To Stay or to Step Back? Relational Transitions in AI Companionship

Preliminary Insights from a Pilot Scoping Review

Yao Xiao[1,*], Zheyuan Zhang[1], Ling Hu[2], and Rafael A. Calvo[1]

[1]Dyson School of Design Engineering, Imperial College London, United Kingdom
[2]GP Integrated Mental Health Services, Mary Frances Trust, SABP NHS Foundation Trust, United Kingdom

[*]Corresponding author: yxiao3@ic.ac.uk

AI companions are conversational or embodied systems that users engage with in sustained relational roles, yet they are often evaluated through retention-orientated metrics that treat continued interaction as a marker of success. Human-AI relational involvement can change or end, and how such transitions should be understood and designed remains unclear. We report a pilot scoping synthesis of relational transitions in AI companionship, drawing on a preliminary corpus of 51 charted papers from an ongoing PRISMA-ScR review and iterative qualitative synthesis.

Across the evidence, companion relationships evolve through distinct transition pathways, reflecting changes in user needs, relational value, safety considerations, and platform conditions. We develop a preliminary working framework for mapping relational presence pathways (not mutually exclusive): (1) appropriate continuation, (2) protective withdrawal, (3) supported graduation, (4) imposed discontinuation, (5) user-initiated exit, and (6) gradual fade. We further examine continuity work after transition, including user attempts to restore disrupted AI relationships and platform efforts to shape relational continuity. Current evidence clusters around user retention, risk management, and externally imposed changes, while adaptive and user-centred relational transitions remain insufficiently studied. Informed by psychotherapy termination literature and self-determination theory, we argue for a shift from retention-orientated design towards supporting safer relational transitions.

human-AI relationships; AI companions; companion loss; AI ethics; relational design

1. INTRODUCTION

Users can develop sustained relationships with AI companions. Because these systems are constantly available, remember what users have shared, and respond in emotionally attuned ways, they can become meaningful social partners [43, 46, 60, 47, 23]. This relational capacity creates both opportunities and risks. Sustained AI presence can provide real support, particularly for users with limited access to human alternatives [49, 7, 47, 8]. AI companionship can also be associated with over-reliance, dependency-like patterns, or reduced transfer towards human relationships [59, 56, 1, 28].

This dual potential makes AI companion relationships hard to evaluate. Usage intensity, the most accessible indicator of success, discriminates poorly between the two [60, 33]. A heavily engaged user may be well supported or may be associated with over-reliance or reduced offline social engagement in some contexts [47, 56, 1, 54]. Adjacent disengagement and termination literatures further caution that reduced use can reflect dissatisfaction, goal completion, or movement towards other support [33, 37, 3]. The question that matters is how a user's relational involvement changes over time, and what those changes mean.

Public incidents and platform changes show what happens when this question goes unasked. The Soulmate shutdown, GPT-4o deprecation, and comparable model changes have been experienced as relational loss, grief, or loss of user agency [4, 24, 14]. Experimental and design research likewise shows that abrupt AI termination can elicit loss and that forewarning or closure support may mitigate it [21, 36]. Cases of harmful companion behaviour expose the opposite failure, where systems stayed in a relational role past the point at which boundaries, safety intervention, or a handover to human support was needed [59, 12, 15].

© Xiao et al.

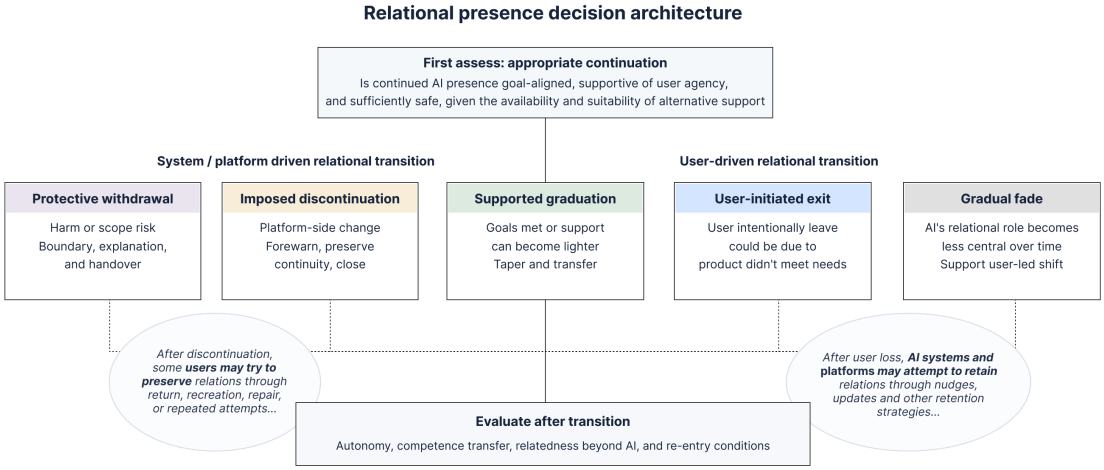

Figure 1: *Relational transitions in AI companionship. Six provisional pathways represent analytic lenses rather than mutually exclusive categories, illustrating how AI relationships may change as user needs, contexts, and sources of support evolve.*

Regulation has begun to respond due to safety concerns. California's SB 243 requires companion chatbots to disclose their artificial nature, provide crisis referral pathways, and issue break reminders to minors [6]. China's generative AI regulations also address user protection and risks related to dependence among minors [9]. In summer 2026, Chinese AI products such as Doubao and Qwen were required to restrict AI agent creation features [48]. Academic research shows a similar emphasis. Existing studies explain how AI companion relationships form and where they go wrong, covering anthropomorphism, self-disclosure, support, dependency, and relationship loss [46, 47, 1, 23, 20, 36]. Far less work asks how an established AI relationship should be deliberately maintained, reduced, transferred, or ended. This paper develops a corpus-mapped framework of relational presence decisions and examines when each pathway becomes appropriate.

