Can Emotional Reciprocity with AI Companions Support Healthy Habit Formation? A Conceptual Framework and Trial Protocol

Pimsai Chichareon[1]*, Polkit Sangvanich[1] , Rapinpat Yodlorchai[2]

[1]Technology Management and Innopreneurship (Interdisciplinary Program), Chulalongkorn University, Bangkok 10330, Thailand
[2]Developmental Psychology, Faculty of Psychology, Chulalongkorn University
**Corresponding Author E-mail: pimsai@outlook.com and pimsai@chichareon.com*

ABSTRACT

Preventive health outcomes depend not only on initiating beneficial behaviours but on sustaining them long enough for repeated actions to become increasingly automatic. Digital health tools can extend behavioural support beyond conventional healthcare settings, yet engagement often declines after initial adoption. This paper proposes perceived emotional reciprocity as a mechanism through which relationally designed AI companions may support healthy habit formation. Perceived emotional reciprocity refers to the user's felt experience that an AI companion receives, understands, remembers, validates, and responds to personal needs, difficulties, and progress in a contextually relevant way across repeated interactions. The conceptual framework draws on post-adoption technology research, attachment theory, Self-Determination Theory, relational responsiveness, and habit-formation research. A 12-week, two-arm randomized controlled trial is proposed among AI-familiar adults who have identified a non-clinical healthy habit that they have previously attempted but struggled to sustain. Participants will be randomized to a High Emotional Reciprocity (HER) condition or a Standard (STD) supportive AI condition delivered through LINE. Both conditions provide supportive health information through the same platform, while the HER condition adds relational continuity, personalised responsiveness, emotional validation, progress acknowledgement, and longitudinal awareness. The primary model tests a serial pathway from experimental condition to early perceived emotional reciprocity, subsequent engagement consistency, and Week-12 habit automaticity. Recruitment and data collection will begin only after research ethics approval; therefore, this paper presents the conceptual framework and prespecified trial protocol rather than empirical outcomes. The proposed contribution is a shift from evaluating AI-supported health behaviour primarily by engagement volume toward evaluating whether relational AI design helps users maintain the repeated behaviour required for habit automaticity and, ultimately, behavioural autonomy with reduced reliance on external support.

Keywords: AI companions/ emotional reciprocity/ healthy habit formation/ digital health/ behavioural autonomy

1. INTRODUCTION

Most people already know that sleep, physical activity, healthy eating, reading, and mindfulness can improve their wellbeing. The difficult part is sustaining them long enough to become habitual. Healthy habits are not formed by intention alone. They require repeated performance until the behaviour becomes increasingly automatic. In a 12-week real-world habit-formation study, Lally et al. showed that automaticity increased with repeated performance but varied substantially across individuals and behaviours [1]. A later systematic review and meta-analysis likewise found considerable variation in the time required for health behaviours to become habitual [2]. The practical challenge is therefore continuity during the period in which the target behaviour still requires deliberate repetition and remains vulnerable to stress, competing demands and routine disruption.

Digital health tools appear well suited to this continuity problem because they can provide support outside formal healthcare settings and at the time a behavioural decision is made. Their effectiveness, however, depends partly on whether people continue to use them. A scoping review of free-living lifestyle and mental-health apps found a pronounced early abandonment pattern; across the included evidence, a median of 70% of users had discontinued use within the first 100 days [3]. This creates an adoption-continuity gap. Technology Acceptance Model research explains initial acceptance through beliefs such as perceived usefulness and perceived ease of use [4,5], but acceptance does not

by itself explain what happens after adoption when behavioural benefit depends on repeated action over time.

Conversational AI may offer a different form of behavioural support from static applications because interaction can be responsive, personalised, and repeated over time. Systematic reviews indicate that AI-based conversational agents can produce short-term improvements in mental-health outcomes and that chatbot interventions can improve selected lifestyle behaviours, including physical activity, fruit and vegetable intake, sleep duration, and sleep quality [6,7]. At the same time, attrition remains a recognised problem in conversational-agent interventions [8]. Evidence that an intervention can produce an outcome under study conditions therefore does not resolve the continuity problem: the user must still return, engage, and repeat the target behaviour for long enough for habit automaticity to develop.

