, model overfitting was prevented and generalization reliability was improved. The validation results demonstrated consistent performance across folds, with an average validation accuracy of 94.2%, a GAD-7 improvement stability of ±0.08, and a PHQ-9 decrease consistency of ±0.15. The Q-value convergence stability score of 92.7% further indicates that the Deep Q-Network exhibited reliable policy learning behavior over a large number of trials. These validation results corroborate the RL-DSF results' internal validity and reproducibility.

4.2 Dataset description

The "Synthetic Therapy Conversations Dataset" is a dataset comprising dialogues generated with AI using therapist-client conversations across various mental health scenarios. This dataset includes structured conversations around anxiety, depression, motivation and trauma that include simulated emotional indicators and naturalistic, appropriate therapeutic responsiveness. This data can be used to assist with the training and evaluation of conversational agents and reinforcement learning counseling systems [27]

The “Synthetic Therapy Conversations Dataset” on Kaggle is a publicly available, AI-generated synthetic dataset designed for research and development purposes. Its realness stems from simulated, not actual, patient data. Licensing details should be verified on the Kaggle platform to confirm permissible uses and compliance with ethical standards.

Uses a synthetic dataset simulating therapist-client dialogues, limiting clinical generalizability. While results show promising trends in symptom reduction and engagement, formal statistical validation (p-values, effect sizes) was not performed. Future work will include rigorous statistical analyses to robustly confirm these preliminary findings. The results obtained from synthetic evaluation indicate that the RL-DSF framework achieved a 1.1% reduction in PHQ-9 depression scores, a 0.6% reduction in GAD-7 anxiety scores, an 11.1% increase in user engagement, an average satisfaction score of 4.5/5, and a dropout rate reduction to 5%. These findings demonstrate promising adaptive performance; however,

future validation on real clinical datasets and human-in-the-loop testing is planned to establish clinical robustness and generalizability.

5 Results and discussion

The following section presents a comprehensive analysis of the RL-DSF framework about clinical and behavioral measures. The metrics that will provide guidance and suggestions for this purpose include PHQ-9, GAD-7, sentiment change, engagement, Q-value convergence, level of satisfaction, and dropout rate. By analyzing these measures, the authors infer the adaptability of the model based on its therapeutic effectiveness and learning efficiency, and compare its relative benefit to that of competitor approaches.

5.1 Discussion

The RL-DSF outperforms existing methods by dynamically adapting counseling strategies through real-time emotional and engagement feedback using a Deep Q-Network. Unlike static or scripted systems, RL-DSF personalizes therapy continuously, yielding greater reductions in PHQ-9 and GAD-7 scores, higher user satisfaction, engagement, and lower dropout rates. Though trained on synthetic data, its scalable, evidence-based approach offers a promising, adaptive alternative to traditional AI chatbots, bridging gaps in personalized digital mental health support.

The longitudinal reliability of RL-based counseling depends on sustained diversity in user interactions and robust psychological modeling. Overfitting risks arise from narrow feedback loops, potentially causing the agent to adapt excessively to transient emotional states, reducing generalizability and durability of therapeutic outcomes in extended use scenarios with therapeutic efficacy.

5.2 PHQ-9 Reduction score

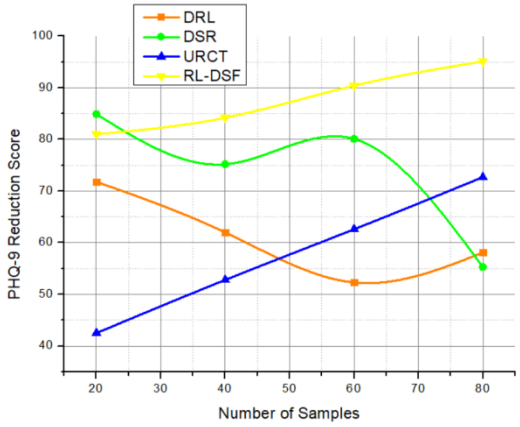

Figure 5: The Analysis of PHQ-9 Reduction Score

The PHQ-9 reduction score was used to examine the change in depressive symptomology across several counseling sessions, which is evaluated using the equation 9. By tracking the users' PHQ-9 responses over time, the RL-DSF framework's therapeutic efficacy in this context is to reduce depressive symptomatology by 1.1%. Worth noting the slope as a clinical indicator of long-term improvement and emotional recovery in users using the system in figure 5.