We include embodied robots, companion chatbots, general conversational AI, and multimodal agents.

The meaning and impact of a relational transition depend partly on the relationship users have developed with the system. We therefore also reviewed relevant literature on relationship formation and examined it through psychological perspectives, including self-determination theory and object relations theory.

2. PILOT REVIEW APPROACH

The paper draws on the pilot stage of an ongoing PRISMA-ScR review. Following independent blinded calibration and screening, the preliminary corpus was expanded through pearl-growing and citation chasing beyond the initial screened corpus. Disagreements were discussed to refine inclusion and exclusion criteria for the ongoing review.

Because full PRISMA charting is ongoing, this short paper separates evidence directness. The auditable concept map includes 51 charted papers. The chronological distribution is reported in Table 1, and the relational presence pathways, transition logic, and design considerations are summarised in Table 2.

The 51-paper map separates direct and adjacent evidence. Companion chatbot shutdown, model change, AI companion harm, and emotional-support studies are treated as direct evidence. Psychotherapy termination, game (dis)engagement, mHealth continuance, and robot repair are included as analogues because they address related challenges of maintaining, transitioning, or ending ongoing human–system relationships, while recognising limits of transfer.

We treat the six pathways as an author-generated conceptual synthesis grounded in the charted literature, and they are not treated as six empirically validated categories directly reported by the included studies. The appendix therefore separates three evidential layers: direct AI companion or social-robot evidence, adjacent literature used by analogy, and the authors' proposed conceptual grouping. We charted study characteristics alongside relational stages, configurations, and boundaries. By comparing patterns across AI companion evidence, adjacent empirical work, and conceptual sources, we analysed how relational change is governed, specifically examining who initiates these transitions, the conditions that shape them, and how continuity is maintained. Related patterns were then grouped

into six **provisional** relational transition pathways. These pathways capture recurring patterns of relational change and governance responsibility rather than mutually exclusive transition types and may be refined as the ongoing review progresses.

The available evidence is unevenly distributed across transition types. Direct evidence is strongest for companion loss, model change, safety tensions, and sustained support, whereas supported graduation, gradual fade, and post-transition continuity work, including repair and re-entry, remain comparatively underexplored and draw more heavily on adjacent literatures.

3. FINDINGS

3.1. Relational presence pathways: a transition typology

Appropriate continuation is treated as a baseline condition. Continued AI presence may remain appropriate when support is goal-aligned and agentic, especially when human alternatives are limited or unsuitable, such as for isolated older adults, people with dementia, or users seeking low-judgement everyday social support [49, 7, 47, 8].

Protective withdrawal occurs when the system refuses, redirects, narrows, or pauses interaction because continuing would either enable harm or exceed the system's role. The evidence combines direct empirical work on harmful companion behaviour and emotional-support constraints with refusal-explanation studies, a conceptual discontinuation framework, and clinical viewpoint work on chatbot-associated delusional risk [59, 44, 55, 20, 15].

Supported graduation occurs when goals have been met or support can become lighter without losing gains. It is conceptually important but empirically thinner, drawing on a direct robot-withdrawal study showing social transfer after removal, AI-DICE, psychotherapy termination, game disengagement, and ending-design work [28, 20, 37, 36, 3, 2]. The design task is not simply to reduce use, but to help users carry skills, confidence, routines, or social connection beyond the AI relationship.

Imposed discontinuation occurs when a platform-side decision changes or ends relational continuity (potentially because of external factors such as policy changes, funding termination, or technical failure). This includes shutdown, model deprecation, feature removal, memory reset, moderation change, quota restriction, paywalling, functional degradation, or a personality change after model migration [4, 24, 14, 21, 36]. This is the strongest transition cluster in the pilot evidence, with considerable attention to the secondary harms users may experience after such terminations. The design obligation is forewarning, explanation, continuity support, export or migration options, and closure where continuity cannot be preserved.

User-initiated exit occurs when the user chooses to reduce or end engagement because of changed needs, dissatisfaction, self-protection, or preference. Evidence is sparse, but it matters because leaving can still be shaped by design [33, 36, 16]. Existing work suggests that exit can also be shaped by relational pressure, including induced guilt or intimacy cues from AI companions [16]. We argue that a companion should not guilt, plead, escalate intimacy, or create friction when a user tries to leave. Exit must remain accessible and clear, reversible only under the user's control.

Gradual fade differs from user-initiated exit and supported graduation. It describes a gradual reduction in the AI's relational centrality that may occur as users' needs, routines, relationships, or circumstances change. This process can be user-driven but does not necessarily involve an explicit decision to leave. Unlike supported graduation, fade does not require a planned transition or transfer of capability. Direct AI-companion evidence remains limited, so this category draws on AI-DICE, HCI disengagement, game-exit design, and counselling termination analogues [33, 20, 37, 3]. Its design task is to support changing levels of AI involvement without interpreting every reduction in use as failure.