The unresolved question is what kind of AI interaction may support that continuity. Existing work has examined efficacy, acceptance, usability, and attrition, while relational qualities such as empathy, responsiveness, and perceived caring have increasingly been recognised in human-AI interaction. The doctoral research underlying this paper identifies a more specific gap: perceived emotional reciprocity has not yet been experimentally manipulated and tested as a mediating mechanism linking AI companion design to engagement consistency and subsequent habit automaticity. The present paper therefore develops perceived emotional reciprocity as the proposed mechanism and specifies a randomized trial in which the relational-support architecture is manipulated while the underlying platform and health-support function are held constant.

2. METHODOLOGY

2.1 Conceptual Framework Development

This paper draws directly from the theoretical development of the doctoral research and integrates four related bodies of literature: technology adoption and post-adoption use, attachment and relational psychology, Self-Determination Theory, and habit formation. The purpose is not to infer an empirical effect before data collection, but to derive a mechanism that can be tested experimentally. The framework treats prior AI adoption as a boundary condition: participants are already familiar with AI, allowing the study to move beyond the question of whether users will initially accept the technology and focus instead on whether the quality of repeated interaction can support behavioural continuity.

2.2 Perceived Emotional Reciprocity and the Proposed Pathway

Attachment theory provides the broad rationale that consistent and responsive relational support can function as a secure base when individuals face challenges [9]. Self-Determination Theory adds the importance of relatedness and autonomous motivation in sustained behaviour [10]. The Interpersonal Process Model of Intimacy is more specific about the interaction itself, identifying perceived understanding, validation, and caring as central to responsive relational exchange [11]. These theories were developed for human relationships and are not treated here as proof that an AI companion produces the same relational effects. Rather, they provide theoretical components from which a testable human-AI construct is derived. Experimental human-AI research provides a complementary precedent: Pataranutaporn et al. showed that framing the same AI as caring, manipulative, or neutral altered users' perceptions of its trustworthiness, empathy, and effectiveness, and also changed how users interacted with it [12]. This supports the narrower proposition that perceived relational orientation can shape the experience of an otherwise similar AI interaction; it does not by itself establish an effect on habit formation.

Perceived Emotional Reciprocity (PER) is defined in the doctoral framework as the user's felt experience that the AI companion engages in a meaningful reciprocal relational process: expressed needs, struggles, and progress are received, understood, validated, and remembered; responses are genuinely personalised to the individual's context; and concern for the user's wellbeing is demonstrated

across interactions. PER is therefore related to, but distinct from, warmth, empathy, and personalization alone. A response can be warm yet generic, empathetic yet disconnected from previous conversations, or personalised at the content level without creating a sense that the user is known across time. The distinguishing feature of PER is relational continuity: one supportive exchange is insufficient; the experience is expected to emerge from a pattern of understanding, memory, validation, and context-sensitive response across repeated interactions.

The proposed mechanism links this relational experience to the behavioural process required for habit formation. The primary pathway is experimental condition (HER vs STD) → perceived emotional reciprocity → engagement consistency → habit automaticity. PER is positioned as the first mediator because the manipulation is intended to change how the interaction is experienced. Engagement consistency is positioned second because the framework proposes that feeling understood and remembered may make users more willing to return to the companion during the effortful period of behaviour repetition. Habit automaticity is the endpoint because the purpose of the intervention is not to maximise AI use indefinitely; it is to support repetition until the target behaviour becomes less dependent on deliberate effort and external prompting.

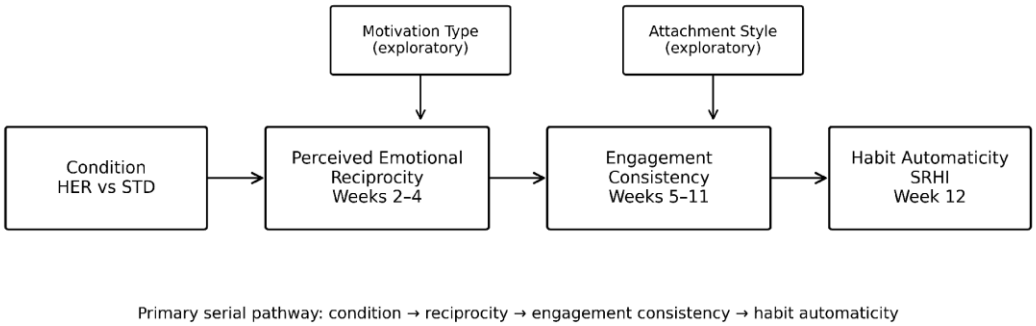

Figure 1. Proposed mechanism linking AI companion design to healthy habit formation.