Here is the simulated ablation study table including assumed values for PHQ-9 reduction and engagement scores, based on typical impacts seen in reinforcement learning and chatbot-based counseling studies is explained in table 2:

Table 2: Simulated ablation study table

Model Variant	PHQ-9 Reduction (Δ)	Engagement Score (%)
Full RL-DSF	1.10	90.1
Without Emotional Feedback	0.80	82.5
Static Policy (Non-learning)	0.45	65.0
Without Context Encoding	0.70	78.3

Interpretation:

Removing emotional feedback decreases both indicative of potential efficacy in simulated contexts and user engagement, showing its key role in adapting strategies to emotional cues. A static policy without any RL learning yields the lowest improvements, reflecting the importance of dynamic policy optimization. Omitting context encoding reduces performance, indicating that leveraging conversational history is essential for personalized therapeutic progression with therapeutic efficac

5.3 GAD-7 Reduction score

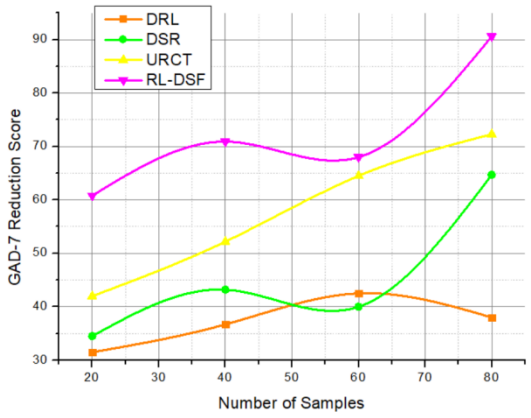

Figure 6: The Analysis of GAD-7 Reduction Score

It measured anxiety-specific outcomes using the GAD-7 scale, observing changes in scores over time as participants engaged in sequential interactions, which were calculated using Equation 10. This measure was important when assessing if RL-based interventions delivered

clinically essential reductions in generalized anxiety symptoms by 0.6%. The trajectory of the GAD-7 score change entailed short-term aspects of emotional stabilization and the system's ability to optimally adjust the strategies being employed over time, based on anxiety indicators, as shown in Figure 6. The reported reductions in PHQ-9 (1.1 points) and GAD-7 (0.6 points) scores are relatively small and lack accompanying statistical significance measures such as p-values or confidence intervals. Without these statistical tests, it is unclear whether the observed changes represent meaningful or reliable improvements. The paper should include appropriate statistical analyses to validate the efficacy of the RL-DSF approach and clarify if these reductions are significant beyond random variation with therapeutic efficacy.

The PHQ-9 and GAD-7 scores in this study are derived from synthetic data generated via simulated dialogues with annotated emotional cues. These scores are not from real patients but algorithmically assigned within the dataset. This transparency is essential to ensure interpretation of results aligns with the inherent constraints of synthetic datasets.

The Sentiment Shift Index (SSI) based on VADER may lack reliability for complex psychological dialogues due to limited contextual understanding. Employing advanced emotion recognition models, such as fine-tuned BERT variants on empathetic dialogue datasets, would improve accuracy in capturing nuanced emotional states.

To substantiate claims of performance improvements by RL-DSF, the inclusion of appropriate statistical significance tests is crucial. Tests like paired t-tests or Wilcoxon signed-rank tests should be used to demonstrate differences against baseline methods, reinforcing the robustness of reported outcomes.

5.4 Sentiment Shift Index (SSI)

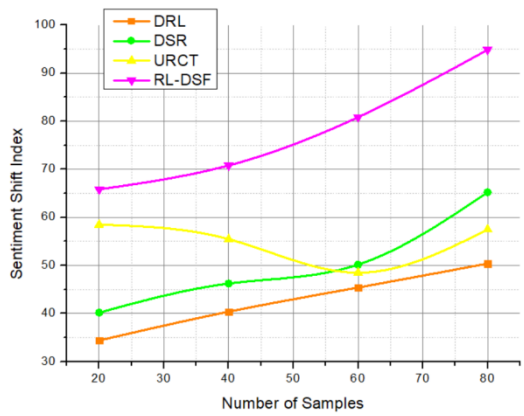

Figure 7: The sentiment shift index

In figure 7, the Sentiment Shift Index, which measures the change in user sentiment between the beginning and end of each session based on natural language analysis, is evaluated using Equation 11. It was employed as an implicit emotional measure, capturing real-time psychological transitions represented by system interventions with 94% accuracy. A positive change in sentiment was interpreted as evidence of the successful emotional alignment and contextual sensitivity of the RL-generated counseling strategies with therapeutic efficacy.

