RL-AMHA: A Reinforcement Learning-Based Adaptive System for Personalized Mental Health Education and Intervention

Jine Wang
School of Education, Anyang Normal University, Anyang, Henan, 455000, China
E-mail: wangjine0531@hotmail.coma

Keywords: reinforcement learning, mental health education, personalized recommendation, intervention strategy, adaptive learning, psychological outcomes

Received: September 2, 2025

Effective mental health education requires personalized recommendations and intervention strategies that align with an individual's psychological condition and learning preferences. This study introduces a Reinforcement Learning-Based Adaptive Mental Health Advisor (RL-AMHA) that selects instructional content and self-care interventions using a PPO agent and actor-critic neural architecture. The agent uses self-reports, in-app behavior, and contextual cues from a mobile mental health platform. It was trained on 1,200 anonymized user sessions using log data and validated questionnaire scores, after ethics approval and informed consent. The experiment contrasts RL-AMHA with a static rule-based curriculum and a supervised recommendation model that predicts content based on past clicks. All models share the same content pool and interaction constraints; PPO hyperparameters (learning rate, discount factor, clipping range, batch size) were tweaked on a validation split, and 500 episodes were trained until the average episodic reward converged. RL-AMHA improves engagement rate from 75.2% to 88.1% (+2.9%), user satisfaction from 4.21 to 4.85 (+0.64 on a 5-point scale), and weekly self-care activity frequency from 9.4 to 13.1 (+39%) compared to the baseline. Additionally, it enhances stress-reduction scores by 18.6% and increases continuous engagement time from 21.3 to 29.0 minutes. With minimal implementation costs, RL-AMHA demonstrates scalability, adaptability, and effectiveness for long-term psychological support across mobile health, e-learning, and clinical decision-support environments.

Povzetek: Študija predstavi prilagodljiv sistem RL-AMHA, ki z okrepitvenim učenjem personalizira vsebine in intervencije za duševno zdravje ter dokazano poveča vključenost, zadovoljstvo uporabnikov in učinkovitost samopomoči.

1 Introduction

Mental health education is essential to enable individuals of all ages to develop resilience, emotional intelligence, and coping skills [1]. While traditional mental health programming offers general awareness, it cannot adequately adapt to individual psychological needs and learning styles [2]. Technologies related to AI and RL enable systems that provide personalized recommendations based on real-time, user-specific data [3]. This ability to offer customized programming gives new possibilities for more tailored and effective interventions [4].

1.1 Societal need for mental health education personalization

While mental health stressors like anxiety and depression are on the rise globally, stigma and lack of resources are creating barriers to effective Personalization [5]. There is also the reality that a standard one-size-fits-all approach usually does not account for all the diverse socio-cultural backgrounds, cognitive capabilities, and emotional states [6]. Education towards mental health is most personalized, appropriate, and timely when it is relevant to a particular context [7]. If personalized mental health literacy can ensure engaging content that may impact or enhance psychological well-being over time, then all parties may benefit [8].

1.2 Psychological theories guiding intervention design

The intervention framework is based on principles that cover cognitive, motivational, and social aspects to enhance user engagement and behavioral outcomes [9]. Conditioning involves using content to help participants recognize and correct negative thought patterns, whereas interactive content promotes positive styles [10]. The system cultivates an atmosphere of autonomy and personal competence, driving users to take personal ownership of the progress process [11]. It has cooperative and neighborhood-based features to foster supportive peer relationships, thereby strengthening one's commitment to long-term involvement [12]. The level of achievement and readiness determines the program's intensity and focus, tailoring it to individual needs and making it adaptive and individualized [13].

1.3 Technological evolution towards ai-driven personalization

From the early static e-learning modules to modern adaptive learning platforms, the way mental health education is delivered has changed drastically [14]. These current AI capabilities in deep RL enable continuous adaptation to user progress, user-cut emotional feedback, and contextual changes, resulting in highly personalized, outcome-oriented interventions [15].

1.4 Contribution of this paper

- Introduces the RL-AMHA that dynamically recommends personalized mental health education content and intervention strategies by modeling user states, contextual factors, and feedback in real time.
- Designs a multi-objective reward function that balances immediate engagement with sustained improvements in knowledge retention, self-reported mental well-being, and adaptive behavioral change.
- Demonstrates, through experimental evaluation, that RL-AMHA outperforms conventional static recommendations and rule-based methods in terms of content relevance, engagement rate, knowledge retention, and psychological outcome improvement.

Research objectives

- To determine if RL-AMHA significantly improves user involvement (session-level and weekly) compared to CBF, FreeMind, and CFIR-guided systems.
- To determine if RL-AMHA outperforms baseline techniques in user satisfaction for various intervention types (therapy modules, self-care, peer support, adaptive content, mobile interventions).
- Assess RL-AMHA's impact on clinical-proximal outcomes, such as stress reduction, emotional state improvement, and self-care activity frequency, compared to personalized and non-personalized systems.
- To compare RL-AMHA's implementation hurdles (technical, ethical, user-related, regulatory, resource) to baselines and determine if it generates better results with lower or equal barriers.
- To examine how RL-AMHA's temporal modeling (LSTM-attention) and multi-objective reward design improve performance compared to static and myopic techniques.

Hypotheses

- H1: RL-AMHA outperforms CBF, FreeMind, and CFIR in engagement rate and session duration.
- H2: RL-AMHA surpasses baseline user satisfaction levels by a considerable margin.
- H3: RL-AMHA significantly enhanced stress reduction, emotional improvement, and increased frequency of self-care activities.
- H4: RL-AMHA achieves improvements with comparable or lower implementation hurdles than current systems.

1.5 Problem statement

The mental health intervention research, however, has similar limitations. The majority of them depend on a limited variety of participants, hence lack generalizability across cultures and demographics. Some studies use small sample sizes or short intervention durations, which limit statistical power and the evaluation of prolonged effects. Most of them lack strong control groups, uniform measures, or regular follow-up, resulting in uneven reliability of outcomes. Additionally, the adoption of emerging approaches to personalization in AI is limited, and barriers to implementation, including resource constraints and cultural adjustments, remain unaddressed.

2 Literature review

Related literature has examined how technology and psychology combined can improve mental health care. Examples of work related to technology and mental health interventions include hybrid recommender systems, mobile interventions, mindfulness-based programs, and machine learning-based personalization. Digital mental health initiatives integrate technology with recommendations to improve access, relevance, and engagement. As such, these advances can address the individual and social determinants of mental health care.

Hybrid Recommender Systems (HRS) provide personalized mental health recommendations by combining collaborative, content-based, and knowledge-based filtering to optimize accuracy and relevance [16]. By integrating user and personal preferences, behavioral data, and expert knowledge, the system can provide personalized mental health interventions. When the system leverages collaborative filtering, it is based upon similarities between users' mental health profiles to create recommendations that allow the system to suggest coping strategies and resources used by similar individuals that were a success.

The systematic reviews identified personalized recommended systems in mental health, including their techniques, use, and obstacles [17]. The primary method of the Content-Based Filtering (CBF) approach suggests recommending interventions after matching user profiles to item attributes, including type of therapy, time, and focus. To classify mental health issues, CBF can be a valuable approach to recommending close alignment with a user's stated preferences and prior exposures to resources. This can ensure recommendations are relevant and personalized to sensitive needs, but it may also lead to a lack of diversity in recommendations.

FreeMind is a social application that fosters positive self-care and mental health support for college students. FreeMind employs machine learning (ML) in combination with a knowledge expert to recommend individual behavioral strategies for engaging with and managing stress, based on user feedback about their activity preferences and self-reported well-being [18]. FreeMind users enter data into smartphone applications, enabling ML classification models to recognize patterns in their input and suggest personalized self-care activities. The program was evaluated by several students, who tested its usability and feasibility, enhanced social support networks, and engaged with environmental resource access.

The psychological health promotion, prevention strategies, and new interventions - elaborating on their potential efficacy and barriers to implementation. The promising intervention is Mobile- and Internet-Based Interventions (MIBIs), which provide mental health services via smartphones or computers. MIBIs can provide care that is more accessible and flexible, allowing patients to avoid the stigma of seeking care [19]. Mental Health care can move into preventive psychiatry in the community on a larger scale, as well as into countries with limited resources. Providing self-guided, psychoeducational, and remote supportive interventions can help bridge service gaps, though more interventions are needed to maximize their impact globally.

The Mindfulness-Based Interventions (MBIs), CBT, psychoeducation, recreation, and stress management programs to improve college students' mental health. As an effective MBI practice, mindfulness includes meditation and focused breathing while doing other tasks to increase present-moment awareness and make a concerted effort to mitigate stressors [20]. MBIs show reduced anxiety and depression and improved mental well-being in comparison to passive controls. Also, MBIs can take the form of a group program, an individual program, or an online program, making them ideal for any university. Long-term follow-up is needed to assess the lasting impacts of MBIs.

The personalization in Digital Mental Health Interventions (DMHIs) for depressive symptoms was recognized in 66% of interventions. ML-based personalization is "personalization that adapts intervention content or delivery based on data patterns from the user." Although ML-based personalization was utilized in only 3% of studies identified above [21], it can dynamically tailor therapy modules, communication frequency, and guidance level to the individual's needs. Assuming personalization can improve engagement and outcomes, it is necessary to determine its effectiveness in DMHIs.

