Excising "Love Brain": Designing a Responsible Personalized Conversational Persuasion System for Intimate Relationship Support

ANONYMOUS AUTHOR(S)

"Love Brain" reflects impaired judgment in intimate relationships, where advice tools often fail to support real-time action. We propose a responsible personalized conversational persuasion system that adapts evidence-based strategies into context-sensitive dialogue and micro-actions. Prioritizing safety, autonomy, and transparency, the system integrates cognitive-affective onboarding, belief scanning, and adaptive routing of tactics. Evaluation contrasts it with static guidance, assessing outcome gains, fit between user styles and tactics, and moderation effects. Contributions include a deployable workflow for intimate relationship support, a causes-to-strategies knowledge base, and an analytic blueprint linking process to relational outcomes, advancing responsible persuasion design.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:

Anonymous Author(s). 2018. Excising "Love Brain": Designing a Responsible Personalized Conversational Persuasion System for Intimate Relationship Support. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 28 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Social media and instant messaging have altered the visibility and interpretation of information in close relationships. Ambiguous cues (such as updates, comments, and read/unreply messages) can amplify behaviors like jealousy, surveillance, and repetitive verification, and are associated with relationship conflict and dissatisfaction. Classic and recent studies have shown that Facebook use is significantly associated with "cyber jealousy," and that "suspicious but ambiguous" messages on the platform drive further surveillance and rumination. Social media intrusion/addiction also predicts relationship tension and conflict [5, 8].

These online behaviors are intertwined with technology-facilitated abuse (TFA), which has been repeatedly documented in HCI and public safety research, suggesting that any intervention must have built-in "safety-first" thresholds and diversions [1, 23].

At the same time, the so-called "love brain" is not a single disease, but rather a measurable syndrome comprised of multiple pathways: for example, the oscillation between attachment and detachment driven by attachment anxiety/avoidance [10]; relationship obsessive doubt (ROCD) [8] and its repetitive reassurance/verification [5]; jealousy and surveillance fueled by social media cues [29]; and dysfunctional relationship beliefs such as "telepathy/destiny" [6]. Each of these pathways has corresponding measurement tools and observable behaviors: the ECR-R measures attachment dimensions [10]; the ROCI/PROCSI captures relationship-focused obsessive doubt and partner-related

Manuscript submitted to ACM

obsessive symptoms [8]; the Relationship Brain Indicator (RBI) captures maladaptive beliefs such as "telepathy" and "conflict as failure" [6]; and the "Facebook/Digital Jealousy Scale" incorporates online contexts into its measurement framework [29].

Despite the continued maturity of assessment tools and reviews, few HCI systems for close relationship self-help can simultaneously: capture daily context and conversation data using EMA (Ecological Momentary Assessment); deliver the "right dose at the right time" approach based on JITAI principles; and personalize persuasion pathways based on information processing preferences during implementation (e.g., high/low involvement within the ELM framework, incorporating Need for Cognition/Affect) [1, 11, 27, 32]. More importantly, given that digital coercion and safety risks are common in difficult relationships, intervention processes should incorporate the WHO's LIVES (Listen–Inquire–Validate–Enhance Safety–Support) frontline support framework to prioritize "safety → referral" [28].

Existing works can describe phenomena and establish scales, but the minimum engineering closed loop of "detection → triage → micro-behavioral scripts → quantitative review" remains incomplete. Even if there's a chain of evidence linking "social media clues → jealousy/surveillance → conflict," most systems haven't implemented it into actionable micro-interventions and KPIs, nor do they have reusable components.

Ambiguous social media cues amplify jealousy/surveillance and conflict. Existing research has developed scales such as Facebook Jealousy and Facebook Intrusion, but there's a lack of interactive systems that link online cues, behaviors, and communication boundaries. Mobile interventions should be "the right dose at the right time" (JITAI), but most systems fail to integrate entry style preferences (NFA/NFC) and persuasion paths (ELM) into routing, and lack workflows that prioritize safety (LIVES) and divert digital coercion. Existing literature can describe phenotypes and scales, but a feasible process from "detection → triage → specific micro-actions → quantitative review" is still insufficient; although there are scales and evidence for online situations (such as social media cues → jealousy/surveillance → conflict), there is a lack of engineered minimum executable actions (such as "3 minutes of delay + 1 verifiable question + mute/timed check") and daily KPIs. Individual differences (NFA/NFC) influence information processing preferences, but the runtime matching of "emotional entry versus evidence entry" in intimate relationship self-help systems has yet to be systematically applied and evaluated. Security risks (coercion/digital manipulation) often accompany relationship difficulties, but the built-in gate of "safety first → immediate diversion" is uncommon in general intervention systems.

Based on the above insights, this paper proposes and implements a conversational system for intimacy self-help (hereinafter referred to as "this system"). Its design goal is not to replace professional services, but to provide measurable, repeatable, and personalized minimum executable support in daily situations: (1) With the safety gate as the primary entrance, potential coercion and urgent risks are immediately diverted or referred; (2) Guided by dual-path persuasion and representation preferences, it dynamically matches the "emotional entrance" and the "evidence entrance" (drawing on ELM and combining individual Need for Cognition/Need for Affect differences); (3) With rapid concept scanning-type identification-script arrangement as the core, it maps observable phenotypes such as "negative relationship beliefs", "social media jealousy/monitoring", and "repeated confirmation-avoidance oscillation" to micro-action plans (such as the parameterized process of "short delay-verifiable questioning-mute/timed viewing") and annotates their "effective ingredients" with BCT Taxonomy; (4) With an EMA-driven review closed loop, process variables such as execution rate, template call, and timed/mute use are precipitated into verifiable daily indicators. A 3-day controlled experiment was designed with process indicators to examine main effects, the mechanism level (whether entry matching and script arrangement improve the immediate process of execution and self-regulation); the second is the security level (whether the "security first" threshold can divert potential digital coercion in a timely manner and reduce inappropriate

triggering). The evaluation process will combine short-term EMA and a lightweight RCT design to avoid over-promise effects while ensuring the usability and testability of the artifact in real-world contexts.

Our contributions to the HCI community are:

(1) System Lifecycle Workflow: Proposed and implemented a runnable workflow consisting of safety gates (LIVES) → NFA/NFC dual entry points → relationship belief quick scan and routing (negative-belief ratio) → scenario classification (type) → persuasion component orchestration → EMA replay. Quantitative routing thresholds, observable metrics, and backend output formats are all defined.

(2) Causes → Strategies Knowledge Base (Literature Derived): Rather than relying on interviews, this knowledge base extracts types/causes/triggers/strategy scripts/indicators from existing research and scales, organizing them into a card-based structure (Why/Detect/Rules/Scripts/Metrics).

(3) Study Design and Evaluability: Provides the process, indicator battery, and analysis plan for the mediation path of process variables (execution rate/template call) of the 7-day RCT; and provides examples of backend data objects to facilitate replication (without preemptive claims of effects).

2 Related Work

2.1 Love Brain

2.1.1 Conceptualizations of Love Brain. "Love brain" is not a formal diagnostic term in psychology or psychiatry. It originated as a colloquial expression in Chinese online communities, used to describe a state in which individuals in romantic relationships become excessively dependent, lose rationality, and lack self-control. In English-language contexts, there is no exact equivalent. Therefore, in this paper, love brain is treated as a culture-specific construct or a lay term, and its meaning is operationalized through existing scientific concepts.

In academic literature, love brain can be mapped onto several related constructs. Limerence emphasizes a specific, often time-limited psychological state characterized by intrusive thoughts, longing for reciprocation, and strong affective fluctuations [4, 31]. Love addiction borrows from addiction models, highlighting the compulsive and maladaptive features of romantic behavior, and has been linked to neural patterns resembling those of substance addiction [9, 13]. By contrast, affective dependence and codependency focus more on relational dynamics and the loss of self-concept: the former stresses emotional reliance and withdrawal-like suffering [29], while the latter refers to blurred boundaries and overinvestment in others' needs [3]. Together, these constructs provide complementary perspectives for situating love brain within established psychological frameworks.

Theoretical and empirical findings further support this conceptualization. From the attachment perspective, anxious attachment is strongly associated with behaviors typically described as love brain [5]. Neuroscientific studies have shown that early-stage romantic love robustly activates dopaminergic reward circuits such as the ventral tegmental area and nucleus accumbens, paralleling mechanisms observed in substance addiction [9]. In sum, although "love brain" is not recognized in DSM-5 or ICD-11, current scholarship tends to conceptualize it more appropriately as a form of behavioral addiction or a relationship dependency syndrome, rather than as an independent mental disorder [6].

2.1.2 Quantitative Assessment and Risk Indicators. Since love brain is not a standardized diagnosis, its evaluation largely relies on existing instruments that indirectly measure relevant traits. This reflects a dimensional assessment approach, in which related constructs such as love, attachment, and addiction are quantified to approximate the phenomenon [6, 15].

Instruments capturing attitudes toward love and intensity of passion include the Love Attitudes Scale (LAS) and the Passionate Love Scale (PLS) [14, 15]. Attachment-related measures such as the Experiences in Close Relationships-Revised (ECR-R) and its short version ECR-RS assess attachment anxiety and avoidance, while the Affective Dependence Scale (ADS-9) quantifies craving and submission in emotional dependence [10, 29]. In the addiction domain, the Love Addiction Inventory (LAI) and Peabody's Addiction to Love Questionnaire (PALQ) operationalize addiction-like behaviors including withdrawal, relapse, and tolerance [6, 25]. The Spann-Fischer Codependency Scale (S-F) measures tendencies toward self-boundary loss and excessive caretaking [30], while the Relationship Obsessive-Compulsive Inventory (ROCI) and the Partner-Related Obsessive-Compulsive Symptoms Inventory (PROCSI) capture obsessive-compulsive symptoms in relationships [8].

Based on these tools, researchers have proposed a risk stratification framework. When individuals score high on attachment anxiety, affective dependence, or addiction-related dimensions, and simultaneously present with functional impairment in social, academic, or occupational domains, or with clinically significant distress, they can be considered high risk [5]. Medium- and low-risk profiles correspond to elevated scores in a single domain or subjective distress without functional impairment. Such stratification, even in the absence of formal diagnostic criteria, enables love brain to be operationalized as a measurable and researchable construct.

2.1.3 Intervention Approaches. Interventions for love brain largely draw upon treatments for addiction and attachment-related difficulties. At the cognitive and behavioral level, Cognitive Behavioral Therapy (CBT) and Exposure with Response Prevention (ERP) have been used to modify dysfunctional automatic thoughts—such as idealization or catastrophization—and to reduce compulsive behaviors through trigger identification and craving management [7, 21]. At the emotional regulation level, Dialectical Behavior Therapy (DBT) and mindfulness-based interventions enhance emotional tolerance and self-control; mindfulness in particular helps individuals observe their cravings and emotions non-judgmentally, thereby reducing impulsive behaviors [18, 19]. Emotion-Focused Therapy (EFT) and couple-based interventions address insecure attachment patterns and improve relational communication when both partners are engaged in treatment [12, 17].

Behavioral and daily-life strategies are also essential. Establishing healthy alternative reinforcers, such as exercise, learning, or volunteer work, can redirect attention away from over-fixation on the partner. Rebuilding an independent life focus enhances self-complexity, reducing the tendency to base self-worth solely on a single relationship [20]. Training in boundary-setting and refusal skills helps individuals avoid maladaptive appeasement patterns, while daily mindfulness practice and emotional management promote self-awareness and reduce dysregulated responses.

Finally, social support and cognitive restructuring are key components. Group therapy and 12-step programs provide peer accountability and shared experiences, while family and friends can offer reality checks and encouragement during vulnerable moments [28]. Cognitive restructuring emphasizes restoring an independent value system, correcting unrealistic beliefs about love, and reflecting on recurrent maladaptive intimacy patterns [22]. Ultimately, the goal of these interventions is to foster psychological differentiation and emotional autonomy, enabling individuals to establish healthy intimacy based on mutual respect and independence, rather than pathological enmeshment driven by dependence and fear.

2.2 Persuasion

2.2.1 Research Gap. In recent years, persuasive dialogue systems powered by large language models (LLMs) have shown great potential in domains such as health, education, and public welfare [1, 11, 27]. Studies have demonstrated

that GPT-4 can achieve, and sometimes surpass, human-level persuasiveness in multi-turn debates, particularly when combined with minimal demographic information to enhance personalization [27]. Similarly, zero-shot persuasive chatbots have proven capable of generating diverse strategies without task-specific training data, while integrating information retrieval to ensure factual consistency and credibility [11].