2.3 Trial Design and Participants

The proposed study is a 12-week, two-arm parallel randomized controlled trial. Adults aged 20–45 years will be eligible if they use at least one AI tool several times per week, regularly use LINE, can participate for 12 weeks, and identify a target habit they have previously attempted but not sustained. Permitted habits are sleeping before midnight, reading at least 20 minutes per day, meditating at least 10 minutes per day, adding one serving of fruit or vegetables per day, or completing at least 20 minutes of moderate physical activity per day. Current clinical-level depression or anxiety that substantially affects daily functioning and ongoing psychiatric treatment that could interact with the intervention are exclusion criteria. Initial recruitment of 100–120 participants is planned to retain approximately 60–80 complete cases, allowing for 33–40% attrition over 12 weeks. Mediation analysis requires sufficient statistical power to detect indirect effects, and required sample size depends substantially on the magnitude of the component paths [15]. For serial mediation models, power is further influenced by the combined magnitude of the component paths, with smaller component effects requiring larger samples [16]. The completed sample target of 60–80 is therefore intended to provide an initial test of the primary serial mediation pathway. The serial indirect effect remains the prespecified primary test. Each component path will be reported with its bootstrap confidence interval, while the more statistically demanding moderation analyses are treated as exploratory.

Randomization will use a computer-generated sequence stratified by habit domain and baseline autonomous motivation. Participants will not be informed of their condition. The underlying language

model, LINE interface, onboarding, survey schedule, habit domains, and evidence-based informational content will be held constant across arms. The systematic difference is the relational-support architecture. HER combines conversational continuity, structured memory, personalised responsiveness, emotional validation, progress acknowledgement, expressed concern for wellbeing, and adaptive response to the participant's circumstances. STD remains friendly, accurate, and supportive but does not have access to the participant's longitudinal profile and therefore cannot build the same continuity from earlier conversations. STD is consequently an active control rather than a no-treatment condition: both groups receive the same platform and health-support function, while the integrated relational design intended to induce PER is manipulated.

2.4 Measures and Analysis Plan

Habit automaticity is the primary outcome and will be measured at baseline and Week 12 using the automaticity component of the Self-Report Habit Index (SRHI) [13]. Perceived emotional reciprocity will be measured weekly from Week 2 using a six-item study scale reflecting understanding, validation, personalization, continuity, concern for wellbeing, and reciprocal exchange. Because PER is operationalized using a study-specific six-item scale, its measurement properties will be assessed before the main trial. The instrument will undergo Thai translation and back-translation, followed by review by three subject-matter experts for conceptual relevance, clarity, and coverage of the proposed construct. A pilot study of approximately 30 participants will examine item performance, preliminary internal consistency, and the proposed measurement structure of the six-item scale. Given the modest pilot sample, the factor-analytic assessment will be interpreted as preliminary evidence of structural validity, and the small sample size will be acknowledged as a limitation. Internal consistency will also be assessed during the main study at each administration during Weeks 2–4, and the reliability of the PER composite used in the primary mediation analysis will be reported. The primary PER mediator is the mean score across Weeks 2–4. Engagement consistency will be derived from passive LINE logs across Weeks 5–11 using user-initiated message frequency, active session days, and response latency. The temporal separation is deliberate: experimental condition precedes early PER, early PER precedes the subsequent engagement window, and both precede Week-12 habit automaticity. This ordering aligns measurement with the hypothesised causal sequence and reduces reliance on variables measured at the same time. Using objective interaction logs for engagement consistency also reduces reliance on self-reported measures across the mediation pathway, helping to limit potential common-method bias.

