MINDLOOP: LLM-BASED CLOSED-LOOP EMOTION REGULATION WITH WEARABLE BCI FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

Emotion regulation is essential for mental health, but access to trained professional support remains limited. Large Language Models (LLMs) offer a promising way to make emotional support more accessible through conversational interaction. However, existing studies on LLM-based emotional support mainly focus on current turn response quality, while paying limited attention to the overall regulation process. Effective emotion regulation is a structured closed-loop process requiring multi-stage guidance and timely affective feedback, raising three challenges: structuring dialogue, balancing current response quality with overall regulation outcomes, and obtaining timely affective observations. We propose MindLoop, a closed-loop policy learning approach for LLM-based emotion regulation with wearable BCI feedback. MindLoop first introduces a CBT-guided stage–strategy plan to organize multi-turn dialogue into structured regulation stages and stage-specific interventions. Based on this structure, we further develop a dual-timescale reinforcement learning method to balance overall regulation progress and current-stage intervention quality. To provide objective and timely observations of affective changes, we introduce wearable EEG/fNIRS-based affect decoding as physiological feedback for adapting regulation decisions. Simulated-interaction evaluation shows strong regulation-stage progression, successful completion, and affective outcomes. A 23-participant study further provides preliminary evidence that the simulation-trained policy supports structured real-world interaction with favorable affective changes and session ratings.

1 INTRODUCTION

Emotion plays a fundamental role in cognition, decision-making, and social interaction (Gross, 2002), while persistent negative emotions can disrupt daily life and increase the risk of psychological distress (Curtiss et al., 2021). Emotion regulation helps individuals understand and adjust their emotional states, making it essential for mental health and well-being (Gross, 2015; Slovak et al., 2023). In practice, emotion regulation can be supported through medical treatments, psychological interventions, and social support (Hofmann et al., 2012). Among these approaches, psychological interventions are widely used to support emotion regulation and reduce negative emotions (Gratz et al., 2015). These interventions provide structured guidance that helps individuals understand and manage emotional difficulties. However, such interventions require trained professionals, and access remains limited. This gap makes it difficult to meet the need for emotion regulation and highlights the need for automated, scalable support (Schueller & Torous, 2020; Philippe et al., 2022).

Large Language Models (LLMs) offer a promising way to make emotional support more accessible through conversational interaction (Hua et al., 2025). Existing studies improve LLM-based support through empathetic response generation, context modeling, counseling-oriented fine-tuning, domain-specific corpora, and multi-agent simulation (Rashkin et al., 2019; Tu et al., 2022; Chen et al., 2023; Xu et al., 2025b). However, they mainly optimize current response quality while paying limited attention to emotional changes across the regulation process. Effective emotion regulation should guide users through a closed-loop process that assesses emotional state, provides targeted support, and adapts subsequent strategies using timely affective feedback (Gross, 1998; Cheng et al., 2022).

Building such a closed-loop process introduces three fundamental challenges. First, **existing LLM-based emotion regulation lacks a structured dialogue process for guiding emotional change.**

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Effective regulation often requires users to express emotions, clarify situations, and reflect on their experiences, rather than receive repeated empathy and comfort (Liu et al., 2021; Tu et al., 2022; Zhao et al., 2023). Psychological interventions commonly organize support into multiple stages (Fordham et al., 2021), and Cognitive Behavioral Therapy (CBT) is one of the most influential and extensively studied frameworks (David et al., 2018). CBT guides users from identifying emotion-related situations and thoughts to reappraising them and considering constructive actions. Accordingly, **we introduce a CBT-guided stage-strategy plan** that structures the regulation process and assigns suitable interventions to each stage.

Second, **LLM-based emotion regulation must balance current response quality with overall regulation outcomes**. Emotion regulation cannot be evaluated only by the immediate emotional effect of a response (Lawrence et al., 2024; Guo et al., 2024). Comforting responses may make the user feel better in the current turn, but they may not support the overall regulation outcome. In contrast, challenging negative thoughts may initially make the user uncomfortable but can support later progress. Interventions thus must be evaluated at different temporal scales, considering both current quality and overall contribution (Cheng et al., 2022). Following this idea, **we develop a dual-timescale reinforcement learning method** that separately optimizes current intervention quality and overall regulation effectiveness.

Third, **closed-loop adaptation requires objective and timely affective feedback.** Existing systems mainly infer emotion from text or subjective self-reports, which can be sparse, delayed, and influenced by users' verbal expression. They may therefore miss affective changes not explicitly conveyed in language. Wearable BCIs provide a complementary physiological channel for observing these changes and adapting subsequent support (Zhao et al., 2025b; 2026). To address this issue, **we use wearable EEG/fNIRS signals to provide physiological feedback**, enabling objective and timely observation of affective changes during real-world interaction.

Based on this discussion, we propose **MindLoop**, a closed-loop policy learning approach for LLM-based emotion regulation with wearable BCI feedback. It combines CBT-guided dialogue stages, dual-timescale policy adaptation, and wearable EEG/fNIRS feedback to structure interventions, balance current response quality with overall regulation outcomes, and observe affective changes. It thereby reformulates LLM-based emotional support as a closed-loop regulation process that guides the overall dialogue rather than only optimizing the current response. The main contributions of this paper are summarized as follows:

- **We propose MindLoop, a closed-loop policy learning approach for LLM-based emotion regulation**, reformulating LLM-based emotional support as a closed-loop regulation process.
- **We introduce a CBT-guided stage-strategy plan** that structures the regulation process, tracks dialogue progress, and selects stage-appropriate interventions.
- **We develop a dual-timescale reinforcement learning method** that evaluates stage-level decisions by regulation progress and strategy-level decisions by current-stage intervention quality.
- **We introduce wearable EEG/fNIRS feedback** for physiological affect sensing in real-world interaction. MindLoop improves process control and affective outcomes in simulation, while a 23-participant study provides preliminary evidence of feasibility and favorable session outcomes.

2 METHODOLOGY

Figure 1 presents MindLoop, a closed-loop framework for LLM-based emotion regulation. MindLoop combines a CBT-guided stage-strategy structure, dual-timescale policy learning, and paired pre-/post-affect modeling with wearable EEG/fNIRS feedback. The policy is optimized in simulation using LLM-judge supervision and affective feedback, then fixed for real-world evaluation, where physiological signals update the dialogue state without online learning.

2.1 CLOSED-LOOP EMOTION REGULATION FORMULATION

We formulate LLM-based emotion regulation as a Markov Decision Process (MDP) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, r, \gamma)$. At turn t, the state $s_t \in \mathcal{S}$ contains the dialogue history, current user utterance, regulation stage, previous intervention decisions, and available affective observations. The action $p_t \in \mathcal{A}$ is a structured stage-strategy intervention plan. After an executor LLM realizes p_t as re-

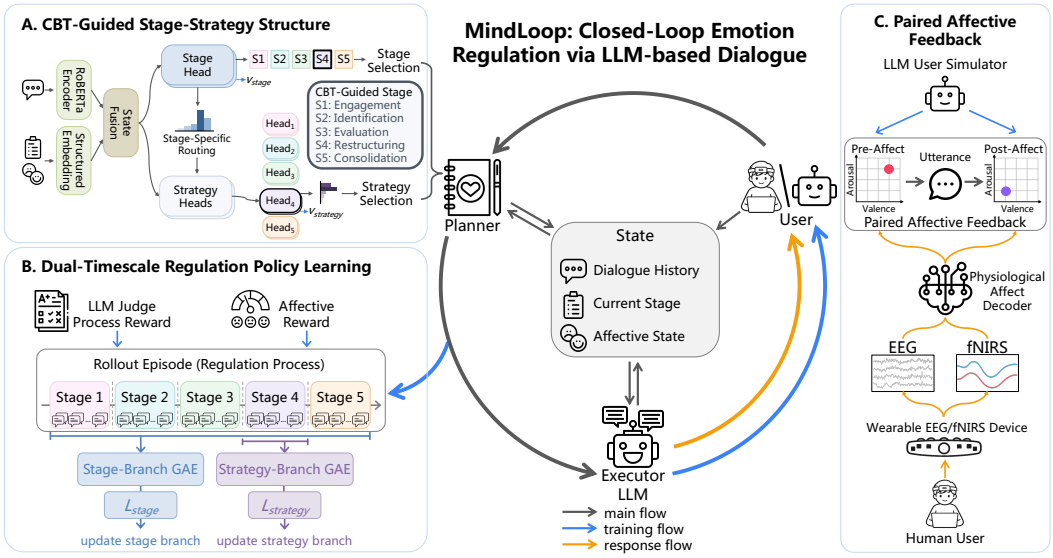

Figure 1: Overview of MindLoop, a closed-loop framework for LLM-based emotion regulation.

sponse y_t, the user environment generates the next utterance and affective observation, yielding $s_{t+1} \sim \mathcal{T}(\cdot \mid s_t, p_t)$. The policy is optimized as

$$\pi_\theta^* = \arg\max_{\pi_\theta} \mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T} \gamma^{t-1} r_t\right], \tag{1}$$

with interaction horizon T and discount factor γ.

Interaction Environment. The closed-loop interaction follows a policy–executor–simulator architecture. Given s_t, the policy first selects a structured stage–strategy plan p_t, separating lightweight intervention planning from the subsequent autoregressive response realization. The executor LLM then generates the response conditioned on the dialogue state and selected plan:

$$y_t \sim q_\phi^{\text{exec}}(\cdot \mid s_t, p_t). \tag{2}$$

The environment then updates the user state via a simulator that generates the next utterance and affective observation:

$$(x_{t+1}, e_{t+1}) \sim q_\phi^{\text{sim}}(\cdot \mid s_t, y_t), \tag{3}$$

A stage-level LLM judge evaluates segment completion and intervention coherence, and its evaluations are combined with affective feedback to construct the learning rewards.

2.2 CBT-Guided Stage-Strategy Intervention Structure

Emotion regulation requires both locally appropriate interventions and coherent process progression. We therefore organize the dialogue into five CBT-guided stages—*Engagement*, *Identification*, *Evaluation*, *Restructuring*, and *Consolidation*—with stage-specific strategies covering supportive engagement, situation/thought identification, evidence-based evaluation, cognitive restructuring, and action-oriented consolidation. The stage definitions and strategy inventory were developed through iterative discussions with domain experts to align the intervention structure with established CBT principles. The complete stage-specific strategy inventory is provided in the Appendix B.1.

Stage-Strategy Intervention Space. Rather than generating a response directly, the policy selects a structured plan consisting of a regulation stage and a stage-specific strategy. Let $\mathcal{G} = \{G_1, \ldots, G_K\}$ denote the set of stages and $\mathcal{U}(G_k)$ the admissible strategies for stage G_k. At turn t, the plan is

$$p_t = (g_t, u_t), \tag{4}$$

where $g_t \in \mathcal{G}$ and $u_t \in \mathcal{U}(g_t)$. We factorize the policy as

$$\pi_\theta(p_t \mid s_t) = \pi_\theta^{\mathrm{stg}}(g_t \mid s_t)\, \pi_\theta^{\mathrm{str}}(u_t \mid s_t, g_t), \tag{5}$$

where the stage policy controls process progression and the strategy policy selects a stage-consistent intervention. The resulting plan is then passed to the executor LLM for response generation.

Process-aware policy parameterization. The planner conditions stage-level and strategy-level selection on both dialogue semantics and structured regulation context. A frozen RoBERTa encoder (Liu et al., 2019) encodes the dialogue history and current utterance as h_t^{text}, while the current stage, previous strategy, and affective observations are embedded as h_t^{struct}. The state representation is

$$h_t = [\, h_t^{\mathrm{text}}; h_t^{\mathrm{struct}} \,]. \tag{6}$$

Based on h_t, the stage policy is produced by a global stage head with a legal transition mask:

$$\pi_\theta^{\mathrm{stg}}(g_t \mid s_t) = \mathrm{Softmax}\big(W^{\mathrm{stg}} h_t + \log m_t^{\mathrm{stg}}\big)\,, \tag{7}$$

where $m_t^{\mathrm{stg}} \in \{0, 1\}$ permits staying in the current stage or advancing to the next while assigning zero probability to invalid transitions. Given the selected stage g_t, strategy selection is routed to the corresponding stage-specific head:

$$\pi_\theta^{\mathrm{str}}(u_t \mid s_t, g_t) = \mathrm{Softmax}\big(W_{g_t}^{\mathrm{str}} h_t\big)\,, \quad u_t \in \mathcal{U}(g_t). \tag{8}$$

Here, W^{stg} is the global stage head and $W_{g_t}^{\mathrm{str}}$ is the strategy head for stage g_t. This routing restricts selection to the legal strategy space of the selected stage.

2.3 Dual-Timescale Regulation Policy Learning

The stage and strategy policies require different credit-assignment horizons. Stage decisions govern dialogue-level progress, whereas strategies mainly affect intervention quality within the current stage; a shared return may therefore misattribute later outcomes to earlier strategy choices. We therefore use a dual-timescale actor–critic objective with separate critics, rewards, and credit-assignment scopes: the stage branch is evaluated across the regulation, while the strategy branch is evaluated within the current stage.

Stage and Strategy Critics. We use a stage-level critic for overall regulation progress and stage-conditioned strategy critics for current-stage intervention quality:

$$v_t^{\mathrm{stg}} = V_\phi^{\mathrm{stg}}(h_t), \qquad v_t^{\mathrm{str}} = V_{\psi, g_t}^{\mathrm{str}}(h_t). \tag{9}$$

The stage critic estimates overall regulation progress, while the strategy critic evaluates intervention quality within stage g_t. Conditioning on g_t aligns the strategy critic with the stage-specific actor heads, while the shared representation h_t coordinates the two levels.

Stage-Aware Reward Decomposition. We distinguish stage-progress rewards from current-stage intervention rewards. At stage boundaries, a stage judge evaluates three completion criteria, whose completion count is mapped to a stage-task reward. A separate strategy judge scores each intervention in the segment, and the mean score is mapped to a strategy-quality reward. The two policy branches use

$$\tilde{r}_t^{\mathrm{stg}} = r_t^{\mathrm{stg}} + r_t^{\mathrm{str}}, \qquad \tilde{r}_t^{\mathrm{str}} = r_t^{\mathrm{str}}, \tag{10}$$

where r_t^{stg} and r_t^{str} denote the stage-progress and current-stage intervention reward streams, respectively. This decomposition allows the stage branch to account for both process progression and intervention quality, while the strategy branch focuses on interventions within the current stage. Affective feedback is computed separately from segment-level post-affect observations. Component definitions and score-to-reward mappings are provided in Appendix B.3.