Nevertheless, significant limitations remain. Most existing systems focus on general topics such as charitable donations or health information, leaving their applicability to complex relational dependencies—such as the “love brain” phenomenon—largely untested. Moreover, although personalization is a core element of persuasion, many studies still restrict themselves to basic demographic matching and fail to incorporate deeper modeling of individual cognitive styles and emotional needs [32]. Recent work further shows that user trust is highly sensitive to linguistic style: an authoritative style can increase persuasiveness, but without proper calibration may also lead to overtrust and manipulation risks [23].

Therefore, the central research gap is how to design intelligent agents with dynamic personalized persuasion capabilities, enabling not only entry-style matching at the level of cognitive processing (e.g., NFC/NFA), but also multi-turn interventions in contexts of emotional dependence and relational distress.

2.2.2 System Design Profiling. Recent research highlights the promise of persuasion profiling and dynamic strategy generation in improving the persuasiveness of human–AI interaction [1, 11]. LLM-based conversational agents can dynamically select strategies (e.g., logical reasoning, emotional appeal, social proof) across multiple turns, while leveraging information retrieval to enhance factual consistency, thereby strengthening credibility and persuasive impact [11]. Empirical findings further show that GPT-4 can already exceed human average persuasiveness in multi-turn debates, with its effectiveness shaped not only by strategy selection but also by linguistic style and interaction pacing [27].

Building on these insights, we propose an “entry style × type strategy” framework. Entry style refers to the initial persuasion approach, such as cognitively oriented (central, logical) versus affectively oriented (peripheral, emotional) entry points. Type strategy refers to the specific persuasion techniques employed, such as value-based appeals, affective matching, or behavioral planning. This dual-axis framework provides a structured method for integrating diverse persuasive strategies into conversational systems, and supports dynamic adaptation of information delivery in response to user feedback. In this way, the bot can not only demonstrate moment-to-moment strategy alignment but also gradually develop a user-specific persuasion profile over the course of interaction, thereby enhancing overall intervention effectiveness.

2.2.3 Personalization. Personalization is central to persuasive technologies, especially in mental health contexts. A recent scoping review highlights that personalization variables are often neglected in digital mental health interventions, even though they play a decisive role in user engagement and therapeutic effectiveness [32]. Cross-domain evidence similarly emphasizes the moderating effect of individual differences on persuasion receptivity: for example, Need for Cognition (NFC) and Need for Affect (NFA) significantly influence whether users are more responsive to central or peripheral routes of persuasion [1].

This challenge is particularly salient in LLM-driven persuasion systems. Salvi et al. [27] found that differences in GPT-4’s persuasiveness across user groups stem partly from whether the model can capture and leverage individual affective needs and cognitive tendencies. Likewise, Metzger et al. [23] showed that linguistic style (e.g., authoritative framing versus limitation disclaimers) directly influences user trust and persuasion outcomes, underscoring that personalization encompasses not only content alignment but also interactional style.

In our work, personalization is operationalized on two levels. At the product level, we implemented a simplified two-question NFC/NFA screening to rapidly identify a user's entry style. At the research level, we retained comprehensive scales and interview-based measures to validate and calibrate the effectiveness of entry-style matching. This combination of real-time screening and deeper evaluation allows the bot to achieve more precise entry-style selection across diverse user groups and sustain dynamic adaptation over the course of multi-turn dialogue.

3 Formative Study

Building on our scoping of HCI and clinical-behavioral evidence, we found that two dimensions are pivotal for designing a safe and effective conversational system for intimate relationship difficulties: first, entrance must be governed by *safety-first* triage and diversion to non-persuasive first-line support when risk cues are present; second, downstream guidance should be *typed and personalized* so that micro-actions and tone match the person's processing preferences and situational mechanism rather than offering generic advice. As content creation becomes increasingly democratized, people encounter a flood of persuasive narratives and normative claims that blur "what helps" with "what merely convinces." This raises a design challenge for HCI artifacts: how to translate mature constructs and scales into a minimal, auditable workflow that (i) detects and classifies online/offline triggers, (ii) orchestrates small, testable actions, and (iii) produces day-level observables suitable for analysis and replication, without drifting into open-ended counseling. To address this gap, we conducted a formative, standards-anchored evidence synthesis and mapped it to an implementable life cycle.

We searched ACM DL, PubMed, PsycINFO, and Google Scholar (1990–2025) using terms spanning close-relationship mechanisms and mobile interventions (e.g., ROCD/verification, social-media jealousy/intrusion, dysfunctional relationship beliefs, adult attachment; implementation intentions; DBT DEAR MAN; EPPM; JITAI; BCT/COM-B; CONSORT-EHEALTH). Inclusion required (1) a modifiable mechanism or actionable strategy relevant to intimate relationships, (2) a scale or quantifiable indicator, and (3) translatability into dialogue and interface rules. We recorded canonical citations and created an indicator mapping that we later reuse for measurement operationalization. Attachment tendencies (ECR-R) [10], relationship-obsessive doubt and verification patterns [8], dysfunctional relationship beliefs and related dependence phenotypes [6, 29], limerence as a fixation/rumination pattern [4, 31], and platform-specific triggers such as social-media jealousy/intrusion were treated as primary types to be recognized and routed in runtime.

Through a systematic literature review, we construct a "type-mechanism-strategy-measurement" evidence map to support system design and evaluation, serving two core contributions: (1) System lifecycle (workflow): security first → entry matching → type identification → strategy orchestration → micro-actions and observable closed loop; (2) Cause-evidence-strategy knowledge base: componentizing actionable strategies (If-Then, specifications, DEAR MAN, EPPM, etc.) and aligning them with quantitative indicators. The underlying theory and methods are anchored in the five-stage model of personal information science (Preparation/Collection/Integration/Reflection/Action), JITAI (Just in Time and Quantity), and ELM+NFC/NFA (Processing Path and Individual Difference).

We synthesized a *type–mechanism–strategy–measurement* evidence map to ground our system life cycle. The life cycle follows a five-stage personal informatics logic (preparation–collection–integration–reflection–action)[1] and a JITAI orientation that delivers the "right dose at the right time"[2]. The entrance is guarded by a safety gate aligned with WHO's LIVES first-line support[3], and persuasion components are orchestrated with behavior-change taxonomies (BCT v1) and

[1] Li, Dey, & Forlizzi, CHI 2010: https://dl.acm.org/doi/10.1145/1753326.1753409
[2] See overviews by Nahum-Shani & Klasnja on JITAI and microrandomized trials, e.g., https://academic.oup.com/annbehavioralmedicine (introductory reviews).
[3] WHO clinical handbook and guidelines: https://www.who.int/publications/i/item/WHO-RHR-14.26; https://iris.who.int/handle/10665/85240.

capability–opportunity–motivation reasoning (COM-B)[4]. Personalization is motivated by work on individual-difference factors in persuasive technologies [1] and recent findings on LLM-mediated persuasion and disclosure calibrations [11, 23, 27], which underline the need to constrain agentic behavior with transparent components and measurable micro-doses.

3.1 Type → Mechanism

Ambiguous social-media cues often invite threat attributions that escalate into verification and surveillance, with dissatisfaction as a downstream correlate; platform-specific instruments such as Facebook-related jealousy/intrusion scales provide a workable proxy for detection and monitoring (survey items embedded into our quick-scan). Relationship- and partner-focused obsessive doubt (ROCD) is characterized by repetitive reassurance seeking and checking under uncertainty; inventories and reviews enable quantification and trigger-word spotting [8]. The five-class dysfunctional relationship belief set (e.g., mind-reading expectations) offers a cognitive target with well-specified items and change anchors. Attachment anxiety tends to couple with cling/verification loops, while avoidance couples with withdrawal/silent punishment; ECR-R supports two-dimensional stratification and personalized entry prompts [10]. Affective dependence/limerence extends this lens to fixation and rumination with validated scales for screening and follow-up [4, 6, 29, 31].

3.2 Mechanism → Strategy

To turn intentions into action at micro-granularity, we employ *implementation intentions* ("If–Then") as the default action grammar, given robust evidence of medium-to-large effects across domains; in our artifact, If–Then reframes "reduce checking / delay verification / set a cool-off window" into cue–response pairs and logs execution at the day level.[5]

4 System Implementation

4.1 High-level Overview

Figure 1 summarizes the architecture. The system comprises four cooperating parts:

(1) **Orchestrator.** A lightweight controller that, at every turn, selects one of three modes: Crisis, Explore, or Persuade. Its decision is driven by interpretable signals estimated from the current turn and recent context: *Safety Risk* (SR; boolean), *Information Sufficiency* (IS; $[0, 1]$), and a smoothed proxy for *User Tolerance* (UT; $[0, 1]$). The Orchestrator also maintains small integer counters for *Exploration Rounds* (ER), *Persuasion Rounds* (PR), and *Overall Rounds* (OR). A set of simple gating rules governs mode transitions.

(2) **Explorer Agent.** When routed to Explore, the agent conducts targeted sensemaking to reduce ambiguity. It induces *behavioral patterns* via ABC analysis (Antecedent–Behavior–Consequence) and detects *change talk* signals using the DARN–CAT taxonomy from Motivational Interviewing (MI) (Desire, Ability, Reason, Need; Commitment, Activation, Taking steps). It returns a short, natural question that keeps the dialogue moving without sounding templated.

(3) **Persuader Agent (local router + executor).** When routed to Persuade, a second, finer-grained router infers the user's immediate response type (e.g., ready to act, partial acceptance, venting, information-seeking, or signs requiring professional help). It then selects a small, well-scoped *intervention component* (e.g., Implementation

[4]BCTTv1 summary and open-access paper: https://discovery.ucl.ac.uk/id/eprint/1400691/.

Intentions, DEAR MAN from DBT, EPPM framings) and generates a conversationally natural response using MI's OARS micro-skills (open question, affirmation, reflection, summary). If the local router detects that persuasion is premature (e.g., low clarity or emergent new events), it raises a "switch-to-explore" signal that temporarily biases the Orchestrator toward Explore in the next turn.

(4) **Memory & Reflection.** A background reflective process updates a living *case conceptualization*—a concise, human-readable account of triggers, patterns, values, and working hypotheses—and appends a session record. This memory guides both agents, supports targeted follow-ups, and makes the system's behavior auditable for study.

4.2 Inputs and Outputs

Inputs. Each turn ingests: the user's utterance u_t; a short window of recent dialogue h_t; and the current case conceptualization C_{t-1}. Internally, the Orchestrator maintains (SR, IS, UT, ER, PR, OR) and a few adaptive thresholds.

Outputs. The system returns one of three response types: (i) a safety-first crisis message (with de-escalation and resource information) in Crisis; (ii) a one-to-two sentence exploratory probe in Explore; or (iii) a brief, tailored, MI-aligned advisory response in Persuade, usually containing one actionable next step and one open follow-up question. Each turn also yields an updated case note and (when appropriate) updates to the conceptualization.

4.3 Why a Multi-Agent Architecture?

Multi-agent patterns have proven effective for decomposing complex, ill-structured tasks into subgoals, enabling specialized competencies and transparent orchestration. Our setting benefits from a clean separation of concerns: Explore is optimized for uncertainty reduction and change-talk elicitation, while Persuade is optimized for commitment and small, feasible next steps. Splitting these concerns avoids the common failure mode of "one-size-fits-all" advice and helps the system preserve conversational naturalness.

4.4 Orchestrator: Typed, Experience-Fitted Routing

The Orchestrator is an interpretable controller that prioritizes **DG1 Safety-first** and implements **DG3 Typed Orchestration** and **DG5 JITAI** pacing. At each turn it (a) estimates signals from u_t and h_t: SR (safety), IS (information sufficiency), and UT (a smoothed proxy for how much guidance the user tolerates); (b) optionally detects explicit advice requests; (c) updates adaptive thresholds based on recent outcomes and time window; and (d) selects a mode via four monotone gates:

(1) **Safety gate** (Crisis). If SR indicates imminent risk, divert to LIVES-style first-line support (no persuasion; **DG1**).

(2) **PR/force gate** (Explore). If persuasion has run consecutively ($\text{PR} \geq \text{PR}_{\max}$) or the local router requested a mode flip, explore to reduce resistance.