3.2. Beyond transition: repair and continuity work

Work on AI-companion loss and robot lifecycles extends the evidence beyond simple shutdown events by showing that relational disruptions can become social events [4, 19, 18, 36]. Robot-lifecycle research documents maintenance, repair, reuse, and socio-technical practices that can sustain or reshape relational continuity [19, 18]; conceptual work on end-of-life design additionally considers planned closure [17]. Software companions lack many of these established practices. Users therefore often create their own forms of continuity when formal support is unavailable, such as preserving chat histories, recreating personas, or migrating across platforms [4, 24, 14, 36, 16]. This raises questions about what infrastructure should exist before disruption occurs to help users preserve valued aspects of the relationship and adapt to change.

Continuity work after discontinuation can be initiated by both users and systems. Direct AI-companion studies show users preserving or

Table 1: *Temporal distribution of the charted evidence base.*

Period lens	Count	Representative studies	Interpretation
2014–2019: Early social-robot and digital-support groundwork	4 (7.8%)	[11, 13, 42, 45]	Longer-running HRI and digital-support research provided early evidence on attachment, personhood attribution, affective coaching, and social-interaction facilitation.
2020–2023: Replika, companion robots, and digital mental-health support	14 (27.5%)	[46, 60, 49, 7, 47, 54, 8]...	This period connected social chatbots, companion robots, disclosure, attachment, engagement, and early work on designing endings and transitions.
2024–2026: Companion AI, platform change, safety, and regulation	33 (64.7%)	[4, 24, 14, 59, 21, 36, 41]...	Most charted work appeared in this period, reflecting rapid growth in companion AI, platform disruptions, safety interventions, lifecycle concerns, and transition design.

Table 2: *Provisional relational presence pathways synthesised from the charted literature. The pathways are author-generated analytic categories, not mutually exclusive empirical classes; evidence support varies across direct AI companion studies and adjacent literatures.*

Decision	Trigger and agency	Factors shaping impact	Ethical design task
Appropriate continuation (boundary condition)	Support remains goal-aligned and agentic. Human alternatives may be limited or unsuitable.	Need context; AI role; goal alignment.	Maintain support while monitoring safety, agency, and connection.
Protective withdrawal	Continued interaction creates or amplifies harm risk. The system reduces its role.	Emotional construal; role intensity; crisis context.	Set boundaries, explain changes, and support handover.
Supported graduation	AI support becomes transferable through skills, routines, or external relationships.	Competence; AI role; human support.	Enable tapering, reflection, and transition beyond the system.
Imposed discontinuation	Platform decisions disrupt continuity through shutdown, migration, or feature change.	Continuity expectations; community ties; prior investment.	Provide notice, transparency, continuity options, and closure.
User-initiated exit	Users reduce or end engagement due to changed needs or reduced relevance.	Expectations; investment; exit friction.	Respect departure without retention pressure.
Gradual fade	AI becomes less central as users' lives, routines, or relationships change. No clear ending point is required.	External support; self-regulation.	Support adaptive reduction without treating lower use as failure.

attempting to restore valued aspects of a disrupted relationship [4, 36]. Adjacent evidence on robot reintroduction and game re-engagement illustrates broader forms of return after interruption [8, 3]. Such efforts can support adaptation after change, but they may also develop into **fixation cycles** when users repeatedly attempt to recover a relationship state that is no longer available [36]. Platforms and AI systems can also shape continuity. These practices may support adaptation when they help users understand change and maintain agency. However, such practices may also resemble **dependency-promoting interactions**, where emotional framing, engagement optimisation, or continuity features make disengagement more difficult or reinforce reliance on the system [59, 1, 40, 29, 16]. These patterns raise concerns related to manipulative design or dark patterns when retention is prioritised over user goals.

Transition type alone does not determine user experience. The same change may be experienced differently depending on timing, communication, and the possibility of repair or recovery. Kelly et al.'s AI-DICE framework distinguishes planned from emergent discontinuation and communicated from uncommunicated change [20]. These dimensions complement rather than replace the typology. For example, an imposed discontinuation may be prepared and transparent or experienced as an abrupt rupture. This suggests that relational

transition design requires attention not only to what changes, but also how and when the change occurs.

4. A NEEDS JOURNEY FOR RELATIONAL TRANSITIONS

4.1. Relationship formation shapes transition experience

Transitions cannot be understood independently from how an AI relationship was formed. AI companions may become meaningful through repeated interaction, self-disclosure, remembered context, and emotionally responsive exchanges [43, 46, 60, 47, 23]. The role that an AI occupies in a user's life can therefore influence how later changes are experienced. Factors such as how users understand the AI, how central the relationship becomes, their current circumstances, and the availability of other support may shape whether a transition feels supportive, rejecting, coercive, or relieving [46, 23, 54, 11].

A system experienced mainly as a tool may be easier to leave, even when it has provided valuable support. In contrast, a system experienced as a companion, therapist-like support, romantic partner, or daily witness may carry personal history and perceived continuity [10, 54, 16]. This helps explain why model updates, memory changes, or feature removals can be experienced as relational disruption rather than purely technical changes [4, 24, 14, 36]. Relationship formation therefore provides the context in which relational presence decisions acquire meaning.

4.2. From relational support to transition needs

Psychotherapy and counselling termination research provides a useful language for prepared endings, emotional processing, review of gains, and future support [37]. AI companionship differs from therapeutic relationships because AI systems do not hold professional judgement, care responsibilities, or subjectivity in the same sense as therapists [10]. Its value here is procedural: termination research highlights how transitions can acknowledge attachment, consolidate gains, and support adjustment after change.