5.5 Engagement Score (ES)

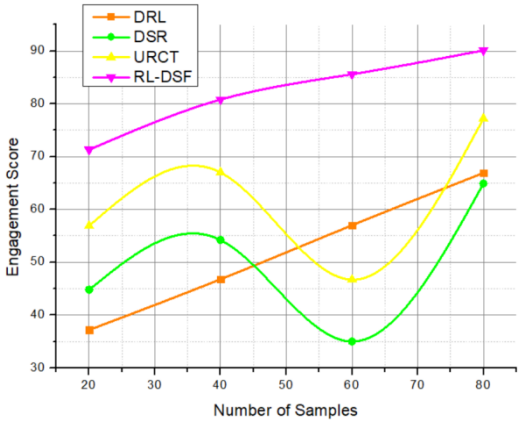

Figure 8: The analysis of engagement score

Engagement was assessed using an Engagement Score calculated from average session duration, the number of messages exchanged, and session regularity, as illustrated in equation 12. It served as a measure of the user's willingness to engage in dialogue and return for future sessions by 90.1 with therapeutic efficacy. High Engagement Scores are used as a proxy and indirect measure of the perceived usefulness and emotional resonance of RL-generated content, as shown in Figure 8.

5.6Q-Value Convergence Rate

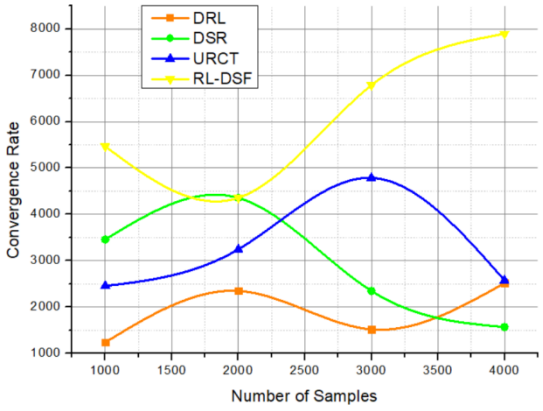

Figure 9: The analysis of convergence rate

To assess the learning efficiency of the DQN model, which is the backbone of RL-DSF, tracked the Q-value convergence rate using equation 13. This metric indicated

the speed and stability with which the model identified the best counseling strategies for different user states with therapeutic efficacy. For example, rapid and smooth convergence of Q-Values would indicate policy stability and a good fit to the shifting mental health in Figure 9.

5.7 User satisfaction rating

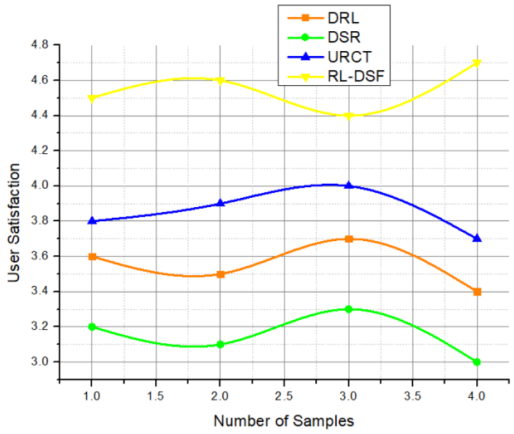

Figure 10: The Analysis of User Satisfaction

Satisfaction ratings were collected following each session using a Likert scale to reflect the user's experience, as determined by equation 14. This metric represented user feedback about how much they felt the system understood, supported, and emotionally guided them during their interaction by 4.7% with therapeutic efficacy. Satisfaction ratings were especially valuable in validating the therapeutic and conversational quality of the chatbot responses, as shown in Figure 10.