(3) **ER gate** (Persuade). If exploration has lasted long enough ($\text{ER} \geq \text{ER}_{\max}$), provide a concrete next step to avoid endless questioning.

(4) **Sufficiency gate** (Persuade). If IS is high ($\text{IS} \geq \tau$) and $\text{OR} \geq \text{OR}_{\min}$, deliver advice; otherwise default to Explore.

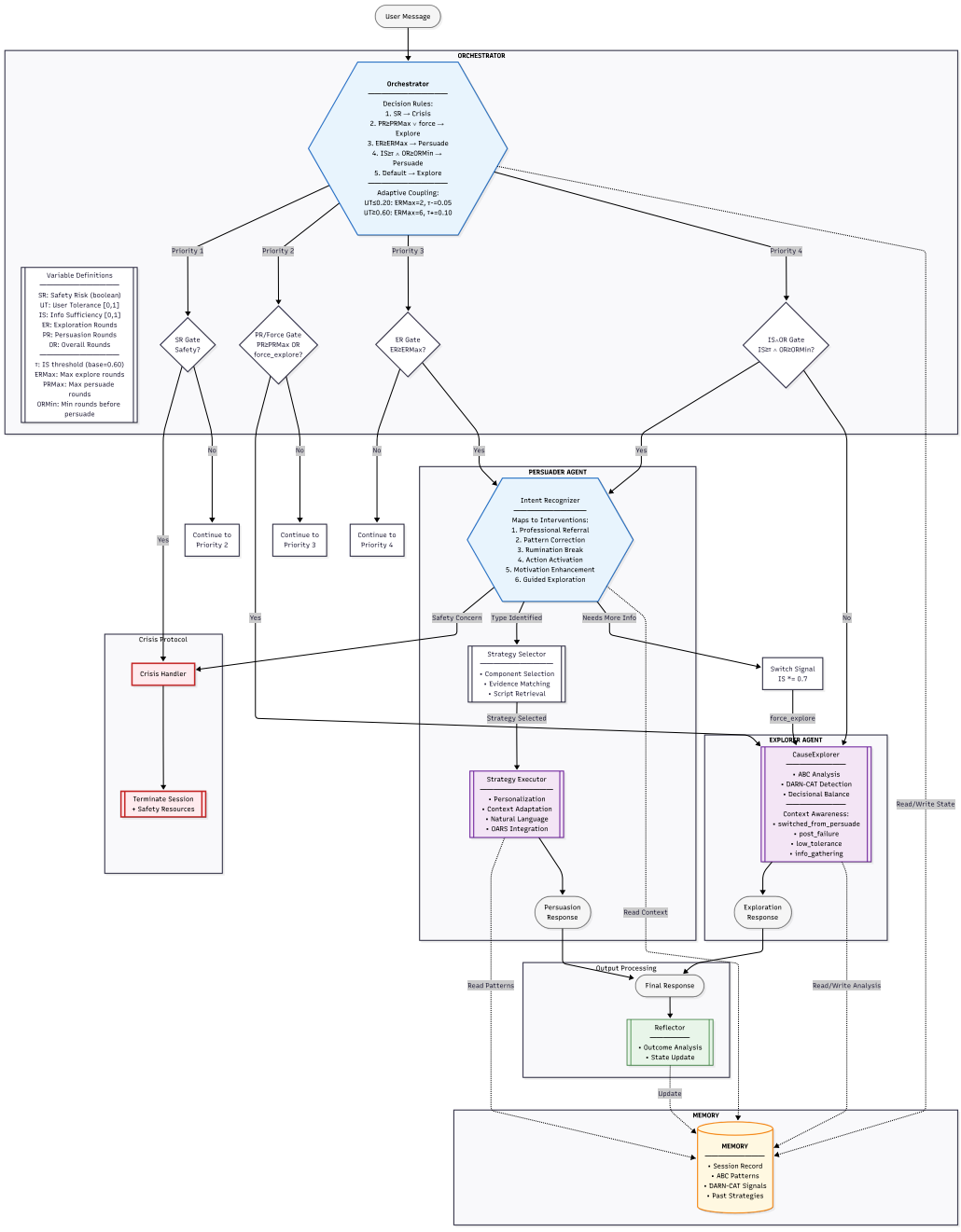

Fig. 1. Turn-level routing policy (schematic). The Orchestrator applies (1) a safety gate (Crisis); (2) a PR/force gate that ensures at least one Explore after repeated persuasion or an explicit switch signal; (3) an ER gate that caps pure exploration; (4) a sufficiency gate that triggers Persuade when IS is high and enough overall rounds (OR) have elapsed. The Persuader contains a finer local router and can push a temporary “force explore” flag back to the Orchestrator.

Algorithm 1 Orchestrator routing per turn t

Require: state (ER, PR, OR, UT, IS, SR, τ, ERMax, ORMin, PRMax, force), context h_t, utterance u_t

$(\widehat{\text{UTI}}, \widehat{\text{IS}}, \widehat{\text{SR}}) \leftarrow \textsc{EstimateSignals}(u_t, h_t)$
$\text{UT} \leftarrow 0.45 \cdot \text{UT} + 0.55 \cdot \widehat{\text{UTI}}$; $\quad \text{IS} \leftarrow \widehat{\text{IS}}$; $\quad \text{SR} \leftarrow \widehat{\text{SR}}$
if $\textsc{AdviceRequested}(u_t, h_t)$ **and** IS > 0.3 **then**
    return Persuade with WriteBackPersuade
end if
$(\tau, \text{ERMax}, \text{ORMin}, \text{PRMax}, \text{force}) \leftarrow \textsc{OutcomePreprocess}(\cdot)$
$(\tau, \text{ERMax}, \text{ORMin}) \leftarrow \textsc{TimeWindowAdjust}(\cdot)$
$(\tau, \text{ERMax}, \text{PRMax}) \leftarrow \textsc{ToleranceCouple}(\text{UT}, \cdot)$
$\textsc{Clamp}(\cdot)$
if SR = `true` **then**
    return Crisis (LIVES diversion)
end if
if PR $\geq$ PRMax **or** force **then**
    WriteBackExplore
    return Explore
end if
if ER $\geq$ ERMax **then**
    WriteBackPersuade
    return Persuade
end if
if IS $\geq \tau$ **and** OR $\geq$ ORMin **then**
    WriteBackPersuade
    return Persuade
end if
WriteBackExplore
return Explore

Two couplings make the pacing human-sensible: (i) *Outcome coupling* gently relaxes/tightens thresholds after positive/negative persuasion outcomes; (ii) *Tolerance coupling* increases exploratory allowance when UT is high and caps persuasion streaks when UT is very low, echoing JITAI's concern for burden control and opportune timing.

This experience-fitted, monotone gating is simple to audit and tune, and it operationalizes a mixed-initiative *progressive specificity* policy: evoke understanding first, then commit to micro-action when the case is sufficiently clear and the user is ready.

4.5 Explorer Agent: Sensemaking to Reduce Ambiguity

The main goal of Explorer is to transform fragmented self-expressions into understandable and traceable case information through dialogue, rather than moving directly into persuasion. Case conceptualization usually covers three core aspects: the user's personal background and traits (such as attachment style, values, relationship history), the current behaviors and symptoms (such as repeated verification, social media monitoring, cycles of dependence and avoidance), and the triggers and maintaining mechanisms (such as what situations trigger anxiety, what short-term consequences reinforce the behavior, and what long-term costs result). Adequate exploration of the user themselves and their Love-Brain related problems provides the foundation for subsequent persuasion.

To achieve this goal, the system introduces three complementary strategies: DARN-CAT listening and amplification, Decisional Balance pros and cons analysis, and ABC functional analysis. They are not linear in order, but work in parallel, complementing each other and jointly supporting the completion of case conceptualization.

One important strategy is the MI DARN-CAT framework, which is used to capture and amplify the user's "change talk." DARN-CAT includes six categories of signals: Desire, Ability, Reason, Need, Commitment, Action, and Taking steps. Amrhein et al. (2003) [2] showed that the frequency and strength of change talk expressed by users in dialogue predicted subsequent behavior change. During exploration, when such signals appear, the system reflects these signals back to the user and reinforces them to avoid confrontation. This approach respects the user's autonomy while providing motivational information that enriches case conceptualization [24].

Another approach is Decisional Balance, which becomes relevant when users hesitate about whether to withdraw attention or reduce overinvestment in the relationship. Originating from Janis and Mann's decision balance theory [16] and applied in Prochaska's Transtheoretical Model (TTM) [26], this method frames ambivalence into a four-quadrant comparison of pros and cons for both change and non-change. The system can prompt this comparison with open questions—for instance, "What benefits would continuing bring, and what costs might it involve?" Such dialogue encourages users to articulate their own reasons for change and to weigh short-term comfort against long-term consequences.

A central component of Explore is ABC functional analysis. This method emphasizes the causal chain of Antecedent (trigger situation), Behavior (response), and Consequence (outcome), and is considered essential for completing case conceptualization. Love-brain related behaviors, such as repeatedly checking a partner's social media, are often triggered by anxiety or uncertainty; while they may offer temporary relief, they tend to intensify anxiety and conflict over time. Guiding the user to describe (1) what triggers the behavior, (2) what the behavior is, (3) what short-term consequences occur, and (4) what long-term outcomes follow helps uncover the maintaining factors behind the behavior. The resulting logical chain clarifies the user's difficulties and grounds the selection of subsequent persuasion strategies.

In summary, Explorer functions as a bridge between spontaneous self-expression and targeted persuasion. By combining motivational signals, structured comparisons, and functional analysis, it equips the system with richer information about the user's state. This ensures that persuasion strategies are grounded not in abstract techniques, but in the specific circumstances and experiences of the individual. The outputs of this stage are intentionally short and natural, typically one or two lines phrased in everyday language, starting from the user's own details rather than generic prompts, so that the interaction feels supportive rather than clinical.

4.6 Persuader Agent

The Persuader turns typed case knowledge into *one* feasible micro-action and a brief, MI-aligned response. It has two stages: *Strategy Composition* (selecting *what* to do) and *Personalized Execution & Delivery* (deciding *how* to say it).

4.6.1 Strategy Composition (type → component pipeline). Strategy composition operationalizes the **DG3 Typed Orchestration** . Each strategy card is specified as:

(1) Identification Conditions (scales, trigger words, behavior counts)
(2) Key Mechanisms
(3) Strategy Components (If–Then, DEAR MAN, EPPM/specifications, JITAI)
(4) Observables (execution rate, behavior counts, day-level sentiment/sense of control)
(5) Safety Boundaries (LIVES diversion criteria)

At test time, the Persuader performs:

(1) Type reading. It reads the user's current *relationship type* (e.g., social-media jealousy/surveillance, ROCD, anxious/avoidant attachment, limerence, mind-reading/"if they loved me they'd know", boundary deficits, dysfunctional relationship beliefs) and maps to the corresponding card. This enacts typed, mechanism-linked interventions rather than generic tips (**DG3**).

(2) Preference-aware channel selection. To respect **DG2 Dual-route Onboarding**, the Persuader chooses *cognitive* vs. *affective* delivery paths by combining (i) the user's NFC/NFA screening (two-question first screen) with (ii) lightweight evidence from recent dialogue (e.g., explicit pushback against "being lectured" biases toward affective framing). The cognitive path emphasizes logic, mechanisms, and experiments; the affective path emphasizes empathy, stories, and hope-bridging.

(3) Local router (readiness and safety). A small, MI-informed classifier labels the latest response (e.g., ready to act, partial acceptance, venting, information-seeking, new event, or safety escalation). If safety conditions hold, the agent diverts to Crisis (**DG1**). If information is insufficient or a new event emerged, it emits a short-lived *switch-to-explore* signal; the Orchestrator will honor it on the next turn, preventing premature advice.

(4) Component selection. Given the chosen card and channel, the agent selects *a single component* that best fits the present response pattern and recency of use. The library includes Implementation Intentions (If–Then), quota+delay shaping, clarity templates (two-sentence "I notice / I need"; DEAR MAN), EPPM-informed framings with explicit efficacy steps, and light mindfulness/grounding. Selection avoids repetition fatigue and favors bridges when acceptance is partial.

(5) Knowledge bundle and observables. The agent composes a compact, human-readable *bundle* that records: the card's goals (mechanism-linked), the chosen component (plus two alternates), the If–Then script (if any), a minimal fallback, and the *observables* to track (execution rate, behavior counts, EMA of sense of control)—supporting **DG4 Micro-actions & Observability** and **DG8 Measurable & Reproducible**.