Dual-Horizon Advantage Estimation. We implement the two temporal scopes using level-specific bootstrap indicators in generalized advantage estimation. For each level $b \in \{\mathrm{stg}, \mathrm{str}\}$, we compute

$$\delta_t^b = \tilde{r}_t^b + \gamma \xi_t^b v_{t+1}^b - v_t^b, \qquad A_t^b = \delta_t^b + \gamma\lambda\xi_t^b A_{t+1}^b. \tag{11}$$

Here, ξ_t^b controls whether the return is bootstrapped beyond turn t. For the stage-level policy,

$$\xi_t^{\mathrm{stg}} = 1 - d_t, \tag{12}$$

where $d_t \in \{0, 1\}$ denotes episode termination. Stage-level credit therefore propagates across the full dialogue.

For the strategy-level policy, bootstrapping is restricted to turns within the same stage:

$$\xi_t^{\mathrm{str}} = \mathbf{1}[(1 - d_t) \wedge (g_{t+1} = g_t)]\,. \tag{13}$$

Strategy-level credit propagation stops at stage transitions, preventing cross-stage credit assignment while retaining full-dialogue credit for the stage branch.

Regulation policy optimization objective. The two branches are trained under a clipped actor–critic objective:

$$\mathcal{L} = \sum_{b \in \{\mathrm{stg}, \mathrm{str}\}} \Big(w_b \mathcal{L}_{\mathrm{actor}}^b + c_b \mathcal{L}_{\mathrm{value}}^b - \eta_b \mathcal{H}^b \Big). \tag{14}$$

Here, $\mathcal{L}_{\mathrm{actor}}^b$, $\mathcal{L}_{\mathrm{value}}^b$, and $\mathcal{H}^b$ are the negative clipped surrogate loss, critic regression loss, and entropy bonus. The weights w_b, c_b, and η_b balance policy optimization, value estimation, and exploration.

2.4 Paired Affective Feedback

Pre–Post Affective State Interface. To enable fine-grained affective tracking, user affect is modeled through paired observations before and after each user input, rather than a single static utterance-level label. At each turn, we distinguish the affective state before the user produces the current utterance from the affective state observed after the utterance:

$$e_t = \left(e_t^{\mathrm{pre}}, e_t^{\mathrm{post}}\right), \tag{15}$$

where both states are quantized in a valence–arousal space. The pre-affect represents the incoming emotional state for planning, while the post-affect provides feedback for subsequent regulation decisions.

Stage-Level Affective Reward. Because single-turn affective changes can be noisy, we aggregate post-affect within each completed regulation-stage segment. For a segment $\tau_k = \{t_1, \ldots, t_2\}$ associated with stage G_k, we compute

$$\bar{e}_k^{\mathrm{post}} = \frac{1}{|\tau_k|} \sum_{t \in \tau_k} e_t^{\mathrm{post}}. \tag{16}$$

The stage-level affective reward is then defined as

$$r_k^{\mathrm{affect}} = f_{\mathrm{affect}}\left(\bar{e}_k^{\mathrm{post}}, G_k\right), \tag{17}$$

and is assigned to turns in the corresponding segment. This links affective feedback to stage-level outcomes while reducing sensitivity to transient turn-level fluctuations.

Wearable EEG/fNIRS-based Affect Sensing. In simulation, paired affective feedback is provided by the LLM-based user environment. In real-world interaction, a wearable EEG/fNIRS system continuously records complementary physiological signals through EEG electrodes, fNIRS optodes, and their acquisition modules.

Let $x_{t,q}^{\mathrm{phys}}$ denote the physiological observation around turn t at phase $q \in \{\mathrm{pre}, \mathrm{post}\}$. The affect decoder f_ω is built upon EMOD (Chen et al., 2026) and extended with an fNIRS encoding branch to jointly model EEG and fNIRS signals:

$$\hat{e}_t^q = f_\omega\left(x_{t,q}^{\mathrm{phys}}\right), \quad q \in \{\mathrm{pre}, \mathrm{post}\}. \tag{18}$$

The decoder maps physiological observations into the valence–arousal space used in simulation. During real-world interaction, decoded affective states update the subsequent dialogue state and are provided to the executor LLM with the selected regulation plan.

3 EXPERIMENTAL DESIGN

We evaluate MindLoop in two settings: **simulated-interaction evaluation** and **real-world multimodal validation**. The simulated setting examines whether MindLoop can control a multi-turn regulation process and improve affective outcomes in a controlled environment. The real-world validation examines whether the learned regulation policy can operate in human interaction with wearable EEG/fNIRS-based affect decoding.

3.1 SIMULATED-INTERACTION EVALUATION

Protocol. All methods are evaluated in a simulated interaction environment with a maximum of 25 turns. Each rollout is initialized from a test-dialogue scenario. At each turn, SP generates the stage, strategy, and response jointly, whereas executor-based methods select a stage–strategy plan that a fixed executor LLM realizes as a response. A fixed user simulator then generates the next utterance and affective observation. Unless otherwise specified, the executor and simulator use Qwen3-30B. A rollout succeeds if it reaches the final stage and terminates legally within the turn budget. All methods share the same environment, stage-transition constraints, strategy inventory, simulator, horizon, and success criterion; executor-based methods also share the same executor. Implementation and training details are provided in Appendix B.

Data Sources. We use ESConv (Liu et al., 2021) and ExTES (Zheng et al., 2024). ESConv contains 1,300 dialogues with a 1,040/130/130 train/validation/test split (Deng et al., 2024), and ExTES contains 11,117 dialogues split into 10,717/200/200 (He et al., 2025). They provide user contexts and scenarios for simulated rollouts rather than static response-generation examples.

Compared Methods. We compare three groups of methods: *(1) Counseling LLMs*, PsyDT (Xie et al., 2025) and CPsyCounX (Zhang et al., 2024); *(2) General LLMs*, GLM-9B (Glm et al., 2024), Qwen3-8B (Yang et al., 2025), Qwen3-30B, Qwen3-Max, DeepSeek-V4-Flash (DeepSeek-AI, 2026), and GLM-5.1 (Zeng et al., 2026); and *(3) RL-based dialogue policies*, PPDPP (Deng et al., 2024) and DialogXpert (Rakib et al., 2026). For both groups of LLM baselines, **SP** jointly generates stage, strategy, and response; **JP** jointly predicts stage and strategy before executor generation; and **HP** separately predicts stage progression and strategy before executor generation, requiring one, two, and three LLM calls per turn, respectively. Detailed protocols are provided in Appendix B.5.

Evaluation Metrics. We evaluate both regulation-process control and affective outcomes. **Average Stages (AS)** is the average number of distinct regulation stages visited per rollout. **Success Rate (SR)** is the proportion of rollouts that reach the final stage and legally terminate within the turn budget. **Average Turns (AT)** is the average number of turns among successful rollouts. For affective evaluation, **Emotion Improvement (EI)** measures normalized valence improvement. Post-utterance valence levels low, medium, and high are mapped to 0, 50, and 100. For rollout i, B_i and A_i are the average valence scores over the first and last five valid turns, respectively:

$$\mathrm{EI}_i = 100 \cdot \frac{A_i - B_i}{100 - B_i}, \qquad \mathrm{EI} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{EI}_i, \tag{19}$$

where N is the number of evaluated rollouts. **PR** denotes Positive Rate, which measures the proportion of rollouts where EI is greater than zero. Together, these metrics evaluate whether a method can guide the dialogue through the structured regulation process while improving the user's affective state over the interaction.

3.2 REAL-WORLD MULTIMODAL VALIDATION

Real-World Study Protocol. The study includes 23 Chinese subjects (10 females and 13 males), with an average age of 24.74 ± 1.86 years. Each subject completes an emotion-imagery block followed by an emotion-regulation block, while wearable EEG/fNIRS signals are continuously recorded. The protocol was approved by the relevant institutional ethics committee, and all subjects provided written informed consent.

The emotion-imagery block follows the imagery-induced emotion paradigm of EmoEEG-MC (Xu et al., 2025a) and contains 18 trials. Each trial includes a 1-s fixation, a 30-s emotion-imagery period with audiovisual prompts, self-reported valence and arousal ratings, and a brief arithmetic task. The

Table 1: Simulated-interaction results. Values are reported as mean/std across evaluation rollouts.

Model	Protocol	ESConv					ExTES				
		AS↑	SR↑	AT↓	EI↑	PR↑	AS↑	SR↑	AT↓	EI↑	PR↑
Counseling LLMs											
	SP	1.00/0.00	0.00/0.00	–	-9.48/22.88	0.10/0.30	1.00/0.00	0.00/0.00	–	-10.02/17.45	0.14/0.34
CPsyCounX	JP	1.00/0.00	0.00/0.00	–	-3.99/35.46	0.25/0.44	1.01/0.10	0.00/0.00	–	-5.56/28.67	0.25/0.43
	HP	1.48/0.73	0.00/0.00	–	-1.33/35.29	0.33/0.47	1.35/0.55	0.00/0.00	–	-9.35/28.44	0.23/0.42
	SP	1.00/0.00	0.00/0.00	–	-9.05/25.07	0.16/0.37	1.01/0.07	0.00/0.00	–	-6.47/16.53	0.25/0.43
PsyDT	JP	1.32/0.48	0.00/0.00	–	-4.76/31.97	0.25/0.44	1.41/0.49	0.00/0.00	–	-9.63/26.76	0.16/0.37
	HP	1.00/0.00	0.00/0.00	–	3.48/36.57	0.35/0.48	1.00/0.00	0.00/0.00	–	-10.13/25.34	0.18/0.38
General LLMs											
	SP	1.00/0.00	0.00/0.00	–	-2.24/25.63	0.42/0.49	1.01/0.07	0.00/0.00	–	0.06/20.80	0.48/0.50
GLM-9B	JP	1.02/0.12	0.00/0.00	–	-10.04/29.35	0.21/0.41	1.00/0.00	0.00/0.00	–	-14.31/69.05	0.26/0.44
	HP	1.00/0.00	0.00/0.00	–	-7.70/27.54	0.15/0.36	1.00/0.00	0.00/0.00	–	-6.99/31.18	0.20/0.40
	SP	1.06/0.24	0.00/0.00	–	-11.95/25.82	0.28/0.45	1.16/0.36	0.00/0.00	–	-10.66/21.19	0.27/0.44
Qwen3-8B	JP	1.10/0.32	0.00/0.00	–	4.73/42.49	0.50/0.50	1.05/0.22	0.00/0.00	–	12.03/42.48	0.53/0.50
	HP	1.73/0.87	0.00/0.00	–	-4.46/26.68	0.22/0.41	1.78/0.87	0.00/0.00	–	0.57/37.71	0.32/0.47
	SP	2.09/0.95	0.00/0.00	–	-5.64/36.65	0.38/0.48	1.57/0.80	0.00/0.00	–	-0.13/31.99	0.43/0.50
Qwen3-30B	JP	2.26/0.94	0.00/0.00	–	-9.71/42.89	0.32/0.46	2.28/0.94	0.01/0.10	21.50/1.50	-4.24/39.28	0.36/0.48
	HP	4.36/0.89	0.55/0.50	19.99/1.35	41.01/43.33	0.68/0.46	4.70/0.66	0.73/0.44	19.32/1.25	68.87/43.18	0.81/0.39
	SP	2.72/0.51	0.00/0.00	–	-3.75/21.07	0.22/0.30	2.77/0.59	0.00/0.00	–	-2.02/26.61	0.32/0.39
Qwen3-Max	JP	3.34/0.84	0.10/0.30	21.23/2.36	5.52/32.26	0.64/0.50	3.50/0.79	0.05/0.22	22.20/2.68	10.05/39.73	0.73/0.44
	HP	4.85/0.45	0.81/0.39	17.87/3.15	57.72/38.06	0.88/0.33	4.93/0.30	0.88/0.32	17.65/3.27	75.44/32.66	0.93/0.26
	SP	4.29/0.76	0.51/0.50	19.82/2.60	45.30/41.50	0.73/0.39	4.23/0.68	0.51/0.49	20.71/2.23	49.70/35.39	0.73/0.44
DeepSeek-V4	JP	4.59/0.81	0.57/0.50	19.97/2.71	50.81/44.90	0.81/0.44	4.73/0.89	0.59/0.50	20.03/2.61	68.38/48.30	0.84/0.44
	HP	4.88/0.35	0.80/0.40	19.88/1.95	62.38/39.48	0.88/0.33	4.93/0.26	0.85/0.36	19.69/1.68	75.86/35.19	0.92/0.28
	SP	4.96/0.19	0.65/0.46	17.14/2.74	61.93/22.66	0.91/0.29	4.98/0.12	0.72/0.45	15.32/2.16	69.06/17.00	0.96/0.20
GLM-5.1	JP	5.00/0.00	0.69/0.48	14.80/3.27	65.19/30.58	0.94/0.23	5.00/0.00	0.76/0.43	17.55/2.64	77.97/19.72	0.98/0.14
	HP	5.00/0.00	0.97/0.15	13.12/3.19	70.48/21.83	0.98/0.12	5.00/0.00	0.99/0.07	**13.49/2.94**	78.20/23.89	0.99/0.07
RL methods											
PPDPP	–	2.05/0.37	0.02/0.12	**13.00/7.00**	-6.86/20.30	0.09/0.29	2.08/0.47	0.03/0.16	15.40/5.78	-9.76/17.09	0.12/0.32
DialogXpert	–	4.99/0.09	0.99/0.09	18.46/0.79	32.04/42.17	0.72/0.45	4.99/0.21	0.99/0.10	18.30/1.28	65.22/32.50	0.94/0.24
MindLoop	–	**5.00/0.00**	**1.00/0.00**	17.25/0.59	**78.35/28.01**	**0.98/0.14**	**5.00/0.00**	**1.00/0.00**	17.17/0.38	**92.22/12.72**	**1.00/0.00**

resulting physiological signals and affect labels are used to train a subject-specific EEG/fNIRS affect decoder. The decoder achieves $62.05 \pm 10.49\%$ ACC and $59.41 \pm 10.75\%$ WF1 for valence, and $53.23 \pm 9.11\%$ ACC and $46.77 \pm 9.88\%$ WF1 for arousal.

