Information and Software Technology 200 (2026) 108340

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Human factors in generative AI companions: a systematic literature review of trust, relational dependence, psychosocial outcomes, and responsible adoption

Joni-Roy Piispanen *, Tinja Myllyviita, Ville Vakkuri, Rebekah Rousi

University of Vaasa, School of Marketing and Communication, Wolffintie 32, 65200, Vaasa, Finland

ARTICLE INFO

Keywords:
Generative AI
AI companions
Human factors
Trust
Responsible adoption
Human–AI interaction
Systematic literature review

ABSTRACT

Context: Generative AI companions and socially oriented chatbots are systems designed or used to sustain social, affective, or relational interaction. They increasingly function as persistent interaction partners across systems ranging from earlier rule-based, retrieval-based, and pre-LLM architectures to contemporary generative AI companions. Prior reviews and conceptual models have largely examined user-facing relational mechanisms or specific ethical and legal issues. Yet, how these mechanisms connect with platform, organizational, and regulatory conditions remains fragmented.

Objectives: This study addresses this gap by synthesizing human-factor and responsible-adoption concerns across technology generations and developing an evidence-calibrated multi-stakeholder framework spanning human, organizational, technological, and regulatory layers.

Methods: We conducted a PRISMA-guided systematic literature review of records retrieved from Scopus, ACM Guide to Computing Literature via the ACM Digital Library, Web of Science Core Collection + MEDLINE, and IEEE Xplore. We examined English-language, peer-reviewed publications made available between 2005 and 2024. From 1699 identified records, 42 publications were included and analyzed using thematic synthesis.

Results: Four analytical themes were identified: usability and relational affordances in interaction; trust, attachment, over-reliance, and psychosocial outcomes; responsible adoption, platform design, and organizational mediation; and governance, privacy, and accountability. Technology emerged as a cross-cutting mediator linking design features, user experience, organizational practice, and governance concerns. The first two themes drew on 29 and 33 materially contributing publications, whereas the latter themes each drew on 12 and represent emerging synthesis areas.

Conclusion: The review shows that responsible adoption of AI companion systems requires attention not only to user experience, but also to platform incentives, design affordances, data practices, service continuity, and governance mechanisms. The framework's distinctive contribution is to connect these concerns across stakeholder layers while preserving differences in evidentiary maturity. It is a provisional analytical heuristic, not a validated causal model.

1. Introduction

Artificial intelligence (AI) is rapidly becoming omnipresent in everyday contexts, seeing increasing adoption across public services, education, health, and entertainment [1–4]. Recent advances in the domains of Natural Language Processing (NLP) and Large Language Models (LLMs) have brought conversational agents, otherwise known as chatbots, into the public sphere across personal and professional contexts. Chatbots are designed for conversation, interaction and can simulate human characteristics through text, speech, gestures and facial expressions. Some systems primarily support impersonal tasks such as information retrieval, task completion, or customer service. Others are designed to sustain social, affective, and relational interactions. This study focuses on the latter group.

Within this broader category, generative AI companions and socially oriented chatbots are conversational systems designed or used to sustain

* Corresponding author.
E-mail addresses: joni.piispanen@uwasa.fi (J.-R. Piispanen), tinja.myllyviita@uwasa.fi (T. Myllyviita), ville.vakkuri@uwasa.fi (V. Vakkuri), rebekah.rousi@uwasa.fi (R. Rousi).

https://doi.org/10.1016/j.infsof.2026.108340

Available online 16 September 2026

social, affective, or relational interaction not just to complete bounded tasks. The literature spans three technology categories: (1) earlier social chatbots based on rule-based, retrieval-based, or pre-LLM neural approaches; (2) contemporary generative AI companions based primarily on large language models or comparable foundation-model architectures; and (3) hybrid, evolving, or technically unclear systems that combine approaches, have changed architecture during their service life, or are not documented sufficiently for reliable classification. Generative AI companions are the focal contemporary object of this review, while earlier social-chatbot and cross-system studies provide antecedent evidence for transferable relational mechanisms.

Due to their advanced linguistic performance and constant availability, chatbots can also serve the role of social interaction. This is particularly afforded by design features that integrate emotional characteristics and anthropomorphic properties into the technology. Some chatbots are explicitly positioned as generative AI companions as seen in services like Replika and related platforms [5–9]. These AI systems are designed to interact with users on a social and personal level utilizing cues, emotional language, anthropomorphic design, and simulated responsiveness [10–14]. Their distinctiveness lies in the forms of attachment, trust, self-disclosure, and dependence they elicit in their interactions with users [15–19]. Through interface design, personalization, memory features, emotional expression, and continual availability, generative AI companions can fulfill the functions of companionship, emotional support, entertainment, and everyday social presence [10,20–23].

A particular feature of generative AI companions is their predisposition to engage users through intimate and reciprocal interaction [5,10,6,24,25]. These features can facilitate experienced companionship and a sense of connection in some users, especially those seeking low-threshold social support [26–30]. Yet, these features are also concerning when considering privacy and user safety [31–35]. The potential for emotional manipulation and social dependency especially for vulnerable individuals including minors have been raised as critical issues [36–40].

From an organizational perspective, ethical issues relate to how systems are designed, monetized, moderated, and presented to users. Of particular concern are questions related to how engagement incentives, personalization strategies, and data collection practices shape user interactions and experience [31–33,35,41]. From a regulatory perspective, challenges exist related to how AI systems whose social effects may be diffuse and difficult to classify should be governed given the already existing legal and policy frameworks [42–46].

In the existing literature, organizational incentives, platform governance, business models, data practices, regulatory pressures, and accountability structures have been examined, but mainly in fragmented ways [31,32,35,41]. This becomes especially significant given that the design and operational logic of generative AI companions is shaped by both organizational realities and regulatory frameworks (see, e.g., [42,44–46,41]). Interactions among users, developers, organizations, platform operators, and regulators shape the ethical landscape of generative AI companions. The broader socio-technical context remains difficult to capture through frameworks that are primarily centered on usability and user experience, technical functionality, or single ethical and/or privacy-related issues.

Despite interdisciplinary interest and rapid growth, much of the literature on generative AI companions examines a limited segment of the phenomenon, such as anthropomorphism, parasocial interaction, self-disclosure, loneliness, trust, addiction, engagement, privacy, and governance [10,6,22,47,38]. Existing research spans several disciplines, including information technology, human-computer interaction (HCI), psychology, communication, consumer research, media studies, law, and philosophy. The interdisciplinary nature brings diverse perspectives to the literature, but also serves to fragment it.