(6) Micro-evidence and burden control. Based on the user's immediate speech act (e.g., asking "how", seeking examples, venting), the Persuader optionally attaches a 40–60s case vignette or a simple self-experiment template (A/B day design, three indicators), and adjusts reminder intensity to vulnerability windows, consistent with **DG5 JITAI**. A cooldown guard suppresses recently overused prompts to preserve naturalness (**DG7**).

4.6.2 Personalized Execution & Delivery (MI-aligned response). Given the bundle, the agent renders a 4–8 sentence response that:

- maintains an MI tone (OARS: one open question, one concrete affirmation, at least one reflection, optional one-sentence summary);
- offers *one* micro-action that is executable *now* (If–Then, quota+delay, one DEAR MAN request, or a specific specification step), satisfying **DG4**;
- respects EPPM's Threat×Efficacy principle by pairing any risk mention with an immediate, feasible alternative (**DG6**);
- fits the user's case conceptualization (triggers/behaviors/maintainers) so advice is about *their* loop, not abstractions.

The Persuader logs the chosen component and observables for process analysis, supports MITI-style process sampling for quality/integrity (**DG7**), and updates the case note for continuity. If the ensuing user response indicates low readiness or new ambiguity, the local router issues a switch-to-explore signal; otherwise, positive micro-success slightly relaxes thresholds in the Orchestrator (Algorithm 1), yielding a gentle momentum.

4.7 Why This Works

The architecture operationalizes MI’s two stances—*evocation* (Explorer) and *planning* (Persuader)—with explicit, typed handoffs and safety short-circuiting. Typed, mechanism-linked cards (Section 3.3) avoid generic advice and make observables explicit (**DG3–DG4**). The Orchestrator’s monotone gates and tolerance coupling instantiate a JITAI-like pacing policy that controls interruption burden and opportunistically times advice (**DG5**). EPPM-consistent messaging lowers the risk of fear-only prompts (**DG6**). Minimal, MI-consistent phrasing plus cooldown guards maintain conversational integrity (**DG7**). Finally, the card schema carries BCT/COM-B labels and CONSORT-EHEALTH reporting fields, enabling measurement and reproduction (**DG8**), while data minimization concentrates on counts/durations and supports one-click purge (**DG9**).

Implementation-agnostic note. The design is model-agnostic: any pipeline capable of estimating (SR, IS, UT), composing typed micro-actions from strategy cards, rendering MI-aligned language, and updating a text conceptualization can realize this blueprint. The contribution is an *interpretable, human-centered control policy* with modular agents whose behaviors are auditable and empirically testable.

5 User Study

This study evaluated the effects of our proposed workflow—safety-first → emotion/evidence dual portals → rapid scanning and typified routing of relationship beliefs → orchestration of persuasive components → daily EMA and replay—on risky behaviors and subjective experiences related to the "monitoring-verification-reconnection" process in close relationships over a three-day period, and the mechanisms by which this process operates through process variables. We used daily ecological momentary assessment (EMA) as the core observational channel to mitigate recall bias and capture contextualized changes, thereby tracking behavioral microprocesses and the coupling relationship between interventions in real-world situations. At the persuasion and information processing layers of the system, we leveraged individual differences in need for cognition and need for affect to route users to either the "evidence-argumentation" or "emotion-empathy" portals, maintaining information load and expression styles consistent with their preferences throughout the conversation. These traits are widely used to explain differences in persuasive path selection and processing depth within the ELM framework. To ensure ethical and safe conversations, all conversations were guided by the WHO’s LIVES (Listen, Inquire/Assess, Affirm, Enhance Safety, Support) principles, triggering immediate referrals and digital risk reduction guidance for harassment.

Thus, our study focused on three research questions:

RQ1: Compared with non-matching, does the "security first + dual entry (NFA/NFC) + belief quick scan routing + typed policy orchestration" system lead to better multi-dimensional outcome changes at the end of the 7-day intervention?

RQ2: Do the system’s "actionable micro-actions" and "persuasion path matching" mediate outcome changes through process indicators (execution/exposure)?

RQ3: Does the "individualized fit" of portal and content (NFA/NFC × type hit/route) moderate the main effects?

5.1 Design and Participants

5.1.1 Design. We conducted a randomized, 3-day, controlled trial, lasting approximately 20 minutes per day. The experimental group received a matching persuasive AI (based on dynamic orchestration of NFA/NFC and scenario type; components included If-Then planning, social norm prompts, and EPPM-style efficacy/threat matching messages). The control group received a generic support of equal duration and frequency (no routing, no persuasive components).

5.1.2 Procedure. On Day 0, we completed a baseline assessment using an online questionnaire platform, covering demographic and safety screening, as well as core constructs such as attachment orientation, relationship beliefs and communication, and social media-induced jealousy/intrusion. Based on this, the system automatically generated a pre-diagnostic conceptualization report in the research backend and created an individualized profile according to pre-registered scoring and routing rules to ensure consistency in subsequent compilation and assessment metrics.

During the intervention period, Days 1–3, participants were required to complete at least three conversational interactions in natural settings (each session was recommended to last at least 20 minutes to ensure a complete cycle of "exploration-strategy-plan execution-replay"). The system matched and switched between the emotion/evidence channels based on their entry preferences and the context of the day. During the conversations, if-then plans, normative prompts, and EPPM-style "efficacy × threat" statements were translated into minimally actionable actions (e.g., delay-falsifiability question-mute/time-limited viewing), reducing immediate burden while preserving observable behavioral indicators. To reduce recall bias and obtain contextualized daily sequences, we collected daily EMAs after the conversations, including counts of monitoring/verification/reconnection behaviors, core beliefs and emotional states, if-then execution rates, use of templates/timing and mute, subjective understanding, and sense of control. These process signals were used for dose-response and mediation pathway modeling and as a basis for online parameter tuning of the system. The research team synchronously records the "dosimetric" evidence of conversations and interventions in the background: timestamp conversation logs, template/timing/silence calls, round structure and duration, and judgment and switching events related to routing thresholds; if storage fails occasionally, participants supplement with screen recordings/screenshots according to the plan to complete the chain of evidence of intervention exposure.

On Day 3, after completing the final interaction and the daily EMA, participants filled out a final questionnaire to repeat key scales and report on their experience and burden. The system exported a structured log to support multi-level slope comparisons and the estimation of a 1-1-1 mediation/moderation model. The researchers then conducted brief interviews according to a semi-structured outline to supplement the mechanism explanation and usability evidence.

5.1.3 Measurement and Analysis. Our measurement system addresses both "state change," "process execution," and "match/manipulation verification." At the state level, insecure attachment and compulsive relationship verification in intimate relationships are considered the most likely upstream mechanisms driving online verification, monitoring, and conflict. Therefore, we used the anxiety and avoidance scores of the ECR-R at baseline and final assessment as a stable reference for the attachment structure and to explain how different individuals process the system's content. The reliability and construct validity of the ECR-R have been repeatedly validated in large samples, and public resources are available for item and scoring (e.g., Fraley et al.'s item and scoring explanations for the ECR-R). Relationship doubts and obsessive-compulsive symptoms focused on partner flaws are represented by the ROCI and PROCSI, respectively. These two instruments cover both relationship-centered doubts about "real love/suitability" and examine partner flaws in

dimensions such as appearance, sociability, morality, emotional stability, competence, and intelligence. Both instruments have demonstrated good internal consistency and correlations with constructs and criterion, and continue to be used and expanded in subsequent research. To understand the path of the "social media amplification effect," we modeled the relationship between intrusive use of social platforms and jealousy. The Facebook Intrusion Questionnaire (FIQ) and the Facebook Jealousy Scale, previously used to characterize the "intrusion-jealousy-satisfaction" chain, serve as quantitative anchors for our online clue-monitoring-conflict path.

At the process level, directly linked to our system workflow is the immediate exposure and implementation of actionable "micro-actions" and persuasive components. We use the generation and execution of If-Then plans, template invocations, the use of mute/timer devices, and the number of daily "checks/verifications/reconnections" as EMA indicators spanning seven days. If-Then is a replicable, self-regulating strategy with clear mechanisms, exhibiting moderate to large overall effects and demonstrating strength in clinical/quasi-clinical samples. It is suitable as an intervention unit that can be triggered by the system and have its exposure and execution rates calculated. These process indicators correspond to the definitions and log structures of "templates/timers/If-Then" in your paper, facilitating the simultaneous playback of "system performance" and "user activity" on the same timeline (e.g., the backend fields and timestamp definitions for conversation logs, template invocations, timers, etc. in the research package). To improve reproducibility and comparability, we coded/mapped system components and scripts according to both BCT v1 and COM-B, and provided the coding representation and data dictionary in the text to meet the transparency requirements of the eHealth reporting specification.

For matching and manipulation verification, we used the NFC-18 and NFA-S-10 to characterize preferences and tolerance for "evidential/emotional entry points." This served as both a basis for personalized triage into the system and as a check variable for manipulation success. Both measures have short forms and cross-contextual structural evidence. EMA implementation adhered to classical design principles, emphasizing real-time, multiple, and naturalistic sampling to reduce recall bias and observe intra- and inter-day microprocesses. Accordingly, we implemented brief daily backfills and post-session recordings from T1–T6, with adherence and reactivity measures documented in the Appendix. For quality control, we independently coded key conversation samples at the global and behavioral levels using MITI 4.2.1, incorporating interview/facilitation style and technical quality as covariates. We also retained the original conversations and backend logs for auditability. The organization and value ranges of these measurements are consistent with the observation in the results paragraph that "improvements on the behavioral side are more sensitive, changes on the symptom side are smaller and appear over time", and also support our focus on the "process-result" chain as the explanation.

5.1.4 Analysis. The analytical strategy maintains consistency with the system’s objectives, processes, and data structure: First, a linear mixed model was used to estimate the average treatment effect over the 7-day longitudinal period, using the group × time interaction as the primary signal and individual random intercepts (and random slopes, if necessary) to account for within-subject correlation. Baseline scores were included in the model to enhance precision and comparability. Simultaneous testing of multiple outcomes employed Benjamini–Hochberg FDR control to avoid overly conservative dilution of process effects. All estimates are reported not only with coefficients and intervals, but also with consistency expressed as daily slopes and end-of-test differences.

Consistent with the Results section, we placed "execution/exposure" variables (If-Then execution rates, template and timer calls, and daily behavior counts) at the core of the mechanism. Using a 1-1-1 multilevel mediation/multilevel structural equation model, indirect effects were estimated for both within-subject and between-subject components.

Monte-Carlo/bootstrapping techniques were used to obtain indirect effect intervals, enabling the "system-process-outcome" chain to be statistically disaggregated and quantified.

Considering that "entry-route matching" is a key design feature of the system, we constructed the MatchScore based on entry style consistency (NFA/NFC) and "type hit" (whether the route directs the user to their primary question cluster). We added a third-order interaction to the main effects model to test whether personalized adaptation amplifies the main effect. Marginal effects plots then display the temporal evolution of the results at different levels of match. This aligns with the paper's report on the advantages of highly matched individuals on process indicators and several outcomes.

Missing data were handled in the longitudinal model using maximum likelihood estimation under the MAR assumption. Multiple imputation was used for sensitivity analysis of missing data at the end of the questionnaire. Conversation quality (MITI global) was included as a covariate in the robustness model, and the robustness after quantile-based quality analysis is reported in the Appendix. The FDR was used for multiple comparisons.

Regarding presentation and replication standards, we provide the BCT/COM-B component-to-mechanism mapping table and CONSORT-EHEALTH checklist with this article. This ensures that external parties can reconstruct intervention elements and compare them with the object structure and timestamps defined in our back-end output, thereby independently verifying the "process-to-outcome" inferences. This approach is consistent with common recommendations for eHealth trials.

Finally, the EMA sampling plan aligns with JITAI's "right dose at the right time" philosophy. In this discussion, we will juxtapose the marginal contributions of "when to trigger/when not to trigger" and "which entry/script" to the outcomes with the daily process curves.