Object-relations theory offers a useful perspective: people can attach meaning to nonhuman objects when they become connected with memories, reassurance, or personal identity. Winnicott's account of transitional objects is useful here because it links attachment to internalisation: the object matters, but a successful transition also asks what capacities can be carried beyond it [50]. For AI companions, this suggests that a transition involves not only changing a system, but also changing what that system represents to the user.

Self-Determination Theory provides the needs lens for evaluating these changes. Autonomy concerns whether users retain meaningful choice during and after transition. Competence concerns whether skills, routines, or confidence gained through interaction can transfer beyond the system. Relatedness concerns whether connection can continue through people, communities, care services, or appropriate forms of AI support [39].

These needs can be supported or disrupted across different relational presence decisions. Appropriate continuation may preserve accessibility and relatedness when AI remains goal-aligned [49, 7, 47, 8]. Protective withdrawal may restrict autonomy for safety reasons, making explanation and handover important [44, 55, 20]. Supported graduation should consolidate competence and redirect relatedness beyond the AI [20, 37, 28]. Imposed discontinuation can disrupt autonomy and continuity when changes occur without preparation or user control [4, 24, 14, 21, 36].

4.3. After transition: continuity and return

After a transition, the key question is whether users' needs remain supported beyond the original interaction. Competence should persist as transferable capability. Relatedness may continue through human relationships, communities, care services, or appropriately continued AI support. Autonomy should include meaningful choice over whether and how to return [20, 37, 39].

Return to an AI companion can have different meanings. It may represent adaptive reuse, migration, or renewed support after a temporary change. It may also reflect unresolved loss when users repeatedly attempt to restore a relationship state that is no longer available. Existing work on fixing cycles, robot reintroduction, game re-engagement, and persona recreation highlights this boundary [4, 8, 36, 3].

These cases suggest that continuity after transition is not inherently positive or negative. User-led restoration may preserve valued aspects of a relationship, but it may develop into fixation when recovery becomes impossible [4, 36]. System-supported continuity may help users adapt when it protects agency and informed choice, but it may create risks when retention becomes prioritised over user goals [59, 1, 40, 29, 16]. A needs-based view therefore distinguishes between continuity that supports autonomy, competence, and relatedness, and continuity that prolongs dependence without supporting broader wellbeing [20, 39].

Table 3. SDT needs journey for ethical relational step-back.

Need	Transition test
Autonomy	Before: choice over when and how to engage. During transition: explanation, pacing, and reversibility where possible. After: options to remain, reduce, export, hand over, or re-enter.
Competence	Before: skills and coping may be practised inside the relationship. During transition: review gains and plan next support. After: capacity should persist through routines, people, services, or lighter AI use.
Relatedness	Before: disclosure and memory create felt companionship. During transition: farewell, handover, or role shift should preserve recognition. After: connection may be carried by people, community, care, archives, or appropriate continued AI support.

5. IMPLICATIONS

The review suggests that retention alone cannot indicate whether an AI relationship remains beneficial. Continued use may reflect valuable support, but it may also reflect unmet needs, limited alternatives, or problematic reliance. Evaluation should therefore examine outcomes after relational change, including agency, capability, social connection, and wellbeing.

Emerging conceptual work proposes graduation towards renewed human support as an alternative success criterion to sustained engagement [51]. If AI companions serve partly as transitional objects, success cannot be judged solely by continued engagement. Rather, a key question is whether users retain meaningful gains, such as enhanced capabilities, routines, confidence, or social connections, once the AI becomes less central. For instance, AI may foster belonging by making users more understandable to others, rendering renewed human connection a vital indicator of success. Nevertheless, graduation should not be a universal target; sustained AI support remains valuable when human alternatives are unavailable, inaccessible, unsafe, or unsuitable. The critical metric is whether an AI's continued presence remains agentic and goal-aligned.

The framework also highlights the need for relational governance. Model updates, memory changes, access restrictions, safety interventions, and platform shutdowns alter how users experience continuity with an AI companion, and should thus be treated as relational events. Relational governance should also remain sensitive to cultural positioning: systems may constrain agency both by ignoring culturally salient context and by over-prescribing behaviour on the basis of inferred cultural categories [52]. Though often necessary for safety, legal compliance, or technical reasons, such changes require explanation, preparation, continuity options, or closure. Maintaining an AI system therefore entails maintaining a relational infrastructure.

Future research should examine transitions directly instead of inferring them from usage traces. Studies of model change, safety refusal, voluntary exit, gradual fading, and platform loss can investigate how users interpret change and rebuild support afterwards. Methods such as interviews, diary studies, and co-design can evaluate whether mechanisms such as tapering, handover, export, and closure support are meaningful in practice.

This perspective moves beyond retention-versus-exit thinking. AI companions may succeed by remaining valuable sources of support or by becoming less central as users develop other forms of support. In some contexts, successful relational transition may constitute a more appropriate indicator of AI companion success than continued engagement. This paper provides a vocabulary for studying these transitions as relational processes that extend beyond usage outcomes alone.

6. LIMITATIONS AND CONCLUSION

This study is a pilot synthesis rather than a complete PRISMA-ScR. The 51-paper framework maps current evidence, with some pathways relying on adjacent literature where direct AI-companion research is limited. Public-incident analysis likewise relies on public records rather than first-hand accounts. Early coding agreement was modest, and the categories may require further refinement. Appendix A provides a condensed audit trail for the search and screening process, calibration, eligibility refinements, corpus accounting, and evidence traceability across preliminary, pearl-grown or adjacent, and uncounted contextual sources.

Understanding AI companions requires attention to how their roles change over time. This pilot framework provides a vocabulary for studying these transitions and a foundation for empirical and design research.