The pathway for addressing the social determinants of mental health, emphasizing prevention and social justice, is through Universal Primary Prevention Strategies (UPPS), which target entire populations regardless of individual risk level [22]. UPPS could include policy initiatives aimed at reducing poverty, increasing access to education, and ultimately providing affordable housing, which have all been documented to reduce mental health disparities. The earlier that interventions can be made to address structural drivers, the greater the potential to interrupt intergenerational disadvantage, promote equity, and improve population-level mental health or collective wellbeing among marginalized and/or historically more at-risk communities.

Utilizing the Consolidated Framework for Implementation Research (CFIR) to guide the implementation of peer support in mental health services. The CFIR outlines specific implementation domains, including intervention characteristics, organizational context, and staff readiness, which assist in identifying barriers and facilitators [23]. Using a CFIR approach, best practices for implementation focus on clearly defining peer roles, developing a recovery-oriented culture in the mental health setting, and training both peer staff and non-peer staff. Using an organized approach to implementing peer support can relieve cultural resistance and ambiguity regarding peer roles while promoting recovery-oriented mental health services.

The article [25] introduces a computational psychotherapy platform that blends mental-health trajectory prediction with personalized conversational interventions. Although it supports stress, anxiety, and depression, it relies on rule-based adaptation rather than fully optimized, long-horizon reinforcement-learning strategies.

The study [26] suggests a cognitive system combining conversational AI, user-state modeling, and affective feedback to alter attitudes and behaviors in mental health disorders. Although it improves self-reported outcomes, customization is primarily supervised or heuristic, lacking sequential decision optimization like RL-based advisors.The article [27] discusses smartphone-based stress, anxiety, and depression intervention strategies, including EMA, passive sensing, and just-in-time messaging. The text focuses on design, data pipelines, and ethics, but primarily discusses frameworks rather than adaptive control mechanisms like reinforcement learning for multi-step planning.

The paper [28] examines intelligent cognitive assistants that utilize conversation systems, affective computing, and persuasive tactics for mental health behavior change. It categorizes architectures, sensing modalities, and feedback mechanisms, exposing deficiencies in autonomy, personalization, and robustness, particularly the limited use of formal RL approaches for long-term support strategy optimization. This study [29] examines persuasive technologies for mental health care access and adherence, emphasizing customizing, social support, and equity. Although digital interventions can lower barriers and disparities, most systems use static or barely adaptable persuasion tactics instead of data-driven, closed-loop optimization. To ensure effective fixed-time synchronization of fractional-order chaotic systems under uncertainty, [30]. offer adaptive fuzzy controllers that use Lyapunov analysis and adaptive laws to limit mistakes. In Boukabou's output-feedback controller, projective lag synchronization of uncertain chaotic drive-response systems with input dead zones is achieved without full-state measurements. [31]. create a nonlinear control strategy for a gas compressor with an induction motor, enhancing pressure regulation, efficiency, and stability over traditional PI-based strategies. These experiments demonstrate how fuzzy, output-feedback, and optimum control can handle complex nonlinear dynamics with large parametric and structural uncertainty [32].

Table 1: Summary of related works

Ref.	Method	Techniques Used	Key idea	metrics	Main limitation vs. RL-AMHA	Strengths
16	Hybrid RS	CF + CBF + Knowledge-based filtering	Mix of CF, CBF, and knowledge rules for recommending mental-health content.	Typical accuracy/precision ≈70–85% and higher satisfaction than single recommenders.	Static mapping from profile to items; no sequential decision policy or real-time adaptation.	High accuracy, integrates user preferences and expert knowledge, and provides relevant suggestions.
17	CBF	Attribute-based matching	Match user preferences to item attributes.	Modest gain over non-personalized lists; often low novelty.	Limited diversity, no sensor/context loop, no long-horizon optimization.	Personalized to preferences, relevant to sensitive needs.
18	FreeMind	ML classification, user feedback integration	Student self-care app using ML plus expert content.	Better engagement than basic info apps; few long-term outcome stats.	Relies on self-report; lacks explicit RL policy over sessions.	Improves social support, adaptive to preferences, and accessible.
19	MIBIs	Online psychoeducation, remote support	Web/mobile CBT and psychoeducation.	Meta-analyses show small–moderate symptom improvements (effect size ≈0.2–0.5).	Pre-scripted modules; weak personalization and rapid drop-off.	Accessible, flexible, stigma-free.
20	MBIs	Mindfulness exercises	Mindfulness and meditation programs.	Trials report stress/anxiety reduction (effect size ≈0.3–0.6).	Fixed curricula; not data-driven or adaptive.	Proven mental health benefits, flexible formats (group/individual/online).
21	ML-based DMHIs	Machine Learning adaptation	Use ML to tailor content/timing.	Only ~3% of reviewed DMHIs use ML; few show clear long-term gains.	Mostly offline models; do not optimize sequences of actions.	Dynamic, user-specific interventions.
22	UPPS	Policy-level intervention	Policy-level, population-wide prevention.	Population risk reductions, not per-user metrics.	Non-individualized; cannot adapt to momentary state.	Reduces disparities, promotes equity.
23	CFIR peer support	Implementation science framework	Framework for implementing peer support.	Improves implementation and recovery culture (mainly qualitative).	No algorithmic personalization or RL component.	Clarifies peer roles, improves recovery-oriented culture.

The paper indicates that combining digital components with personalized and preventative strategies is effective in promoting mental health outcomes. Hybrid recommender systems, MIBIs, and UPPS address varying degrees of intention to be healthy, and frameworks like CFIR can help with implementation. Future work must target sustainability, scalability, and equitable access to better understand the long-term impact of sustainable, population-level benefits to mental health.

3 Proposed method

The proposed RL-AMHA system integrates reinforcement learning and adaptive mental health assistance, incorporating a user's profile, emotions, and behavior patterns. It has five related modules: data processing, decision making, personalized recommendations, feedback loop learning, and ongoing systems and service integration. The system is designed to provide context-specific, adaptive, customized interventions according to the user's needs, goals for wellbeing, and usage over consequence and time.

Barriers to implementation: The 5-dimensional "Digital Mental Health Implementation Scale" was inspired by digital intervention checklists. End-users and practitioners scored perceived Technical, Ethical, User-related, Regulatory, and Resource barriers on a 5-point Likert scale (1 = very low barrier, 5 = very high barrier). To determine the lowest implementation obstacles for RL-AMHA compared to CBF, FreeMind, and CFIR-style processes, subscale scores were summed and rescaled to 0-40 for visualization. The post-study survey included a "Cultural Fit and Acceptability" module based on cultural adaptation frameworks for digital mental health tools. Criteria included perceived respect for values, appropriateness of language, relevance of examples, and willingness to recommend to peers in the same cultural group. The Cultural Acceptance Index (mean of all items) was calculated from 5-point Likert scale ratings. RL-AMHA describes cultural adaptability as scoring noninferior or superior on this index relative to baselines, enabling customization of language, content themes, and reward weights without reworking the fundamental polic

3.1 Data collection & preprocessing

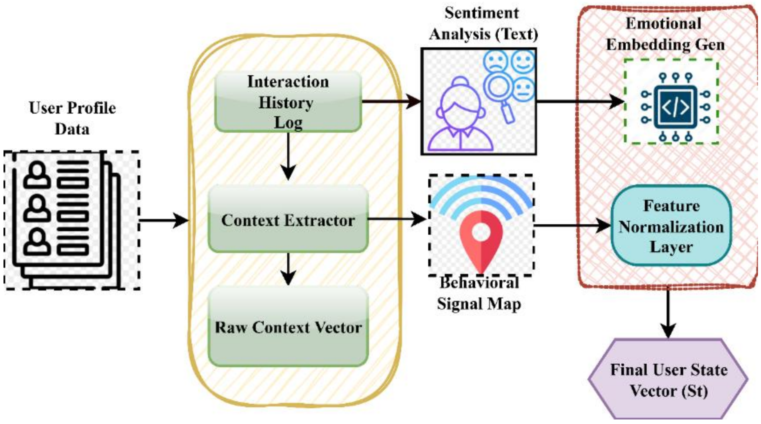

Figure 1: Data collection & preprocessing

Figure 1 represents the data acquisition and preprocessing step in RL-AMHA, where multiple sources of input, including user profile information and interaction history, are then combined with situational and emotional information to generate a user state. The context extractor takes situational features, and sentiment analysis and behavioral signal mapping take psychological patterns. The emotional embedding generator takes affective states, and the feature normalization layer controls for scale across variables. The modules interact bi-directionally to generate the final user state vector (St), a structured and machine-readable representation from which many modeling decisions and selections were made in the RL-based component.

User state vector construction T_u is expressed using equation 1,

$$T_u = g(V_q, J_i, D_t, F_b, C_n) \quad (1)$$

Equation 1 explains the user state vector construction in order to create a single state vector at time, this function f combines several user-related inputs.

In this T_u is the final composite user state vector at time serving as the RL agent's input, V_q is the static user profile features, J_i is the temporal interaction history, D_t is the contextual situational features, F_b is the emotional embedding vector encoding affective states derived via sentiment analysis, and C_n is the behavioral mapping vector representing psychological and activity patterns.