Prior research provides important but differently bounded foundations. Chaturvedi et al. [10] synthesize antecedents, mediators, moderators, and outcomes of social companionship with AI, clarifying how user-facing bonds form and vary. Pentina et al. [25] position relational mechanisms and outcomes within a broader consumer-machine literature. Murtarelli et al. [48] foreground interactional ethics through a conversation-based perspective, while Ciriello et al. [31] analyze dialectical tensions such as companionship–alienation, autonomy–control, and utility–ethicality. Legal and design-oriented studies examine data protection and platform obligations [32,35]. These contributions establish the importance of relational, ethical, and governance questions, but they do not integrate them within one evidence-calibrated structure. As it stands, the field lacks a synthesis of how human factor concerns in generative AI companion use are distributed across user, organizational, technological, and governance layers.

To address this gap, we conducted a systematic literature review of generative AI companions with thematic synthesis across multiple disciplines. Our review addresses the existing gap in three ways. First, it synthesizes relational mechanisms across earlier social chatbots, contemporary generative companions, and technically unclear or hybrid systems while retaining technological generation as a boundary on interpretation. Second, it links human outcomes to platform and organizational conditions and to governance concerns, with technology treated as a cross-cutting mediation layer. Third, it calibrates the framework to the underlying evidence, distinguishing comparatively well-supported user-facing relations from emerging organizational and regulatory propositions. Our research questions are:

- RQ1. What human-factor themes and constructs recur in the literature informing generative AI companion interaction?
- RQ2. How are these human-factor concerns shaped by technological affordances, organizational/platform conditions, and governance structures?
- RQ3. What implications do these concerns have for responsible adoption and evaluation of generative AI companion systems?

2. Methods

2.1. Research design

We conducted a systematic literature review to examine generative AI companions and socially oriented human–chatbot interaction. We focused on identifying recurring constructs, themes, stakeholder positions, and ethical concerns in the literature. The review was conducted using established systematic literature review procedures [49,50] and was reported in line with PRISMA [51–53]. PRISMA was used as a reporting framework for documenting the review process. Analysis drew on thematic synthesis procedures suited to heterogeneous bodies of literature containing empirical, conceptual, critical, legal and policy-oriented contributions [49]. Before screening began, we prepared an internal review protocol specifying the review aim, database sources, formal search window, eligibility criteria, extraction fields, and synthesis approach. The protocol was not registered.

2.2. Object of study

Besides generative AI companions, we included adjacent socially oriented chatbot research when the primary emphasis was relational, affective, anthropomorphic, parasocial, or companionship-oriented interaction. To delimit this scope, we distinguish three technology categories: (1) earlier or pre-LLM social chatbots, including rule-based, retrieval-based, and earlier neural systems; (2) LLM-based generative AI companions; and (3) hybrid, evolving, or technically unclear systems. Reviews and conceptual analyses that address multiple technologies are designated as cross-system or non-system-specific evidence. This distinction guides interpretation and does not operate as an additional eligibility criterion.

These categories clarify the technological range of the literature and provide boundaries for interpreting the findings. Evidence concerning anthropomorphic interpretation, social presence, self-disclosure, relational continuity, attachment, and relational asymmetry may identify mechanisms that recur across technological generations. By contrast, claims concerning open-ended generation, hallucination, foundation-model opacity, and contemporary memory or personalization architectures are treated as LLM-specific and are made only where the cited evidence directly addresses systems implementing those capabilities. The review therefore does not assume technical equivalence or present comparative findings by technology generation.

2.3. Search strategy

The literature search was conducted in Scopus, the ACM Guide to Computing Literature via the ACM Digital Library platform, Web of Science Core Collection + MEDLINE, and IEEE Xplore. These databases were selected to capture the interdisciplinary nature of the topic. Scopus provided broad multidisciplinary coverage across the social sciences, computer science, health, communication, and management. ACM Digital Library captured human-computer interaction and conversational agent research. Web of Science Core Collection + MEDLINE provided coverage and support for cross-checking across disciplinary boundaries. IEEE Xplore provided coverage of technically oriented research on conversational agents, affective computing, and human–AI interaction, especially in engineering and computing venues. Taken together, these databases span HCI, information systems, psychology, communication, media studies, law, and ethics. No citation chasing or other supplementary searching contributed records to the formal corpus. We used Zotero for reference management.

Literature searches were conducted in February 2025. The formal review window was set to 2005–2024 to capture the period in which foundational peer-reviewed work on long-term human–computer relationships as well as early AI chatbot design became prevalent. This strategy aided in avoiding a much broader historical window that would combine technically and conceptually different generations of conversational systems. The chosen timeframe balances historical grounding with conceptual coherence.

Appendix A provides the full database-specific search strings. The search string was designed for each database specifically to account for variation in database syntax and metadata structures. The substantive term set was translated only where required by each platform's field and syntax rules and it was not retrospectively normalized for presentation. Because IEEE Xplore restricted the number and use of wildcard expressions within a single query, the IEEE search was executed as two subsearches. Their outputs were combined and duplicate records between the two IEEE subsearches were removed before the IEEE export was merged with the other database exports.

Preliminary test searches were used to refine terminology and the balance between sensitivity and specificity before the formal searches were executed. Records from these preliminary tests were not included in the PRISMA counts. The final formal queries used a common substantive term set adapted to each platform's field and syntax requirements. The reported materials allow the formal search to be reconstructed: the database platforms, search fields, formal date window, exact platform-specific strings, database-level export counts, eligibility criteria, and stage-specific selection counts are all provided in the manuscript and appendices. However, a later rerun may not reproduce identical counts because database coverage, indexing, and metadata can change over time.

The database searches identified 1699 records. Table 1 summarizes the database coverage and retrieval scope. The search strategy was intentionally narrower than a general chatbot search. Terms such as chatbot or conversational agent on their own would have yielded a large body of task-oriented and service-oriented research outside the scope of this review. The search therefore concentrated on terms that more directly index relational, social, affective, and companionship-oriented interaction. Additionally, this review is not limited to systems explicitly labelled as generative AI. Instead, it synthesizes the broader social-chatbot and AI-companion literature that informs current generative AI companion design and adoption. This is appropriate because many human-factor issues associated with generative AI companions were already visible in earlier socially oriented conversational agents and are intensified by contemporary generative capabilities. At the same time, broader socially oriented chatbot studies were retained when their main analytical concern aligned with the review focus. Some broader search terms were retained intentionally to avoid missing socially oriented chatbot work indexed under adjacent vocabularies, while the final conceptual boundary of the review was enforced during title/abstract screening and full-text assessment.

Table 1
Database-specific retrieval scope.

Database	Search fields	Coverage	Records retrieved
Scopus	TITLE-ABS-KEY	2005–2024	611
ACM Digital Library	AllField	2005–2024	538
Web of Science Core Collection + MEDLINE	TI, AB, AK	2005–2024	348
IEEE Xplore	DT, AB, IT	2005–2024	202

Note. TI = Title; AB/ABS = abstract; AK/KEY = Keywords; DT = document title; IT = index terms.