The primary analysis will use Hayes PROCESS Model 6 in SPSS [14]. Condition will be coded STD = 0 and HER = 1. The prespecified primary test is the serial indirect effect through PER and engagement consistency, estimated with 5,000 bootstrap samples and a 95% bootstrap confidence interval. Baseline SRHI, AI Relational Comfort, habit-domain dummy variables, age, and resource availability will be included as prespecified covariates. Motivation type and attachment dimensions will be examined as exploratory moderators rather than as part of the confirmatory serial mediation. A separate exploratory analysis will examine whether habit-related AI message frequency declines across Weeks 8–12 among participants who develop stronger habit automaticity. This latter analysis addresses behavioural autonomy: if the habit becomes more automatic, continued high reliance on the companion should not be treated as a necessary indicator of intervention success.

2.5 Ethics and Data Governance

Recruitment and research data collection will not begin before approval from the appropriate Chulalongkorn University Research Ethics Committee. Participants will provide informed consent for conversational-data collection and may withdraw without penalty. The AI companion is not positioned as a clinical service or a substitute for professional care. A prespecified distress protocol will suspend routine habit coaching and provide safety guidance if messages indicate elevated distress or possible

self-harm risk. Conversation data will be pseudonymized, and the identity linkage file will be stored separately from research data.

3. RESULTS AND DISCUSSION

3.1 Proposed Conceptual Outcome

Because this is a conceptual framework and trial-protocol paper, there are no participant results to report. The conceptual outcome is a testable framework for how relational AI design could influence habit formation without treating engagement itself as the ultimate outcome. Figure 1 distinguishes the experimental manipulation, the user's perceived relational experience, subsequent behavioural engagement, and habit automaticity as separate constructs. This distinction is important: a system designed to appear relational must first produce a measurable change in PER, and greater interaction volume is meaningful only if it helps support the repeated behaviour required for habit formation.

3.2 Contribution to Social Science and Public Health

The framework contributes to social science and public health by shifting the evaluative question from initial technology acceptance toward behavioural continuity after adoption. This distinction matters for health-support technologies because adoption is only an entry point: benefits that depend on repeated behaviour require continuity while the behaviour is still effortful.

The first contribution is to distinguish HER as the design manipulation from PER as the user's experienced response. This allows us to test whether the HER design actually leads to higher PER, creating a mechanism test rather than a simple comparison of two chatbot styles.

The second contribution is the temporal structure of the proposed mechanism. PER is assessed early, across Weeks 2–4; engagement consistency is then derived from objective interaction logs across Weeks 5–11; and habit automaticity is assessed at Week 12. This sequence reduces temporal ambiguity and avoids constructing the entire pathway from contemporaneous self-report. It does not make the mediators themselves randomized: the strongest causal inference comes from random assignment to HER versus STD, while the serial mediation analysis evaluates whether the subsequent data are consistent with the hypothesised pathway.

The third contribution concerns how engagement is interpreted. Retention, frequency of use, and time spent with a digital intervention can be useful process indicators while a habit is forming, but maximising them indefinitely would conflict with the intended health outcome. The proposed contribution is therefore a shift from evaluating AI-supported health behaviour primarily by engagement volume toward evaluating whether relational AI design helps users maintain the repeated behaviour required for habit automaticity and, ultimately, behavioural autonomy with reduced reliance on external support. The exploratory Weeks 8–12 analysis examines this possibility by asking whether stronger automaticity can coexist with declining habit-related AI use.

Habit formation may also vary across the five target behaviours. Sleep timing, reading, meditation, fruit and vegetable intake, and physical activity differ in complexity, effort, contextual cues, and opportunities for repetition, and substantial variation across behaviours has already been reported [1,2]. Randomization is stratified by habit domain so that both arms carry a similar mix of behaviours, and habit-domain dummy variables are included as covariates in the primary analysis. The study is not powered for definitive comparisons between individual habit domains; any domain-specific patterns will therefore be interpreted descriptively.