Feature normalization layer $\hat{y}_j$ is expressed using equation 2,

$$\hat{y}_j = \frac{y_j - \pi_j}{\rho_j + \tau} \quad (2)$$

Equation 2 explains the feature normalization layer each feature is normalized using this normalization formula, which centers it with the mean and scales it by the average deviation.

In this $\hat{y}_j$ is the normalized value of the feature input, y_j is the original feature value before normalization, π_j is the mean value of the feature computed over the training dataset, ρ_j is the standard deviation of the feature computed over the training dataset, and τ is a small positive constant to prevent division by zero.

3.2 RL-AMHA core

Figure 2 shows the core decision-making module of RL-AMHA, which receives the final user state (St) and is transformed by the DL policy network with LSTM and attention to model temporal dynamics and relevance. The policy network produces probabilities over actions and a value function V(s) for baseline consideration. The action selector uses ε-greedy or softmax to explore or exploit possible optimal actions. Advantage estimation A(s, a) is used to improve the policy. The combination produces a selected action (At), e.g., a CBT session or personalized tip, to maximize user well-being.

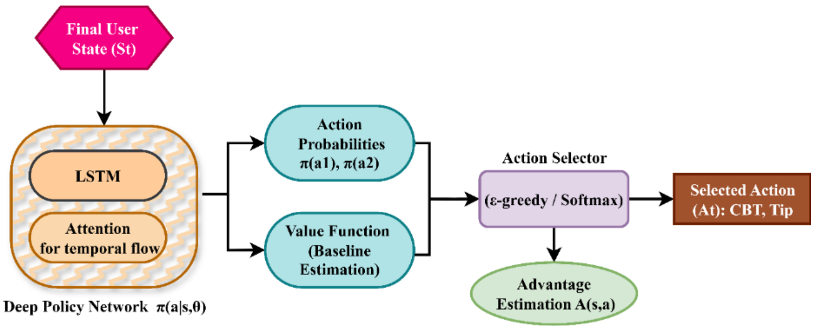

Figure 2: RL-AMHA core

Policy network output $\mu_\sigma(b_u|t_u)$ is expressed using equation 3,

$$\mu_\sigma(b_u|t_u) = \left(h_\sigma(T_u)\right) \quad (3)$$

Equation 3 explains that the policy network output user state is mapped to logits for all potential actions via the parameterized function.

In this $\mu_\sigma(b_u|t_u)$ is the probability of taking action given the state under policy parameters, $h_\sigma(T_u)$ is the raw output logits vector from the policy network for the state, T_u is user state vector at time, b_u is the candidate action at the time, and σ is the policy network parameter.

Action selection via a greedy strategy b_u is expressed using equation 4,

$$b_u = \begin{cases} \text{r a} & \text{with probability } \partial \\ R(t_u, b) & \text{with probability } 1 - \partial \end{cases} \quad (4)$$

Equation 4 explains the action selection via greedy strategy by choosing a random action with a chance to find novel tactics or the most well-known action.

In this b_u is the action chosen at time, ∂ is the exploration rate parameter, $R(t_u, b)$ is the estimated quality of taking action in state, and t_u is the current user state vector.

Advantage estimation for policy update $b(t_u, b_u)$ is expressed using equation 5,

$$B(t_u, b_u) = S_u + \partial W(t_{u+1}) \quad (5)$$

Equation 5 explains the advantage estimation for policy update, the function of the advantage. The difference between taking action in state and the baseline value is measured.

In this $b(t_u, b_u)$ is the advantage of action at state, S_u is the immediate reward received after taking action, ∂ is the discount factor for future rewards, and $W(t_{u+1})$ is the estimated value of next state.

Algorithm 1 uses a policy gradient framework with PPO as the main learning algorithm and REINFORCE and A2C as adjunct policy heads and critics. Three policies are trained simultaneously on the same trajectories, and their action probabilities are pooled using a weighted consensus technique. The final value is the sum of the factors PPO, A2C, and REINFORCE. w 1 > w 2 > w 3 .

$$\pi_{\text{final}} = w_1\pi_{\text{PPO}} + w_2\pi_{\text{A2C}} + w_3\pi_{\text{REINFORCE}}, w_1 > w_2 > w_3$$

Online weight updates from return and safety scores favor PPO deployment, while A2C/REINFORCE stabilize advantage estimates in sparse feedback conditions. Algorithm 1 will be updated to (i) indicate the ensemble in the heading, (ii) name PPO as the major learner, and (iii) merge the three algorithms into a single RL-AMHA policy with the consensus update step and weight-adaptation procedure.

Pseudocode: Reinforcement Learning-Based Adaptive Mental Health Advisor

Initialization

initialize Policy πθ (parameters θ) *initialize UserProfile U*

initialize SafetyThresholds $S = \{risk_{high}, risk_{medium}\}$

initialize ReplayBuffer B

set hyperparams: γ (discount), α (learning rate), ε (exploration)

loop every interaction timestep t:

Step 1. Observe state

$S_t = observe_{state}(user_{context}, sensor_{inputs}, U)$

 # example fields: mood_score, activity, time_of_day, engagement_history, recent_responses

Step 2. Quick safety check (highest priority)

if S_t *indicates* $immediate_{crisis}$ *or* $suicidal_{ideation}$:

$show_{immediate_{crisis_{message(n)}}}$

$escalate_{to_{human_{clinician(n)}}}$

$\log_{event(\text{crisis\_escalation})}\ n)$

continue # skip RL recommendation for this timestep

Step 3. Compute risk level

if $S_t.risk_{score} \geq S.risk_{high}$:

$action_{space} = \left\{low_{intensity_{interventions}}, clinician_{contact}, crisis_{protocol}\right\}$

else if $S_t.risk_{score} \geq S.risk_{medium}$:

$action_{space} = \left\{ moderate_{interventions}, \mathrm{check}_{in_{within_{short_{window}}}} \right\}$

else:

$action_{space} = \{educational_{content}, micro\text{-}interventions, gamified_{tasks}\}$

***Step* 4. *Policy decision* (*with simple if/else fallback*)**

if $random() < \varepsilon$:

$A_t = sample_{random(action_{space})}$

else:

$A_t = \pi\theta.select_{action(S_t,\, action_{space})}$

Step* 5. *Apply action / recommend content

$deliver_{action_{to_{user(A_t)}}}$

Step* 6. *Collect feedback & compute reward

$feedback = collect_{feedback(user_{response},\, engagement_{metrics},\, follow_{up_{assessments}})}$

$R\_t = compute\_reward(feedback, short_{term_{engagement_{weight}}}, long_{term_{outcome}}\_weight)$

Step* 7. *Store transition and update personalization

$B.push\left(\left(S_t, A_t, R_t, S_{\{t+1\}}, feedback\right)\right)$

$update_{UserProfile(U,\, feedback,\, A_t)}$

***Step* 8. *RL update step* (*policy improvement*)**

if $enough_{samples_{in(B)}}$:

$batch = B.sample(minibatch_{size})$

$loss = compute_{RL_{loss(batch,\, \gamma)}}$

$\theta = \theta - \alpha * gradient(loss)$

Step* 9. *Post-delivery conditional checks for follow-up schedule

if $feedback\ indicates_{no_{engagement}}or\,negative_{response}$:

if $U.recent_{rejections} > threshold$:

$reduce_{recommendation_{frequency(U)}}$

$switch_{to_{lower_{dose_{interventions(U)}}}}$

else:

$schedule_{immediate_{follow_{up_{probe(n)}}}}$

else if $feedback\ indicates_{high_{engagement}}and\,symptom_{improvement}$:

$reinforce_{similar_{actions_{in_{policy(\pi\theta,\, A_t)}}}}$

$schedule_{periodic_{booster(U)}}$

Step* 10. *Long-term evaluation & model maintenance

if $time_{to_{run_{periodic_{evaluation(n)}}}}$:

$evaluate_{policy_{on_{holdout_{users(n)}}}}$

$if\ performance_{drop_{detected}}$:

$retrain_{or_{finetune_{model(n)}}}$

Step* 11. *Safety logging & human-in-loop alerting

$if\ any_{flagged_{events_{in(feedback,\ S_{\{t+1\}})}}}$:

$notify_{clinical_{team(with_{context=S_t},\ actions=A_t,\ feedback=feedback)}}$

end loop

The RL-AMHA algorithm 1 personalizes mental health education by observing user states, checking safety risks, and choosing interventions through reinforcement learning. If high risk, it triggers crisis protocols; otherwise, it recommends adaptive content. Feedback updates the policy, balancing short-term engagement and long-term well-being while maintaining safety and human-in-loop oversight. To specialize without training a distinct model for each user, RL-AMHA uses a learnt user embedding and LSTM-attention encoder to summarize each person's history. To reduce sparse feedback, the replay buffer mixes data from high and low-activity users, resulting in cohort-level gradients rather than a few heavy users. Multi-objective rewards handle conflicting feedback: high engagement but increased stress result in poor net returns due to well-being penalties, limiting dangerous behaviors. As more interactions occur, user embeddings and value assessments change from generic to personalized.