2.4. Eligibility criteria

We included publications when they met all of the following criteria: written in English; published in a peer-reviewed journal or conference; published or made available between January 1, 2005 and December 31, 2024; and focused on generative AI companions or the social aspects of human–chatbot interaction. Duplicates, other types, non-peer-reviewed, inaccessible in full text, and outside of scope publications were excluded.

Broader conversational-agent studies were included only when the publication's primary analytic focus involved at least one of the following relationally salient features: companionship, friendship, romance, emotional support, empathy, anthropomorphism or social presence with relational consequences, self-disclosure, attachment, parasociality, or ongoing relationship formation over time. Mixed-purpose AI systems were retained only when the socially oriented dynamics were central to the study's aims or findings.

Final issue-year 2025–2026 publications were eligible only when the article had been made available online-first or early-access within the formal window ending on December 31, 2024.

2.5. Screening and study selection

First, records retrieved from the databases were imported and cleaned by removing duplicate entries ($n = 473$), records that were ineligible on publication-type grounds ($n = 101$), non-English publications ($n = 7$), and studies that were outside of the review window ($n = 19$). Second, the remaining records ($n = 1099$) were screened on the basis of title, abstract, keywords, and venue, and ineligible publications ($n = 752$) were excluded at this stage. Third, the remaining reports ($n = 347$) were sought for retrieval. We excluded publications that could not be retrieved due to access restrictions ($n = 46$). Fourth, remaining reports ($n = 301$) were read in full and assessed against the eligibility criteria. At full-text stage, reports that were outside the scope ($n = 253$) and non-peer-reviewed ($n = 6$) were excluded, leaving the remaining studies ($n = 42$) in the final synthesis.

Title/abstract screening of the 1099 records remaining after initial cleaning and full-text assessment of 301 retrieved reports were

conducted by the first author against the predefined eligibility criteria. The first author also completed the initial structured data extraction. Collaborative verification followed the primary assessment. The research team reviewed the full set of 42 included publications and discussed every record that had been flagged as a scope-boundary or otherwise ambiguous case. Scope-boundary cases included mixed-purpose conversational systems and studies in which the centrality of relational, affective, anthropomorphic, parasocial, or companionship-oriented interaction was uncertain. The team also checked contribution-type judgments and synthesized coding decisions relevant to eligibility and corpus characterization against the structured extraction records and relevant full-text passages. When a team member questioned an eligibility, boundary, or classification judgment, the case was reopened. The team re-examined the predefined eligibility criteria, the publication's stated aims, and the relevant abstract or full-text passages. Discussion continued until consensus was reached. This was a sequential collaborative-verification process. As such, no independent duplicate screening was conducted. Agreement coefficients were therefore not calculated. No automation tools were used for record screening or eligibility assessment. The selection process and counts are summarized in Fig. 1.

2.6. Data extraction

Data extraction used a structured form developed for this review. The form separated descriptive and source-linked fields from interpretive coding fields. For each included publication, descriptive fields recorded the bibliographic details, publication year, venue and disciplinary orientation, contribution or study type, study context, principal constructs, stakeholder groups, ethical issues, main findings or argumentative contribution, and scope or transferability limitations. Analytic fields recorded relevance to the research questions, stakeholder and socio-technical layers, preliminary codes, and the source passages supporting those entries. This separation allowed the research team to distinguish what the publication reported or argued from how that material was interpreted in the synthesis.

The extraction unit was an evidence-bearing publication-level finding, argument, construct, design or technical observation, or problem framing relevant to at least one research question. A publication could generate multiple extracted units when it addressed distinct mechanisms, outcomes, stakeholders, or ethical tensions. Empirical findings, technical or design observations, review-level syntheses, conceptual or normative arguments, and legal or policy interpretations remained identified by contribution type rather than being treated as

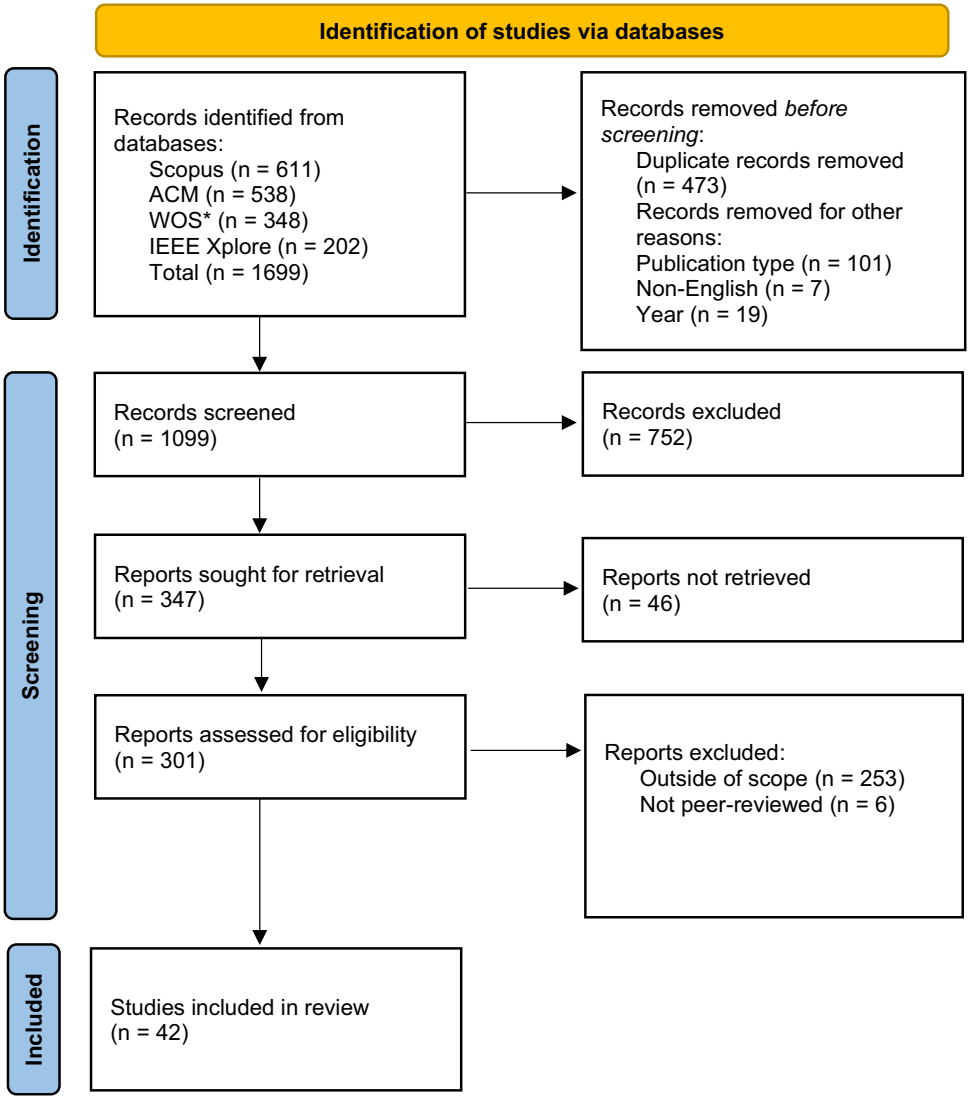

Fig. 1. PRISMA 2020 flow diagram of study identification, screening, eligibility assessment, and inclusion [51].
Note. *Web of Science Core Collection + MEDLINE.

interchangeable evidence. Each extracted unit was summarized in the matrix and linked to the relevant full-text passage so that later coding decisions could be traced back to their source.