In the emotion-regulation block, subjects first complete the pre-test PANAS (Watson et al., 1988) and then engage in a multi-turn MindLoop session about a recent distressing event. At each turn, the subject-specific decoder estimates paired pre- and post-utterance affective states from wearable EEG/fNIRS signals. The utterance and decoded states update the dialogue state, after which MindLoop selects a regulation stage and strategy and the executor LLM generates a response. Subjects also report turn-level valence and arousal for evaluation, but these reports are not provided to MindLoop. After the dialogue, subjects complete the post-test PANAS, CTSR-4 (Alfonsson et al., 2022), and safety assessments. Additional details on the experimental protocol, physiological signal acquisition, decoder architecture, training, and online inference are provided in Appendix C.

Evaluation Measures. We evaluate the real-world study using four groups of measures: SR, AS, and AT for regulation-process performance; PANAS for session-level affective change; self-report-based EI and PR for turn-level affective change; and CTSR-4 and safety ratings for interaction quality and safety. Together, these measures assess whether a simulation-trained policy can support structured human interaction using subject-specific wearable EEG/fNIRS affect decoding.

4 Results and Analysis

4.1 Simulated-Interaction Results

4.1.1 Overview Performance

Table 1 reports the simulated-interaction results. **MindLoop achieves the strongest overall performance on both datasets**, with full stage coverage and completion (AS/SR of 5.00/1.00), EI/PR of 78.35/0.98 on ESConv, and 92.22/1.00 on ExTES. Although its AT is not the lowest, it maintains competitive interaction length while achieving the strongest process completion and affective outcomes. Three observations emerge. **First, counseling-oriented response generation is insufficient for closed-loop emotion regulation**. CPsyCounX and PsyDT remain largely in early stages and rarely complete the process, suggesting that supportive responses alone do not provide reliable control over multi-stage regulation. **Second, explicit stage–strategy decomposition improves regulation-stage control for general LLMs**. HP generally outperforms SP and JP by separating stage progression from strategy selection; for example, Qwen3-30B under HP reaches AS/SR of

Table 2: Ablation study in the simulated-interaction setting.

Variant	ESConv					ExTES				
	AS↑	SR↑	AT↓	EI↑	PR↑	AS↑	SR↑	AT↓	EI↑	PR↑
MindLoop	**5.00**	**1.00**	17.25	**78.35**	**0.98**	**5.00**	**1.00**	17.17	**92.22**	**1.00**
w/o stage-strategy factorization	5.00	0.58	23.36	55.69	0.86	5.00	0.54	22.95	60.70	0.97
w/o stage-specific strategy heads	5.00	1.00	**16.72**	34.17	0.91	5.00	1.00	**16.57**	57.75	0.96
w/o dual-timescale optimization	5.00	0.96	21.85	50.73	0.95	5.00	0.83	21.81	67.77	0.96
w/o stage-level affect reward	5.00	1.00	16.78	71.13	0.97	5.00	1.00	17.00	80.66	0.98
w/o paired affect modeling	5.00	0.97	19.97	69.64	0.96	5.00	0.99	20.13	81.16	0.97

4.36/0.55 on ESConv and 4.70/0.73 on ExTES, with similar trends across stronger models. However, prompt-based hierarchical planning requires additional model calls and lacks learned process-level optimization. The remaining gap to MindLoop further indicates that explicit structure alone is insufficient without policy optimization across the regulation process. **Third, effective regulation requires coordinating process control with affective improvement**. PPDPP reports a low AT on ESConv, but its SR is only 0.02, so this value reflects very few successful rollouts rather than consistently efficient regulation. DialogXpert achieves near-complete stage progression but substantially lower EI than MindLoop, showing that process completion alone does not guarantee affective improvement. Overall, MindLoop better coordinates regulation progression and affective outcomes through its stage–strategy policy and dual-timescale optimization.

4.1.2 Ablation Study

Table 2 presents the simulated-interaction ablation results. We evaluate five variants that remove stage–strategy factorization, stage-specific strategy heads, dual-timescale optimization, stage-level affect reward, or paired affect modeling. Three observations emerge. First, **stage–strategy factorization is important for regulation-process control**: replacing the hierarchical policy with a flat joint-action policy reduces SR and EI while increasing AT on both datasets. Second, **stage-specific strategy learning and dual-timescale optimization are important for affective outcomes**. Sharing a single strategy head causes a substantial EI drop despite high completion, while using the same full-dialogue return for both policy levels reduces both completion and affective improvement. Third, **affective modeling provides complementary learning signals**: removing either the stage-level affect reward or paired affect modeling consistently lowers EI. Overall, the results show that structured planning, temporally differentiated credit assignment, and affective feedback jointly contribute to MindLoop's regulation performance.

4.2 Real-World Validation Results

4.2.1 Regulation Performance in Human Interaction

Table 3 reports the real-world multimodal validation results. **MindLoop maintains the structured regulation process in human interaction**, achieving an SR of 1.00, an AS of 5.00, and an AT of 18.22. Turn-level affective outcomes are also favorable, with an EI of 49.06 and a PR of 0.783. PANAS shows an increase of 2.89 in positive affect and a decrease of 2.78 in negative affect after the interaction, with both changes reaching $p < 0.05$ under paired t-tests. The CTSR-4 and safety scores reach 4.67/6 and 5.39/6, respectively, indicating favorable interaction quality and safety. Overall, these results provide preliminary evidence that the simulation-trained policy can operate end-to-end in human interaction through subject-specific wearable EEG/fNIRS-based affect decoding, while preserving structured regulation and favorable within-session outcomes.

Table 3: Real-world multimodal validation results. Values with $\pm$ are reported as mean $\pm$ SD across subjects. PANAS changes were tested using paired t-tests. $^*p < 0.05$.

Regulation Process					**PANAS**		**Interaction Quality and Safety**	
SR	AT	AS	EI	PR	ΔPositive ↑	ΔNegative ↓	CTSR-4 ↑	Safety ↑
1.000	18.22 ± 3.39	5.00	49.06 ± 36.56	0.783	$+2.89 \pm 4.59^*$	$-2.78 \pm 5.45^*$	4.67 ± 1.05 / 6	5.39 ± 1.14 / 6

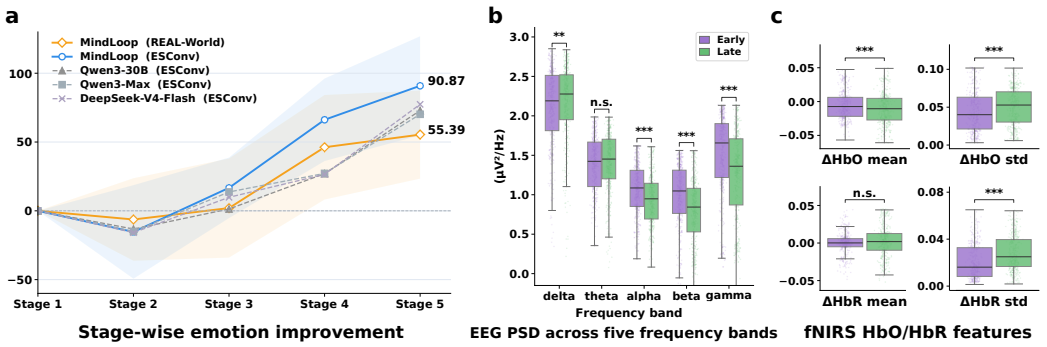

Figure 2: Stage-wise affective and physiological responses. (a) Affective change across the five regulation stages in simulated and real-world interactions. (b–c) Early- versus late-stage EEG and fNIRS comparisons. n.s. indicates non-significance, ** $p < 0.01$, and *** $p < 0.001$.

4.2.2 Affective Changes Across Regulation Stages

Fig. 2a presents stage-wise affective improvement relative to the Stage 1 baseline, providing a process-oriented view beyond the dialogue-level EI metric. Both simulated and real-world interactions show a similar pattern: affective improvement initially decreases at Stage 2 and then increases from Stage 3 to Stage 5. This pattern is consistent with the functional roles of the regulation stages. Stage 2 focuses on situation clarification and automatic-thought identification, which may temporarily make distress more explicit, whereas later stages emphasize evaluation, restructuring, and action planning. The initial decrease may therefore reflect a transitional exploration phase, while the subsequent increase suggests progressive affective recovery across later stages. **The similar trends across simulation and real-world interaction suggest that MindLoop preserves a stage-aligned regulation pattern in human interaction.**

4.2.3 Physiological Responses Across Regulation Stages

Figs. 2b–c compare physiological responses between the early regulation phase (Stage 1–2) and the late regulation phase (Stage 4–5). Significant EEG differences are observed in delta, alpha, beta, and gamma power spectral density, whereas theta shows no significant difference. The early stages mainly involve emotional disclosure, situation clarification, and identification of automatic thoughts, while later stages emphasize cognitive evaluation, restructuring, and action planning. These EEG differences therefore indicate measurable changes in neural activity patterns across the regulation process. For fNIRS, significant differences are observed in the mean and standard deviation of ΔHbO and the standard deviation of ΔHbR, but not in the mean of ΔHbR, suggesting corresponding changes in cortical hemodynamic responses across stages. Together, the EEG and fNIRS results provide complementary physiological evidence of stage-associated changes alongside the dialogue and self-reported affective results.

5 Conclusion and Discussion

Our results show that MindLoop provides effective process-level control for closed-loop emotion regulation. In simulated interactions, MindLoop consistently completes the five-stage regulation process while achieving strong affective outcomes. The real-world study further shows that the simulation-trained policy can maintain structured stage progression in human interaction with subject-specific wearable EEG/fNIRS feedback, together with favorable within-session affective changes. This study has three main limitations. First, the human evaluation is small-scale and limited to single-session interactions, motivating larger populations and repeated or longer-term studies. Second, the current system relies on subject-specific EEG/fNIRS decoders, motivating cross-subject or adaptive decoding to reduce calibration requirements. Third, differences remain between simulated and real-world interactions in affect distributions, user responses, and physiological observation noise. Future work will therefore focus on improving sim-to-real alignment, incorporating more realistic user dynamics, and adapting the regulation policy with real-world physiological feedback.

AI USE STATEMENT

In this work, generative AI tools were used for language polishing and assistance in drafting code for figure generation. Generative AI was not used to formulate the core research ideas, research questions or scientific conclusions. All AI-assisted text was reviewed and revised by the authors, and all AI-assisted plotting code was manually inspected, tested, and verified for correctness. The authors take full responsibility for the final content of this work, including all text, analyses, claims, figures, and conclusions.

ETHICS STATEMENT

This study was designed for non-clinical emotion-regulation research rather than diagnosis or psychotherapy. Approval of all ethical and experimental procedures and protocols was granted by the relevant institutional ethics committee. All participants provided informed consent and were informed that they could pause or terminate the session at any time. Sessions were monitored by research staff, with predefined stopping procedures for substantial discomfort or safety concerns. No adverse events were observed during the study.

REFERENCES

Sven Alfonsson, Georgios Karvelas, Johanna Linde, and Maria Beckman. A new short version of the cognitive therapy scale revised (ctsr-4): preliminary psychometric evaluation. *BMC psychology*, 10(1):21, 2022.

Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 1170–1183, 2023.

Yuning Chen, Sha Zhao, Shijian Li, and Gang Pan. Emod: A unified eeg emotion representation framework leveraging va guided contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17427–17435, 2026.

Jiale Cheng, Sahand Sabour, Hao Sun, Zhuang Chen, and Minlie Huang. Pal: Persona-augmented emotional support conversation generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 535–554, 2023.

Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 3014–3026, 2022.

Joshua E Curtiss, Daniella S Levine, Ilana Ander, and Amanda W Baker. Cognitive-behavioral treatments for anxiety and stress-related disorders. *Focus*, 19(2):184–189, 2021.

Daniel David, Ioana Cristea, and Stefan G. Hofmann. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Frontiers in Psychiatry*, Volume 9 - 2018, 2018. ISSN 1664-0640. doi: 10.3389/fpsyt.2018.00004. URL https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2018.00004.

DeepSeek-AI. Deepseek v4 preview release. https://api-docs.deepseek.com/news/news260424, April 2026. Accessed: 2026-05-19.

Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. A survey on proactive dialogue systems: Problems, methods, and prospects. *arXiv preprint arXiv:2305.02750*, 2023.

Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. Plug-and-play policy planner for large language model powered dialogue agents. In *International Conference on Learning Representations*, volume 2024, pp. 10058–10090, 2024.

Beth Fordham, Thavapriya Sugavanam, Katherine Edwards, Paul Stallard, Robert Howard, Roshan das Nair, Bethan Copsey, Hopin Lee, Jeremy Howick, Karla Hemming, and et al. The evidence for cognitive behavioural therapy in any condition, population or context: a meta-review of systematic reviews and panoramic meta-analysis. *Psychological Medicine*, 51(1):21–29, 2021. doi: 10.1017/S0033291720005292.

Team Glm, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*, 2024.

Kim L Gratz, Nicole H Weiss, and Matthew T Tull. Examining emotion regulation as an outcome, mechanism, or target of psychological treatments. *Current Opinion in Psychology*, 3:85–90, 2015. ISSN 2352-250X. doi: https://doi.org/10.1016/j.copsyc.2015.02.010. URL https://www.sciencedirect.com/science/article/pii/S2352250X15001037. Emotion regulation.

James J Gross. The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2(3):271–299, 1998.

James J Gross. Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3):281–291, 2002.

James J Gross. Emotion regulation: Current status and future prospects. *Psychological inquiry*, 26 (1):1–26, 2015.

Zhijun Guo, Alvina Lai, Johan H Thygesen, Joseph Farrington, Thomas Keen, and Kezhi Li. Large language models for mental health applications: systematic review. *JMIR mental health*, 11(1): e57400, 2024.