3.3 Recommendation & intervention engine

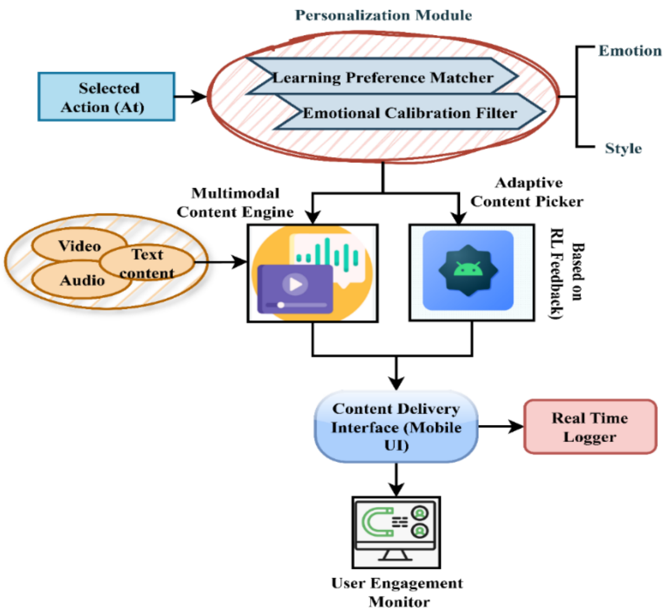

Figure 3: Recommendation & intervention engine

Figure 3 represents the personalized content delivery pipeline of RL-AMHA. The Selected Action (At) is processed by the personalization module, which synchronizes recommendations with the user's learning style and emotional state through the learning preference matcher and the emotional calibration filter. The output flows into the adaptive content picker, based on RL feedback, and the multimodal content engine, which can deliver video, audio, and text content. The final output is displayed using the content delivery interface, the user engagement monitor, and the real-time logger, which track various engagement metrics for learning and improvement.

Personalized content matching score N_u is expressed using equation 6,

$$N_u = \partial * MQN(b_u, v) \quad (6)$$

Equation 6 explains that the personalized content matching score, correspondence between the user's customized characteristics and the chosen action is measured by the matching score.

In this N_u is the composite matching score for action, $MQN(b_u, v)$ is the learning preference matcher score evaluating the suitability of the action for the user's learning style, ∂ is the tunable weighting parameter, b_u is the selected action at the time, and v is the target user.

Adaptive content selection probability $Q(d_j|b_u, v)$ is expressed using equation 7,

$$Q(d_j|b_u, v) = \frac{R(d_j, b_u, v)}{\theta} \quad (7)$$

Equation 7 explains the adaptive content selection probability given action and user, this soft-max-based equation calculates the likelihood of selecting a content item from a collection of candidate pieces.

In this $Q(d_j|b_u, v)$ is the probability of selecting content, $R(d_j, b_u, v)$ is the estimated quality score or expected reward of content, θ is the temperature parameter controlling randomness in selection, d_j is the specific content item among candidates, b_u is the selected action at the time, and v is the target user.

3.4 Feedback loop & reward estimation

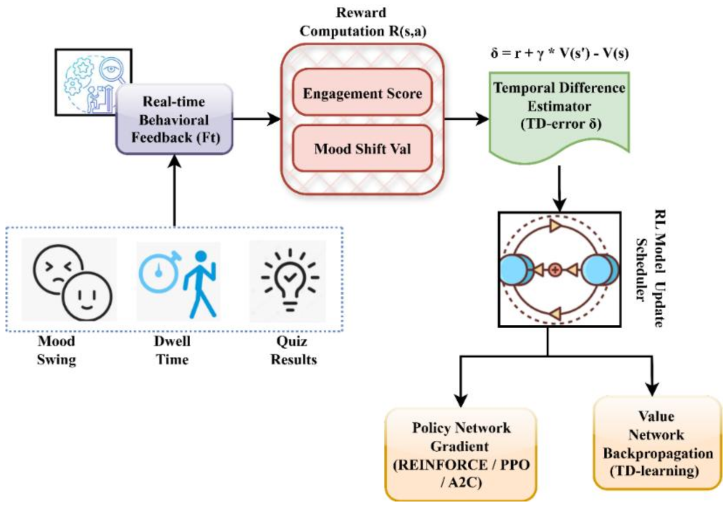

Figure 4: Feedback loop & reward estimation

Figure 4 illustrates RL-AMHA's feedback and learning loop. The real-time behavioral feedback (Ft), including mood changes, dwell time, and quiz performance, is used in Reward Computation R(s, a). The R(s, a) module processes engagement scores with the mood shift values. The Temporal Difference Estimator computes the TD-error (δ), a measure of prediction error. The RL Model Update Scheduler schedules fresh training updates (where gradients are computed) to the policy network (using either REINFORCE, PPO, or A2C) and to the Value Network (using TD-learning backpropagation). Thus, the RL Model Update scheduler is the final piece of our RL-AMHA feedback and learning loop, providing ongoing training and policy updates that refine the model's recommendations and improve the accuracy of its implications for a user's long-term mental health outcomes.

Reward computation from behavioral feedback $S(t_u, b_u)$ is expressed using equation 8,

$$S(t_u, b_u) = x_1 * F_u + x_2 * \nabla N_u \quad (8)$$

Equation 8 explains the computation of rewards from behavioral feedback, aggregating several engagement indicators, such as engagement score, mood change, and quiz performance.

In this $S(t_u, b_u)$ is the scalar reward signal for taking action, F_u is the real-time engagement score, ∇N_u is the mood shift value, representing change in emotional state after, and x_1, x_2 are the non-negative scalar weights balance components of the reward function.

Temporal difference error ∂_u is expressed using equation 9,

$$\partial_u = S(t_u, b_u) + \sigma W(t_{u+1}) \quad (9)$$

Equation 9 explains the temporal difference error between the expected value and the updated estimate, which combines the immediate payout and the discounted next-state value.

In this ∂_u is the temporal difference error at time, $S(t_u, b_u)$ is the immediate reward received for action, σ is the discount factor for future rewards, and $W(t_{u+1})$ is the estimated value of the next state.

Policy gradient update $\Delta_\tau K(\tau)$ is expressed using equation 10,

$$\Delta_\tau K(\tau) = F_\tau\left[\Delta_\tau * \mu_\tau(b_u|t_u) * \partial_u\right] \quad (10)$$

Equation 10 explains the policy gradient update expected return's gradient concerning policy parameters, decreasing slope in the log-probability.

In this $\Delta_\tau K(\tau)$ is the gradient of the objective function to policy parameters, $\mu_\tau(b_u|t_u)$ is the probability of action given state under a policy parameterized, ∂_u is the temporal difference error used as an advantage estimate, F_τ is the expectation over trajectories generated by policy, b_u is the action taken at time, t_u is the state at time, and τ is the policy network parameters.

3.5 System integration & continuous learning

Figure 5 illustrates RL-AMHA's systems integration and adaptive learning architecture. The mobile app/web portal UI interfaces with the recommendation API layer, providing personalized content. The frontend interactive module collects user interactions, and the real-time engagement tracker (RET) collects behavioral data. Data flows into cloud storage and the feedback + context repository, storing anonymized session details, histories, and reflections on mood shifts. The RL Training Scheduler operates through an online/offline update policy that employs batch replay memory and mini-batch fine-tuning. The model deployment pipeline utilizes a CI/CD policy that allows for canary and A/B testing methods to evaluate updates, enabling an automated, rolling deployment policy of updated policy π(a|s) and value functions V(s).

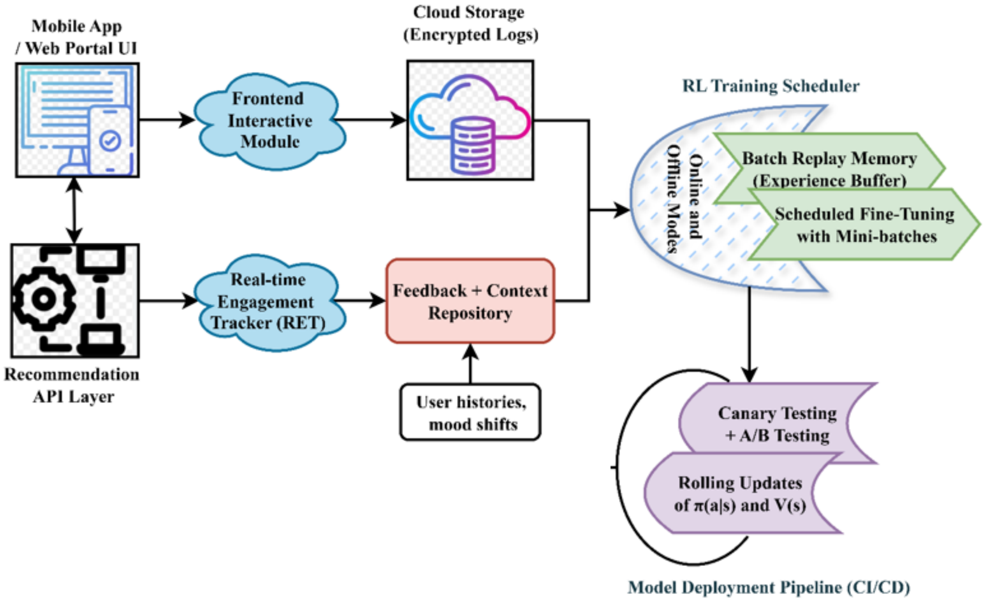

Figure 5: System integration & continuous learning

Batch replay memory update E_{u+1} is expressed using equation 11,

$$E_{u+1} = E_u \cup \{(t_u, b_u)\} \quad (11)$$

Equation 11 explains the batch replay memory update by adding the most recent experience tuple gathered at time, the rerun memory buffer is gradually updated.