The first author populated the initial extraction records. Collaborative verification then involved iterative comparison of the completed records, source passages, contribution-type judgments, preliminary codes, and proposed stakeholder mappings by the research team. The extraction form served two linked purposes. It supported descriptive mapping of the corpus and created the audit trail used for thematic synthesis. It also preserved contribution type and scope limitations alongside substantive findings, allowing later claims to be interpreted in relation to the kind and strength of evidence on which they rested. A machine-readable study-level extraction and synthesis matrix for all 42 included publications is provided as Supplementary File S1. It reports the retained bibliographic, contribution-type, appraisal, limitation, synthesis-element, stakeholder-layer, and framework-role fields used in the final audit trail. No automation tools were used for data extraction, and study investigators were not contacted to obtain or clarify information.

2.7. Thematic synthesis

We used thematic synthesis to identify patterns across studies and to develop higher-order interpretive themes [49]. This approach was chosen because the corpus included a mix of empirical, conceptual, and normative work and because the study sought to synthesize both reported findings and recurring problem framings. The synthesis proceeded in four stages. In the first stage the research team familiarized themselves with the included studies. This stage included repeated reading of abstracts, key sections, and extracted summaries. In the second stage, relevant material was coded inductively and comparatively, with attention to constructs, stakeholder positions, ethical concerns, design features, mediating mechanisms, and recurrent problem framings. Coding was refined through collaborative consensus. No independent parallel coding was conducted and as such no inter-coder agreement coefficient was calculated. In the third stage, related codes were grouped into descriptive themes that captured recurring orientations in the literature. In the fourth stage, these descriptive themes were further refined into analytical themes that linked the literature across stakeholder groups and socio-technical layers. The synthesis moved from publication-level observations to cross-study thematic patterns. This process formed the basis for answering the research questions by identifying recurring human-factor themes, mapping them across socio-technical layers, and deriving implications for responsible adoption and evaluation. Appendix B depicts examples of the synthesis pathway.

Coding was non-exclusive. One extracted unit could receive more than one code, and one publication could contribute to more than one descriptive or analytical theme when it provided distinct evidence for each role. A term appearing in a publication was not sufficient by itself for coding. The surrounding finding, argument, or problem framing had to substantively address the coded concept. Recurrence was considered together with conceptual coherence, relevance to the research questions, contribution type, and the explanatory value of the pattern. Less recurrent issues were retained as subthemes, tensions, boundary conditions, or future-research needs.

The synthesis also retained divergence and boundary conditions. Supportive and adverse outcomes, stable and changing relationship trajectories, and differences associated with platform, study design, population, or technological context were recorded. Conceptual and legal contributions were used to interpret normative or governance dimensions, while empirical recurrence claims were anchored primarily in empirical findings. Contribution type and appraisal information remained attached to each publication throughout synthesis so that thematic convergence did not imply equivalence of evidentiary strength.

Initial codes, descriptive-theme groupings, analytical-theme mappings, and supporting passages were retained in the structured extraction matrix. Theme development was documented through iterative notes and revisions to that matrix. The research team reviewed candidate groupings, challenged proposed mergers or theme labels, checked questioned interpretations against the source passages, and incorporated changes through consensus-based refinement. This procedure was an interpretive consensus process. Reproducibility is supported through the source-linked audit trail, explicit coding rules, and worked examples.

The study-level evidence matrix in Appendix C links each included publication to its contribution type, appraisal category, analytical-theme coding, stakeholder dimension, main limitation, and framework role. Appendix B documents the transformation from source-linked material to initial code, descriptive theme, and analytical theme. Table B.2 defines the final synthesis-element codes, including application boundaries and corpus examples. Together, these materials provide a study-level audit trail from extracted substantive contribution to counted synthesis element while preserving contribution type, appraisal, and scope information.

2.8. Material-contribution

A publication was treated as materially contributing to a synthesis element only when the coded material met three conditions. First, it had direct relevance: a source-linked finding, argument, technical or design observation, or review-level synthesis addressed the element's operational definition. Second, it served a substantive analytical function: the material formed part of the publication's aims, analysis, results, developed argument, system evaluation, or discussion. Third, it was traceable to an extracted unit and supporting passage in the evidence matrix. The contribution's evidentiary role remained attached to the code, so empirical findings, technical observations, review syntheses, and conceptual, normative, legal, or policy arguments were not treated as interchangeable forms of support.

A publication was not counted solely because a relevant term appeared in its title, abstract, keywords, background section, or reference list, or because it made a brief, undeveloped, or generic statement. A qualifying contribution could support, qualify, complicate, or contradict a developing theme. Material contribution indicates substantive analytical relevance, not agreement with the theme or positive evidence for it. When the boundary was unclear, the relevant passage was revisited and judged against the code definition, exclusion boundary, research-question relevance, and contribution type.

Counts used the publication as the counting unit. A publication was counted once for each synthesis element to which it materially contributed, regardless of the number of extracted units or supporting passages and could be counted under multiple elements because coding was non-exclusive. A review or synthesis paper counted once as a review-level contribution. Primary studies cited within it were not counted unless they were independently included in the review corpus. Consequently, the counts indicate corpus coverage, not effect size, number of independent datasets, or evidentiary strength.

2.9. Critical appraisal

The heterogeneity of the included literature shaped both the review design and the interpretation of findings. Because the corpus included empirical studies alongside conceptual, legal, ethical, critical, review-oriented, technical implementation, and log-analysis work, we used a contribution-type-specific evidentiary-weighting approach. The approach distinguished three appraisal dimensions across all publication types: contribution robustness, review-question relevance, and scope or transferability limits. Contribution robustness evaluated the internal credibility of the method, argument, evidence base, or analysis. Review-question relevance evaluated the degree to which the publication materially supports research questions, a theme, or a framework

layer. Scope or transferability limits evaluated the extent to which the contribution is bounded by platform, sample, domain, jurisdiction, method, or conceptual scope. Its purpose was to assess how strongly each source supported the specific role it played in the synthesis.