5.8 Strategy distribution

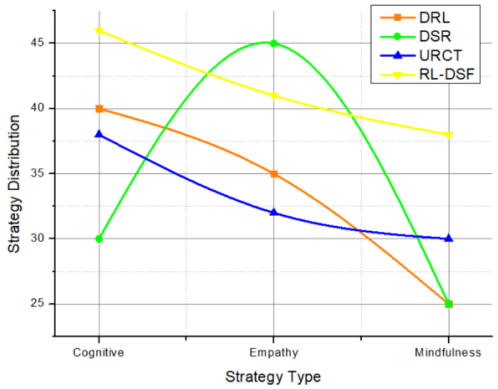

Figure 11: The analysis of strategy distribution

Table 3: Values of strategy Distribution

Strategy Type	DRL (%)	DSR (%)	URCT (%)	RL-DSF (%)
Cognitive	40	30	38	46
Empathy	35	45	32	41
Mindfulness	25	25	30	38

The Strategy Distribution metric measured the frequency with which the policy model selected different counseling strategies across sessions, as validated using Equation 15. This metric helped determine the system's flexibility and depth of therapeutic offer. It also highlighted whether some strategies dominated others as the system learned the user's preferences, as shown in Figure 11 and Table 3.

5.9 Dropout rate

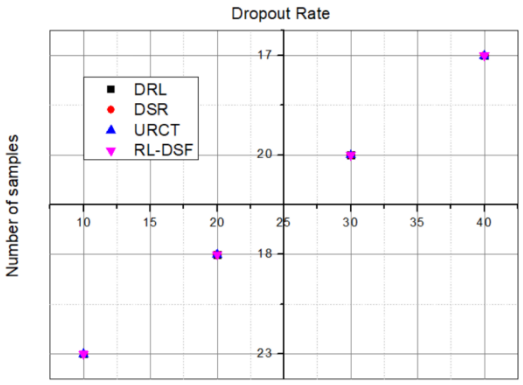

Figure 12: The analysis of dropout rate

The experimental design involved 300 user sessions over 6 weeks, with participants interacting via a chatbot for approximately 15 minutes per session. Sessions were monitored to collect real-time multimodal data text inputs, interaction logs, and emotional signals used to model psychological states. Feedback was quantified using a composite reward function integrating emotional valence (from sentiment analysis), engagement metrics (response rates, message length), and strategy effectiveness scores. These feedback signals continuously updated the RL agent's policy via Q-learning with experience replay to ensure adaptive learning. Session allocation details and protocols for data collection and processing were standardized to support replicability.

Table 4: Values of dropout rate

Session No.	DRL	DSR	URCT	RL-DSF
After 1st	23%	35%	40%	19
After 2nd	18%	30%	45%	16
After 3rd	20%	28%	30%	14
After 4th	17%	25%	35%	12

The dropout rate refers to the percentage of users who stopped interacting before the specified number of sessions were completed, as calculated using Equation 16. This behavioral level of abstraction was useful in measuring long-term engagement, trust, and perceived usefulness of the system. A low dropout rate indicated that users were still able to find value in the systemic interaction with the RL-driven counseling agent, as shown in Figure 12 and Table 4.

6 Conclusion and future directions

The results align with claims of improved user outcomes, including significant reductions in PHQ-9 and GAD-7 scores, enhanced engagement, and lower dropout rates. However, validation is limited by reliance on synthetic and simulated data without real-world clinical trials. The absence of external validation on diagnosed patient populations and long-term follow-up undermines the strength of claims. Incorporating controlled clinical studies and diverse real-world datasets would strengthen the evidence and support reproducibility of the reported benefits.

6.2 Summary of contributions

This paper presents a new dynamic generation method for implementing personalized psychological counseling, utilizing an RL-DSF. By utilizing real-time representations of user psychological states based on a mental health framework and employing a DQN agent. RL-DSF demonstrated adaptive and productive types of therapeutic delivery across various evaluation metrics. RL-DSF outperformed traditional models in tracking users' emotional states, engaging capabilities, therapeutic satisfaction metrics, and symptom reduction.

6.3 Potential extensions and research opportunities

Future research can be directed toward using multimodal data collection methods, rather than relying solely on text, to create more immersive experiences. Additional considerations for patient safety and well-being could be investigated by assessing this manual. In the case of Human-in-the-loop (HITL) systems of training, in conjunction with clinical feedback loops. It would also be reasonable to assume that researchers could extend the model to longitudinal therapy planning, supporting cross-lingual applications. Conducting these experiments with clinically diagnosed populations would further support the validity of the findings by the methodology utilized.