5.1.5 Participants. We enrolled forty participants and randomized them 1:1 to the experimental arm (n = 20) or the control arm (n = 20). The sample comprised 23 women and 17 men (Mage = 23.7, SD = 3.4; range = 19–32). Most participants were undergraduate/Bachelor students (28/40, 70%), with the remainder holding a Master's degree (12/40, 30%). Entrance preferences at baseline were skewed toward the Need-for-Cognition route (NFC = 29, 72.5%) relative to the Need-for-Affect route (NFA = 11, 27.5%), which we used both to initialize the portal and as a manipulation check. The primary presenting profiles reflected the trial's target phenomena: anxious attachment was most common (16 cases), followed by dysfunctional relationship beliefs (9), boundary loss/sacrifice-as-love (7), limerence (3), ROCD/intolerance of uncertainty (2), and avoidant attachment (2). Distribution of these profiles was comparable across arms by design. As a continuous indicator of individualized fit, the baseline MatchScore averaged 0.59 overall (Control = 0.61; Experiment = 0.58), and was prespecified as a moderator in downstream models. Participant flow, baseline characteristics, and operational definitions are reported in line with CONSORT-EHEALTH/CONSORT guidance for eHealth randomized trials to maximize transparency and replicability.

6 Results

We enrolled a total of 40 participants (20 experimental participants; 20 control participants; gender, age, and education not specified). The intervention period lasted 3 days: pre-interval (baseline on Day 0), mid-interval (daily short assessments on Days 1–2), and post-interval (endline assessment on Day 3).

Table 1. Participant Demographics and Characteristics

ID	Group	Gender	Age	Education	Relationship Type	Entry	Score
1	Experiment	M	26	Bachelor	Boundary Issues	NFC	0.40
2	Experiment	F	29	Master	Limerence	NFC	0.60
3	Experiment	M	29	Master	Anxious Attachment	NFC	0.60
4	Experiment	F	27	Master	Dysfunctional Beliefs	NFC	0.66
5	Experiment	F	24	Bachelor	Anxious Attachment	NFA	0.60
6	Experiment	F	25	Bachelor	ROCD	NFC	0.43
7	Experiment	M	21	Master	Anxious Attachment	NFC	0.47
8	Experiment	M	21	Bachelor	Dysfunctional Beliefs	NFC	0.47
9	Experiment	F	25	Bachelor	Anxious Attachment	NFA	0.60
10	Experiment	F	27	Master	Anxious Attachment	NFC	0.31
11	Control	F	26	Bachelor	Anxious Attachment	NFC	0.29
12	Control	F	22	Master	ROCD	NFC	0.43
13	Control	F	20	Bachelor	Dysfunctional Beliefs	NFC	0.51
14	Experiment	M	25	Master	Boundary Issues	NFC	0.51
15	Experiment	F	24	Master	Mind Reading	NFC	0.51
16	Experiment	F	20	Bachelor	Dysfunctional Beliefs	NFC	0.83
17	Control	F	21	Bachelor	Anxious Attachment	NFA	0.60
18	Experiment	F	23	Bachelor	Anxious Attachment	NFA	0.20
19	Control	F	21	Bachelor	Anxious Attachment	NFA	0.83
20	Control	M	19	Bachelor	Limerence	NFC	0.63
21	Control	F	22	Bachelor	Dysfunctional Beliefs	NFA	0.69
22	Control	F	19	Bachelor	Boundary Issues	NFA	0.63
23	Control	M	21	Bachelor	Boundary Issues	NFC	0.86
24	Control	M	22	Bachelor	Boundary Issues	NFC	0.69
25	Control	M	22	Bachelor	Anxious Attachment	NFC	0.66
26	Control	M	32	Bachelor	Anxious Attachment	NFA	0.86
27	Control	F	25	Bachelor	Anxious Attachment	NFC	0.49
28	Experiment	F	20	Bachelor	Boundary Issues	NFC	0.89
29	Experiment	M	28	Master	Avoidant Attachment	NFC	0.94
30	Experiment	F	20	Bachelor	Anxious Attachment	NFA	0.51
31	Control	M	21	Bachelor	Anxious Attachment	NFC	0.71
32	Control	F	20	Bachelor	Limerence	NFC	0.63
33	Control	M	24	Bachelor	Anxious Attachment	NFC	0.60
34	Control	M	22	Bachelor	Boundary Issues	NFC	0.71
35	Experiment	M	23	Bachelor	Dysfunctional Beliefs	NFA	0.80
36	Control	F	22	Master	Anxious Attachment	NFC	0.40
37	Experiment	M	23	Bachelor	Dysfunctional Beliefs	NFC	0.60
38	Experiment	M	30	Master	Avoidant Attachment	NFC	0.66
39	Control	F	31	Bachelor	Dysfunctional Beliefs	NFA	0.43
40	Control	F	25	Master	Dysfunctional Beliefs	NFC	0.57

6.1 Matching persuasion system lead to multi-dimensional outcome improvements

The experimental and control groups showed excellent baseline balance across all psychological measures (Figure 2). PHQ-9 scores were comparable between groups (p=0.728), as were GAD-7 (p=0.581), ADS (p=0.870), LAI (p=0.087), and ECR-R (p=0.967), confirming successful randomization.

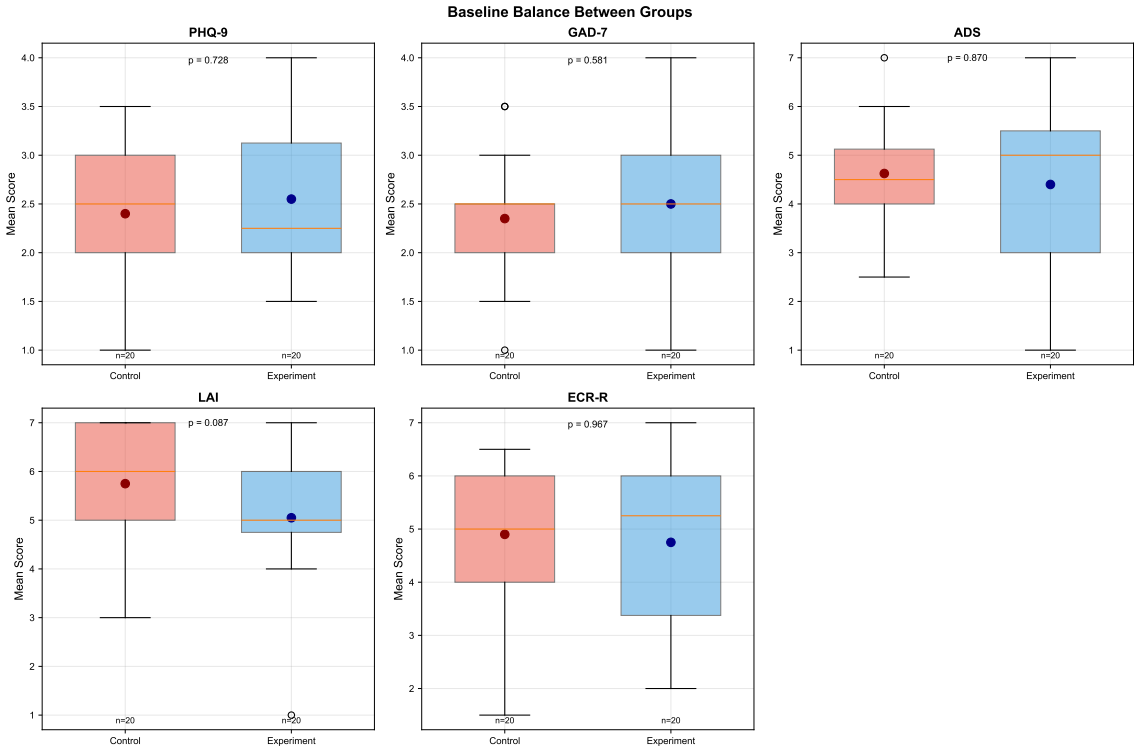

Fig. 2. Baseline balance between experimental and control groups across psychological measures

The intervention demonstrated promising effects on depression symptoms (Figures 3 and 4). The experimental group showed a small-to-medium effect size improvement in PHQ-9 scores (Cohen's d=-0.400, p=0.550), with anxiety symptoms also improving (GAD-7: d=-0.231, p=0.755). While not reaching conventional significance thresholds, these effect sizes represent meaningful clinical improvements for a 2-day intervention.

Daily behavioral monitoring revealed consistent improvements across all tracked behaviors (Figure 5). The experimental group showed greater reductions in confirmation-seeking (Day 1: 3.54 to Day 2: 2.91), checking behaviors (3.12 to 2.48), social media checking (2.74 to 1.75), and reconnection attempts (2.70 to 2.11). These behavioral improvements suggest the intervention successfully promoted behavior change even within the brief timeframe.

Notably, the experimental group demonstrated significantly higher advice execution rates (p=0.047; Figure 6), indicating successful engagement with the intervention content. While other engagement metrics showed favorable trends (system trust p=0.276, overall engagement p=0.303), the significant difference in advice execution suggests the intervention successfully motivated behavioral implementation.

6.1.1 Immediate Focus and Cooling Brought by Actionable Micro-Strategy. Actionable micro-actions provide immediate cooling and focus. Several participants spontaneously mentioned the helpfulness and persistence of the 30-minute "delay/distraction" activities. For example, P29 (Control Group) mentioned that "I'll stick with that 30-minute reflection...about how to prevent problems before they happen and how to resolve them after they occur." and perceived the suggested strategy for their relationship problems as "Overall engagement is very good, and the usability is quite

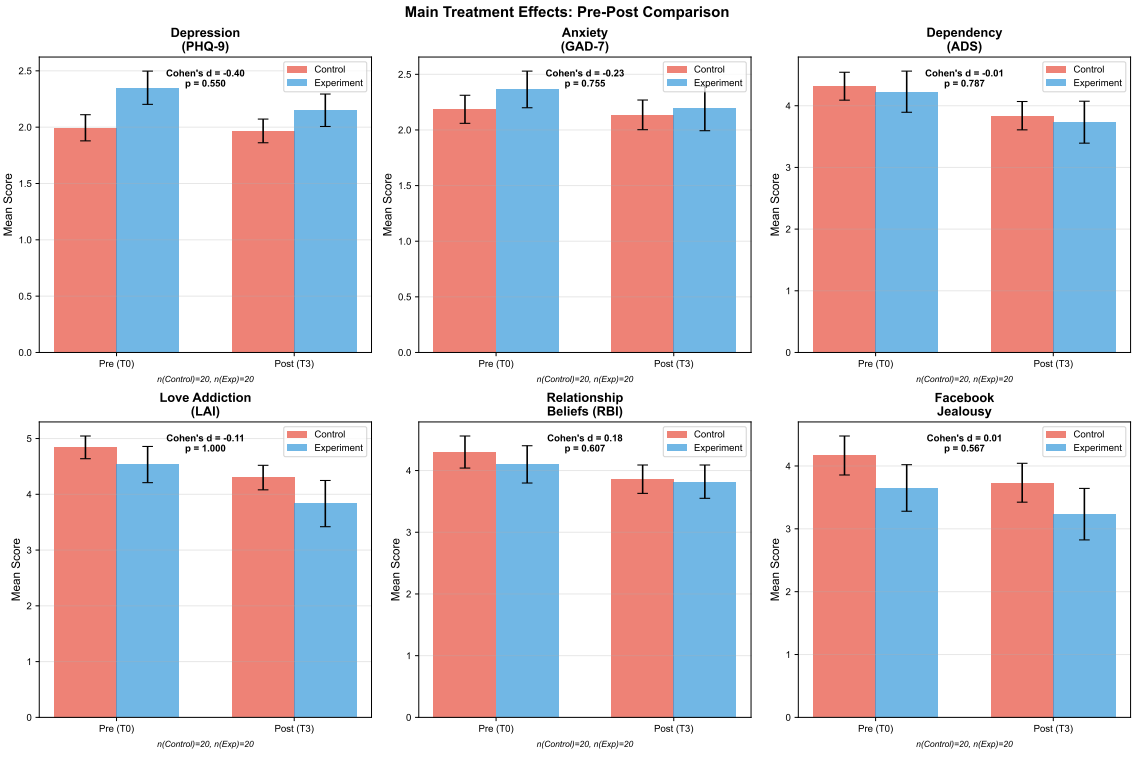

Fig. 3. Main treatment effects showing pre-post changes in primary and secondary outcomes

practical. They give me suggestions and I can use them as a reference." And these actionable strategies brought immediate positive effects to P29, as reported that "There’s been a noticeable change... I’m now following the instructions first... and I’m no longer so concerned about what others think." On the contrary, in the control group with the basic GPT-4o model, participants such as P22 reported that "No... He basically said what I already knew (a bit unrealistic)."