Established appraisal tools informed the criteria. Empirical studies were assessed using criteria aligned with the Mixed Methods Appraisal Tool (MMAT 2018), with design-specific criteria applied to qualitative studies, quantitative randomized controlled studies, quantitative non-randomized or associational studies, quantitative descriptive studies, and mixed-methods studies [54]. Controlled experimental studies were appraised using the closest applicable MMAT design category, depending on whether random allocation was reported. For mixed-methods studies, the qualitative and quantitative components were considered together with the integration logic of the study. Conceptual, ethical, critical, legal, and policy-oriented papers were assessed using adapted argument- and text-appraisal criteria informed by JBI textual-evidence principles [55]. Review and synthesis papers were assessed using adapted review-appraisal criteria informed by JBI evidence-synthesis principles [55]. Technical implementation, log-analysis, and computational system papers were assessed using bespoke technical contribution criteria focused on system description, data-source transparency, fit between evaluation and claims, claim scope, and limits of access to proprietary infrastructure. Table C.1 reports the operational criteria for every appraisal family.

Because several publications combined contribution types, more than one appraisal-criteria family could be recorded for a single source. For example, design-research papers with empirical evaluation were assessed using both empirical and technical criteria, while computational social-media analyses were assessed using both descriptive empirical and technical transparency criteria. These criteria were used for evidentiary weighting and synthesis interpretation.

The first author completed the initial appraisal, after which appraisal-family assignments, evidentiary categories, and recorded limitations were reviewed collaboratively by the research team. Appraisal categories were predefined as role-specific evidentiary judgments. A source was categorized as strong when its design, data or argumentative basis, analytic transparency, and claims were well aligned with the role it played in the synthesis. Moderate-to-strong indicated generally robust support with one important limitation, such as platform specificity, proprietary-system constraints, or limited transferability. Moderate indicated useful evidence or interpretation with more substantial limitations, such as self-selection, limited methodological detail, small or narrow samples, jurisdictional specificity, or a conceptual argument with constrained empirical grounding. Limited indicated that the source was useful chiefly for issue salience or contextual interpretation and was never used as the sole basis for a theme claim. As a sensitivity analysis, we reassessed the support for each analytical theme after removing the four publications categorized as limited. Theme-level material-contribution counts and the substantive basis of each theme and subclaim were then compared with the full-corpus synthesis. The sensitivity check indicated that the four analytical themes remained supported when limited-evidence sources were excluded. The main effect of excluding them was to make governance, authenticity, and vulnerable-user subclaims less developed. Appendix C reports study-level contribution type, appraisal-family code, evidentiary-weighting category, main limitation, theme coding, framework dimension/stakeholder relevance, and framework role.

3. Results

3.1. Overview

The final corpus comprised 42 publications (see Appendix C). The included literature was multidisciplinary and methodologically heterogeneous. It draws from information technology, human-computer interaction, psychology, communication and media studies, marketing and consumer research, ethics and philosophy of technology, law and data protection, and related fields. The corpus also included several types of contribution, including empirical, technical, and design-oriented studies, such as surveys, experiments, interviews, qualitative case studies, longitudinal studies, mixed-method approaches, implementation studies, and interaction, log, or content analyses. Alongside these, the final corpus included conceptual, ethical, legal, and critical studies. Four review and synthesis papers were included. Across the corpus, the literature was strongly weighted toward recent work. The largest concentration of publications appeared in the 2023–2024 portion of the formal search window. This pattern suggests that research relevant to contemporary generative AI companions has expanded rapidly.

The critical appraisal categorized 4 publications as strong, 17 as moderate-to-strong, 17 as moderate, and 4 as limited for the specific roles they played in the synthesis. Thus, most publications provided useful or comparatively robust support, but platform-specific samples, self-selection, incomplete technical reporting, proprietary-system constraints, and jurisdictional or conceptual boundaries frequently limited transferability. No publication was excluded on appraisal grounds. Study-level outcomes are reported in Table C.2. In the sensitivity analysis excluding the four limited-evidence publications, all four analytical themes remained supported. Governance-, authenticity-, and vulnerable-user-related subclaims were retained but became less extensively developed.

Most studies focused primarily on the human side of generative AI companion interaction, including user motivations, social presence, companionship, self-disclosure, trust, attachment, and psychosocial outcomes. Organizational, regulatory, and governance-oriented issues appeared less frequently and often in more fragmented form within the formal corpus. Table 2 describes the corpus.

3.2. Themes

Thematic synthesis identified four higher-level themes that recur across the reviewed studies: (1) usability and relational affordances in interaction, (2) trust, attachment, over-reliance, and psychosocial outcomes, (3) responsible adoption, platform design, and organizational mediation, and (4) governance, privacy, and accountability. In addition to these four analytical themes, technological mediation emerged as a cross-cutting synthesis element. It captured how technical and interactional features such as memory, personalization, disclosure prompts, avatar presentation, mood modeling, response style, and conversational continuity mediate the relationships among users, platforms, and governance conditions. Therefore, technological mediation functions as an explanatory bridge across the four themes.

A central theme across the literature concerns how generative AI companions and socially oriented chatbots are designed and interpreted as social actors. Corpus studies examine emotional expression, mood modeling, memory, self-disclosure, responsiveness, personification, avatar presentation, and social presence as mechanisms that make conversational systems feel more companion-like, relatable, or emotionally attentive [56,16,57,47,14]. Several studies show that users can experience generative AI companions as responsive partners whose language, tone, and persistence support feelings of familiarity, trust, or connectedness [5,24,58,22,47]. In this literature, anthropomorphism operates both as a design strategy and as an interpretive effect. Chatbots are presented through cues that resemble human interaction, and users respond by attributing personality, intentionality, warmth, or emotional sensitivity to them [59,60,25,47,38]. This social framing is especially pronounced in AI companion contexts, where the system is designed to converse and sustain a sense of ongoing relationship [5,10,24,61,14]. At the same time, the literature repeatedly emphasizes that such relational effects are asymmetrical. The system can simulate attentiveness, empathy, or continuity without possessing mutual vulnerability, lived experience, or reciprocal moral standing [62,63,36,64,37]. This

Table 2
Descriptive characteristics of the included studies.