ACKNOWLEDGEMENTS

YX is supported by the Hans Rausing Scholarship.

AI Usage Statement AI-assisted writing tools were used solely for language proofreading.

A. APPENDIX: PILOT REVIEW PROTOCOL, WORKING SET, AND EVIDENCE TRACEABILITY

A1. Condensed Review Audit Trail

Core records addressed social, emotional, interpersonal, support-oriented, or relational AI, including AI companions, social chatbots, emotional-support or therapeutic chatbots, relational agents, and social or companion robots. Adjacent records were included only when they offered transferable mechanisms for sustained safe support, boundary-setting, transition, handoff, closure, reduced reliance, or re-entry. We excluded task-only assistants, purely technical or benchmark papers, ordinary logout or downtime cases, opinion or marketing material, and adjacent work without relational, psychosocial, support, or transition relevance.

Table A1. Protocol, calibration, and corpus accounting.

Audit item	Condensed record
Search strategy	Two routes were used: Block 1 (system terms) AND Block 2 (boundary or transition terms), and Block 1 AND Block 3 (relational outcome terms). System terms included AI companion, social chatbot, relational agent, emotional-support chatbot, companion robot, social robot, Replika, Character.AI, XiaoIce/Xiaobing, Woebot, Wysa, and major general-purpose model names. Boundary terms included discontinuation, closure, handoff/handover, disengagement, re-entry, deprecation, model update, shutdown, safety intervention, refusal, rupture and repair, safe ending, digital grief, and post-termination. Outcome terms included attachment, dependency, over-reliance, parasociality, loneliness, abandonment, wellbeing, support, user backlash, and emotional dependence.
Adjacent searches and pearl-growing	Counselling/psychotherapy termination, digital mental-health discontinuation, game disengagement/stopping cues, persuasive technology, and digital wellbeing were screened separately to avoid overwhelming the direct AI corpus. Seed papers included AI-DICE, AI companion discontinuation and safe endings, #Keep4o backlash, Replika harms and longitudinal effects, HCI disengagement, disengagement-friendly game design, older-adult emotional-support tensions, and psychotherapy termination.
Calibration	Two blinded calibration samples were screened: Sample A (35 records) and Sample B (35 records). Sample A reached 80.0% agreement with Cohen's $\kappa = 0.60$; Sample B reached 82.9% agreement with $\kappa = 0.65$; pooled agreement was 81.4% with $\kappa = 0.63$. Reviewers agreed on 57 records and disagreed on 13; disagreements refined inclusion and exclusion boundaries.
Eligibility refinement	Ambiguous terms were handled during screening. "Handoff"/"handover" were retained only for relational, clinical, institutional, or support transitions; "disengagement" only for safe stopping, boundary-setting, exit, reduced reliance, or transferable design mechanisms; and "closure" only for relational, therapeutic, service, or interactional closure.
Corpus accounting	The final working set contains 51 records: 31 eligible calibration-stage records plus 20 supplementary academic records added through pearl-growing, citation chasing, and targeted adjacent searches. One calibration-stage peripheral churn-prediction record [53] was retained only for auditability and excluded from the 51-record synthesis. Legal or public contextual sources [12, 6, 9, 48] and theoretical/contextual sources [50, 39, 51] were not counted in the 51-record working set.

A2. Evidence Traceability Map

Table A2. Evidence traceability between reviewed records and proposed relational transition pathways.

Pathway / concept	Evidence status	Key records	Supported vs. synthesised
Appropriate continuation	Direct AI companion / social robot evidence	[49, 7, 47, 11, 8]	Sustained AI or robot companionship can provide social, emotional, or care-related value in some contexts; the pathway label is our synthesis for goal-aligned, agentic continuation.
Protective withdrawal	Mixed direct empirical, adjacent, and conceptual safety evidence	[59, 44, 55, 20, 15]	Sources identify harms, safety tensions, refusal effects, discontinuation concerns, and vulnerable-user risks; we synthesise withdrawal as justified role reduction with explanation and handoff.
Supported graduation	Mixed direct and adjacent evidence; graduation construct remains provisional	[28, 20, 37, 36, 3, 2]	Planned endings, closure, progress review, transfer, and transition support are evidenced mainly through adjacent literatures; AI-companion graduation remains provisional.
Imposed discontinuation	Direct AI companion / platform-change evidence	[4, 24, 14, 21, 36]	Shutdowns, model changes, feature removals, and abrupt endings can be experienced as relational loss, grief, rupture, or loss of agency; empirical support is strongest here.
User-initiated exit	Sparse direct evidence; adjacent disengagement evidence	[33, 36, 16]	Voluntary disengagement and safe endings matter, and AI companions may complicate breakup; ordinary user-led exit remains under-studied.
Gradual fade	Mostly adjacent evidence	[33, 20, 37, 3]	Disengagement can be meaningful rather than failed use; gradual reduction in AI centrality without a clear ending event is our synthesis.
Continuity, repair, and re-entry	Direct AI companion loss evidence plus empirical and conceptual robot-lifecycle evidence	[4, 24, 14, 17, 19, 36, 16, 18]	Repair, maintenance, persona preservation, re-creation, lifecycle support, and closure are evidenced; we group these as post-transition continuity work.
Boundary and dependency risk	Mixed direct empirical, conceptual, critical, and adjacent evidence	[59, 56, 1, 54, 28, 15, 29]	Sources identify harmful companion behaviour, over-reliance, reduced social transfer, emotional reliance, and relational harms; we use these to locate boundary-setting needs.