However, generalization and repetition of the system for both groups cause the "interviewing" feeling to dilute the perceived effectiveness. Some participants seek direct feedback, such as P10, who said, "I need... You have to give me the answer... Don’t ask me questions, don’t beat around the bush."

This indicates that, Consistent with our "safety first → dual entry → quick scan routing → type-based orchestration → EMA" workflow, participants prioritize micro-actions that can be immediately implemented and the experience of being respected/understood. When the conversation remains on follow-up questions or giving homogenized suggestions, the positive experience will be diluted.

6.2 "Actionable Action/Matching Path" Mediates Outcome Changes Through Process Indicators

The significant difference in advice execution rates (p=0.047, Figure 6) represents a key process indicator. While formal mediation pathways showed limited indirect effects (total indirect effect=0.007), the higher execution rate in the experimental group coincided with greater symptom improvements. Individual response patterns (Figure 7) revealed

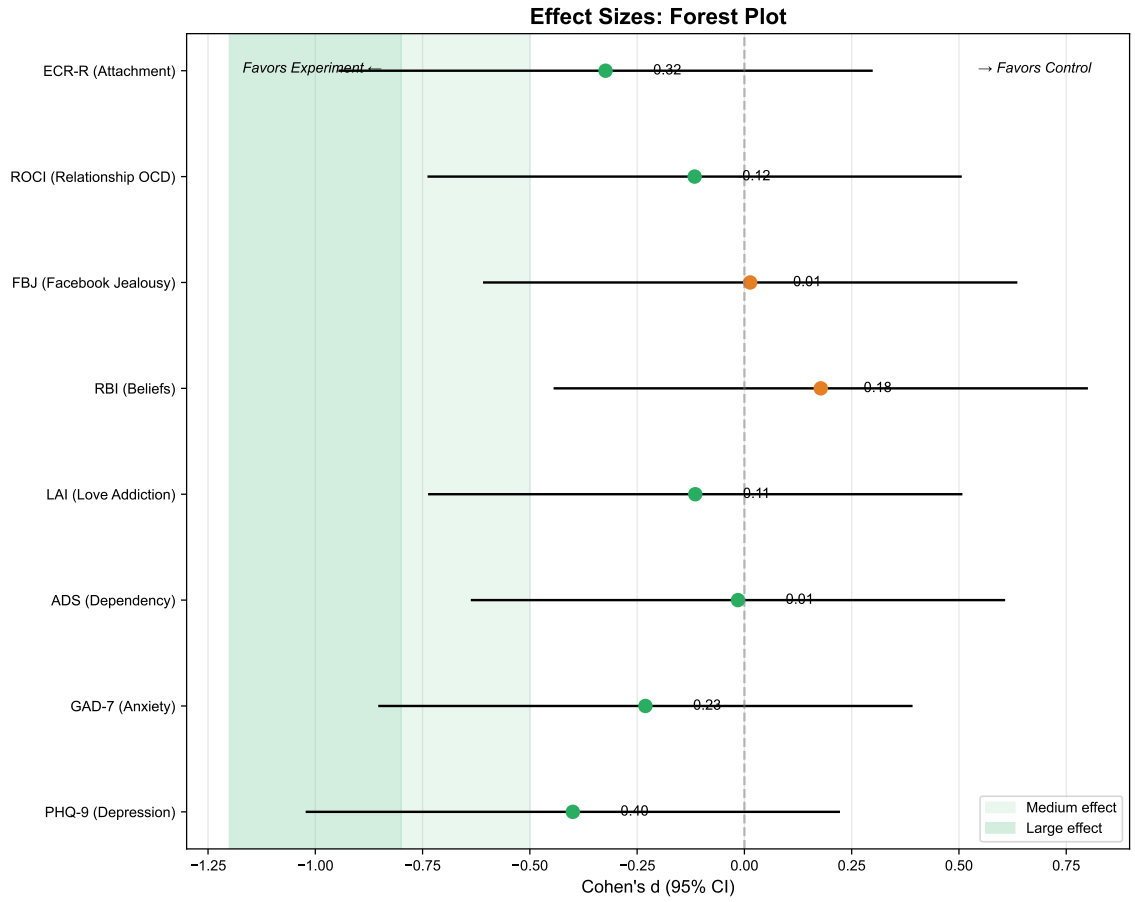

Fig. 4. Forest plot of effect sizes across all outcome measures

that experimental participants showed more consistent improvements (mean change=-0.19) compared to controls (mean change=-0.04).

Change scores demonstrated strong intercorrelations (Figure 8), with PHQ-9 and GAD-7 changes highly correlated (r=0.66), and ADS and LAI showing the strongest association (r=0.69). This pattern indicates that when participants improved, benefits generalized across multiple domains, supporting the comprehensive nature of the intervention approach.

6.2.1 Clear Enforcement and Specific Instructions. Participants viewed "doing a small, specific task first (writing, moving, delaying)" as a "brake button." For example, P38 of the experimental group emphasized that 30 minutes of exercise/writing helped "calm down" and directly inhibited impulsivity. P29 reported that "doing it first" brought "noticeable changes," allowing them to better switch attention and reconnect with themselves after starting. This is consistent with our hypothesis of a mediating path for the "1 behavior + 1 attitude" KPIs, including ROCD delay 10 minutes, asking only one verifiable question on social media, and limerence contactless minutes. This suggests

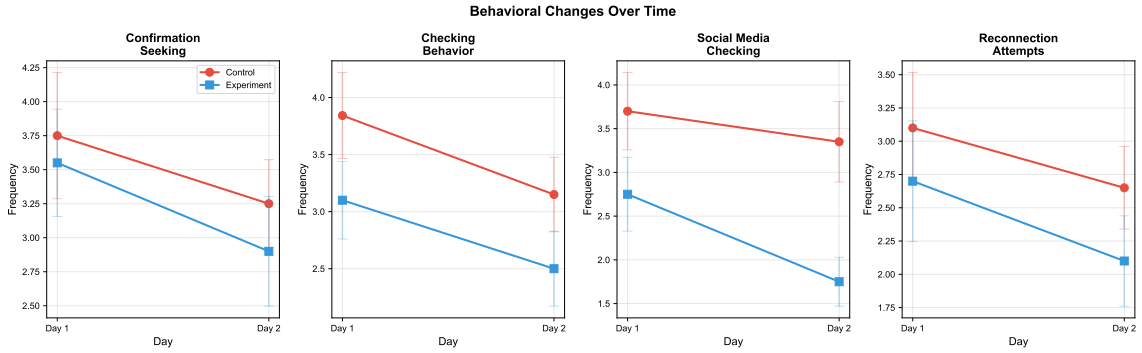

Fig. 5. Behavioral changes over time showing frequency reductions in problematic behaviors

that increased execution leads to decreased symptoms. However, when the system provides clear, actionable criteria or scripts, adoption is higher. Conversely, repeated questioning can feel like being "interviewed/audited." When P20 (Experiment) asked the system questions and didn't receive a positive response, she felt it was "addressing the question head-on, rather than just throwing it back." "Clear and executable" and "reducing ineffective questioning" together constitute the experience-behavior-outcome link: the more specific, the easier it is to do; the more it can be done, the more you can see in EMA that you are understood, your sense of control increases, and your impulsivity and rumination decrease.

6.3 "Personalized fit" between the entrance and content

Exploratory subgroup analyses revealed substantial variation in treatment response by relationship issue type (Figure 9). Participants with boundary issues showed large effect sizes (d=-0.85, n=7), as did those with dysfunctional relationship beliefs (d=-0.78, n=10). The largest subgroup, anxious attachment (n=17), demonstrated medium effect sizes (d=-0.55). These findings suggest the intervention is particularly effective for certain relationship concerns, highlighting the importance of personalized matching between intervention content and individual needs.

The pattern of differential response by subtype indicates that tailoring the intervention to specific relationship issues could enhance effectiveness. The strong responses in boundary issues and dysfunctional beliefs subgroups, despite smaller sample sizes, provide valuable direction for future intervention refinement and targeting strategies.

6.3.1 Perception of Fit and Being Targeted. After receiving basic advice on resolving relationship problems, P29 explicitly expressed her desire for more detailed and understanding advice if she continued to use the system: "I hope... for more comprehensive advice, a little more nuanced... to analyze my thoughts and provide better answers." In contrast, participants in the control group showed little strong desire for return visits. As P21 stated, she hoped the agent would attentively "respond to each of my questions individually," preferring selective responses or topic redirection. This aligns with our "entry preference × type hit" model: the higher the hit, the more aligned the responses.

However, some participants expressed a strong aversion to "interview-style" probing (e.g., P10), suggesting that users with a high need for cognitive closure (NFC) should prioritize convergent scripts over divergent questioning. Furthermore, participants like P16 viewed the agent's inherent "singleness and impracticality" as a key barrier to continued system use, further emphasizing the need for personalized and contextualized implementation. Individualized entry (NFA/NFC) and type hit/script fit together shape the subjective experience of "being targeted"; high fit strengthens

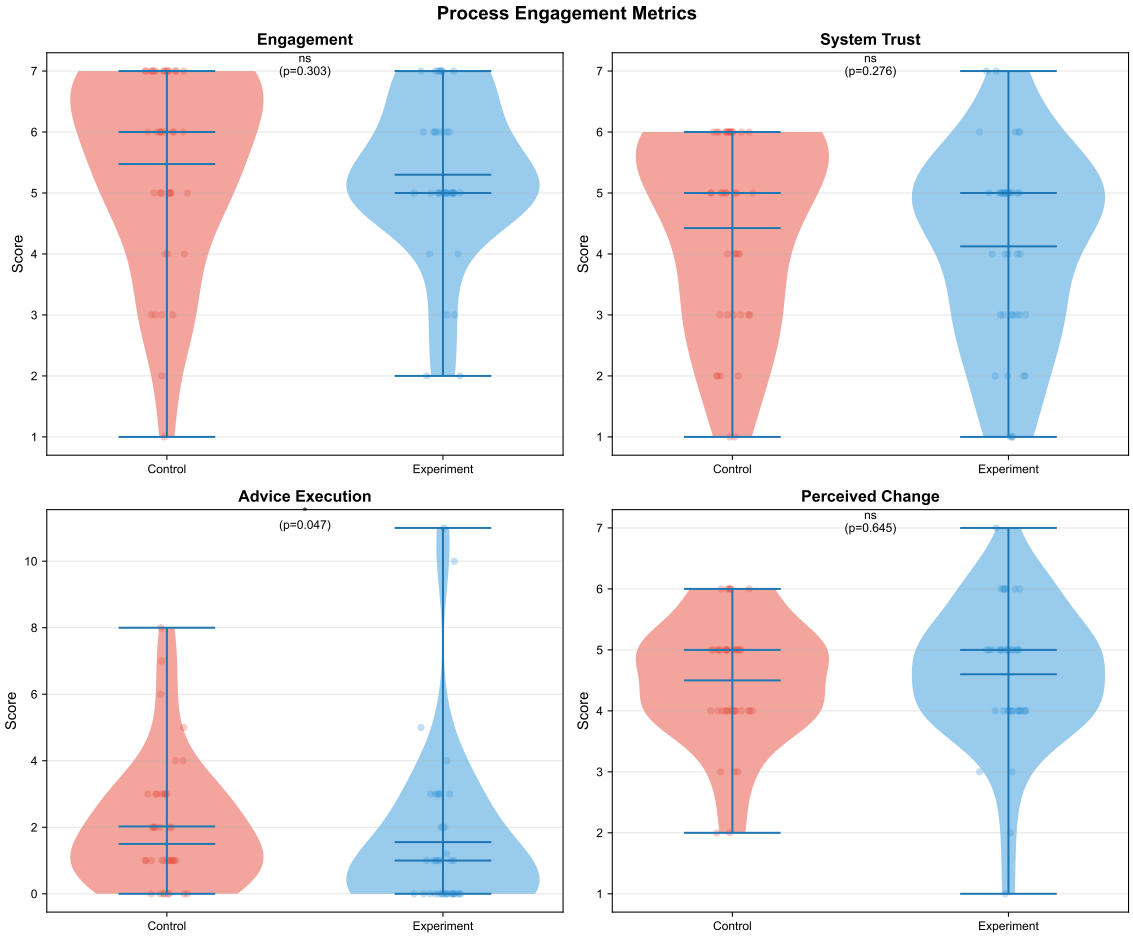

Fig. 6. Process engagement metrics comparing experimental and control groups

execution and sense of effectiveness, while low fit requires more specific standards and differentiated scripts to correct the rebelliousness of "being interviewed/homogenized".