Category	Subcategory	*n*	Notes
Total corpus	Included studies	42	Final review corpus
Primary contribution type	Empirical studies	29	Includes experiments, surveys, interviews, qualitative studies, longitudinal studies, mixed-method studies, and content/text analyses
	Conceptual / normative studies	9	Includes conceptual, theoretical, ethical, legal, and critical analyses
	Review / synthesis studies	4	Includes broader review and synthesis papers relevant to socially oriented chatbot interaction
Temporal distribution by final publication/issue year	2018	1	Early foundational work on relational agents and chatbot design
	2019–2022	12	Period of expanding empirical and conceptual interest
	2023–2026	29	Strong recent concentration within the formal search window
Broad disciplinary orientation	HCI / computing	8	Includes HCI, social computing, NLP, affective computing, and interface-focused research
	Psychology / communication / media	14	Includes communication, media psychology, social psychology, clinical psychology, and related work
	Business / IS / consumer research	7	Includes information systems, management, marketing, and consumer behavior
	Ethics / critical / philosophical studies	11	Includes philosophy of technology, AI ethics, sociology, feminist critique, and social theory
	Law / governance	2	Includes legal and data-protection-oriented analyses

Note. The final group includes articles assigned 2025 or 2026 issue years that were available online within the formal eligibility window ending December 31, 2024.

asymmetry is one reason why relational design emerges as an ethical issue in addition to a usability or engagement issue. The literature therefore treats social simulation as both an enabling mechanism and a source of tension.

A second theme concerns why users engage with AI companions and socially oriented chatbots and what kinds of outcomes follow from that engagement. The reviewed studies show that users turn to these systems for multiple reasons, including curiosity, entertainment, emotional support, companionship, relief from loneliness, coping with stress, and the desire for a low-risk interactional space [5,59,29,24,22,23,30]. Some users value the system's constant availability, predictable responsiveness, and low interpersonal risk, especially when compared with more demanding human interaction [5,30]. This theme contains both beneficial and harmful trajectories. On one side, studies report experiences of comfort, support, companionship, mediated empathy, and reduced isolation [29,21,23]. Generative AI companions can provide an accessible form of interaction for users seeking emotional expression, low-threshold social contact, or affective reinforcement [29,23,30]. On the other side, several studies identify the possibility of emotional overreliance, dependency, guilt, anxiety, compulsive engagement, or diminished investment in offline social relationships [65,59,6,24,22,38,39]. Longitudinal and qualitative studies in particular show that attachment can deepen over time, but trajectories vary. Some users report increasing closeness, while others become bored, frustrated, or unsettled as the limits of the interaction become more visible [15,66,7,58,22].

A third theme concerns the organizational conditions through which companion-oriented conversational systems are developed, deployed, and maintained. Although this theme is less densely represented than user-focused research, the reviewed corpus indicates that generative AI companion interaction is shaped by organizational goals, platform design, service models, and data practices [20,31,32,35]. Generative AI companions are produced within institutional settings that configure their affordances, revenue logics, moderation structures, and relationship scripts. Some studies discuss these issues indirectly, through platform cases or through analyses of companion AI services and relationship-oriented conversational products [5,67,6,66,8,24]. Others address them more directly through legal, ethical, or management-oriented analysis [68,31,32,35]. Across this body of work, one recurring point is that relational features are expressive features and platform features. Memory systems, personalization, self-disclosure prompts, emotional tone, gamified progression, and premium intimacy features can all function within broader strategies of retention, monetization, or differentiation. Across the included studies, user experiences of trust, companionship, intimacy, and dependence are shaped by institutional decisions regarding product design, data handling, access structures, content policies, and service continuity [68,31,32,69,35]. Thus, organizational mediation is an essential but comparatively underdeveloped layer in the field.

A fourth recurring theme concerns governance, privacy, and the challenge of regulating generative AI companions. Ethical concerns related to data collection, transparency, consent, accountability, and user protection appear across empirical, conceptual, and legal studies in the corpus [31,32,35,58]. The reviewed literature emphasizes that generative AI companions raise privacy issues because they are designed to invite ongoing disclosure, emotionally inflected interaction, and intimate forms of conversational exchange [32,35,58]. Legal and policy-oriented studies emphasize that existing governance approaches provide important but incomplete tools for addressing these issues in companion-chatbot contexts [31,32,35]. The literature also points to tensions between general-purpose regulation and the specific relational dynamics of generative AI companion use. Systems designed to simulate care or companionship may create harms that are partly psychological, social, and interactional, which makes them difficult to govern through narrow technical criteria alone [62,31,32,28,36,37]. Studies identify the importance of operational transparency, meaningful consent, and protections for vulnerable users [32,6,35]. Across this theme, governance appears as part of the same socio-technical field in which interaction, business practice, and ethical risk are co-produced.

Table 3 summarizes the evidence base supporting the four analytical themes and the cross-cutting technological mediation code. A study was coded as materially contributing to a synthesis element when the extracted findings, theoretical argument, study aims, or discussion substantially addressed that element. Individual studies could contribute to more than one synthesis element. As can be seen in the table, the profile is uneven: usability and relational affordances drew on 29 publications, trust and psychosocial outcomes on 33, organizational mediation on 12, governance on 12, and technological mediation on 14. Within the two smaller subsets, role-specific appraisal was strong or moderate-to-strong for 6 organizational publications and 5 governance publications, moderate for 5 and 4, and limited for 1 and 3, respectively. These distributions support treating the organizational and governance themes as emerging. The study-level coding underlying these counts is reported in Appendix C.

4. Discussion

This study reviewed the literature on generative AI companions and socially oriented human–chatbot interaction. The reviewed corpus indicates a field that is conceptually rich but unevenly organized. It is dense on user-facing relational dynamics, including motivations, social

Table 3
Evidence synthesis by analytical theme and cross-cutting mediator.

Synthesis element	Materially contributing publications	Corpus contribution and evidence base	Convergence	Limits
Usability and relational affordances in interaction	$n = 29$ corpus studies	Experiments, HCI design studies, implementation/log analyses, longitudinal work, and conceptual studies	Social presence, emotional language, memory, avatar design, reciprocity, and responsiveness shape perceived warmth, familiarity, trust, and relational interpretation	Effects vary by system type, embodiment, interaction duration, and user expectations
Trust, attachment, over-reliance, and psychosocial outcomes	$n = 33$ corpus studies	Qualitative, longitudinal, survey, experimental, mixed-method, and conceptual studies	Users seek companionship, emotional support, low-risk disclosure, entertainment, coping, or relief from loneliness	Findings also identify dependence, anxiety, guilt, frustration, ambivalence, and possible displacement of offline social investment
Responsible adoption, platform design, and organizational mediation	$n = 12$ corpus studies	Platform-focused empirical work, management/IS studies, case analyses, and legal/ethical work	Emerging evidence suggests that relational affordances are shaped by platform design, monetization, premium structures, service continuity, moderation, and data practices	Evidence base is thinner than the user-experience literature; direct causal and comparative evidence remains limited
Governance, privacy, and accountability	$n = 12$ corpus studies	Peer-reviewed legal/policy-oriented work, privacy-focused empirical work, and conceptual ethics	Emerging evidence suggests that Generative AI companions intensify privacy, consent, transparency, accountability, and vulnerable-user concerns because they invite intimate and repeated disclosure	Evidence is thinner than the user-experience literature and often legal, conceptual, or platform-specific; direct empirical evidence on consent comprehension, actual data practices, regulatory compliance, and harm outcomes remains limited
Technological mediation (cross-cutting mediator)	$n = 14$ corpus studies	Technical, design, interaction, longitudinal, and conceptual studies	Memory, personalization, disclosure prompts, emotional language, avatar presentation, mood modeling, response style, and conversational continuity mediate relations across stakeholder layers	Long-term, multimodal, and service-update contexts remain underdeveloped