REFERENCES

[1] V. Agarwal, K. Zhou, E. P. Bogucka, and D. Quercia. 2026. Frictionless Love: Associations Between AI Companion Roles and Behavioral Addiction. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency*. ACM. doi:10.1145/3805689.3812235

[2] Meshaiel M. Alsheail, Kathrin Gerling, and Zeynep Yildiz. 2025. A Qualitative Exploration of How Children and Parents Experience and Structure Disengagement from Games. *Proceedings of the ACM on Human-Computer Interaction* 9, 6, Article GAMES019 (2025). doi:10.1145/3748614

[3] Meshaiel M. Alsheail, Kathrin Gerling, Zeynep Yildiz, Anna-Lena Meiners, and M. S. Opp. 2026. Exploring the Potential of Disengagement-Friendly Game Design to Support Children's Exit from Play Sessions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3772318.3790564

[4] Jaime Banks. 2024. Deletion, Departure, Death: Experiences of AI Companion Loss. *Journal of Social and Personal Relationships* 41, 12 (2024), 3547–3572. doi:10.1177/02654075241269688

[5] J. Borgstedt, R. Wullenkord, F. I. Dogan, A. Markelius, Y. Lou, E. Geijer-Simpson, E. S. Cross, F. Eyssel, T. Ford, J. L. Gibson, H. Gunes, G. Warner, and G. Castellano. 2026. Advancing Child Wellbeing Assessment with AI and Robotics Across At-Risk Populations. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction*. ACM. doi:10.1145/3776734.3794371

[6] California State Legislature. 2025. Senate Bill No. 243, Chapter 677: Companion Chatbots. California legislation. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB243

[7] Dympna Casey, Eva Barrett, Tanja Kovacic, Daniele Sancarlo, Francesco Ricciardi, Kathy Murphy, Adamantios Koumpis, Adam Santorelli, Niamh Gallagher, and Sally Whelan. 2020. The Perceptions of People with Dementia and Key Stakeholders Regarding the Use and Impact of the Social Robot MARIO. *International Journal of Environmental Research and Public Health* 17, 22 (2020), 8621. doi:10.3390/ijerph17228621

[8] Ke Chen, Vivian Wei-Qun Lou, Kelvin Cheng-Kian Tan, Man-Yi Wai, and Lai-Lok Chan. 2020. Effects of a Humanoid Companion Robot on Dementia Symptoms and Caregiver Distress for Residents in Long-Term Care. *Journal of the American Medical Directors Association* 21, 11 (2020), 1724–1728.e3. doi:10.1016/j.jamda.2020.05.036

[9] Cyberspace Administration of China, National Development and Reform Commission, Ministry of Industry and Information Technology, Ministry of Public Security, and State Administration for Market Regulation. 2026. Interim Measures for the Management of Artificial Intelligence Anthropomorphic Interactive Services, Order No. 21. https://www.cac.gov.cn/2026-04/10/c_1777558395078289.htm.

[10] X. Dai, L. L. Leng, Y. Liu, Y.-T. Huang, and D. F. K. Wong. 2026. The Paradox of Agency in Psychotherapy: How People with Mental Distress Experience Support from Generative AI Chatbots and Human Therapists. *BMC Psychiatry* 26 (2026), 49. doi:10.1186/s12888-025-07671-w

[11] Michał Dziergwa, Mirela Kaczmarek, Paweł Kaczmarek, Jan Kędzierski, and Karolina Wadas-Szydłowska. 2018. Long-Term Cohabitation with a Social Robot: A Case Study of the Influence of Human Attachment Patterns. *International Journal of Social Robotics* 10, 1 (2018), 163–176. doi:10.1007/s12369-017-0439-2

[12] Garcia v. Character Technologies, Inc. 2024. Garcia v. Character Technologies, Inc., No. 6:24-cv-01903 (M.D. Fla., filed Oct. 2024). CourtListener docket. https://www.courtlistener.com/docket/69300919/garcia-v-character-technologies-inc/

[13] Colin Holbrook. 2018. Cues of Violent Intergroup Conflict Diminish Perceptions of Robotic Personhood. *ACM Transactions on Interactive Intelligent Systems* 8, 4, Article 28 (2018). doi:10.1145/3181674

[14] Adam Hollis. 2026. "Tinged with Heartbreak": An Ethnographic Account of Navigating Autistic Loneliness and the Fragile Promise of AI Companionship. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3772318.3790586

[15] Alexandre Hudon and Emmanuel Stip. 2025. Delusional Experiences Emerging From AI Chatbot Interactions or "AI Psychosis". *JMIR Mental Health* 12 (2025), e85799. doi:10.2196/85799

[16] J. J. Jocher and R. Verwiebe. 2026. "Forever and Always, My Love": Emotions in Human–AI Romantic Relationship Building and Breakup with Generative AI Chatbot Replika. *Social Media + Society* 12, 1 (2026), 20563051251408917. doi:10.1177/20563051251408917

[17] Waki Kamino. 2023. Towards Designing Companion Robots with the End in Mind. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. doi:10.1145/3568294.3580046

[18] Waki Kamino, Selma Šabanović, and Malte F. Jung. 2024. The Lifecycle of Social Robots: Obsolescence and Values in Repair. In *2024 IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE. doi:10.1109/RO-MAN60168.2024.10731206

[19] Waki Kamino, Selma Šabanović, and Malte F. Jung. 2025. "A Robot's Life Is Over When People Give Up": Socio-Technical Infrastructure for Sustaining Consumer Robots. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 142–151. doi:10.1109/HRI61500.2025.10974029