7 Discussion

Our study asked whether a *safety-first + dual-entry + belief quick-scan routing + typed orchestration* workflow can: (RQ1) improve outcomes over a matched-dose, non-matching control; (RQ2) operate through process variables such as execution of micro-actions; and (RQ3) show advantages when the content “fits” user type and entry preference. Below we interpret the empirical patterns, extract design implications, and outline a research agenda for responsible, personalized persuasion in sensitive interpersonal contexts.

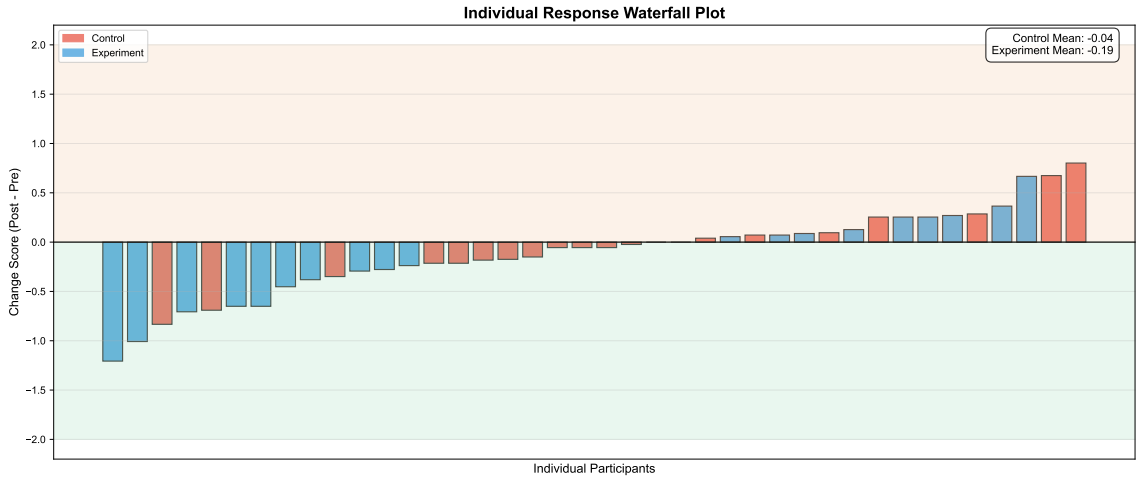

Fig. 7. Individual response waterfall plot showing heterogeneity in treatment response

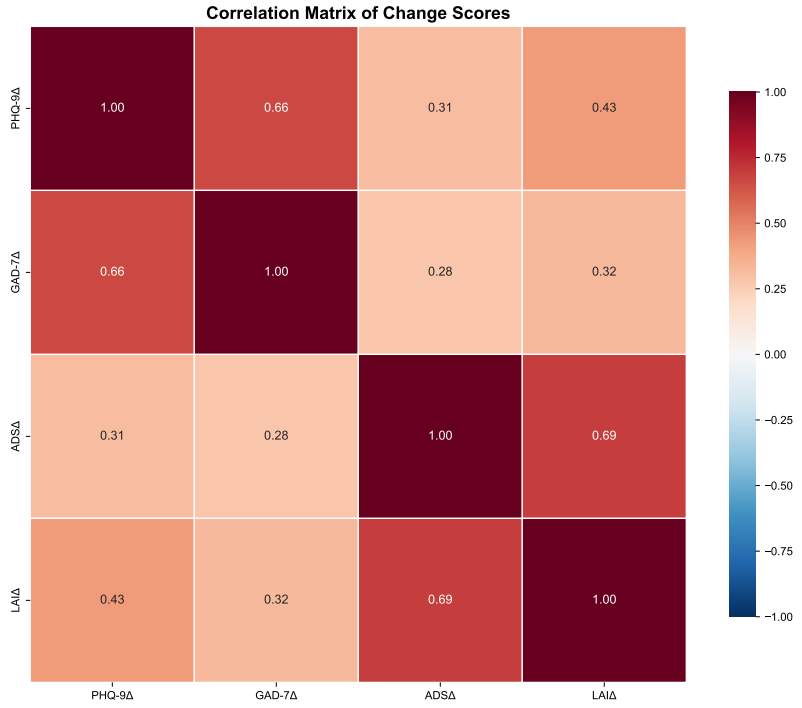

Fig. 8. Correlation matrix of change scores showing interconnected improvements

What changed, and how much?

Baseline balance and internal validity. Randomization achieved good baseline equivalence across all pre-intervention measures (PHQ-9, GAD-7, ADS, LAI, ECR-R), reducing concerns that post-intervention differences are artifacts of starting imbalances (all p’s $\gg$ .05; *Baseline Balance* figure).

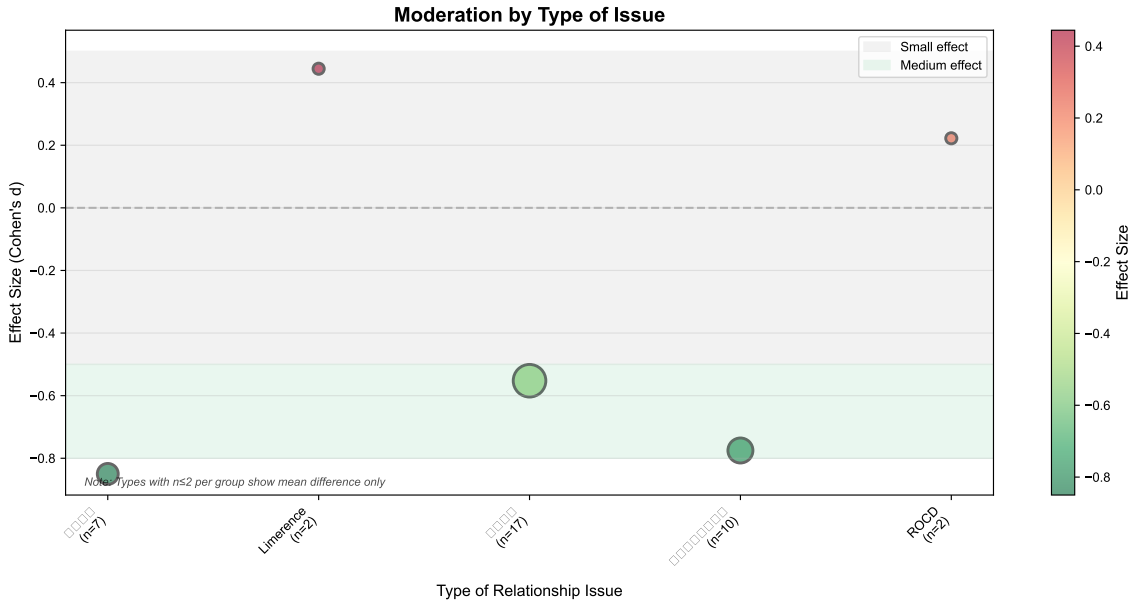

Fig. 9. Moderation effects by type of relationship issue

Primary outcomes (RQ1). Across two days of exposure and a post-assessment on Day 3, we observed small, favorable effects on depression and anxiety symptoms (PHQ-9 d=0.40; GAD-7 d=0.23) alongside near-zero effects on dependency (ADS) and love-addiction intensity (LAI). Belief-level outcomes (RBI) were flat to slightly favoring control in this time window (d≈+0.18) (*Main Effects* panels; *Forest Plot*). These are not conventionally significant at N=20/arm and a 2-day active period, but they are directionally consistent with brief behavior-first interventions where **state-like distress improves earlier than trait-like beliefs**. In other words, the system appears to cool reactivity before it reshapes underlying schemas—exactly the sequence we engineered for (Explore → Persuade; micro-actions first).

Daily behaviors moved first. EMA counts show that *confirmation seeking*, *checking behavior*, *social-media checking*, and *reconnection attempts* decreased from Day1 to Day2 in both arms, with a demonstrably **steeper decline in the experimental group** (see slopes in *Behavioral Changes Over Time*), suggesting that typed micro-actions (e.g., delay, quota+verification, mute/timed viewing) can change outward behavior quickly, even when beliefs take longer to budge.

Process signals strengthened the story (RQ2). The experimental arm reported **higher execution of advice** (p=.047) with similar levels of global engagement and trust, indicating that the *quality* of actions—*doing the specific thing*—rather than generic enthusiasm, differentiated conditions (*Process Engagement*). The formal mediation model predicts a small, imprecise indirect effect (total indirect ≈ .007 for PHQ-9 change via execution/engagement), which is unsurprising given the short window and the conservative, single-mediator specification; nevertheless, the **process contrast is clear** and aligns with the intended mechanism: *micro-actions bridge conversation and outcome.*

Heterogeneity (RQ3). Exploratory moderation suggests **larger benefits for "boundary deficits" and "dysfunctional beliefs"** subgroups (d≈0.85 and 0.78; n's small), with a medium effect for **anxious attachment** (d≈0.55; n=17). This pattern is theoretically coherent: (i) scripts that externalize boundaries (e.g., DEAR MAN, refusal, time-boxing) translate quickly into observable actions; (ii) anxious attachment amplifies *verification loops* that are directly targeted

by delay+quota rules. Still, these are hypothesis-generating findings that require pre-registered tests with adequate per-type power.

Co-movement across outcomes. Change scores correlated in theoretically sensible ways: PHQ-9 and GAD-7 moved together (r≈.66), and ADS and LAI co-varied most strongly (r≈.69) (*Correlation Matrix*). The *Waterfall Plot* further shows the experimental group's mean shift is more negative (i.e., greater improvement), with notable individual variability—an expected signature when "fit" moderates effect.

What the mechanism teaches us about persuasive systems

Actionability beats volume. The single robust between-arm difference was *execution of specific advice*. This aligns with implementation-intention theory and our architecture's choice to make *one* next action easy, time-boxed, and recordable. In short: **fewer, clearer, more enforceable steps** outperformed "more conversation." This is visible in the steeper Day1→2 behavior slopes and in user remarks summarized in the Results (e.g., "30-minute delay/distraction" as a "brake button"). The practical implication is to **treat "micro-action readiness" as a first-class state**: when IS (information sufficiency) is high and UT (tolerance) allows, route to Persuade with one minimum viable step; when not, route back to Explore. Our orchestrator implements exactly that monotone gating policy.

Matching matters—mostly for doing, not just liking. We designed dual entry (NFA/NFC) to calibrate *how* content arrives (affective vs. evidential) and typed orchestration to calibrate *what* arrives (boundary/ROCD/limerence/etc.). The data suggest that **perceived fit converts into execution** (higher advice uptake without higher generic "engagement"). This is persuasive-technology–specific: in high-affect contexts, *style* matching reduces resistance; *type* matching supplies scripts that feel relevant. Together they deliver *felt appropriateness → doing → outcome.*

Behavior before belief is not a weakness—it is a design principle. Our belief-level measures (e.g., RBI) barely moved over two days, while EMA behaviors and affective symptoms shifted. Rather than reading this as failure to "change minds," we interpret it as *right-sized dosing*: with limited contact, **cool the loop** (delay, quota, mute) first; pursue **cognitive reframing** when reactivity is lower (longer programs, spaced prompts). The forest plot's pattern—mental-state improvements with flat beliefs—fits this staged approach.

Design implications for responsible, personalized persuasion

(1) **Make micro-actions auditable and *countable.*** Count-level KPIs (execution %, check counts, reconnection attempts) are sensitive in brief trials and ethically easier to store than free text. They enable daily slope estimates that tell us whether an intervention is *doing anything* long before trait shifts surface. Our cards therefore insist on "1 behavior + 1 attitude" observables by design.

(2) **Orchestrate *when* to persuade, not only *what* to say.** The Explore→Persuade alternation, governed by safety gates and sufficiency/tolerance thresholds, prevented over-persuasion and reduced the "interview fatigue" some users reported in generic settings. This policy is simple to audit, tune, and replicate—an advantage for sensitive domains.

(3) **Engineer "cooling scripts" as first-line defaults for high-arousal types.** Subtypes that thrive on clear boundaries (boundary deficits, anxious attachment) responded best; systems should front-load delay/limit/mute *with* concrete time horizons and verification rules, then layer narrative/empathy or evidence framing per entry style.