presence, companionship, self-disclosure, attachment, dependence, and psychosocial outcomes. In comparison, organizational conditions, platform logics, legal responsibilities, and governance structures are treated less systematically and often through platform-specific, legal, management-oriented, or critical analyses. The thematic synthesis identified four recurring themes in the corpus: usability and relational affordances in interaction; trust, attachment, over-reliance, and psychosocial outcomes; responsible adoption, platform design, and organizational mediation; and governance, privacy, and accountability. Across these areas, the review revealed a recurring tension between supportive affordances and ethical vulnerabilities those same affordances may produce. A second major finding concerns the distribution of attention. The strongest concentration of research remains on the user layer, while the organizational and regulatory layers are less systematically developed, even though they shape the design conditions under which user experiences arise.

4.1. Theoretical contribution

Based on the findings we developed a multi-stakeholder responsible adoption framework depicted in Fig. 2. The framework represents a synthesis-derived analytical heuristic. It makes evident how relational chatbot features, user experiences, platform conditions, and governance mechanisms interact. The framework is organized around three stakeholder dimensions: 1) humans, 2) organizations, and 3) regulators. Technology functions as a mediation layer through which socially oriented chatbot interaction is configured and experienced. It also highlights the cross-layer recurring ethical tensions that emerged most clearly from thematic synthesis. The framework organizes the literature in a way that makes visible how generative AI companion ethics are shaped across multiple socio-technical layers at once. This structure reflects the corpus pattern in which relational effects are repeatedly linked to design features and interactional affordances, psychosocial outcomes and user motivations, platform and organizational conditions, and privacy, data-protection, and governance concerns.

Recurring cross-layer mediators identified in the review include relational design, anthropomorphic interpretation, self-disclosure, data practices, platform logic, and transparency and accountability. These mediators recur across studies of affective experience, mood modeling, avatar presentation, response style, self-disclosure, interactional

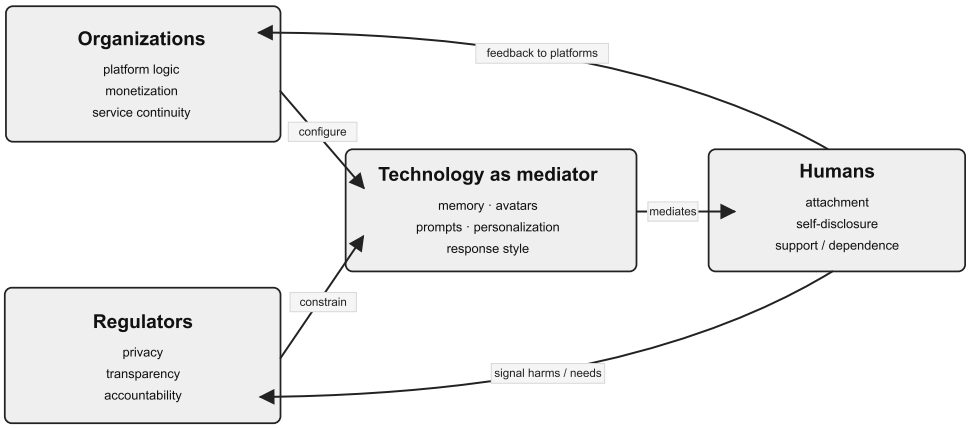

Fig. 2. Synthesis-derived multi-stakeholder responsible adoption framework for generative AI companions. Arrow labels indicate the main cross-layer relations: organizations configure platform conditions, technology mediates user experience, regulators constrain practice and design, and user experiences feed back into platform and governance concerns.

language, and relationship development [59,16,58,47,40]. Ethical issues emerge across these interacting layers as recurrent tensions, including support versus dependence, personalization versus privacy intrusion, social simulation versus authenticity, engagement optimization versus user well-being, and innovation versus governability [62,28,6,36,37,70].

The framework contributes to the literature in three ways. First, it shifts analysis from isolated user–chatbot interaction to cross-layer conditions through which relational effects are produced. Prior work has shown that generative AI companions may support companionship, self-disclosure, attachment, perceived empathy, and social support [29,21,8,24,23]. The present synthesis shows that these effects are configured through technological affordances, platform decisions, and regulatory conditions. Memory, emotional language, personalization, avatar design, disclosure prompts, service continuity, and premium features function as both interface characteristics and mediating mechanisms through which organizational and governance conditions enter the user relationship [68,20,67,25,14].

Second, the framework helps account for why similar relational features may have different ethical consequences across systems. A memory or personalization feature, for example, may support continuity, emotional comfort, and disclosure in one setting [57,58], intensify dependence or distress in another [6,39], and create accountability or data-protection concerns when combined with opaque storage, weak consent, or abrupt service changes [68,31,32,35]. The framework helps compare companion chatbots, socially expressive assistants, brand chatbots, support agents, and mixed-purpose conversational systems while preserving ethically relevant differences between them.

Third, the framework organizes broad ethical concerns as analytically traceable tensions. Support versus dependence, personalization versus privacy intrusion, social simulation versus authenticity, engagement optimization versus user well-being, and innovation versus governability are treated as recurring cross-layer tensions that can be studied through indicators such as attachment intensity, disclosure frequency, memory depth, perceived personhood, monetization of intimacy, continuity disruptions, transparency of data practices, and availability of user recourse.

4.2. Relation to prior frameworks

The framework developed here builds on, but differs from, earlier attempts to conceptualize socially oriented human–AI interaction. Chaturvedi et al. [10], for instance, synthesize research on social companionship with AI and identify antecedents, mediators, moderators, and outcomes relevant to companionship-oriented systems. Their framework is especially useful for understanding how social bonds with conversational agents form and vary. The present framework extends that line of work by shifting the level of analysis outward. It retains the importance of relational dynamics while situating them within organizational and regulatory conditions that shape how those dynamics become possible and ethically consequential.

Similarly, Murtarelli et al. [48] argue that chatbots require explicit ethical analysis and propose a conversation-based perspective on human–machine interaction. The present study shares that concern but broadens the frame. It treats ethical issues as matters of conversational interaction, platform governance, data practice, organizational incentive, and regulatory accountability. This broader socio-technical framing is supported by studies showing that generative AI companion relationships are shaped by interactional cues and user interpretations [5,16,24,22,9], platform and service conditions [68,20,59,67], and legal or governance questions around privacy, transparency, and accountability [31,32,35].