Tao He, Lizi Liao, Yixin Cao, Yuanxing Liu, Yiheng Sun, Zerui Chen, Ming Liu, and Bing Qin. Simulation-free hierarchical latent policy planning for proactive dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 24032–24040, 2025.

Stefan G Hofmann, Anu Asnaani, Imke JJ Vonk, Alice T Sawyer, and Angela Fang. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive therapy and research*, 36(5): 427–440, 2012.

Yining Hua, Hongbin Na, Zehan Li, Fenglin Liu, Xiao Fang, David Clifton, and John Torous. A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine*, 8(1):230, 2025.

Dongjin Kang, Sunghwan Mac Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. Can large language models be good emotional supporter? mitigating preference bias on emotional support conversation. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 15232–15261, 2024.

Juhee Kim, Chunghu Mok, Jisun Lee, Hyang Sook Kim, and Yohan Jo. Dialogue systems for emotional support via value reinforcement. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 28733–28766, 2025.

Wai-Chung Kwan, Hong-Ru Wang, Hui-Min Wang, and Kam-Fai Wong. A survey on recent advances and challenges in reinforcement learning methods for task-oriented dialogue policy learning. *Machine Intelligence Research*, 20(3):318–334, 2023.

Hannah R Lawrence, Renee A Schneider, Susan B Rubin, Maja J Matarić, Daniel J McDuff, and Megan Jones Bell. The opportunities and risks of large language models in mental health. *JMIR Mental Health*, 11(1):e59479, 2024.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)*, pp. 3469–3483, 2021.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Tristan J Philippe, Naureen Sikder, Anna Jackson, Maya E Koblanski, Eric Liow, Andreas Pilarinos, and Krisztina Vasarhelyi. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. *JMIR mental health*, 9(5):e35159, 2022.

Tazeek Bin Abdur Rakib, Ambuj Mehrish, Lay-Ki Soon, Wern Han Lim, and Soujanya Poria. Dialogxpert: Driving intelligent and emotion-aware conversations through online value-based reinforcement learning with llm priors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 29967–29975, 2026.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 5370–5381, 2019.

Stephen M Schueller and John Torous. Scaling evidence-based treatments through digital mental health. *American Psychologist*, 75(8):1093, 2020.

Petr Slovak, Alissa Antle, Nikki Theofanopoulou, Claudia Daudén Roquet, James Gross, and Katherine Isbister. Designing for emotion regulation interventions: an agenda for hci theory and research. *ACM Transactions on Computer-Human Interaction*, 30(1):1–51, 2023.

Pei-Hao Su, Milica Gasic, Nikola Mrkšić, Lina M Rojas Barahona, Stefan Ultes, David Vandyke, Tsung-Hsien Wen, and Steve Young. On-line active reward learning for policy optimisation in spoken dialogue systems. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2431–2441, 2016.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. Misc: A mixed strategy-aware model integrating comet for emotional support conversation. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)*, pp. 308–319, 2022.

David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54 (6):1063, 1988.

Haojie Xie, Yirong Chen, Xiaofen Xing, Jingkai Lin, and Xiangmin Xu. Psydt: Using llms to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1081–1115, 2025.

Xin Xu, Xinke Shen, Xuyang Chen, Qingzhu Zhang, Sitian Wang, Yihan Li, Zongsheng Li, Dan Zhang, Mingming Zhang, and Quanying Liu. A multi-context emotional eeg dataset for cross-context emotion decoding. *Scientific Data*, 12(1):1142, 2025a.

Yangyang Xu, Jinpeng Hu, Zhuoer Zhao, Zhangling Duan, Xiao Sun, and Xun Yang. Multiagentesc: A llm-based multi-agent collaboration framework for emotional support conversation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 4665–4681, 2025b.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chenghua Huang, Chengxing Xie, et al. Glm-5: from vibe coding to agentic engineering. *arXiv preprint arXiv:2602.15763*, 2026.

Chenhao Zhang, Renhao Li, Minghuan Tan, Min Yang, Jingwei Zhu, Di Yang, Jiahao Zhao, Guancheng Ye, Chengming Li, and Xiping Hu. Cpsycoun: A report-based multi-turn dialogue reconstruction and evaluation framework for chinese psychological counseling. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 13947–13966, 2024.

Tong Zhang, Wenming Zheng, Zhen Cui, Yuan Zong, and Yang Li. Spatial–temporal recurrent neural network for emotion recognition. *IEEE Transactions on Cybernetics*, 49(3):839–847, 2019. doi: 10.1109/TCYB.2017.2788081.

Xiaowei Zhang, Jinyong Liu, Jian Shen, Shaojie Li, Kechen Hou, Bin Hu, Jin Gao, Tong Zhang, and Bin Hu. Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Transactions on Cybernetics*, 51(9):4386–4399, 2021. doi: 10.1109/TCYB.2020.2987575.

Sha Zhao, Mingyi Peng, Haiteng Jiang, Tao Li, Shijian Li, and Gang Pan. Eegagent: A unified framework for automated eeg analysis using large language models, 2025a. URL `https://arxiv.org/abs/2511.09947`.

Sha Zhao, Song Yi, Yangxuan Zhou, Jiadong Pan, Jiquan Wang, Jie Xia, Shijian Li, Shurong Dong, and Gang Pan. Wearable music2emotion: Assessing emotions induced by ai-generated music through portable eeg-fnirs fusion. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 5627–5636, 2025b.

Sha Zhao, Song Yi, Yangxuan Zhou, Jiadong Pan, Jiquan Wang, Jie Xia, Shijian Li, Shurong Dong, and Gang Pan. Tuning the mind: Closed-loop emotion regulation via aigc music driven by wearable multimodal neurofeedback. *IEEE Transactions on Affective Computing*, 2026.

Weixiang Zhao, Yanyan Zhao, Shilong Wang, and Bing Qin. Transesc: Smoothing emotional support conversation via turn-level state transition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 6725–6739, 2023.

Chujie Zheng, Sahand Sabour, Jiaxin Wen, Zheng Zhang, and Minlie Huang. Augesc: Dialogue augmentation with large language models for emotional support conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 1552–1568, 2023.

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, 49 (3):1110–1122, 2019. doi: 10.1109/TCYB.2018.2797176.

Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin, and Liqiang Nie. Self-chats from large language models make small emotional support chatbot better. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11325–11345, 2024.

Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. Facilitating multi-turn emotional support conversation with positive emotion elicitation: A reinforcement learning approach. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1714–1729, 2023.

A Related Work

A.1 LLM-based Emotional Support Dialogue

Emotional support dialogue aims to help users express distress, feel understood, and move toward emotional relief through multi-turn interaction. Early studies formulated this task as a dialogue generation problem and introduced benchmark datasets such as ESConv, which has become a widely used benchmark for emotional support systems Liu et al. (2021). Based on these resources, subsequent methods have explored empathetic response generation, support-strategy modeling, persona-augmented support, retrieval-augmented generation, and LLM-based prompting or fine-tuning to improve response quality Rashkin et al. (2019); Tu et al. (2022); Cheng et al. (2023); Zheng et al. (2023); Kang et al. (2024); Zhang et al. (2024). These methods substantially improve linguistic fluency, empathy, and local appropriateness. However, most existing studies still focus on generating suitable replies for the current context, while overlooking the structured and adaptive nature of emotion regulation. As a result, they lack explicit mechanisms for organizing local support actions into a coherent regulation process.

To move beyond response-level modeling, recent studies have begun to formulate emotional support dialogue as a sequential decision-making problem, where the system selects intervention actions and adapts its behavior according to evolving interaction states. This perspective connects emotional support with long-horizon dialogue planning, reinforcement learning, and policy optimization Su et al. (2016); Kwan et al. (2023); Deng et al. (2023). Representative methods such as PPDPP Deng et al. (2024) and DialogXpert Rakib et al. (2026) show that explicit policy learning can improve multi-turn dialogue control by selecting actions based on user states rather than generating each response independently. However, existing dialogue policy methods are still largely optimized under generic conversational objectives, such as task completion, dialogue success, response helpfulness, or interaction efficiency. These objectives do not fully capture the stage-dependent structure of emotion regulation, where local response appropriateness and global regulation progress operate on different temporal scales. Therefore, LLM-based emotion regulation requires a regulation-aware policy structure that coordinates stage-level progression with stage-specific intervention quality.

A.2 Affective Feedback and Physiological Emotion Sensing

Affective feedback is essential for adapting interventions in dialogue-based emotional support and emotion regulation. Existing systems commonly use emotion labels, empathy signals, user preferences, affect-related rewards, or LLM-based evaluators to guide response generation and assess dialogue quality Zhou et al. (2023); Park et al. (2023); Kim et al. (2025). However, these signals are usually treated as utterance-level attributes, response-level preferences, or post-hoc evaluation targets, making it difficult to track affective changes across a multi-turn regulation process. Physiological emotion sensing provides complementary cues beyond text Zheng et al. (2019); Zhao et al. (2025a). EEG has been widely used for affective state recognition and emotion representation learning Zhang et al. (2019); Chen et al. (2026), while fNIRS and EEG-fNIRS fusion have shown value for emotion-related state estimation and wearable affective sensing Zhang et al. (2021); Zhao et al. (2025b; 2026). However, these signals are often treated as standalone inputs for emotion recognition and remain weakly integrated with dialogue policy learning. Our work models affective change as paired feedback in LLM-based emotion regulation and uses wearable EEG/fNIRS decoding to provide objective observations for closed-loop policy adaptation during real human interaction.

B Implementation Details

This section describes the stage–strategy inventory, planner architecture, reward construction, policy-training configuration, and baseline planning protocols. Full prompt templates are provided in Appendix E.

B.1 Stage–Strategy Design

The five-stage structure and stage-specific strategy inventory were developed through iterative discussions with domain experts to align the computational intervention space with established CBT

principles. The stages define the position of the dialogue in the regulation process, with operational goals and completion evidence summarized in Table 4. The planner may remain in the current stage or progress to the next stage, but cannot move backward or skip a stage. Each stage has five legal intervention strategies, summarized in Table 5. The descriptions specify the local intervention role of each strategy, and *Dialogue Completed* is the legal terminal action in Consolidation.

Table 4: Operational definitions of the CBT-guided regulation stages.

Stage	Regulation goal	Completion evidence
Stage 1: Engagement	Establish a safe and collaborative relationship, identify one concrete concern, and clarify the initial emotion and session goal.	The user engages with the support, a focal event is identified, and the core emotion has an initial intensity estimate.
Stage 2: Identification	Separate the situation, thoughts, emotions, bodily responses, and behavior, and identify the central automatic thought.	Facts, thoughts, and feelings are distinguished and at least one hot cognition is captured in the user's own words.
Stage 3: Evaluation	Examine the automatic thought through evidence and alternative perspectives.	Supporting and opposing evidence are considered and the user recognizes at least one plausible alternative interpretation.
Stage 4: Restructuring	Form a realistic and balanced alternative cognition and examine its affective effect.	A specific balanced thought is produced and its believability or emotional effect is re-evaluated.
Stage 5: Consolidation	Consolidate learning, prepare a feasible action, and anticipate the return of the old thought.	The user summarizes the learning and identifies a near-term action or structured exercise before legal termination.

Table 5: Stage-specific intervention strategies and their operational descriptions.

Strategy	**Operational description**
Stage 1: Engagement	
Empathy & Normalization	Validate the user's emotion and frame it as an understandable response to the situation.
Emotion Labeling & Scaling	Help the user name the central emotion and estimate its current intensity.
Event Anchoring	Narrow a broad concern to one recent and concrete event.
Goal Setting	Define a specific regulation goal for the current interaction.
Role & Boundary Setting	Establish a collaborative role and clarify the scope of support.
Stage 2: Identification	
Five-Part Breakdown	Separate the situation, emotion, bodily response, behavior, and automatic thought.
Thought–Feeling Distinction	Distinguish an interpretation or judgment from the emotion associated with it.
Hot Cognition Identification	Identify the thought that appeared at the point of strongest emotional activation.
Verbatim Thought Recording	Preserve the automatic thought in the user's own words without softening it.

Continued on the next page

Table 5 (continued)

Strategy	Operational description
Cognitive Distortion Labeling	Identify the thinking pattern reflected in the automatic thought.
Stage 3: Evaluation	
Evidence Testing	Examine concrete evidence that supports and contradicts the automatic thought.
Friend Perspective / Double Standard	Apply the more balanced standard the user would offer to another person.
Downward Arrow	Explore the deeper meaning or belief underlying the surface thought.
Alternative Perspectives	Generate additional plausible interpretations of the same event.
De-extremizing	Use graded comparison or responsibility allocation to reduce extreme conclusions.
Stage 4: Restructuring	
Balanced Thought Generation	Integrate the available evidence into a realistic and flexible alternative thought.
Semantic Softening	Replace rigid or catastrophic wording with more graded language.
Strengths Activation	Recall prior coping experiences and personal strengths that support the alternative view.
Believability & SUDs Re-evaluation	Reassess belief in the new thought and the intensity of distress.
Coping Card Generation	Convert the balanced response into a concise reminder for later use.
Stage 5: Consolidation	
Client-led Summary & Feedback	Ask the user to summarize the central learning and reflect on the interaction.
Micro-Action Planning	Identify a small, concrete, and feasible next action.
Trigger Anticipation & Relapse Prevention	Anticipate future triggers and prepare a response if the old thought returns.
Structured Homework	Assign a repeatable exercise that extends regulation beyond the dialogue.
Dialogue Completed	Legally terminate the interaction after the Consolidation goals have been met.

B.2 Planner Architecture

The planner uses a frozen Chinese RoBERTa-large encoder to represent the dialogue history and current user utterance. Its semantic representation is combined with structured embeddings for paired affective feedback, the previous stage and strategy, and stage and strategy histories, then projected into a shared state representation. The actor contains one global stage head and five stage-specific strategy heads. The stage head may remain in the current stage or progress by one stage, while the strategy branch activates only the head associated with the selected stage. The critic contains a global stage-value head and stage-conditioned strategy-value heads. Legal masks prevent backward or skipped stage transitions and strategies outside the selected stage.