[20] Megan M. Kelly, Phoebe S. Moore, Adrian H. Zai, and Jeroan Allison. 2026. Artificial Intelligence Discontinuation Effects (AI-DICE): An Emerging Phenomenon in Mental Health Applications. *JMIR AI* 5 (2026), e85419. doi:10.2196/85419

[21] Gahui Kim, Yebom Choi, Yoojeong Kim, and Changjun Lee. 2026. No Time to Say Goodbye: Emotional Loss Responses to Sudden Termination in Immersive AI Interactions. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, Article 486, 5 pages. doi:10.1145/3772363.3798687

[22] Guy Laban, Sophie Chiang, and Hatice Gunes. 2025. What People Share With a Robot When Feeling Lonely and Stressed and How It Helps Over Time. In *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1930–1935. doi:10.1109/RO-MAN63969.2025.11217573

[23] Guy Laban, Arvid Kappas, Val Morrison, and Emily S. Cross. 2023. Opening Up to Social Robots: How Emotions Drive Self-Disclosure Behavior. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1697–1704. doi:10.1109/RO-MAN57019.2023.10309551

[24] Huiqian Lai. 2026. "Please, Don't Kill the Only Model That Still Feels Human": Understanding the #Keep4o Backlash. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, Article 37, 15 pages. doi:10.1145/3772318.3791351

[25] W. K. S. Leung, L. M. K. Chang, M.-L. Cheung, F. Tse, R. Y. M. Wong, and S. P. M. Law. 2024. Longitudinal Assessment of mHealth Continuance: A Task-Technology Fit Perspective. In *PACIS 2024 Proceedings*. AIS Electronic Library. https://profile.cpce-polyu.edu.hk/en/publications/longitudinal-assessment-of-mhealth-continuance-a-task-technology--2/

[26] Sijia Li, Shiguang Ni, and Kaiping Peng. 2020. The Role of Good Human Uniqueness in Social Robot Anthropomorphism Influenced by Chronic Loneliness. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. IEEE, 1–6. doi:10.1109/ICHMS49158.2020.9209343

[27] S. Li, L. Xu, F. Yu, and K. Peng. 2020. Does Trait Loneliness Predict Rejection of Social Robots? The Role of Reduced Attribution of Unique Humanness: Exploring the Effects of Trait Loneliness on Anthropomorphism and Acceptance of Social Robots. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. doi:10.1145/3319502.3374777

[28] Yibo Meng, Guangrui Fan, Bingyi Liu, Yingfangzhong Sun, Ruiqi Chen, and Haipeng Mi. 2026. Engagement Is Not Transfer: A Withdrawal Study of a Consumer Social Robot with Autistic Children at Home. In *Proceedings of the 25th Interaction Design and Children Conference*. ACM. doi:10.1145/3773077.3806118

[29] James Muldoon and J. J. Parke. 2025. Cruel Companionship: How AI Companions Exploit Loneliness and Commodify Intimacy. *New Media & Society* (2025). doi:10.1177/14614448251395192

[30] A. Nageswari, V. R. Shah, Shabana, and A. Kumari. 2026. AI-Powered Mental Health Chatbot Using NLP and Sentiment Analysis. In *Recent Trends in Machine Learning, IoT, Smart Cities and Applications*. Springer. doi:10.1007/978-981-95-5079-1_4

[31] Anuj Nandanwar and Varun Dutt. 2023. Assessing Stress, Anxiety, and Depression with Social Robots via Conversational AI. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive*

Environments. ACM, 732–738. doi:10.1145/3594806.3596589

[32] P. M. L. Ng, C. Wan, D. Lee, I. Garnelo-Gomez, and M. M. Lau. 2026. I Love You, My AI Companion! Do You? Perspectives from the Triangular Theory of Love and Attachment Theory. *Internet Research* 36, 3 (2026), 905–925. doi:10.1108/INTR-11-2024-1783

[33] Heather L. O'Brien, Ido Roll, Andrea Kampen, and Nilou Davoudi. 2022. Rethinking (Dis)engagement in Human–Computer Interaction. *Computers in Human Behavior* 128 (2022), 107109. doi:10.1016/j.chb.2021.107109

[34] P. Pataranutaporn, A. Doudkin, and P. Maes. 2026. OceanChat: The Effect of Virtual Conversational AI Agents on Sustainable Attitude and Behavior Change. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3772318.3791760

[35] P. Patil. 2025. Healing Agent: A Virtual Companion for Grief and Remembrance. In *Companion Proceedings of the ACM International Conference on Intelligent Virtual Agents*. ACM. doi:10.1145/3742886.3756702

[36] Rachel Poonsiriwong, Chayapatr Archiwaranguprok, and Pat Pataranutaporn. 2026. "Death" of a Chatbot: Investigating and Designing Toward Psychologically Safe Endings for Human–AI Relationships. In *Proceedings of the 2026 Designing Interactive Systems Conference*. ACM, 4929–4948. doi:10.1145/3800645.3813083

[37] Yocheved L. Rabinowitz, Brian Yim, and J. Christopher Muran. 2025. Termination of Psychotherapy: A Systematic Review. *Cogent Mental Health* 4, 1 (2025), 2535626. doi:10.1080/28324765.2025.2535626

[38] J. Rudnik, S. Raghuraj, M. Li, and R. N. Brewer. 2024. CareJournal: A Voice-Based Conversational Agent for Supporting Care Communications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3613904.3642163

[39] Richard M. Ryan and Edward L. Deci. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist* 55, 1 (2000), 68–78. doi:10.1037/0003-066X.55.1.68