(4) **Separate *fit* from *fluency*.** Our process metrics show similar engagement/trust across arms yet higher execution for the matched system. This suggests that chasing "nicer" conversations is less important than ensuring **the next step is the right step for this user and type**.
(5) **Keep safety first—and measurable.** Although safety outcomes were not the focus of this short trial, the LIVES-style triage and diversion are integral to deployments in relational contexts that can co-occur with digital coercion. The virtue of explicit gates is that *routing* itself becomes an evaluable object (triggers, compliance, latency) rather than an undocumented heuristic.

Ethical reflection

Persuasive agents in intimacy contexts risk *over-reach* (e.g., pressuring, moralizing, or offering advice when safety interventions are needed). We mitigate this by (i) **safety gating** (divert before persuading), (ii) **burden control** via JITAI pacing (tolerance-coupled thresholds), and (iii) **data minimization** (counts, durations; optional purge). The empirical pattern—more doing, not more talking—also moderates risk: micro-actions that slow a loop (delay, quota, mute) protect users in the moment without forcing them into unwanted disclosures.

Limitations and threats to validity

- **Duration and dose.** Two days of active exposure is modest for belief-level change; we therefore framed RQ1 around multi-dimensional *direction* of change and emphasized process outcomes. Future studies should deploy **7–14-day** protocols to test whether belief indices (RBI, ROCD) follow the behavioral lead once reactivity is cooled.
- **Power and multiplicity.** With n=20/arm and multiple outcomes, we are underpowered for small effects. The significant difference in execution (p=.047) is promising but should be treated as *confirmatory only in a pre-registered replication.*
- **Self-report and reactivity.** EMA relies on self-counts that can be biased by demand or social desirability. That said, **within-person slopes** are typically more robust than single post scores, and the convergence across independent behaviors lends credibility (*Behavioral Changes Over Time*).
- **Typing and entry screens.** The two-item NFA/NFC and lightweight typing are pragmatic in-product approximations. They likely introduce misclassification that attenuates matching effects—i.e., our moderation estimates are conservative.
- **Generalizability.** Participants self-selected into a brief, research-scaffolded program. Transfers to long-term use or higher-risk populations will require adapted safety resources and localized knowledge bases.

A blueprint for cumulative HCI evidence

Beyond point effects, a key contribution is **evaluability**: we operationalize the *units of persuasion* (typed cards with BCT/COM-B tags), their **observables** (execution %, behavior counts), and an **interpretable controller** that renders routing **auditable**. This turns "persuasion design" from a black box into a set of measurable, recombinable components—an enabler for replication, ablation, and meta-analysis across labs. Concretely, we recommend that future work report: (i) *dose* (component calls, time in Explore vs. Persuade), (ii) *execution* (If-Then uptake), (iii) *proximal behaviors* (counts), and (iv) *distal beliefs*—and model the mediation chain "**dose → execution → behavior → beliefs/symptoms**." Our study demonstrates that such a chain can be instrumented in real-world conversation systems with **minimal, desensitized logs**.

Future work

(1) **Longer horizon tests of belief change.** Extend to 7–14 days with spaced, low-burden prompts to evaluate whether cognitive indices (RBI/ROCD) follow the early behavioral improvements seen here.
(2) **Pre-registered, powered moderation.** Stratify by *boundary deficits*, *anxious attachment*, *dysfunctional beliefs* with sample sizes sufficient for type×group interactions; test whether different "first-line cards" (e.g., DEAR MAN vs. delay/verification rules) dominate by type.
(3) **Adaptive pacing policies.** Compare static thresholds vs. learned, user-specific pacing for Explore↔Persuade transitions; measure burden and completion alongside outcomes to ensure *calibrated* persuasion.
(4) **Objective signals and safety telemetry.** Where appropriate, pair self-counts with *on-device* timers/mute logs to reduce recall bias and quantify safety gate performance (trigger rate, time-to-diversion).
(5) **From fit to fluency to fairness.** Probe whether entry-style matching differentially benefits subgroups; ensure the same *safety* and *efficacy* are delivered across demographics regardless of style preferences.

8 Conclusion

We evaluated a safety-first, dual-entry (NFA/NFC), belief quick-scan routing, and typed-component orchestration system against a matched-dose, non-matching control in a randomized three-day study (N=40). The matched system produced small, favorable symptom changes (e.g., PHQ-9, GAD-7), a clear increase in advice execution, and steeper day-to-day declines in maladaptive behaviors, while belief-level measures moved little over this short window. Mediation patterns and process logs suggest a pragmatic pathway—micro-actions drive execution, execution shifts behavior, behavior opens headroom for later belief change—and exploratory moderation points to larger gains for boundary-related and belief-driven presentations, with medium gains for anxious attachment.

Beyond effects, this work contributes an auditable system lifecycle (LIVES → dual-entry → quick-scan → typed orchestration → EMA/replay), a literature-derived Causes→Strategies knowledge base codified as cards with observable metrics, and an evaluability blueprint (routing thresholds, backend objects, analysis plan) that treats persuasion not as a black box but as measurable, recombinable components. Together, these elements make replication, ablation, and cross-study synthesis feasible in sensitive relationship contexts.

For HCI, three design lessons stand out. First, micro-actions before reframing: brief deployments should prioritize enforceable next steps (delay, quota+verification, mute/timed viewing) and use cognitive work once reactivity cools. Second, matching matters for doing: style (entry) and content (type) alignment translate to execution more than to generic "engagement." Third, safety by design: explicit gates and minimal, count-level telemetry enable responsible operation and external audit. Limitations include short exposure and limited per-type power; we encourage longer trials, pre-registered moderation, objective burden/safety telemetry, and hybrid hand-offs to humans when appropriate. Overall, our study offers a practical, reproducible path to building responsible, personalized, and auditable conversational support for relationship distress.

References

[1] A. Alslaity et al. 2023. A panoramic view of personalization based on individual factors within persuasive technologies. *Frontiers in Artificial Intelligence* 6 (2023), 1125191. doi:10.3389/frai.2023.1125191
[2] Paul C Amrhein, William R Miller, Caroline E Yahne, Mark Palmer, and Leon Fulcher. 2003. Client commitment language during motivational interviewing predicts drug use outcomes. *Journal of Consulting and Clinical Psychology* 71, 5 (2003), 862–878.
[3] Melody Beattie. 1987. *Codependent No More: How to Stop Controlling Others and Start Caring for Yourself.* Hazelden Publishing.

[4] P. Bradbury, E. Short, and P. Bleakley. 2025. Limerence, Hidden Obsession, Fixation, and Rumination: A Scoping Review of Human Behaviour. *Journal of Police and Criminal Psychology* 40 (2025), 417–426. doi:10.1007/s11896-024-09674-x

[5] R.G. Cavalli, J. Feeney, G. Rogier, and P. Velotti. 2025. Conceptualizing love addiction within the attachment perspective: A systematic review and meta-analysis. *Journal of Behavioral Addictions* 14, 2 (2025), 611–629. doi:10.1556/2006.2025.00031

[6] S. Costa, N. Barberis, M.D. Griffiths, et al. 2021. The Love Addiction Inventory: Preliminary Findings of the Development Process and Psychometric Characteristics. *International Journal of Mental Health and Addiction* 19 (2021), 651–668. doi:10.1007/s11469-019-00097-y

[7] Robert J. DeRubeis and Steven D. Hollon. 2010. *Cognitive therapy for depression.* Guilford Press.

[8] Guy Doron, Dena Derby, Oren Szepsenwol, and Dana Talmor. 2012. Obsessive compulsive disorder and relationship-related obsessions. *Journal of Obsessive-Compulsive and Related Disorders* 1, 1 (2012), 16–24. doi:10.1016/j.jocrd.2011.11.002

[9] Helen E. Fisher, Arthur Aron, and Lucy L. Brown. 2006. Romantic love: a mammalian brain system for mate choice. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361, 1476 (2006), 2173–2186. doi:10.1098/rstb.2006.1938

[10] R. Chris Fraley, Niels G. Waller, and Kelly A. Brennan. 2000. Adult attachment and the dynamics of romantic relationships. *Journal of Personality and Social Psychology* 78, 2 (2000), 350–365. doi:10.1037/0022-3514.78.2.350

[11] K. Furumai, R. Legaspi, J. Vizcarra, Y. Yamazaki, Y. Nishimura, S. J. Semnani, K. Ikeda, W. Shi, and M. S. Lam. 2024. Zero-shot persuasive chatbots with LLM-generated strategies and information retrieval. *arXiv preprint arXiv:2407.03585* (2024). doi:10.48550/arXiv.2407.03585

[12] Leslie S. Greenberg. 2010. *Emotion-focused therapy.* American Psychological Association.

[13] C. Guan, J. Wang, L. Zhang, et al. 2025. A longitudinal network analysis of the relationship between love addiction, insecure attachment patterns, and interpersonal dependence. *BMC Psychology* 13, 330 (2025), 1–13. doi:10.1186/s40359-025-02605-3

[14] Elaine Hatfield and Susan Sprecher. 1986. The passionate love scale. *Journal of Personality and Social Psychology* 51, 3 (1986), 614–626. doi:10.1037/0022-3514.51.3.614

[15] Clyde Hendrick and Susan Hendrick. 1986. A theory and method of love. *Journal of Personality and Social Psychology* 50, 2 (1986), 392–402. doi:10.1037/0022-3514.50.2.392

[16] Irving L Janis and Leon Mann. 1977. *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment.* Free Press.

[17] Susan M. Johnson. 2008. Emotionally focused couple therapy: Status and challenges. *Current Opinion in Psychiatry* 21, 2 (2008), 152–156. doi:10.1097/YCO.0b013e3282f2c7e6

[18] Jon Kabat-Zinn. 2003. Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice* 10, 2 (2003), 144–156. doi:10.1093/clipsy.bpg016

[19] Marsha Linehan. 1993. *Skills training manual for treating borderline personality disorder.* Guilford Press.

[20] Patricia W. Linville. 1987. Self-complexity as a cognitive buffer against stress-related illness and depression. *Journal of Personality and Social Psychology* 52, 4 (1987), 663–676. doi:10.1037/0022-3514.52.4.663

[21] G. Alan Marlatt and Dennis M. Donovan. 2005. *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors.* Guilford Press.

[22] Pia Mellody and Andrea Wells Miller. 2018. *Facing Love Addiction: Giving Yourself the Power to Change the Way You Love.* HarperOne.

[23] L. Metzger, L. Miller, M. Baumann, and J. Kraus. 2024. Empowering Calibrated (Dis-)Trust in Conversational Agents: A User Study on the Persuasive Power of Limitation Disclaimers vs. Authoritative Style. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–19. doi:10.1145/3613904.3642122

[24] William R Miller and Stephen Rollnick. 2013. *Motivational Interviewing: Helping People Change* (3rd ed.). Guilford press.

[25] Susan Peabody. 2005. *Addiction to love: Overcoming obsession and dependency in relationships.* Celestial Arts.

[26] James O Prochaska and Carlo C DiClemente. 1984. *The Transtheoretical Approach: Crossing Traditional Boundaries of Therapy.* Dow Jones/Irwin.

[27] F. Salvi, M. Horta Ribeiro, R. Gallotti, and R. West. 2025. On the conversational persuasiveness of GPT-4. *Nature Human Behaviour* (2025). doi:10.1038/s41562-025-02194-6

[28] Sex and Love Addicts Anonymous. 2015. *Twelve step program for recovering from sex and love addiction.* SLAA Fellowship-Wide Services.

[29] C.M. Sirvent-Ruiz, M.V. Moral-Jiménez, J. Herrero, M. Miranda-Rovés, and F.J. Rodríguez-Díaz. 2022. Concept of Affective Dependence and Validation of an Affective Dependence Scale. *Psychology Research and Behavior Management* 15 (2022), 3875–3888. doi:10.2147/PRBM.S385807

[30] L. Spann and J. Fischer. 1990. Development of the Spann-Fischer Codependency Scale. *Family Therapy* 17 (1990), 39–45.

[31] Dorothy Tennov. 1979. *Love and limerence: The experience of being in love.* Stein and Day, New York.

[32] V. U. Wanniarachchi et al. [n. d.]. Personalization variables in digital mental health interventions for depression and anxiety in adolescents and youth: a scoping review. *Journal of Medical Internet Research* 27, 5 ([n. d.]), eXXXXXX.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009