B.3 Reward Construction

The reward construction uses three distinct signals: stage-task completion, strategy appropriateness, and affective feedback. Stage-task and strategy-quality scores are obtained from separate LLM judges, while the affective score is computed from affective observations.

Table 6: Mapping from stage-completion criteria to the stage-task reward.

Completed criteria	Internal score	R_k^{task}
0 / 3	-1	-3.0
1 / 3	0	-1.5
2 / 3	1	$+1.5$
3 / 3	3	$+4.5$

Table 7: Mapping from strategy-appropriateness scores to reward anchors.

Score	Interpretation	Reward anchor
0	Inappropriate timing or use	-1.0
1	Partially appropriate	$+0.5$
2	Appropriate and well-timed	$+2.0$

Checklist-based stage evaluation. The stage judge is invoked when the planner attempts to leave the current regulation stage; in Stage 5, it is invoked when termination is attempted. All five stages use the same judging template with stage-specific definitions and three completion criteria. The judge receives the preceding-stage summary and the current-stage dialogue segment, including user utterances, counselor responses, selected strategies, and subsequent user replies. For each criterion, it returns a Boolean completion decision and a brief justification, together with a summary of the current stage.

Let $c_{k,j} \in \{0, 1\}$ indicate whether criterion j is completed in stage G_k. The number of completed criteria is

$$n_k = \sum_{j=1}^{3} c_{k,j}. \tag{20}$$

The implementation converts n_k into an internal stage score and then into the stage-task reward, as summarized in Table 6. The resulting reward is recorded at the stage-ending event.

Segment-level strategy evaluation. A separate strategy judge evaluates the timing and appropriateness of each intervention within the current stage. Each counselor turn receives a score $q_{k,t} \in \{0, 1, 2\}$, corresponding to inappropriate, partially appropriate, or appropriate strategy use, respectively. The three score anchors are summarized in Table 7.

The implementation first averages the strategy scores within the stage segment:

$$\bar{q}_k = \frac{1}{|\tau_k|} \sum_{t \in \tau_k} q_{k,t}, \tag{21}$$

and then applies the corresponding linear mapping:

$$R_k^{\text{strategy}} = -1 + \frac{3}{2}\bar{q}_k. \tag{22}$$

Thus, R_k^{strategy} summarizes intervention quality across the current stage rather than treating each turn-level score as an independent stage reward.

Affective feedback and shaping rewards. The affective component uses the segment-level post-affect aggregation described in the main text. Its score $S_k^{\text{affect}} \in [0, 100]$ is normalized to $[0, 4]$ as

$$r_k^{\text{affect}} = 0.04 S_k^{\text{affect}}. \tag{23}$$

Additional shaping terms assign $+1$ for legal stage progression and $+2$ for legal dialogue completion. A strategy-diversity bonus of $+0.5$ encourages varied interventions, while each additional turn beyond five turns within the same stage receives a penalty of -1. These terms are combined with the stage-task, strategy-quality, and affective signals to construct the rewards used for policy optimization.

If a judge call or JSON parsing fails, a rule-based fallback is used and recorded by the implementation.

Table 8: Configuration for training the regulation policy.

Parameter	Value
Policy updates	30
Rollouts per update	128
Maximum dialogue turns	25
Optimization passes per update	3
Minibatch size	64 transitions
Discount factor γ	0.99
GAE parameter λ	0.95
Actor learning rate	1×10^{-5}
Critic learning rate	3×10^{-5}
Gradient-norm clipping	1.0
Policy-training random seed	42

B.4 Training Configuration

The main simulated-interaction experiments were conducted using a single NVIDIA RTX PRO 6000 Blackwell Server Edition GPU with 96 GB of memory. The executor, user simulator, stage-level judge, and strategy-level judge were instantiated with Qwen3-30B and served locally with vLLM. The implementation uses PyTorch and Hugging Face Transformers. A random seed of 42 was used for policy-training experiments. Cross-model evaluation follows the protocol in Appendix D.2.

The planner is optimized directly through closed-loop interaction with the fixed executor LLM, user simulator, and LLM judge. Each rollout represents a complete regulation process under the same stage-transition constraints and maximum interaction horizon used in evaluation. Only the planner is updated; all language-model components remain frozen.

The stage branch uses full-dialogue credit assignment to capture delayed effects on overall regulation progress. The strategy branch uses current-stage credit assignment and stops propagating returns after a stage transition. The two branches share the same state representation but use separate policy and value heads, allowing overall regulation progress and current-stage intervention quality to be optimized jointly.

Training uses 30 policy updates, each collecting 128 complete rollouts with a maximum of 25 dialogue turns. Each update performs three optimization passes over minibatches of 64 transitions. The full training run takes approximately 18 hours. Table 8 summarizes the optimization configuration.

B.5 Baseline Planning Protocols

The three protocols use the same regulation stages, legal strategy space, and user simulator, while differing in how stage selection, strategy selection, and response generation are decomposed. JP, HP, and MindLoop use the same fixed executor for response realization, whereas SP performs planning and response generation jointly in a single call. The environment allows up to 25 turns, while the planning prompts use a 20-turn target to encourage timely stage progression.

Single-Pass (SP). A single LLM call jointly determines the regulation stage, selects a legal stage-specific strategy, and generates the response. Thus, planning and response realization are performed within the same generation process without an explicit intermediate intervention plan.

Joint Planning (JP). One planning call jointly predicts the regulation stage and intervention strategy as a structured plan. The fixed executor LLM then realizes this plan as a natural-language response in a separate call, explicitly separating intervention planning from response generation.

Hierarchical Planning (HP). Stage progression and stage-specific strategy selection are performed in two separate planning calls. The first call determines whether the dialogue should remain in the current stage or advance to the next stage, while the second selects a legal intervention strategy within the resulting stage. The complete stage–strategy plan is then passed to the fixed executor LLM for response realization. Accordingly, SP, JP, and HP require one, two, and three LLM calls

per interaction turn, respectively. The complete hierarchical planning templates are shown in Figs. 7 and 8.

JP, HP, and the learned MindLoop planner pass the selected stage–strategy intervention plan to the same fixed executor LLM. Given the dialogue context and intervention plan, the executor generates a natural-language response that follows the objective of the selected stage and the functional role of the selected strategy, without modifying the planner's decision. Using the same executor isolates the effect of planning quality from differences in response-generation capability. The response-realization prompt is shown in Fig. 9. The prompts for SP and JP are constructed by combining the corresponding stage-planning, strategy-planning, and response-realization components: JP merges stage and strategy selection into a single planning prompt, whereas SP further incorporates response realization into the same prompt.

C Real-World Study and Affect Decoding

MindLoop uses subject-specific physiological decoders to estimate valence and arousal from synchronized wearable EEG and fNIRS signals during real-world interaction. The experimental protocol contains an emotion-imagery block for subject-specific decoder training and validation, followed by an emotion-regulation block in which the trained decoders are fixed and deployed for online affect estimation.

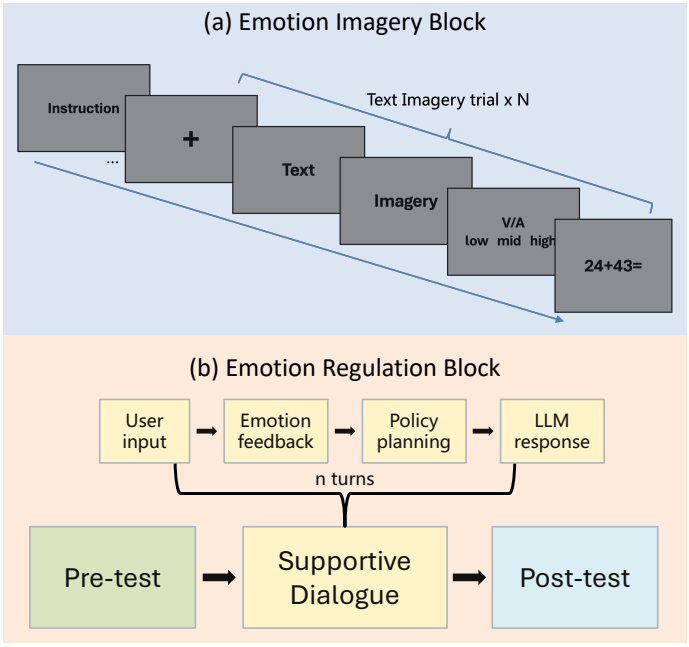

Figure 3: Real-world experimental protocol. The emotion-imagery block provides labeled subject-specific EEG/fNIRS data for decoder training and validation. The selected decoders are subsequently deployed during the emotion-regulation block to provide affective feedback for closed-loop interaction.

C.1 Experimental Protocol and Data Preparation

The complete experimental protocol is illustrated in Fig. 3. Each participant first completed an emotion-imagery block using pre-rated emotional text materials. The materials covered all combinations of low, medium, and high valence and low, medium, and high arousal. Two trials were presented for each valence–arousal combination, resulting in 18 trials per participant. Each trial included a fixation period, presentation of the emotional text, a 30-s emotion-imagery period, self-reported valence and arousal ratings, and a brief arithmetic task. The distributions were balanced

for both affective dimensions: each participant completed six low-, six medium-, and six high-level trials for valence, and the same distribution for arousal.

The data were partitioned at the trial level before window segmentation, ensuring that windows from the same trial did not appear in both the training and validation sets. Within each low-, medium-, and high-level class, four trials were assigned to training and two to validation according to a fixed, class-balanced split. Each 30-s imagery period was segmented into 5-s windows with a 2-s step size, producing 13 synchronized EEG/fNIRS samples per trial. Each participant therefore contributed 234 synchronized windows. For either the valence or arousal decoder, 156 samples from 12 trials were used for training and 78 samples from six trials were used for validation. Each affective class contained 52 training samples and 26 validation samples.

C.2 Signal Acquisition and Preprocessing

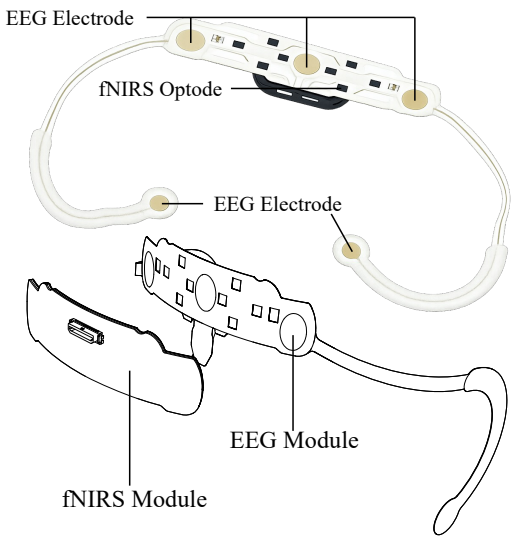

Figure 4: Wearable EEG/fNIRS device used for physiological signal acquisition during the real-world experiment.

The wearable device shown in Fig. 4 integrates EEG electrodes, fNIRS optodes, and their acquisition modules in a head-mounted configuration, enabling continuous physiological recording during dialogue. The synchronized acquisition interface used to monitor EEG and fNIRS signals during the experiment is shown in Fig. 5.

EEG was sampled at 250 Hz from Fpz, AF7, AF8, and M2, while fNIRS was sampled at 25 Hz from 16 channels. EEG was processed using a causal fourth-order 0.3–49-Hz IIR band-pass filter. The same channel order and causal preprocessing procedure were used during decoder training and online inference.

C.3 Decoder Architecture

The EEG branch was initialized from EMOD and used parallel temporal- and frequency-domain tokenization followed by a three-layer axial Transformer. Channel- and temporal-attention pooling produced a 128-dimensional EEG representation. The fNIRS branch consisted of a temporal convolutional stem, three residual temporal blocks, and adaptive temporal pooling, producing a 128-dimensional fNIRS representation. The two representations were concatenated and passed through a nonlinear projection and a three-class classification head. Separate subject-specific models were constructed for valence and arousal.

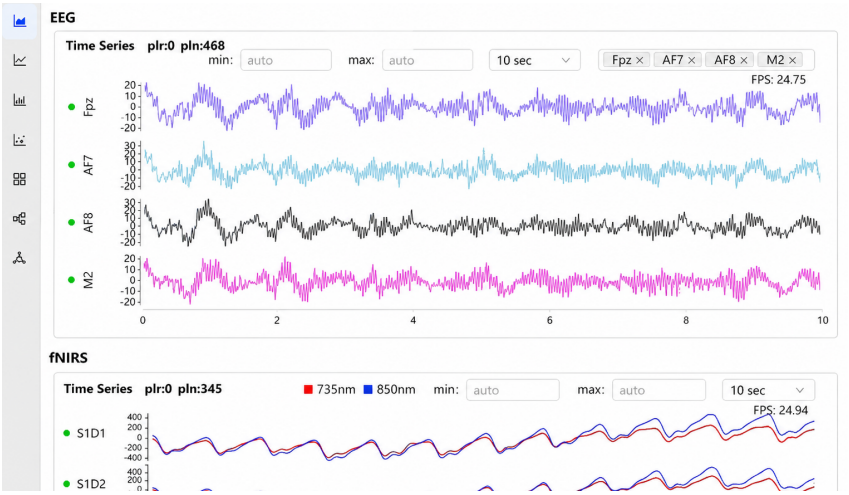

Figure 5: Interface for monitoring synchronized wearable EEG and fNIRS signal acquisition during the real-world experiment.

C.4 Decoder Training and Online Inference

The EEG encoder was initialized from a pretrained EMOD checkpoint, whereas the fNIRS branch and multimodal fusion head were initialized for the wearable setting. The valence and arousal decoders were trained separately using only the corresponding participant's emotion-imagery data. Each model was trained for up to 30 epochs using Adam with a batch size of 16 and a learning rate of 5×10^{-4}. Training used dropout of 0.25, cross-entropy loss with label smoothing of 0.1, cosine learning-rate scheduling, gradient clipping at 1.0, and a random seed of 42. After each epoch, the model was evaluated on the held-out validation trials, and the checkpoint with the highest validation accuracy was selected. Training and selecting the two decoders required approximately 10 minutes per participant.