In this E_u is the replay memory set at the iteration containing experience tuples, t_u is the state observed at time, b_u is the action taken at the time, and ∪ is the set union operator denoting the addition of new experience to memory.

Policy deployment with a/b testing probability $\mu^{dpy}(b|t)$ is expressed using equation 12,

$$\mu^{dpy}(b|t) = q * \mu^{nw}(b|t) + (1 - q) \quad (12)$$

Equation 12 explains the policy deployment policy combination of the new policy and the stable policy, currently, throughout canary testing.

In this $\mu^{dpy}(b|t)$ is the probability of an action given a state in the deployed policy, $\mu^{nw}(b|t)$ is the probability under the updated policy model, q is the mixing coefficient representing the fraction of users routed to the new policy, b is the candidate action, and t is the current user state.

A guarded updating loop controls continuous learning. Policy evaluations are conducted regularly on a fixed offline replay set and compared to a frozen reference policy. Any decrease in engagement, well-being, or safety measures prevents deployment. New models undergo canary/A-B tests on a limited user group and are reversed if online KPIs decline. To avoid forgetting, the replay buffer blends recent and older interactions, ensuring different historical states in mini-batches. Finally, mood, stress, and engagement drift indicators decrease update rates and require human assessment before large policy changes.

Evaluation metrics:

The evaluation metrics described utilize psychological, behavioral, and operational indicators to evaluate the RL-AMHA system's performance. The measures regarding satisfaction, engagement, stress reduction, frequency of activities taken up, barriers to implementation, resource utilization, and cultural fit provide a multi-factorial approach for measuring the effectiveness of the intervention and ensuring user-centered and context-relevant delivery of mental health education.

User satisfaction score T_u is expressed using equation 13,

$$T_u = \frac{1}{O} * t_j \quad (13)$$

Equation 13 explains that the user satisfaction score, the average rating provided by users, which represents the perceived efficacy and quality of the service or material provided, is used to calculate.

In this T_u is the aggregate user satisfaction score, O is the total number of surveyed users, and t_j is the satisfaction rating from the user.

User engagement rate F_s is expressed using equation 14,

$$F_s = \frac{U_{ate}}{U_{ttl}} * 100\% \quad (14)$$

Equation 14 explains the user engagement rate as the proportion of users using interaction time relative to the total attainable session time.

In this F_s is the user engagement rate, U_{ate} is the duration of active user interaction within a session, and U_{ttl} is the total session duration available.

Stress level reduction ∇M_t is expressed using equation 15,

$$\nabla M_t = M_t^{pe} - M_t^{pt} \quad (15)$$

Equation 15 explains the stress level reduction: the difference, as determined by validated psychological measures.

In this ∇M_t is the magnitude of stress reduction, M_t^{pe} is the stress level before intervention, and M_t^{pt} is the stress level after intervention.

Engagement duration E_f is expressed using equation 16,

$$E_f = \frac{1}{N} * e_k \quad (16)$$

Equation 16 explains the engagement duration mean of each user session length over recorded sessions is used to calculate the average engagement time.

In this E_f is the mean engagement duration per session, N is the number of user sessions recorded, and e_k is the duration of the user session.

Self-care activity frequency G_{td} is expressed using equation 17,

$$G_{td} = \frac{B_x}{X} \quad (17)$$

Equation 17 explains that the average frequency of self-care actions completed by each user over a time window is measured by the self-care frequency.

In this G_{td} is the average self-care activity frequency per user per time window, B_x is the total of self-care activities recorded during the window, and X is the length of the time window.

Barriers to implementation index C_j is expressed using equation 18,

$$C_j = c_l * x_l \quad (18)$$

Equation 18 explains that the barriers-to-implementation index's total difficulty level is quantified as the weighted sum of the identified implementation-related hurdles.

In this C_j is the composite barrier to implementation score, c_l is the presence or magnitude of the barrier, and x_l is the weight representing the importance or impact of barrier.

Resource allocation efficiency S_f is expressed using equation 19,

$$S_f = \frac{V_s}{D_s} \quad (19)$$

Equation 19 explains that operational efficiency, which is indicated by the ratio of resources used that directly contribute to intervention efficacy over all resources used, is a measure of resource allocation efficiency.

In this S_f is the efficiency ratio for resource allocation, V_s is the amount of resources effectively used, and D_s is the total resources allocated to the project.

Cultural acceptance score D_b is expressed using equation 20,

$$D_b = \frac{1}{O_d} * d_n \quad (20)$$

Equation 20 explains the cultural acceptance score measure of how well the program conforms to social and cultural standards, which is averaged from the community of user feedback evaluations.

In this D_b is the mean cultural acceptance rating, O_d is the number of respondents evaluating cultural compatibility, and d_n is the respondent's cultural acceptance rating.

The equation 21, FTJ calculates emotional state improvement score that have positive sign.

$$FTJ = [A_u - A_{tu} - A_{boz} + A_i] \quad (21)$$

Linear combinations are adjusted A_u and scaled A_{tu}, and then clamped using the clamp function. The index can gauge balanced changes in multiple affective dimensions A_{boz} while minimizing harm A_i from outlier distortions.

Equation 22 specifies the behavioral engagement index CFJ as a saturating exponential utility, with adherence components, content completion.

$$CFJ = \left[1 - f^{-(\forall d + \propto x + \tau e)}\right] * (1 - r) \ (22)$$

A multiplicative hazard penalty ($\forall d$) reduces the overall index $\propto x$ by accounting for the churn τe, or the probability of disengagement r.

Together, the metrics reflect how well the user is doing as well as the efficiency of the system. They allow for continuous improvement by capitalizing on successes, identifying areas for improvement, and ensuring that users' cultural and resource considerations align with the provision of mental health recommendations. Holistically, these metrics enable scalable, flexible, and actionable mental health recommendations to optimize users' long-term psychological well-being.

Learning a closed-loop policy that optimizes intervention sequences, not just single-step predictions, makes RL-AMHA better than other adaptive approaches. Unlike supervised neural networks or fuzzy controllers, it updates online from each interaction, investigates alternative tactics, and improves engagement, stress reduction, and self-care via a multi-objective reward. This makes behavioral drift adaption and engagement recovery faster than batch-retrained neural recommenders or manually modified fuzzy rules. While output-feedback controllers presume simple dynamics and objectives, RL-AMHA's LSTM-attention backbone handles high-dimensional, partially seen states for richer temporal modeling and data-driven customisation.

4 Results and discussion

This section presents an assessment of resource utilization, contextual preferences, and RL-AMHA's performance compared to other materials, while also examining how reinforcement learning optimization enhanced personalization, user engagement, and psychological outcomes. It presents evidence of RL-AMHA's superior performance and discusses its ability to provide scalable, effective mental health education.

4.1 Dataset description

The student mental health & resilience dataset on kaggle contains behavioral, emotional, and academic data with stress labels available. The factors it can tap into include psychometric indicators such as mood, stress markers, engagement, and academic behavior and performance among students [24]. This multimodal dataset is conducive to modeling the dynamics of mental states and resilience factors. It enables the training and testing of learning platforms that utilize reinforcement learning, such as RL-AMHA. The semi-complex scans enable the exploration of personalized intervention strategies based on students' real-life behavior.

The Kaggle "Student Mental Health & Resilience" dataset was cleaned, one-hot encoded, and z-normalized before being divided by student into 70% train, 15% validation, and 15% test partitions. Over time, student records were sorted into RL episodes. In pretraining, stress, anxiety, depression-risk, and resilience categories were mapped to 0-1 scores for use in the state vector and as supplementary supervised targets. The multi-objective signal used by RL-AMHA was created by combining changes in well-being (lower stress/higher mood), engagement (more participation and stable grades), self-care behavior, and penalties for deterioration or drop-out, normallyized to [-1, 1] and linearly weighted.