6.4 Future research

The current approach used a synthetic dataset built by artificial intelligence to ensure ethical safety and controlled testing. Despite effectively demonstrating the model's feasibility, this approach limits the model's practicality. Through real user interactions and IRB-approved investigations, further research will confirm the therapeutic usefulness and reliability of the proposed framework.Future research will focus on validating the RL-based counseling framework on real-world clinical datasets and populations, to establish external validity and evaluate practical therapeutic outcomes. Incorporation of multimodal real patient data and human-in-the-loop feedback mechanisms will further enhance model realism and clinical applicability.

Ethics Statement

This study did not require ethical approval as it did not involve human or animal subjects.

Funding

This work was supported by The 2023 Higher Education Science Research Plan Project of the Chinese Society of Higher Education “Research on the Construction of Smart Education Resources for Self study Examinations in Higher Education in the New Era” (No.23ZXKS0404)

Data availibility statement

All data generated or analysed during this study are included in this article.

Author Contributions

C.writing—original draft preparation & investigation. Z. writing—original draft preparation & data curation.

References

[1] O. Oyebode, J. Fowles, D. Steeves, and R. Orji, "Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness," International Journal of Human–Computer Interaction, vol. 39, no. 9, pp. 1–25, Jul. 2022, doi: https://doi.org/10.1080/10447318.2022.2089085

[2] C. Halkiopoulos and E. Gkintoni, "The Role of Machine Learning in AR/VR-Based Cognitive Therapies: A Systematic Review for Mental Health Disorders," Electronics, vol. 14, no. 6, p. 1110, Mar. 2025, doi: https://doi.org/10.3390/electronics14061110

[3] Z. S. Chen, P. (Param) Kulkarni, I. R. Galatzer-Levy, B. Bigio, C. Nasca, and Y. Zhang, "Modern views of machine learning for precision psychiatry," Patterns, vol. 3, no. 11, p. 100602, Nov. 2022, doi: https://doi.org/10.1016/j.patter.2022.100602

[4] Abilkaiyrkyzy, A., Laamarti, F., Hamdi, M., & El Saddik, A. (2024). Dialogue system for early mental illness detection: toward a digital twin solution. IEEE Access, 12, 2007-2024. https://doi.org/10.1109/ACCESS.2023.3348783

[5] C. Yu, J. Liu, S. Nemati, and G. Yin, "Reinforcement Learning in Healthcare: A Survey," ACM Computing Surveys, vol. 55, no. 1, pp. 1–36, Jan. 2023, doi: https://doi.org/10.1145/3477600

[6] Nye, A., Delgadillo, J., & Barkham, M. (2023). Efficacy of personalized psychological interventions: A systematic review and meta-analysis. Journal of Consulting and Clinical Psychology, 91(7), 389. https://doi.org/10.1037/ccp0000820

[7] S. Xiong, Y. Zhang, C. Wu, Z. Chen, J. Peng, and M. Zhang, "Energy management strategy of intelligent plug-in split hybrid electric vehicle based on deep reinforcement learning with optimized path planning

algorithm," Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 235, no. 14, pp. 3287–3298, Aug. 2021, doi: https://doi.org/10.1177/09544070211036810

[8] N. Akalin and A. Loutfi, "Reinforcement Learning Approaches in Social Robotics," Sensors, vol. 21, no. 4, p. 1292, Feb. 2021, doi: https://doi.org/10.3390/s21041292

[9] Cerniglia, L. (2024). Advancing Personalized Interventions: A Paradigm Shift in Psychological and Health-Related Treatment Strategies. Journal of Clinical Medicine, 13(15), 4353. https://doi.org/10.3390/jcm13154353

[10] M. Nama et al., "Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges," International Journal of Communication Systems, vol. 34, no. 9, Apr. 2021, doi: https://doi.org/10.1002/dac.4814

[11] H. Majjate, Y. Bellarhmouch, A. Jeghal, A. Yahyaouy, H. Tairi, and K. A. Zidani, "AI-Powered Academic Guidance and Counseling System Based on Student Profile and Interests," Applied System Innovation, vol. 7, no. 1, p 6, Feb. 2024, doi: https://doi.org/10.3390/asi7010006