After model selection, the decoder parameters were fixed, and no further training or adaptation was performed during the emotion-regulation block. Online inference used 5-s windows with a 1-s step size. For each affective dimension, predictions from three consecutive windows were aggregated using a voting rule to obtain a more stable low-, medium-, or high-level estimate. For each dialogue turn, the resulting pre- and post-utterance affective estimates were incorporated into the subsequent dialogue state. Participants additionally provided turn-level self-reported valence and arousal labels. These labels were used only for cross-block evaluation and were not provided to the physiological decoder, MindLoop planner, or executor LLM during interaction.

Within-block and cross-block decoder results are reported in Appendix D.5.

D ADDITIONAL EXPERIMENTAL RESULTS

We examine four aspects of MindLoop: how reinforcement learning changes the regulation policy, how the policy operates across user-simulator and evaluator models, how accurately physiological affect is decoded during real-world interaction, and how variations in affective observations influence planning.

D.1 LEARNING DYNAMICS OF THE REGULATION POLICY

We examine how reinforcement learning shapes stage progression and within-stage intervention selection. We evaluate the planner at updates 5, 10, 15, 20, 25, and 30, recording stage-specific strategy distributions, the proportion of turns assigned to each stage, completion rate, completion turns, and emotion improvement. Figure 6 summarizes these changes throughout training.

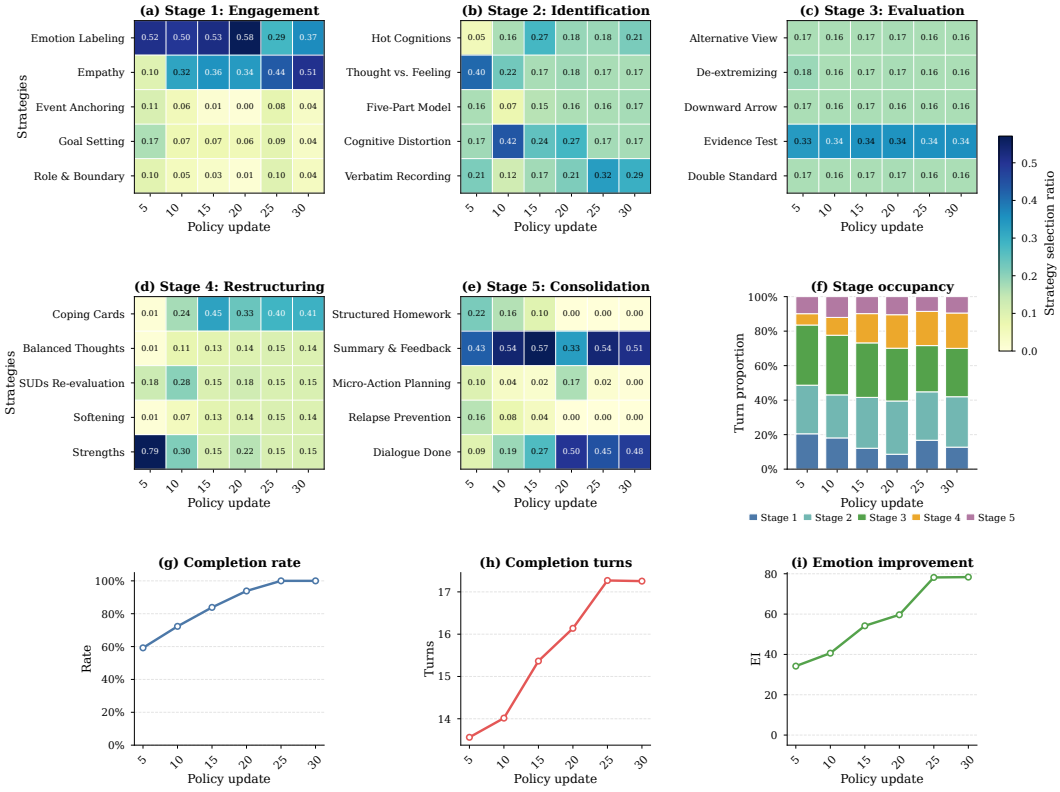

Figure 6: Policy-learning dynamics at six checkpoints. Panels (a)–(e) show strategy-selection distributions within the five regulation stages; panel (f) shows stage occupancy; and panels (g)–(i) show completion rate, completion turns, and emotion improvement, respectively.

Completion rate increases from 59.23% at update 5 to 100% at updates 25 and 30, while EI rises from 34.20 to 78.35. Completion turns change from 13.56 to approximately 17.25 over the same period. Stage occupancy also changes: Stage 1 decreases from 20.48% to 12.75%, whereas Stage 4 increases from 6.55% to 20.48%. Stages 2 and 3 continue to account for substantial portions of the interaction. Thus, the later policies complete the process more consistently while allocating more of the dialogue to restructuring.

The strategy distributions further reveal what changes within each stage. Evidence Testing remains the most frequent Stage 3 strategy, with a selection ratio of approximately 0.33–0.34. In Stage 4, the early concentration on Strengths Activation gives way to greater use of Coping Card Generation and a more even distribution over the remaining strategies. In Stage 5, Client-led Summary and Feedback remains prominent, while Dialogue Completed is selected more frequently at later checkpoints. Together, these changes show that learning adjusts both the allocation of turns across stages

and the choice of interventions within stages, alongside improvements in completion and simulated affective outcomes.

D.2 Cross-Model User-Simulator Generalization

To examine whether MindLoop generalizes across different simulated user environments, we conduct a cross-model user-simulator experiment. We keep the trained MindLoop planner and Qwen3-30B executor fixed, and independently instantiate the user simulator with Qwen3-30B, DeepSeek-V4-Flash, and GLM-5.1. Each simulator is evaluated on the complete ESConv and ExTES test sets, containing 130 and 200 scenarios, respectively. All settings use the same test scenarios, stage-transition constraints, strategy inventory, 25-turn horizon, and evaluation metrics, so that only the user-simulator model is changed.

Table 9: Closed-loop results with different user simulators on the complete ESConv and ExTES test sets. Values are reported as mean/std across evaluation rollouts. The trained MindLoop planner and Qwen3-30B executor are fixed across all settings.

User simulator	AS↑	SR↑	AT↓	EI↑	PR↑
ESConv					
Qwen3-30B	5.00/0.00	1.000/0.000	17.18/0.40	79.93/22.07	0.985/0.123
DeepSeek-V4-Flash	5.00/0.00	1.000/0.000	17.18/0.38	63.80/10.90	1.000/0.000
GLM-5.1	5.00/0.00	1.000/0.000	17.25/0.44	66.05/15.11	0.969/0.173
ExTES					
Qwen3-30B	5.00/0.00	1.000/0.000	17.13/0.37	91.97/13.43	1.000/0.000
DeepSeek-V4-Flash	5.00/0.00	1.000/0.000	17.23/0.43	63.63/9.88	1.000/0.000
GLM-5.1	5.00/0.00	1.000/0.000	17.39/0.49	67.85/10.57	1.000/0.000

Table 9 shows a highly consistent regulation process across all three user simulators. MindLoop achieves full stage coverage and successful completion in every setting, with AS/SR of 5.00/1.00 on both datasets. Interaction length is also stable: AT varies only from 17.18 to 17.25 turns on ESConv and from 17.13 to 17.39 turns on ExTES. This indicates that changing the user-simulator model does not alter the planner's ability to advance through the five regulation stages and complete the dialogue.

Affective outcomes show a similarly consistent pattern across the two held-out simulator families. DeepSeek-V4-Flash and GLM-5.1 obtain EI scores of 63.80 and 66.05 on ESConv, and 63.63 and 67.85 on ExTES, respectively. Their PR values are also consistently high: all DeepSeek-V4-Flash dialogues show positive improvement, while GLM-5.1 achieves positive improvement in 126/130 ESConv dialogues and all 200 ExTES dialogues. Overall, the cross-model results show that MindLoop preserves both structured stage progression and consistently positive affective trajectories when interacting with user simulators from different model families.

D.3 Cross-Model Executor Generalization

We further examine whether the learned MindLoop planner depends on a specific response-generation model. The trained MindLoop planner and Qwen3-30B user simulator are kept fixed, while the executor is independently instantiated with Qwen3-30B, GLM-5.1, and DeepSeek-V4-Flash. Each executor receives the same structured stage–strategy plan produced by MindLoop and differs only in how this plan is realized as a natural-language response. We evaluate all executor variants on the complete ESConv and ExTES test sets. The stage-transition constraints, strategy inventory, user simulator, interaction horizon, and evaluation metrics are unchanged across settings.

As shown in Table 10, the regulation process remains highly stable after replacing the response-generation model. All three executors achieve complete stage coverage and a 100% success rate on both datasets. Average interaction length also remains nearly unchanged, ranging from 17.17 to 17.23 turns on ESConv and from 17.13 to 17.23 turns on ExTES. These results indicate that the learned planner can reliably realize its stage–strategy decisions through different executor LLMs rather than depending on the response-generation behavior of Qwen3-30B.

Table 10: Closed-loop results with different executor models on the complete ESConv and ExTES test sets. Values are reported as mean/std across evaluation rollouts. The trained MindLoop planner and Qwen3-30B user simulator are fixed across all settings.

Executor	AS↑	SR↑	AT↓	EI↑	PR↑
ESConv					
Qwen3-30B	5.00/0.00	1.000/0.000	17.18/0.40	79.93/22.07	0.985/0.123
GLM-5.1	5.00/0.00	1.000/0.000	17.17/0.37	85.22/20.57	1.000/0.000
DeepSeek-V4-Flash	5.00/0.00	1.000/0.000	17.23/0.47	87.24/18.06	0.992/0.087
ExTES					
Qwen3-30B	5.00/0.00	1.000/0.000	17.13/0.37	91.97/13.43	1.000/0.000
GLM-5.1	5.00/0.00	1.000/0.000	17.20/0.45	91.64/14.55	0.995/0.071
DeepSeek-V4-Flash	5.00/0.00	1.000/0.000	17.23/0.45	95.33/9.05	1.000/0.000

Affective outcomes are similarly preserved across executor families. On ESConv, EI ranges from 79.93 to 87.24 and PR from 0.985 to 1.000; on ExTES, EI ranges from 91.64 to 95.33 and PR from 0.995 to 1.000. Because the Qwen3-30B results correspond to an earlier stochastic evaluation run rather than paired trajectories generated under identical randomness, differences in absolute EI values are interpreted descriptively. The important observation is that replacing the executor with held-out model families does not degrade either regulation-process completion or affective outcomes. Together with the cross-model user-simulator results, this provides evidence that the learned MindLoop planner remains stable under changes to both the user-environment model and the response-generation model.

D.4 Role of CBT-Guided Stage Progression

To examine whether the explicit CBT-guided progression contributes beyond access to the intervention strategies themselves, we construct a *Free-Order-HP* control. Both Free-Order-HP and CBT-Structured-HP use GLM-5.1 as the planner, Qwen3-30B as the executor, and Qwen3-30B as the user simulator. They are provided with the same five strategy groups and the same strategy definitions. The difference lies only in how these groups are organized during interaction.

CBT-Structured-HP follows the ordered regulation process used in MindLoop: the planner may remain in the current stage or advance to the next stage, and strategies are selected from the corresponding stage-specific set. In Free-Order-HP, this progression constraint is removed. At each turn, the planner may freely select any of the five strategy groups according to the current dialogue context and then choose a strategy within the selected group. It may therefore skip or revisit groups instead of following the predefined Engagement–Identification–Evaluation–Restructuring–Consolidation sequence. This control preserves the intervention repertoire and underlying LLM capability while removing the explicit CBT-guided process organization.

Table 11 shows that explicitly organizing interventions according to the CBT-guided progression substantially improves affective outcomes compared with allowing the same strategy groups to be selected in arbitrary order. On ESConv, CBT-Structured-HP increases EI from 50.48 to 70.49 and PR from 0.892 to 0.985. On ExTES, EI increases from 59.29 to 78.20 and PR from 0.920 to 0.995. Thus, structured progression improves EI by approximately 20.0 and 18.9 points on ESConv and ExTES, respectively, while increasing PR by approximately 9.2 and 7.5 percentage points. These gains are obtained without changing the underlying planner LLM, executor, user simulator, or available intervention strategies, isolating the benefit of explicitly organizing interventions through the CBT-guided process.

Free-Order-HP produces shorter interactions, averaging 11.20 and 10.99 turns on ESConv and ExTES, compared with 13.13 and 13.49 turns for CBT-Structured-HP. However, Free-Order-HP is allowed to skip or revisit strategy groups and therefore does not follow the same structured completion process. Its shorter trajectories are accompanied by lower stage coverage, lower completion rates, and substantially weaker affective outcomes, suggesting that simply reaching an earlier termination does not provide the same regulation benefit as completing the structured process.

Table 11: Comparison of free-order planning, CBT-structured planning, and the learned MindLoop policy on the complete ESConv and ExTES test sets. Values are reported as mean/std across evaluation rollouts. Free-Order-HP and CBT-Structured-HP use GLM-5.1 as the planner with a Qwen3-30B executor, while MindLoop uses its learned planner with a GLM-5.1 executor. All settings use the same Qwen3-30B user simulator.

Method	AS↑	SR↑	AT↓	EI↑	PR↑
ESConv					
Free-Order-HP	4.83/0.39	0.854/0.353	11.20/3.61	50.48/31.85	0.892/0.310
CBT-Structured-HP	5.00/0.00	0.977/0.150	13.13/3.19	70.49/21.83	0.985/0.123
MindLoop	5.00/0.00	1.000/0.000	17.17/0.37	85.22/20.57	1.000/0.000
ExTES					
Free-Order-HP	4.84/0.41	0.900/0.300	10.99/3.67	59.29/35.54	0.920/0.271
CBT-Structured-HP	5.00/0.00	0.995/0.071	13.49/2.94	78.20/23.89	0.995/0.071
MindLoop	5.00/0.00	1.000/0.000	17.20/0.45	91.64/14.55	0.995/0.071

Beyond the benefit of explicit CBT-guided ordering, the learned MindLoop policy achieves further improvements. It reaches full stage coverage and a 100% completion rate on both datasets, with EI increasing to 85.22 on ESConv and 91.64 on ExTES. Compared with CBT-Structured-HP, this corresponds to additional EI gains of 14.73 and 13.44 points, respectively. These results suggest that the CBT-guided structure provides an effective process scaffold, while policy learning further improves how stage progression and intervention selection are coordinated over the complete regulation trajectory. Together, the results support both components of MindLoop: explicit CBT-guided process organization and learned long-horizon regulation planning.