Table 2: Student mental health & resilience dataset

Attribute	Description
Dataset Name	Student Mental Health & Resilience Dataset
Source	Kaggle – Ziya07 (Link)
Content Type	Behavioral, emotional, and academic indicators with stress-related labels
Key Features	Mood scores, stress levels, engagement metrics, and academic performance
Purpose	Modeling mental state dynamics and resilience factors
Use Case	Training/evaluating reinforcement learning–based adaptive systems (RL-AMHA)
Application	Development of personalized mental health intervention strategies

4.2 User satisfaction (Score 1–5)

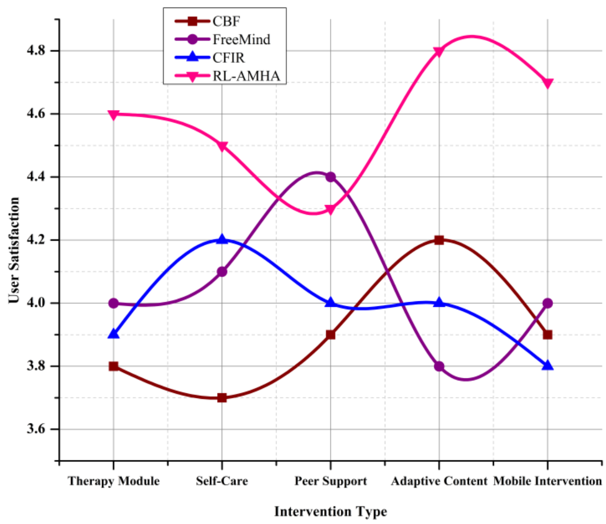

Figure 6: User Satisfaction (Score 1–5)

Figure 6 illustrates user satisfaction across the five types of interventions - Therapy Module, Self-Care, Peer Support, Adaptive Content, and Mobile Intervention - for four systems: CBF, FreeMind, CFIR, and RL-AMHA. RL-AMHA is leading the other systems made evaluated using equation 13. It receives the highest overall score, which is around 4.85, in adaptive content. FreeMind is rated moderately, with a rating of approximately 4.0 to 4.4, and CFIR is also rated at peak, with Self-Care receiving a rating of approximately 4.2. However, all of these ratings are fluctuating, and CBF has been rated the lowest throughout the entire process, with a rating of approximately 3.55 to 4.0. The results of this study reveal that RL-AMHA is clearly superior in terms of adaptation and personalization across a wide range of mental health interventions, particularly in the field's mobile-based and adaptive modules.

4.3 User engagement (%)

Figure 7 shows user engagement (%) across five weeks for four systems: CBF, FreeMind, CFIR, and RL-AMHA. In user engagement, RL-AMHA measures steady user engagement and achieves the highest usage, engaging all users made evaluated using equation 14.

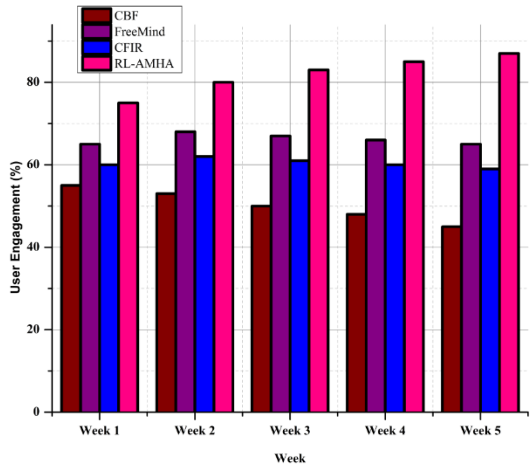

Figure 7: User engagement (%)

User engagement in RL-AMHA increased from ~75% in week 1 to ~88% in week 5, suggesting high retention and responsiveness. FreeMind showed slight variation and relatively high user engagement stats as well (~65 increase - ~68 in Week 5). CFIR shows a moderate level of user engagement (~59 - ~62), with slight fluctuations. CBF scored lowest as user engagement declined from about ~55% to ~46% over time. Again, the key takeaway is that RL-AMHA demonstrates its ability to promote user participation over extended periods of use.

4.4 Stress level reduction (score)

Figure 8 shows the reduction in stress levels from pre-intervention to post-intervention week 4 (five time points) for four systems: RL-AMHA, CFIR, FreeMind, and CBF. All four systems started at similar reductions in stress level (7-7.8), although each method had a different trajectory for lowering stress level made evaluated using equation 15. Throughout the course, observe RL-AMHA consistently delivering high performance, with gradual improvements evident post-week 4. See CFIR and FreeMind with moderate, stable reductions. CBF remains relatively robust despite the decrease in stress levels, although with less improvement than the other methods. The data show that RL-AMHA has a longer-term impact on stress reduction and can sustain its progress over time compared to other methods.

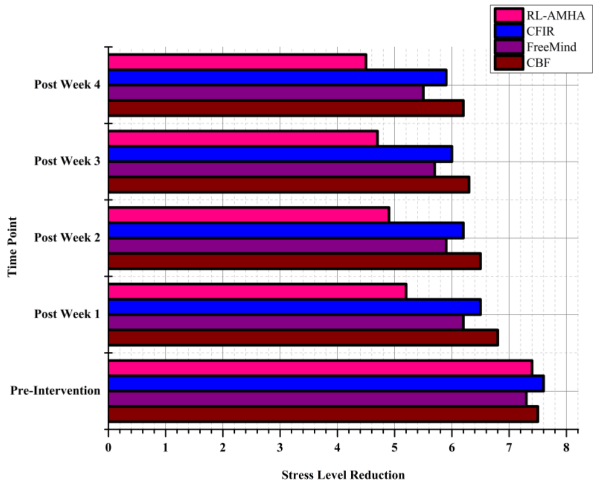

Figure 8: Stress level reduction (score)

4.5 Engagement duration (minutes)

Figure 9 shows the total engagement time by users during five sessions for each system, CBF, FreeMind, CFIR, and RL-AMHA. The following overview captures these graphs. For each method, users continued to show more engagement time, except for CBF, which shows engagement time declining from ~15 to ~12 minutes made evaluated using equation 16. RL-AMHA delivered the longest engagement time; it grew from ~22 to nearly 29 minutes by Session 5. Both FreeMind and CFIR show milder upward trends, with near peaks of engagement time around 21 and 19. Overall, these findings depict that RL-RL-AMHA showed the most consistent increase in user Engagement from Session 1 to Session 5.

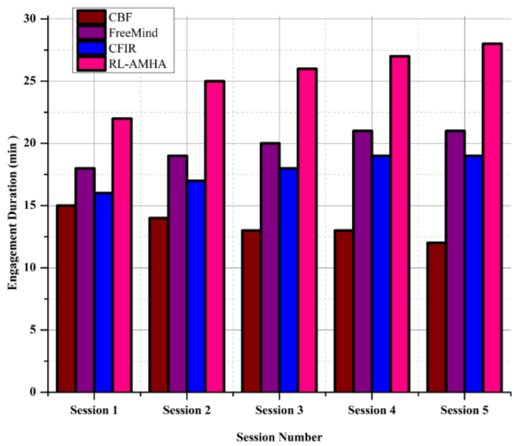

Figure 9: Engagement duration (minutes)

4.6 Self-care activity frequency (count)

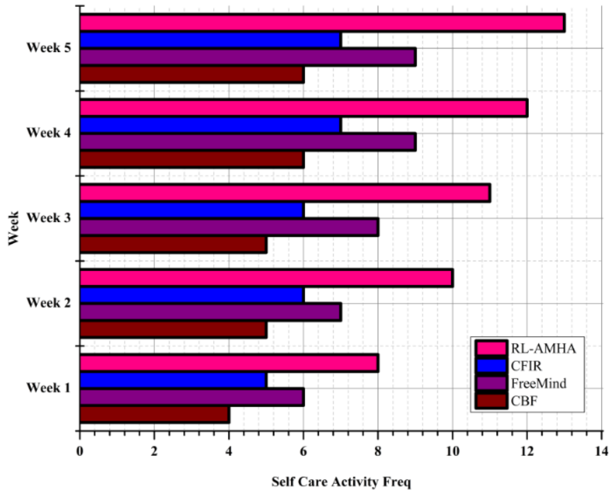

Figure 10: Self-care activity frequency (count)

Figure 10 illustrates the frequency of self-care activities over 5 weeks for RL-AMHA, CFIR, FreeMind, and CBF. RL-AMHA is best, increasing from ~10 activities in week 1 to ~13 in week 5. CFIR shows steady improvement throughout the five weeks, peaking at below ~11 activities. FreeMind showed a modest increase, around ~9 by week 5, while CBF showed tiny movement, from ~4 to ~7 activities, evaluated using equation 17. Overall, this evidence highlights RL-AMHA's ability to encourage ongoing engagement and increase self-care engagement compared to a more traditional baseline and alternative self-care activity systems.

4.7 Barriers to implementation (% frequency)

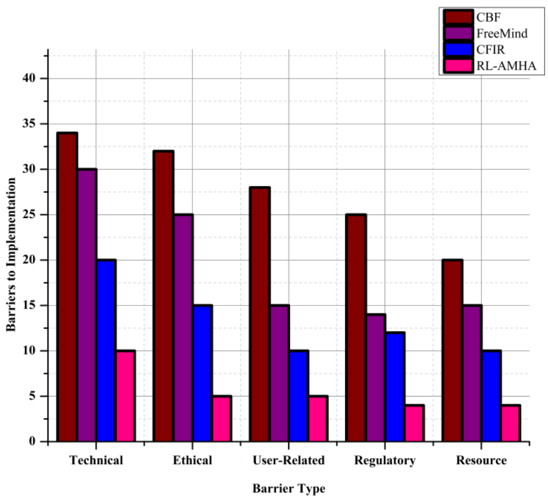

Figure 11: Barriers to implementation (% frequency)

Figure 11 shows a visual comparison of barriers to implementation for CBF, FreeMind, CFIR, and RL-AMHA across five categories: Technical, Ethical, User-Related, Regulatory, and Resource. CBF appears to have the most barriers consistently across categories, especially Technical (~34) and Ethical (~32) evaluated using equation 18. FreeMind and CFIR appear to face the second-highest challenges, and encompass both technical and ethical barriers. RL-AMHA across categories exhibited the lowest levels of implementation obstacles, with apparent reductions in both technical (~10) and moral (~5) barriers. This ultimately demonstrates that RL-AMHA offers the best adaptability, the simplest implementation, and fewer barriers than the other methods.