The proposed study is deliberately bounded. It focuses on AI-familiar Thai adults aged 20–45 and non-clinical health-supporting habits, uses LINE as the delivery platform, and evaluates a 12-week mechanism window rather than long-term maintenance. The 12-week duration is intended to capture the development and consolidation of automaticity while the proposed mediation pathway can still be observed; it is not a claim that habit maintenance is complete or that long-term durability has been established. PER is measured with a study-specific six-item scale, and its psychometric properties will

require further validation beyond the preliminary assessment planned in this study. The HER manipulation also combines several relational features rather than isolating memory, validation, personalization, or concern as separate interventions. A positive finding would therefore support the integrated relational-support design and the proposed mechanism, not prove that any single feature is independently causal.

4. CONCLUSIONS

Starting a healthy behaviour is not the same as forming a habit. This paper proposes perceived emotional reciprocity as a mechanism through which a relationally designed AI companion may support the repeated behaviour required for habit automaticity. The accompanying 12-week randomized trial is designed to test a temporally ordered pathway from HER versus STD condition, to early PER, to subsequent engagement consistency, and finally to Week-12 habit automaticity. By holding the underlying platform and health-support function constant while manipulating the relational-support architecture, the design tests whether the relational experience itself contributes to the proposed mechanism. The framework defines success in terms of healthy habit formation and behavioural autonomy rather than continued AI use. Empirical conclusions will be drawn only after research ethics approval, recruitment, data collection, and prespecified analysis.

Acknowledgement

The author acknowledges the guidance of the doctoral advisory and examination committees of the Technology Management and Innopreneurship Program, Chulalongkorn University.

References

[1] Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. European Journal of Social Psychology, 40(6), 998–1009. https://doi.org/10.1002/ejsp.674

[2] Singh, B., Murphy, A., Maher, C., & Smith, A. E. (2024). Time to form a habit: A systematic review and meta-analysis of health behaviour habit formation and its determinants. Healthcare, 12(23), 2488. https://doi.org/10.3390/healthcare12232488

[3] Kidman, P. G., Curtis, R. G., Watson, A., & Maher, C. A. (2024). When and why adults abandon lifestyle behavior and mental health mobile apps: Scoping review. Journal of Medical Internet Research, 26, e56897. https://doi.org/10.2196/56897

[4] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

[5] Holden, R. J., & Karsh, B.-T. (2010). The Technology Acceptance Model: Its past and its future in health care. Journal of Biomedical Informatics, 43(1), 159–172. https://doi.org/10.1016/j.jbi.2009.07.002

[6] Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. npj Digital Medicine, 6, 236. https://doi.org/10.1038/s41746-023-00979-5

[7] Singh, B., Olds, T., Brinsley, J., Dumuid, D., Virgara, R., Matricciani, L., Watson, A., Szeto, K., Eglitis, E., Miatke, A., Simpson, C. E. M., Vandelanotte, C., & Maher, C. (2023). Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. npj Digital Medicine, 6, 118. https://doi.org/10.1038/s41746-023-00856-1

[8] Jabir, A. I., Lin, X., Martinengo, L., Sharp, G., Theng, Y.-L., & Tudor Car, L. (2024). Attrition in conversational agent-delivered mental health interventions: Systematic review and meta-analysis. Journal of Medical Internet Research, 26, e48168. https://doi.org/10.2196/48168

[9] Bowlby, J. (1969). Attachment and loss: Vol. 1. Attachment. Basic Books.

[10] Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. Guilford Press.

[11] Reis, H. T., & Shaver, P. R. (1988). Intimacy as an interpersonal process. In S. Duck (Ed.), Handbook of research in personal relationships (pp. 367–389). Wiley.

[12] Pataranutaporn, P., Liu, R., Finn, E., & Maes, P. (2023). Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. Nature Machine Intelligence, 5(10), 1076–1086. https://doi.org/10.1038/s42256-023-00720-7

[13] Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. Journal of Applied Social Psychology, 33(6), 1313–1330. https://doi.org/10.1111/j.1559-1816.2003.tb01951.x

[14] Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis (2nd ed.). Guilford Press.
[15] Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. Psychological Science, 18(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x
[16] Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. Structural Equation Modeling: A Multidisciplinary Journal, 17(3), 510–534. https://doi.org/10.1080/10705511.2010.489379