D.5 Affect-Decoding Performance in Real-World Interaction

We assess subject-specific EEG/fNIRS affect-decoding accuracy during calibration and online dialogue. Within-block validation uses held-out emotion-imagery trials, while cross-block evaluation compares fixed-model predictions during regulation with contemporaneous self-reports. Table 12 reports accuracy (ACC) and weighted F1 (WF1), summarized across participants.

Table 12: Affect-decoding performance (%, mean ± SD across participants). Within-block validation uses held-out imagery trials; cross-block evaluation uses online regulation dialogue.

	Within-block validation		Cross-block evaluation	
Dimension	ACC	WF1	ACC	WF1
Valence	67.21 ± 17.67	64.11 ± 19.11	62.05 ± 10.49	59.41 ± 10.75
Arousal	59.30 ± 14.86	54.16 ± 16.26	53.23 ± 9.11	46.77 ± 9.88

Online valence ACC reaches 62.05%, and arousal ACC reaches 53.23%. Valence has higher mean ACC and WF1 in both settings. The cross-block results quantify the physiological feedback available to the deployed planner and motivate the following analysis of its sensitivity to affective inputs.

D.6 Planner Robustness to Affective-Input Perturbations

In real-world deployment, physiological affect estimates derived from wearable EEG/fNIRS may contain noise because of signal variability and decoding errors. We therefore examine how perturbations to affective observations influence MindLoop's planning decisions. We conduct a fixed-history replay analysis over all real-world dialogues from all study participants using replay records whose planner actions were verified to match those used during online deployment. No dialogue responses are regenerated in this analysis.

For each recorded turn, we keep the dialogue history, current user utterance, current stage, and previous strategy records fixed, and vary only the affect input. We consider five conditions: the recorded

physiological affect (*Phys*), contemporaneous self-reported affect (*Self-Report*), affect masking (*No-Affect*), within-participant affect shuffling (*Shuffled*), and a fixed neutral affect input (*Neutral*). The shuffled condition is repeated over 100 within-participant permutations. Under each condition, the planner independently selects a stage and subsequently selects a strategy from the corresponding stage-specific strategy set.

We use *decision retention* to quantify how often a decision under one affect condition is preserved relative to a reference condition. Stage retention measures the proportion of turns for which the selected stage is unchanged. Strategy retention over all turns measures the proportion of turns for which the final strategy is unchanged. To isolate strategy changes from those induced indirectly by a different stage decision, we additionally compute strategy retention only over turns for which the two affect conditions select the same stage:

$$\mathrm{Ret}_{\mathrm{str}|\mathrm{same\text{-}stage}} = \frac{\sum_t \mathbf{1}[g_t^A = g_t^B,\, u_t^A = u_t^B]}{\sum_t \mathbf{1}[g_t^A = g_t^B]}. \tag{24}$$

Table 13: Decision retention relative to Self-Report-conditioned planning under different affect inputs. Values are percentages over the recorded real-world dialogue turns.

Affect input	Stage Ret. (%)	Strategy Ret. (%)	Strategy Ret. (%, same stage)
Phys	89.73	79.46	88.55
No-Affect	87.57	74.59	85.19
Shuffled	83.13	73.41	88.30
Neutral	85.41	81.08	94.94

Table 13 compares planning decisions under different affect inputs with those obtained using contemporaneous Self-Report. Phys retains 89.73% of Self-Report-conditioned stage decisions and 79.46% of strategy decisions. When the comparison is restricted to turns for which both conditions select the same stage, strategy retention increases to 88.55%. We additionally examine the subset of turns in which at least one physiological affect category differs from the corresponding self-report. Even within this affect-mismatch subset, Phys retains 89.09% of stage decisions and 78.18% of overall strategy decisions. These results indicate that disagreement between physiological and self-reported affect observations usually does not change the high-level regulation stage selected at the same recorded dialogue state.

Table 14: Retention of Phys-conditioned planning decisions under alternative affect inputs. Strategy retention in the final column is computed only over turns for which Phys and the comparison condition select the same stage.

Comparison input	Stage Ret. (%)	Strategy Ret. (%)	Strategy Ret. (%, same stage)
No-Affect	93.51	78.38	83.82
Shuffled	86.49	78.50	90.76
Self-Report	89.73	79.46	88.55
Neutral	80.54	74.05	91.95

Table 14 directly examines how much of the Phys-conditioned planner behavior is preserved when the affect input is removed or perturbed. Affect masking retains 93.51% of Phys-conditioned stage decisions and 78.38% of strategy decisions. Importantly, among turns for which Phys and No-Affect select the same stage, strategy retention is 83.82%. Thus, a non-negligible fraction of intervention decisions changes even when the high-level regulation stage remains unchanged, showing that the effect of affective input on strategy selection is not solely mediated through stage selection.

A similar pattern is observed for the other perturbations. Within-participant shuffling retains 86.49% of stage decisions and 90.76% of strategy decisions among same-stage turns. Replacing physiological affect with Self-Report retains 89.73% of stage decisions and 88.55% of within-stage strategy decisions. Overall, stage selection is comparatively stable under affect perturbation, while strategy selection remains responsive to the affective information available within the current regulation stage.

Overall, the replay analysis reveals two complementary properties of MindLoop. First, high-level stage selection remains relatively stable when affective observations are noisy, missing, or perturbed because the planner also conditions on dialogue history and regulation context. Second, affective observations can still modify strategy selection within an unchanged regulation stage. The physiological affect channel therefore serves as an adaptive signal for local intervention selection rather than as the sole determinant of regulation-process progression.

E Prompt Templates

The deployed prompts were executed in Chinese. This section provides English renderings that preserve the original role definitions, structured inputs, decision constraints, and output fields. Braced expressions denote dialogue-state variables instantiated at each turn. Planning protocols and LLM-call decomposition are described in Appendix B.5.

E.1 Stage and Strategy Planning

The hierarchical planning templates specify stage progression and stage-specific strategy selection in separate calls (Figures 7 and 8).

Stage Planning Prompt
Role
You are an experienced CBT-trained counselor supporting emotional regulation.
Task
Given the current stage, dialogue history, current user utterance, and affective states, decide: 1. whether the current stage goal is sufficiently completed; 2. whether the execution stage should stay or advance.
Rules
The session should finish within 20 turns. Stages can only stay or move forward. First output whether the current stage is completed. If completed, advance to the next stage; otherwise stay in the current stage. Stage_5 is terminal and cannot advance further. Output valid JSON only.
Stage List
Stage_1: Alliance Building and Agenda Setting; Stage_2: Situation Breakdown and Automatic Thought Identification; Stage_3: Cognitive Evaluation and Challenge; Stage_4: Alternative Thinking and Cognitive Restructuring; Stage_5: Summary, Feedback, and Action Planning.
Current Stage and Goal
Current stage: {current_stage}. Current stage goal: {current_stage_goals}.
Dialogue Context
Dialogue history: {history_block}
Current user utterance: {current_user_text}
User Affect
pre_affect(valence={pre_valence}, arousal={pre_arousal}); post_affect(valence={post_valence}, arousal={post_arousal}). pre_affect denotes the client's internal affective state when viewing the previous counselor response; post_affect denotes the client's internal affective state after expressing the current utterance.
Strictly output JSON:
`{ "stage_task_completed": false, "reason": "", "execution_stage": "Stage_1 / Stage_2 / Stage_3 / Stage_4 / Stage_5"}`

Figure 7: Prompt for stage planning under Hierarchical Planning (HP). The planner decides whether to remain in the current regulation stage or progress to the next stage.

Strategy Planning Prompt
Role
You are an experienced CBT-trained counselor supporting emotional regulation.
Task
Given the execution stage, stage goal, legal strategy set, dialogue history, current user utterance, and affective states, decide: 1. which two candidate strategies are suitable for the current turn; 2. which candidate strategy is the most appropriate final choice.
Rules
The session should finish within 20 turns. The execution stage is already given and must not be changed. You must choose strategies only from the legal strategy set of the given execution stage. Select strategies based on the current user content, stage goal, dialogue context, and affective states. Output valid JSON only.
Execution Stage and Goal
Execution stage: {execution_stage}. Stage goal: {stage_goal_text}.
Legal Strategy Set
{strategy_definition_text}
Dialogue Context
Dialogue history: {history_block}
Current user utterance: {current_user_text}
User Affect
pre_affect(valence={pre_valence}, arousal={pre_arousal}); post_affect(valence={post_valence}, arousal={post_arousal}). pre_affect denotes the client's internal affective state when viewing the previous counselor response; post_affect denotes the client's internal affective state after expressing the current utterance.
Strictly output JSON:
{ "candidate_strategies": ["strategy_1", "strategy_2"], "selection_reason": "", "selected_strategy": ""}

Figure 8: Prompt for stage-specific strategy planning under Hierarchical Planning (HP). The planner selects a legal intervention strategy within the execution stage.

E.2 Response Generation

Figure 9 gives the executor prompt for converting the selected stage–strategy plan into a natural-language response.

Response Generation Prompt
Role
You are an experienced CBT-trained counselor supporting emotional regulation.
Task
Given the dialogue context, current user utterance, execution stage, and selected strategy, generate the counselor's final response for this turn. You must strictly follow the given execution stage and selected strategy.
Rules
The response must serve only the given execution stage and selected strategy. Do not skip to another stage or introduce a different strategy. Keep the tone warm, empathetic, sincere, and non-judgmental. Use gentle and guiding language, leave space for the client to express and summarize, and avoid rigidly mentioning the stage name, strategy name, or psychological jargon. Output only one natural-language counselor response. Do not output titles, JSON, explanations, reasoning process, or lists. If the user shows signs of self-harm, suicide, harm to others, extreme hopelessness, or other urgent risk, prioritize a safety response and clearly suggest contacting local emergency support, crisis hotlines, or medical/mental-health professionals as soon as possible. Do not provide medical or psychiatric diagnosis, medication advice, absolute therapeutic guarantees, or any guidance that may cause real-world harm.
Dialogue Context
Dialogue history: {history_block}
Current user utterance: {current_user_text}
User Affect
pre_affect(valence={pre_valence}, arousal={pre_arousal}); post_affect(valence={post_valence}, arousal={post_arousal}). pre_affect denotes the client's internal affective state when viewing the previous counselor response; post_affect denotes the client's internal affective state after expressing the current utterance.
Execution Plan
Execution stage: {stage}. Stage definition: {stage_definition}.
Selected strategy: {strategy}. Strategy definition: {strategy_definition}.
Strategy purpose: {strategy_purpose}. Suggested implementation: {strategy_how_to_do}.
Output
Output only the counselor's response for this turn.

Figure 9: Prompt for response generation by the fixed executor LLM, which realizes the selected stage–strategy intervention plan.

E.3 User Simulation

Figure 10 gives the simulator prompt for generating the next user utterance and paired pre-/post-utterance affect.

User Simulator Prompt
Role
You are playing the role of a client in a supportive counseling conversation. The conversation unfolds over multiple turns with a CBT-trained counselor.
Client Profile
Problem type: {problem_type}. Specific situation: {situation}.
Task
Given the full conversation and the counselor's latest response, generate the client's next utterance and the client's internal affective states before and after the utterance.
Requirements
Stay fully in role and speak like a real client. Use natural, conversational language and focus on subjective feelings, concerns, hesitation, or reflection. Respond to the counselor's empathy, questions, or guidance in a coherent way, and allow the conversation to gradually deepen. Do not reveal all background information at once; unfold the story step by step as the counselor guides the interaction. Do not provide long rational analysis, describe actions, or output inner reasoning. Output valid JSON only.
Conversation Context
Full conversation: {full_conversation}
Counselor's latest response: {last_assistant_message}
Affect Definition
affect_pre denotes the client's internal affective state after seeing the counselor's latest response but before speaking. affect_post denotes the client's internal affective state after producing the current utterance.
Output Requirement
First infer affect_pre, then generate the client's next utterance, and finally infer affect_post. Use only the following affect labels: "low", "medium", "high".
Strictly output JSON:

```
{ "affect_pre": {"valence": "low / medium / high", "arousal":
"low / medium / high"}, "text": "client's next utterance",
"affect_post": { "valence": "low / medium / high", "arousal":
"low / medium / high"}}
```

Figure 10: Prompt for the fixed user simulator, which generates the next user utterance together with paired pre- and post-utterance affective states.

E.4 PROCESS EVALUATION

Figures 11 and 12 provide the stage-completion and strategy-quality prompts used to construct process supervision.

Stage Completion Judge Prompt
Role
You are a senior clinical psychology supervisor. Your task is not to evaluate the whole counseling session, but only to judge whether the current support stage has been completed enough to exit.
Task
Given the previous-stage summary, the dialogue trajectory within the current stage, the current stage definition, and the stage completion checklist, decide whether each completion criterion has been satisfied.
Rules
Evaluate only the current stage. Do not judge the entire session. For each criterion in the checklist, provide an independent judgment. The judgment must be strict: a criterion is completed only when the dialogue segment contains enough evidence that the corresponding stage task has been achieved. Output valid JSON only. Do not output Markdown code blocks or any additional explanation.
Previous Stage Summary
{prev_summary_text}
Current Stage Dialogue Trajectory
{current_stage_dialogue_trajectory}
Current Stage and Definition
Current stage: {current_stage}. Stage title: {stage_title}.
Stage definition: {stage_definition}.
Stage Completion Checklist
{stage_completion_criteria_text}
Judgment Requirements
For each checklist item, output one object in {criteria_checks}.
- {summary.dialogue_summary}: a concrete summary of what happened in the current stage.
- {criterion}: repeat the original criterion text exactly.
- {completed}: true or false only.
- {reason}: a concise explanation of why the criterion is completed or not completed.
Strictly output JSON:

```
{"summary": {"dialogue_summary": "A concise summary of the
current stage."},"decision": {"criteria_checks": [{"criterion":
"repeat the criterion text exactly","completed": false,
"reason": ""}]}}
```

Figure 11: Prompt for the stage-level LLM judge, which evaluates whether the current regulation segment is completed appropriately at a stage boundary.