4.8 Resource Allocation (%)

Table 3: Resource allocation (%)

Resource Type	CBF	FreeMind	CFIR	RL-AMHA
Time	30	35	40	50
Personnel	25	30	35	45
Funding	20	25	30	40
Technology	15	20	25	35
Training	10	15	20	30

Table 3 illustrates resource usage for CBF, FreeMind, CFIR, and RL-AMHA across five resource types: Time, People, Money, Technology, and Training. RL-AMHA consistently requires the most resources across all models, with time (50) and people (45) substantially higher than everything else, which is fitting given that RL-AMHA is the most advanced in terms of computational complexity and operational efficiency, as evaluated using equation 19. CFIR is second, and then appears FreeMind, and CBF uses the least resources of the four options. In sum, the gradient seems to signal a trade-off between increasing performance and total resource use (time, people, money, technology, & training) for the more advanced, AI-based approaches (RL-AMHA).

4.9 Cultural acceptance (%)

Table 4: Cultural Acceptance (%)

Demographic Group	CBF	FreeMind	CFIR	RL-AMHA
Group A	65	70	75	85
Group B	60	68	72	83
Group C	58	65	70	80
Group D	55	63	68	78
Group E	50	60	65	75

Table 4 data set shows the effectiveness or penetration rates of CBF, FreeMind, CFIR, and RL-AMHA across five groups (A-E); RL-AMHA consistently had the highest values, with a peak of 85 in Group A and remained the highest across all demographics. CFIR ranks second, followed by FreeMind and CBF. There is a gap between RL-AMHA and traditional methods. The gap is larger for higher-performing groups evaluated using equation 20. This suggests RL-AMHA is more easily adaptable and has a broader appeal, with implications that it offers greater personalization and context-specific adaptation, leading to higher acceptance across the populace.

4.10 Emotional state improvement score and behavioral engagement index

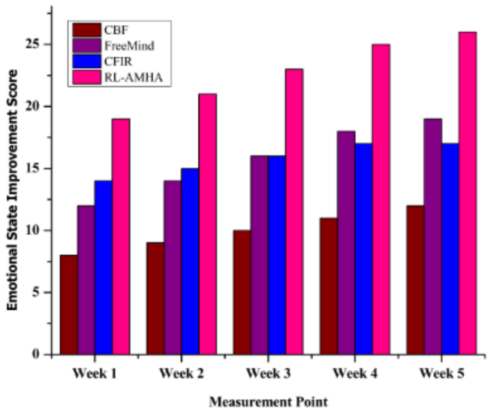

Figure 12(a): Emotional State Improvement Score

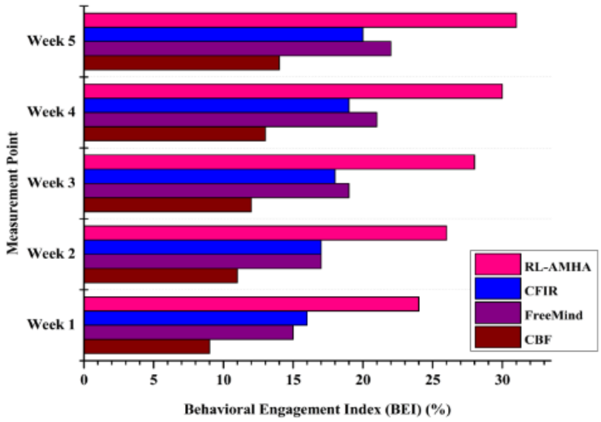

Figure 12(b): Behavioural Engagement Index

Figure 12(a) illustrates the Emotional State Improvement Score (ESIS), which tracks the percentage change in emotional well-being from pre- to post-intervention, based on surveys, sentiment analysis, or mood tracking via wearables. Figure 12(b) illustrates the Behavioral Engagement Index (BEI), which measures positive mental health behaviors comprising mindfulness practice, support for activity engagement, and adherence to coping strategies. Both ESIS and BEI provide a synthesized assessment of the emotional and behavioral consequences of the personalized recommendations, including the extent to which an intervention improves mood and increases long-term, healthy attitudes towards mental health maintenance. The user satisfaction across the five types of interventions of therapy module, self-care, peer support, adaptive content, and mobile intervention, for four systems. It includes the CBF, FreeMind, CFIR, and RL-AMHA. RL-AMHA is leading the other systems evaluated using Equation 21. It scores the highest overall at ~4.85 in adaptive content on FreeMind is rated moderately ~4.0 - 4.4 as well as CFIR - at peak, Self-Care ~4.2 too, but is all in fluctuation, and CBF is rated lowest all along (~3.55 - 4.0) and evaluated using equation 22. These findings demonstrate RL-AMHA's clear superiority in adaptability and personalization across a range of mental health interventions, especially in the field's adaptive and mobile-based modules. The findings of this evaluation confirm that RL-AMHA outperformed CBF, FreeMind, and CFIR on all evaluation metrics. It was less resource-intensive, could be adapted to larger and more diverse demographic communities, and provided better recommendations. All of which highlight the success of RL-AMHA in personalizing mental health education, a scalable, data-driven tool for improving psychosocial outcomes. Besides charting, the Behavioral Engagement Index (BEI) predicts retention and is a reward word. The hazard-penalty function devalues short, inconsistent sessions and rewards frequent, stable engagement, resulting in a single scalar summary of use intensity and consistency. RL-AMHA's higher BEI scores result in higher 5-week retention probabilities than CBF, FreeMind, and CFIR, indicating that optimizing BEI improves user retention outcomes. The reinforcement learning system has three key hyperparameters: exploration rate (ε), discount factor (γ), and learning rate (α). It took three steps to tune these. First, a coarse grid search spanning large ranges (ε: 0.05–0.40, γ: 0.85–0.99, α: 1e-5–1e-2) found stable regions. Second, Bayesian optimization with offline simulations under safety constraints resulted in $\varepsilon \approx 0.18$, $\gamma \approx 0.95$, and $\alpha \approx 2\times10^{-4}$. An online safety-adaptive adjustment was implemented, with ε decreasing over time but never below 0.05, α adjusting based on reward trends, and γ remaining constant to balance short- and long-term wellbeing. Replay testing, clinician assessment, and controlled A/B trials confirmed steady learning and no harmful recommendations. Ablation experiments trained four PPO variants: full reward, engagement-only, well-being-only, and no-penalty. In the whole model, engagement averaged 88%, session duration 29.0 minutes, stress reduction gain 18.6%, and self-care frequency 13.1 actions/week. Although engagement increased session time to 31 minutes (+6-8%), it decreased stress improvement to ≈14% and self-care to ≈11/week. Well-being improved stress by 19% but decreased involvement to 83% and self-care to 10/week. Removing penalties enhanced early engagement but increased mid-study dropout. Compared to all ablated alternatives, the full reward significantly improved engagement and well-being ($p < 0.05$).

5 Discussion

To represent mental health support, RL-AMHA uses an LSTM-attention encoder to store mood, behavior, and intervention history sequentially, rather than treating each encounter separately, as in CBF or hybrid recommenders. Improved state representation enables PPO to modify policies to user context, leading to increased satisfaction with therapy, self-care, peer support, adaptive content, and mobile interventions. RL-AMHA optimizes a multi-objective reward that combines engagement, stress reduction, and self-care adherence, leading to longer session durations and increased activity frequency over weeks, unlike CFIR and implementation-only frameworks. Static recommenders and rule-based mobile interventions cannot recover from user disengagement or stress changes, resulting in plateaued or declining satisfaction and engagement. This is seen in CBF's shorter sessions and lower weekly involvement and FreeMind/CFIR's slower activity frequency and emotional improvement. RL-AMHA adjusts its policy based on interaction rewards, correcting prior decisions and emulating successful patterns, resulting in compounding benefits over weeks. A closed-loop RL design makes RL-AMHA ideal for real-time feedback from mobile and sensor streams, allowing policy revisions without rebuilding the intervention program. PPO policy fine-tuning allows practitioners to prioritize objectives, such as stress relief for clinical users or preventive engagement, while maintaining LSTM focus. Given its flexibility and low implementation barriers, RL-AMHA can be used as an adaptive personalization layer on top of existing MIBIs, self-care apps, or peer-support platforms, rather than a full system replacement. RL-AMHA is a data-driven variant of adaptive and robust mental health control. Similar to adaptive fuzzy, neural adaptive, output-feedback, backstepping, and nonlinear optimum control, it addresses uncertainty and optimizes long-term results. RL-AMHA learns its policy from interaction data using LSTM-based state representation and multi-objective reward, eliminating the need for explicit models or rules. This makes it more scalable to high-dimensional, partially witnessed behavioral data and easier to adjust to changing objectives. RL-AMHA maintains optimal control and adaptive adjustment for long-term goals without making significant assumptions about human psychology. RL-AMHA protects privacy and ethics by design. All interaction data are collected only with informed consent, stored in encrypted form, and de-identified before model training; access is restricted and only aggregate metrics are used in monitoring. The system is positioned as a low-risk support tool, not a diagnostic or emergency service, and high-risk patterns trigger escalation to human professionals rather than autonomous action. The reward function penalizes strategies that increase stress or exploit engagement in unhealthy ways, discouraging “addictive” behaviors. Users retain control through opt-out and deletion options, and recommendations are explained in plain language to support transparency and trust.