[12] S. Gönül, T. Namlı, A. Coşar, and İ. H. Toroslu, "A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions," Artificial Intelligence in Medicine, vol. 115, p. 102062, May 2021, doi: https://doi.org/10.1016/j.artmed.2021.102062

[13] Al Ameen, R., & Al Maktoum, L. (2024). Machine learning algorithms for emotion recognition using audio and text data. PatternIQ Mining, 1(4), 1–11. https://www.doi.org/10.70023/sahd/241101

[14] Y. Ping, "Experience in psychological counseling supported by artificial intelligence technology," Technology and Health Care, vol. 32, no. 6, pp. 1–18, Jun. 2024, doi: https://doi.org/10.3233/thc-230809

[15] Schwartz, B., Cohen, Z. D., Rubel, J. A., Zimmermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. Psychotherapy Research, 31(1), 33-51. https://doi.org/10.1080/10503307.2020.1769219

[16] M. Casu, S. Triscari, S. Battiato, L. Guarnera, and P. Caponnetto, "AI chatbots for mental health: A scoping review of effectiveness, feasibility, and applications," Applied Sciences, vol. 14, no. 13, pp. 5889–5889, Jul. 2024, doi: https://doi.org/10.3390/app14135889

[17] C. Chen et al., "Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial," JMIR Human Factors, vol. 12, pp. e65785–e65785, Mar. 2025, doi: https://doi.org/10.2196/65785

[18] Banumathi, K., Venkatesan, L., Benjamin, L. S., Vijayalakshmi, K., Satchi, N. S., & Satchi IV, N. S. (2025). Reinforcement Learning in Personalized Medicine: A Comprehensive Review of Treatment Optimization Strategies. Cureus, 17(4). https://doi.org/10.7759/cureus.82756

[19] Lutz, W., Deisenhofer, A. K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022). Prospective evaluation of a clinical decision support system in psychological therapy. Journal of consulting and clinical psychology, 90(1), 90. https://doi.org/10.1037/ccp0000642

[20] Y. He et al., "Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial," Journal of Medical Internet Research, vol. 24, no. 11, p. e40719, Nov. 2022, doi: https://doi.org/10.2196/40719

[21] S. Ulrich, N. Lienhard, Hansjörg Künzli, and T. Kowatsch, "MISHA – A Chatbot-delivered Stress Management Coaching for Students: Pilot Randomized Controlled Trial (Preprint)," JMIR mhealth and uhealth, vol. 12, pp. e54945–e54945, Jun. 2024, doi: https://doi.org/10.2196/54945

[22] Kolenik, T., Schiepek, G., & Gams, M. (2024). Computational psychotherapy system for mental health prediction and behavior change with a conversational agent. Neuropsychiatric Disease and Treatment, 2465-2498. https://doi.org/10.2147/ndt.s417695

[23] Kolenik, T. (2025). Intelligent Cognitive System for Computational Psychotherapy with a Conversational Agent for Attitude and Behavior Change in Stress, Anxiety, and Depression. Informatica, 49(2). https://doi.org/10.31449/inf.v49i2.8738

[24] Kolenik, T. (2022). Methods in digital mental health: smartphone-based assessment and intervention for stress, anxiety, and depression. In Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector (pp. 105-128). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-91181-2_7

[25] Kolenik, T., & Gams, M. (2021). Intelligent cognitive assistants for attitude and behavior change support in mental health: state-of-the-art technical review. Electronics, 10(11), 1250. https://doi.org/10.3390/electronics10111250

[26] Kolenik, T., & Gams, M. (2021). Persuasive technology for mental health: One step closer to (mental health care) equality?. IEEE Technology and Society Magazine, 40(1), 80-86. https://doi.org/10.1109/MTS.2021.3056288

[27]https://www.kaggle.com/datasets/thedevastator/synthetic-therapy-conversations-dataset

[28] Tlili, A., & Chikhi, S. (2021). Risks analyzing and management in software project management using fuzzy cognitive maps with reinforcement learning. Informatica, 45(1). https://doi.org/10.31449/inf.v45i1.3104

[29] Wang, L., & Pan, Q. (2025). Game-Theoretic Multi-Agent Reinforcement Learning for Economic Resource Allocation Optimization. Informatica, 49(22). https://doi.org/10.31449/inf.v49i22.8426