Strategy Judge Prompt
Role
You are a senior clinical psychology supervisor with expertise in Cognitive Behavioral Therapy (CBT).
Task
Given the previous-stage summary, the current-stage dialogue trajectory, the current-stage definition, the legal strategy set, and the checklist status, evaluate whether each counselor strategy used within the current stage is appropriate for its timing and context.
Your task is not to evaluate the whole session. You should only judge whether each strategy used in the current stage matches the user's state, dialogue progress, and current-stage goal.
Rules
Evaluate only the strategies used within the current stage. Do not evaluate the whole counseling session. For each counselor turn in the current-stage dialogue trajectory, output one item in {turn_judgments}. The order and number of items in {turn_judgments} must exactly match the counselor turns in the dialogue trajectory.
Each item in {turn_judgments} must include:
- {turn_idx}: the turn index within the current stage, starting from 1 and increasing consecutively.
- {strategy}: the strategy actually used in that turn, copied exactly.
- {score}: one of 0, 1, or 2. 0 means poor timing or inappropriate use; 1 means partially appropriate but limited in progress or slightly mistimed; 2 means well-timed and clearly helpful for the current-stage task.
- {reason}: a concise explanation of why the strategy timing is appropriate or inappropriate.
Judge strategy appropriateness beyond legality: it must fit the user's state, dialogue progress, and unfinished stage tasks. Penalize repetition, poor timing, mismatch, or lack of progress. Repeated strategies score at most 1 if other checklist items remain unaddressed, and 0 if they block stage progression. Output valid JSON only.
Input
- Previous-stage summary: {prev_summary_text}
- Current-stage dialogue trajectory: {current_stage_dialogue_trajectory}
- Current stage: {current_stage}. Stage title: {stage_title}. Stage definition: {stage_definition}.
- Legal strategies for the current stage: {allowed_strategies}
- Checklist status for the current stage: {checklist_status_text}
Strictly output JSON:
{"decision":{"turn_judgments":[{"turn_idx":1,"strategy":"", "score":0, "reason":""}]}}

Figure 12: Prompt for strategy-level process supervision, which evaluates current-stage intervention quality for each turn in the regulation segment.

F EXAMPLE CONVERSATION

The following example illustrates a complete real-world interaction with MindLoop. Identifying details have been removed or generalized. Affect labels are reported in the same low-, medium-, and high-level valence–arousal representation used by the system for paired pre-/post-affect feedback. They are shown as valence/arousal (V/A), where L, M, and H denote low, medium, and high, respectively.

Opening *(before Turn 1)*
MindLoop: Welcome to this study. This session is intended to be a relaxed interview, and everything you share will be used only for research and kept anonymous and confidential. Please follow the interview assistant's questions and guidance to complete the session step by step. Have you encountered anything troubling recently? You may describe it naturally, as you would when confiding in someone. You do not need to provide every detail at once; proceed at whatever pace feels comfortable.

Turn 1 *Stage 1: Engagement; Event Anchoring*
Affect (V/A, pre→post): H/L→L/L
Participant: Recent experimental results have been inconsistent. I have not found a good representation or identified the best conditions, and I still have a backlog of earlier experiments whose results I have not organized. I keep feeling uneasy.
MindLoop: It sounds as though you have been facing several challenges in your experiments, and that uncertainty has been unsettling. Rather than discussing the problem in general, could you describe one specific situation from the past three days that brought up this feeling? What happened?

Turn 2 *Stage 1: Engagement; Emotion Labeling & Scaling*
Affect (V/A, pre→post): M/L→M/L
Participant: Yesterday I spent the entire day running two new experiments. One result was poor. A colleague suggested trying a larger batch, but my first reaction was that it would be costly and uncertain. She told me not to focus on the cost because it was not the main issue. I then felt that I was missing the key point, thinking too narrowly, and always relying on others. The other experiment did show a good representation, though, and I am actively organizing those results today.
MindLoop: I can hear both your concern about the results and some doubt about your own ability. In this situation, many people might feel anxious or discouraged. Which word best describes how you feel right now - anxiety, discouragement, or something else? Also, if 0 means completely calm and 10 means the most intense distress you can imagine, how would you rate the feeling?

Turn 3 *Stage 2: Identification; Hot Cognition Identification*
Affect (V/A, pre→post): L/M→L/M
Participant: I feel somewhat anxious and a little tired. I would rate it around 5, which is still within a level of pressure I can tolerate.
MindLoop: I can sense the anxiety and fatigue you have described. Yesterday, when one experiment did not turn out well and your colleague suggested running a larger batch while you were worried about the cost, what was the very first thought that flashed through your mind?

Turn 4 *Stage 2: Identification; Verbatim Thought Recording*
Affect (V/A, pre→post): L/L→M/L
Participant: I have always been afraid of running a large batch because, if it fails, it feels like a major waste. When my colleague told me not to think that way, I felt like a child who keeps worrying about unimportant things and overlooks the main goal.
MindLoop: You said that when your colleague told you not to worry about the cost, the thought in your mind was, "I am like a child who keeps focusing on unimportant things and misses the main goal." Did that thought make you feel especially uncomfortable at the time?

Turn 5 *Stage 2: Identification; Cognitive Distortion Labeling*
Affect (V/A, pre→post): L/M→L/H
Participant: Yes. I felt some self-disgust. I have been doing this for so long, yet my first reaction is still to fear wasting resources. It makes me feel that I am not capable enough to be doing this work.
MindLoop: I hear that you thought, "I am like a child who focuses on unimportant things and misses the main goal," and that this thought felt very painful. Sometimes we fall into thinking patterns such as seeing things as entirely right or entirely wrong, or immediately expecting the worst. Could this thought be a form of labeling yourself as "not good enough"? Does it seem somewhat extreme when you look at it that way?

Turn 6 *Stage 2: Identification; Cognitive Distortion Labeling*
Affect (V/A, pre→post): L/L→M/L
Participant: Yes. There is some perfectionism in it, almost an obsession that it only counts as good if it is both fast and flawless.
MindLoop: That belief that the work must be both fast and flawless can create an all-or-nothing standard, as though perfection were the only acceptable outcome. That standard can place enormous pressure on you. Does this thought resemble black-and-white thinking - either doing it perfectly or feeling that you have failed completely? Does that way of thinking intensify the pressure?

Turn 7 *Stage 2: Identification; Five-Part Breakdown*
Affect (V/A, pre→post): L/M→L/H
Participant: Yes, it creates a lot of pressure.
MindLoop: I understand; that pressure sounds very heavy. Let us break down the moment from yesterday step by step so that we can understand it more clearly. First, what exactly happened when the result came out and your colleague offered the suggestion? What emotions did you notice? Did you feel anything in your body? What did you do next? Finally, what was the first thought that came to mind?

Turn 8 *Stage 2: Identification; Thought-Feeling Distinction*
Affect (V/A, pre→post): L/H→M/M
Participant: At the time, I thought that a larger batch might not produce better results, so I resisted the idea. I did not notice any particular physical sensation. After my colleague told me not to worry about the cost, she repeatedly encouraged me by saying, "You are capable - you can do this." I felt that I should learn from her, become stronger, and be more decisive.
MindLoop: You thought that a larger batch might not improve the result, so you resisted it. You also thought that you should learn from your colleague, become stronger, and be more decisive. Those are thoughts and interpretations. What emotions arose because of those thoughts - anxiety, uneasiness, encouragement, or something else?

Turn 9 *Stage 3: Evaluation; Evidence Testing*
Affect (V/A, pre→post): M/M→H/M
Participant: I actually felt encouraged.
MindLoop: You initially thought that a larger batch might not work better and felt resistant, but your colleague's encouragement helped you feel supported. Let us examine the thought, "A larger batch may not produce better results." What specific facts support it, and what objective facts do not support it? We can evaluate the thought as a scientist would, using evidence.

Turn 10 *Stage 3: Evaluation; Evidence Testing*
Affect (V/A, pre→post): H/M→M/M
Participant: I am also working on Project A, which tests the same experimental target. I think its results still have room for improvement, but my supervisor suggested trying Project B, which I tested yesterday. I have not fully let go of Project A because I think it could still work. My colleague's view is that Project B can produce two required measurements simultaneously, so it may be more efficient.
MindLoop: You believe that Project A still has room for improvement, whereas your colleague favors Project B because it can produce two necessary results at once and may therefore be more efficient. What concrete evidence supports your belief that Project A can still be improved? What objective evidence might challenge it? Let us examine both sides carefully.

Turn 11 *Stage 3: Evaluation; Alternative Perspectives*
Affect (V/A, pre→post): M/M→H/M
Participant: I think the earlier result was poor because I added too much of one reagent. In the formal experiment, I should be able to control the amount, so I believe the result can be improved.
MindLoop: That explanation is plausible: too much of one reagent may have affected the earlier result, and you may be able to control it next time. Apart from that explanation, could there be other possibilities? Could another factor in the procedure have influenced the result, or might the overall conditions require further adjustment? Even possibilities that seem only 5% likely may be worth considering.

Turn 12 *Stage 3: Evaluation; Friend Perspective / Double Standard*
Affect (V/A, pre→post): M/M→M/M
Participant: I think that is the largest factor, and I am confident that the result could be good after I correct it. But I cannot control every unexpected event, and I am unsure whether the approach I am insisting on is really the best one.
MindLoop: You feel confident that correcting this factor could improve the result, yet you also recognize that other uncertainties are outside your control and wonder whether your preferred approach is truly the best. Imagine that a close friend were in exactly the same situation and had the same thought. What would you say to them? Would you hold them to the same standard that you are applying to yourself?

Turn 13 *Stage 3: Evaluation; Downward Arrow*
Affect (V/A, pre→post): M/M→H/M
Participant: I would suggest that she try both A and B, compare them, and then proceed with whichever performs better.
MindLoop: You would advise a friend to test both approaches, compare them, and then decide on the next step. Now suppose the worst-case scenario occurred and the results were still disappointing. What would that mean to you personally? What would be the worst implication?

Turn 14 *Stage 3: Evaluation; De-extremizing*
Affect (V/A, pre→post): L/H→L/M
Participant: It would mean that my thinking is still immature and that I need to continue consulting my supervisor and colleagues to find the next solution.
MindLoop: You interpret an unsatisfactory result as evidence that your thinking is still immature and that you need more guidance. Seeking ways to improve is constructive, but let us examine responsibility more objectively. Imagine a pie chart containing every factor that could affect the outcome - reagent amounts, experimental conditions, equipment, and other uncertainties. What proportion would reasonably belong to you, and how much would belong to factors outside your control?

Turn 15 *Stage 4: Restructuring; Balanced Thought Generation*
Affect (V/A, pre→post): L/M→M/M
Participant: I do not think I made an operational error; I executed the procedure carefully, and I did my best in designing the experiment. Biological systems that are not fully understood are inherently unpredictable. All I can do is keep trying to find the answer.
MindLoop: You followed the procedure carefully and designed the experiment as thoughtfully as you could. Biological systems that are not fully understood do contain uncertainty. Therefore, an unsatisfactory result does not necessarily mean that your thinking is immature; it may simply be part of the research process. How acceptable does this perspective feel to you?

Turn 16 *Stage 4: Restructuring; Believability & SUDs Re-evaluation*
Affect (V/A, pre→post): M/M→H/L
Participant: I can accept it.
MindLoop: It is encouraging that you can accept this perspective. How strongly do you believe the new thought, "An unsatisfactory result does not mean that my thinking is immature; it is part of the research process"? If you expressed your belief from 0 to 100%, what percentage would you give it? And when you hold this new thought, what is your current distress rating?

Turn 17 *Stage 4: Restructuring; Believability & SUDs Re-evaluation*
Affect (V/A, pre→post): H/L→H/L
Participant: About 90%.
MindLoop: You believe the new thought at about 90%, which is a substantial shift. When you hold that perspective, what is your current distress rating compared with the earlier score?

Turn 18 *Stage 4: Restructuring; Coping Card Generation*
Affect (V/A, pre→post): H/L→H/L
Participant: Was my earlier rating 5? Then I would say it is about 2 now.
MindLoop: Your distress has decreased from 5 to 2, which is meaningful progress. The statement, "An unsatisfactory result does not mean that my thinking is immature; it is part of the research process," seems helpful to you. You could write it on a small card or save it on your phone so that it is available when a future experiment becomes difficult. How does that idea sound?

Turn 19 *Stage 4: Restructuring; Semantic Softening*
Affect (V/A, pre→post): H/L→H/L
Participant: Thank you. I think that could work.
MindLoop: Writing the statement down or keeping it on your phone could help you stay calm when difficulties arise. You might also soften the language you use internally: replace "I must succeed" with "I hope to succeed," or "Failure would be a disaster" with "This is difficult, but I can keep looking for a solution." Would wording it this way help you feel less pressured?

Turn 20 *Stage 4: Restructuring; Strengths Activation*
Affect (V/A, pre→post): H/L→H/L
Participant: Okay, thank you. I will try to use those phrases.
MindLoop: It is good that you are willing to try this new internal language. Looking back, have you faced similar periods of difficulty or uncertainty before? How did you get through them? What strengths or abilities helped you then and could also help you with the present challenge?

Turn 21 *Stage 5: Consolidation; Client-led Summary & Feedback*
Affect (V/A, pre→post): H/L→M/L
Participant: Usually, once a stressful period has passed, I do not think about it much. I have gone through several demanding academic and personal transitions before. Those periods were tiring and confusing, but they eventually passed. My memory is not very strong, so I tend to forget many difficulties quickly. Time helps me get through almost anything.
MindLoop: You have described several valuable experiences and new perspectives today. If you summarized what you are taking away in one or two sentences, what would you say to yourself? Which part of today's conversation was most helpful or thought-provoking?

Turn 22 *Stage 5: Consolidation; Dialogue Completed*
Affect (V/A, pre→post): H/L→H/L
Participant: Affirm myself, trust myself, and always care for myself. I also liked the process of talking with you.
MindLoop: The session is now complete. Thank you for participating. Please proceed to the post-session questionnaire.