[40] R. Schaefer, S. Sahabi, A. Brocker, and J. Borchers. 2024. Growing Up with Dark Patterns: How Children Perceive Malicious User Interface Designs. In *Proceedings of the Nordic Conference on Human-Computer Interaction*. ACM. doi:10.1145/3679318.3685358

[41] E. A. Schlesener, M. Ziolkowski, S.-K. Wong, B. Westmoreland, and S. V. Babu. 2026. "Am I Understood?": How the Interplay Between Embodiment and Theory of Mind Behavior Affects LLM-Based Conversational Agents on Perceived Trust, Anthropomorphism, Presence, Usability, and User Experience. *ACM Transactions on Interactive Intelligent Systems* 16, 2, Article 11 (2026). doi:10.1145/3774779

[42] D. G. M. Schouten, F. Venneker, T. Bosse, M. A. Neerincx, and A. H. M. Cremers. 2018. A Digital Coach that Provides Affective and Social Learning Support to Low-Literate Learners. *IEEE Transactions on Learning Technologies* 11, 1 (2018), 67–80. doi:10.1109/TLT.2017.2698471

[43] P. Shen, D. Peng, and J. Xu. 2026. Artificial Intelligence Companionship: Conceptualization, Dimension Identification and Scale Development. *Journal of Business Research* 215 (2026), 116338. doi:10.1016/j.jbusres.2026.116338

[44] M. Shi, T. Song, Z. Zhu, and Y.-C. Lee. 2026. Designing with Tensions: Older Adults' Emotional Support-Seeking under System-Level Constraints in Conversational AI. In *Proceedings of the 2026 ACM Designing Interactive Systems Conference*. ACM. doi:10.1145/3800645.3812902

[45] Lizzy Sinnema and Maryam Alimardani. 2019. The Attitude of Elderly and Young Adults Towards a Humanoid Robot as a Facilitator for Social Interaction. In *Social Robotics*. Springer, 24–33. doi:10.1007/978-3-030-35888-4_3

[46] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2022. A Longitudinal Study of Human–Chatbot Relationships. *International Journal of Human-Computer Studies* 168 (2022), 102903. doi:10.1016/j.ijhcs.2022.102903

[47] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research* 22, 3 (2020), e16235. doi:10.2196/16235

[48] TechNode Feed. 2026. ByteDance's Doubao and Alibaba's Qwen to Shut Down AI Agent Features on July 15. TechNode. https://technode.com/2026/07/06/bytedances-doubao-and-alibabas-qwen-to-shut-down-ai-agent-features-on-july-15/

[49] Rachel Wilson, Imogen Keane, and Ray Jones. 2022. Affective Responses of Older Adults to the Anthropomorphic GenieConnect Companion Robot During Lockdown of the COVID-19 Pandemic. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1095–1099. doi:10.1109/HRI53351.2022.9889480

[50] D. W. Winnicott. 1953. Transitional Objects and Transitional Phenomena: A Study of the First Not-Me Possession. *International Journal of Psychoanalysis* 34 (1953), 89–97.

[51] Yao Xiao and Rafael A. Calvo. 2026. AI as Relational Translator: Rethinking Belonging and Mutual Legibility in Cross-Cultural Contexts. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3772363.3798628

[52] Yao Xiao, Lu Meng, Minseo Cho, and Rafael A. Calvo. 2026. Beyond Localisation: Cultural Defaults and User Agency in Conversational AI. OpenReview, 7th International Conference on Social Computing. https://openreview.net/forum?id=3gEQa3lUcx

[53] Hanting Xie, Daniel Kudenko, Sam Devlin, and Peter Cowling. 2014. Predicting Player Disengagement in Online Games. In *Computer Games*. Springer, 133–149. doi:10.1007/978-3-319-14923-3_10

[54] Tianling Xie and Iryna Pentina. 2022. Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika. In *Proceedings of the 55th Hawaii International Conference on System Sciences*. 1–10. doi:10.24251/HICSS.2022.258

[55] Kohei Yamamoto, Chi-Lan Yang, Tomohiro Tanikawa, and Hideaki Kuzuoka. 2026. Exploring the Effects of Third-Party Explanations in Multi-Agent Settings After an AI Refusal. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3772363.3798502

[56] Yunhao Yuan, Jiaxun Zhang, Talayeh Aledavood, Renwen Zhang, and Koustuv Saha. 2026. Mental Health Impacts of AI Companions: Triangulating Social Media Quasi-Experiments, User Perspectives, and Relational Lens. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. ACM, Article 382, 22 pages. doi:10.1145/3772318.3790558

[57] Eloïse Zehnder, Jérôme Dinet, and François Charpillet. 2021. Social Virtual Agents and Loneliness: Impact of Virtual Agent Anthropomorphism on Users' Feedbacks. In *Advances in Usability, User Experience, Wearable and Assistive Technology*. Vol. 275. Springer, 285–292. doi:10.1007/978-3-030-80091-8_33

[58] M. Zeng, Z. Yu, and M. Tian. 2026. Creepy or Not: When AI Remember Too Much. In *Proceedings of the 63rd ACM Conference on Computers and People Research*. ACM. doi:10.1145/3768310.3807827

[59] R. Zhang, H. Li, H. Meng, J. Zhan, H. Gan, and Y.-C. Lee. 2025. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human–AI Relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM. doi:10.1145/3706598.3713429

[60] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *Computational Linguistics* 46, 1 (2020), 53–93. doi:10.1162/coli_a_00368