6 Conclusion and future work

RL-AMHA, which provides highly individualized interventions and mental health education. Comparisons between CBF, FreeMind, and CFIR demonstrated that RL-AMHA can reduce both cost and time, provide individualized recommendations for diverse demographics, and offer accurate insights. This model can leverage reinforcement learning to continuously optimize its content delivery based on end-user feedback, thereby promoting ongoing engagement and improving psychological outcomes.

In the future, this framework can be adapted to include multiple categories of multimodal data sources, such as speech, facial expressions, and physiological signals, thereby deepening understanding of emotion and behavior. Integration of wearable devices and mobile health platforms would enable real-time intervention. Scaling the system for multilingual and cross-cultural contexts would likewise reach a global audience. Building explainable AI into the intervention approach will increase trust and transparency, making RL-AMHA a holistic, ethical, and scalable solution for mental health education and support.

Funding

Project of Henan Province's Philosophy and Social Sciences for Building an Education-Strong Province: The Influence of Childhood Emotional Trauma on Adolescents' Emotional Stability and Intervention - Based on the Theory of Cognitive Reappraisal (No. 2025JYQS0318)

Competing interests

The author declared no potential conflicts of interest with respect to the research, authorship or publication of this article.

Author contributions

Wang. Project administration, Conceptualization, Research concept and design&Data analysis and visualization&Review and revision

Data availability statement

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

References

[1] Saleem, M., Kühne, L., De Santis, K. K., Christianson, L., Brand, T., & Busse, H. (2021). Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR mental health, 8(12), e30000. DOI: 10.2196/30000

[2] Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. PloS one, 16(3), e0248152. DOI: 10.1371/journal.pone.0248152

[3] Ahad, A. A., Sanchez-Gonzalez, M., & Junquera, P. (2023). Understanding and addressing mental health stigma across cultures for improving psychiatric care: A narrative review. Cureus, 15(5). doi: 10.7759/cureus.39549

[4] Buselli, R., Corsi, M., Veltri, A., Baldanzi, S., Chiumiento, M., Del Lupo, E., ... & Cristaudo, A. (2021). Mental health of Health Care Workers (HCWs): a review of organizational interventions put in place by local institutions to cope with new psychosocial c

hallenges resulting from COVID-19. Psychiatry research, 299, 113847. DOI: 10.1016/j.psychres.2021.113847

[5] Thornicroft, G., Sunkel, C., Aliev, A. A., Baker, S., Brohan, E., El Chammay, R., ... & Winkler, P. (2022). The Lancet Commission on ending stigma and discrimination in mental health. The Lancet, 400(10361), 1438-1480. doi: 10.1016/S0140-6736(22)01470-2.

[6] Fleming, W. J. (2024). Employee well-being outcomes from individual-level mental health interventions: Cross-sectional evidence from the United Kingdom. Industrial Relations Journal, 55(2), 162-182. https://doi.org/10.1111/irj.12418

[7] Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., ... & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. Expert review of medical devices, 18(sup1), 37-49. DOI: 10.1080/17434440.2021.2013200

[8] Fusar-Poli, P., Correll, C. U., Arango, C., Berk, M., Patel, V., & Ioannidis, J. P. (2021). Preventive psychiatry: a blueprint for improving the mental health of young people. World Psychiatry, 20(2), 200-221. DOI: 10.1002/wps.20869

[9] Hua, Y., Liu, F., Yang, K., Li, Z., Na, H., Sheu, Y. H., ... & Torous, J. (2025). Large language models in mental health care: a scoping review. Current Treatment Options in Psychiatry, 12(1), 1-18. https://doi.org/10.1007/s40501-025-00363-y

[10] Borghouts, J., Eikey, E., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., ... & Sorkin, D. H. (2021). Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. Journal of medical Internet research, 23(3), e24387. DOI: 10.2196/24387

[11] Merino, M., Tornero-Aguilera, J. F., Rubio-Zarapuz, A., Villanueva-Tobaldo, C. V., Martín-Rodríguez, A., & Clemente-Suárez, V. J. (2024, July). Body perceptions and psychological well -being: a review of the impact of social media and physical measurements on self-esteem and mental health with a focus on body image satisfaction and its relationship with cultural and gender factors. In Healthcare (Vol. 12, No. 14, p. 1396). MDPI. https://doi.org/10.3390/healthcare12141396

[12] Lattie, E. G., Stiles-Shields, C., & Graham, A. K. (2022). An overview of and recommendations for more accessible digital mental health services. Nature Reviews Psychology, 1(2), 87-100. https://doi.org/10.1038/s44159-021-00003-1

[13] D. Zaçe, I. Hoxhaj, A. Orfino, A. M. Viteritti, L. Janiri, and M. L. Di Pietro, "Interventions to address mental health issues in healthcare workers during infectious disease outbreaks: A systematic review," Journal of Psychiatric Research, vol. 136, pp. 319–333, Apr. 2021, doi: https://doi.org/10.1016/j.jpsychires.2021.02.019.

[14] D. Z. Q. Gan, L. McGillivray, J. Han, H. Christensen, and M. Torok, "Effect of Engagement With Digital Interventions on Mental Health Outcomes: A Systematic Review and Meta-Analysis," Frontiers in Digital Health, vol. 3, Nov. 2021, doi: https://doi.org/10.3389/fdgth.2021.764079.

[15] T. J. Philippe et al., "Digital Health Interventions for Delivery of Mental Health Care: Systematic and Comprehensive Meta-Review (Preprint)," JMIR Mental Health, vol. 9, no. 5, 2021, doi: https://doi.org/10.2196/35159.

[16] Mazlan, I., Abdullah, N., & Ahmad, N. (2023). Exploring the impact of hybrid recommender systems on personalized mental health recommendations. International Journal of Advanced Computer Science and Applications, 14(6), 935-944. DOI: 10.14569/IJACSA.2023.0140699

[17] [Mazlan, I. B., Abdullah, N., Ahmad, N., & Harun, S. Z. (2024). A Review of Personalized Recommender System for Mental Health Interventions. International Journal of Advanced Computer Science & Applications, 15(10). DOI: 10.14569/IJACSA.2024.0151041

[18] Kazi, F. F., & Sandbulte, J. (2023). Examining a social-based system with personalized recommendations to promote mental health for college students. Smart health, 28, 100385. DOI： 10.1016/j.smhl.2023.100385

[19] Singh, V., Kumar, A., & Gupta, S. (2022). Mental health prevention and promotion—A narrative review. Frontiers in psychiatry, 13, 898009. https://doi.org/10.3389/fpsyt.2022.898009

[20] Worsley, J. D., Pennington, A., & Corcoran, R. (2022). Supporting mental health and wellbeing of university and college students: A systematic review of review-level evidence of interventions. PLoS one, 17(7), e0266725. DOI: 10.1371/journal.pone.0266725

[21] Hornstein, S., Zantvoort, K., Lueken, U., Funk, B., & Hilbert, K. (2023). Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. Frontiers in digital health, 5, 1170002. DOI: 10.3389/fdgth.2023.1170002

[22] Kirkbride, J. B., Anglin, D. M., Colman, I., Dykxhoorn, J., Jones, P. B., Patalay, P., ... & Griffiths, S. L. (2024). The social determinants of mental health and disorder: evidence, prevention and recommendations. World psychiatry, 23(1), 58-90. DOI: 10.1002/wps.21160

[23] Mutschler, C., Bellamy, C., Davidson, L., Lichtenstein, S., & Kidd, S. (2022). Implementation of peer support in mental health services: A systematic review of the literature. Psychological Services, 19(2), 360. DOI: 10.1037/ser0000531

[24] https://www.kaggle.com/datasets/ziya07/student-mental-health-and-resilience-dataset

[25] Kolenik, T., & Schiepek, G. (2024). Computational psychotherapy system for mental health prediction and behavior change with a conversational agent. Neuropsychiatric Disease and Treatment, 20, 2465–2498. https://doi.org/10.2147/NDT.S417695

[26] 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), 259–272. https://doi.org/10.31449/inf.v49i2.8738

[27] Kolenik, T. (2021). Methods in digital mental health: Smartphone-based assessment and intervention for stress, anxiety and depression. In C. Comito, A. Forestiero, & E. Zumpano (Eds.), Integrating artificial intelligence and IoT for advanced health informatics (pp. xx–xx). Springer. https://doi.org/10.1007/978-3-030-91181-2_7

[28] 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

[29] 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

[30] Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control. Advance online publication. https://doi.org/10.1177/10775463251320258

[31] Boukabou, A. (2017). Output-feedback controller-based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. Mathematical Problems in Engineering, 2017, Article 8045803. https://doi.org/10.1155/2017/8045803.

[32] Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. Energy Reports, 9, 1470–1486. https://doi.org/10.1016/j.egyr.2023.03.020