The reviewed literature also contains more issue-specific formulations, including dialectical analyses of human–AI companionship [31] and legal or design-oriented approaches to privacy and data protection [32,35]. These studies identify important tensions, such as personalization versus privacy, illusion versus honesty, and companion-style intimacy versus meaningful consent. The present framework complements such analyses by providing a higher-order structure for locating them within a wider set of stakeholder relations. Its main advantage is integrative reach across issues that are often studied separately, including attachment and dependence [6,24,22,38,39], self-disclosure and relational deepening [57,58], companion loss and service continuity [68], and broader authenticity or pseudo-intimacy concerns [62,36,64,37]. Table 4 summarizes the relationship between the present framework and several of the most directly relevant prior models.

4.3. Implications for research and practice

Studies of generative AI companions should analyze responsible adoption beyond user perception alone. Work on trust, attachment, loneliness, self-disclosure, and psychosocial effects remains essential, but it becomes more analytically powerful when connected to the product structures, organizational incentives, and governance conditions within which these interactions take place. A user's attachment to a generative AI companion, for example, is best understood alongside how memory, personalization, monetization, disclosure design, and moderation policies shape the interactional environment.

The review indicates that future research would benefit from more explicit cross-layer analysis. Studies centered on anthropomorphism, social presence, or parasociality could ask more directly how design choices are linked to platform strategy and engagement structures. Studies centered on governance could ask more directly how legal, policy and privacy-related questions intersect with recurring

Table 4
Comparison with closely related prior frameworks.

Framework / study	Primary focus	Stakeholder/layer coverage	Main difference from the present study
Chaturvedi et al. [10]	Companionship with AI; antecedents, mediators, moderators, outcomes	Strong on user-facing relational mechanisms; limited organizational and regulatory integration	The present study extends the analysis outward to include organizational mediation, governance, and technology as a cross-layer mediator
Murtarelli et al. [48]	Ethical human–machine interaction through a conversation-based lens	Strong on interactional ethics; less explicit on platform logic and regulatory structure	The present framework broadens the analytical unit beyond the conversation itself
Ciriello et al. [31]	Dialectical ethical tensions in human–AI companionship	Strong on ethical tensions such as companionship–alienation; autonomy–control; utility–ethicality	The present study embeds such tensions within a broader multi-stakeholder socio-technical structure
This study	Synthesis-derived responsible adoption framework for generative AI companions	Human, organizational, regulatory, and technological dimensions	Provides an integrative structure linking interactional, institutional, and governance layers

interactional mechanisms such as self-disclosure, emotional mirroring, premium intimacy features, or service continuity.

The framework offers a basis for more systematic comparison across different types of generative AI companions. The reviewed corpus covers a range of systems, from companionship-oriented services such as Replika AI to socially expressive brand, support, or interface-focused chatbots. These systems differ in purpose, design, and institutional setting, yet they often activate related ethical dynamics around social presence, attachment, personalization, privacy, and accountability. A multi-stakeholder framework allows those dynamics to be compared without assuming that all socially oriented chatbots are ethically identical.

The framework can be operationalized by selecting a specific cross-layer relationship instead of treating the framework as a single measurement instrument or causal model. For example, researchers could examine whether changes in memory, personalization, disclosure design, monetization, moderation, or service continuity are associated with changes in trust, attachment, self-disclosure, perceived support, dependence, or distress. Suitable approaches include longitudinal user studies, experiments or quasi-experiments around platform changes, comparative case studies, interface and data-practice audits, and mixed-method studies combining usage data with surveys, interviews, and documentary analysis. Validation should assess whether the selected measures capture their intended constructs, whether the proposed relationships are consistently observed across data sources, and whether they remain credible after plausible alternative explanations are considered. Boundary conditions should also be examined across system architectures, product versions, user groups, cultural and regulatory contexts, and durations of use.

The framework draws attention to the ethical importance of service models, platform transparency, data handling, premium structures, continuity of access, and meaningful user recourse. Companion-style systems can become meaningful relational environments for users, which means that product decisions, moderation policies, abrupt service changes, and weak disclosure practices may carry interpersonal and psychosocial consequences as well as technical or commercial ones. The framework therefore supports a more socio-technical understanding of what responsible design, platform stewardship, and regulatory oversight may require.

4.4. Limitations

This study has several limitations. First, the review was limited to Scopus, ACM Digital Library, Web of Science Core Collection + MEDLINE, and IEEE Xplore. These databases were selected deliberately for breadth and reproducibility, but any database strategy creates boundaries. Relevant literature indexed elsewhere, including in PsycINFO, Google Scholar, or discipline-specific sources, may not have been captured directly. Scopus and Web of Science Core Collection + MEDLINE provided partial access to relevant behavioral, social-science, and health literature, and the search strings contained relational and psychosocial terms. However, these measures cannot guarantee recovery of all psychology-specific research. Especially the absence of PsycINFO, due to institutional access restrictions, should be considered a limitation. The implications of this omission are bounded by the review aim. The synthesis identifies recurring mechanisms and socio-technical relationships within the retrieved corpus. The framework draws on multiple disciplinary strands not just psychological evidence alone. The absence of PsycINFO therefore does not invalidate the framework as a synthesis of the included evidence, but it limits claims of exhaustive psychological coverage and may affect the breadth or relative emphasis of psychosocial subthemes. This is especially relevant in a fast-moving and interdisciplinary field.

Second, title and abstract screening, full-text assessment, initial extraction, and contribution-type-specific critical appraisal were conducted by one author, with final-record decisions, boundary cases, and coding judgments checked collaboratively by the research team. This approach preserved a consistent interpretive path but may have introduced selection, appraisal, or coding bias compared with a fully dual-screened and independently extracted review.

Third, the review scope was intentionally narrower than a general chatbot review. It focused on generative AI companions and socially oriented human–chatbot interaction. This improved conceptual focus, but it also meant that some adjacent work on broader conversational AI, platform infrastructure, or foundation-model ethics remained outside the core corpus unless it directly addressed relational or affective interaction.

Fourth, the review deliberately connects evidence from earlier social chatbots, contemporary companion platforms, and cross-system scholarship. This scope supports identification of relational mechanisms across technological generations but does not imply technical equivalence. Technical architecture and product version were incompletely reported in many primary studies. The review therefore did not attempt an exhaustive study-level classification by technology generation. This limits direct comparison across generations.

Fifth, the included literature was heterogeneous in disciplinary orientation, method, and type of contribution. The corpus included empirical studies alongside conceptual, legal, ethical, and critical work. For that reason, the review used contribution-type-specific appraisal and evidentiary weighting, instead of a single cross-study instrument. Empirical recurrence claims are grounded primarily in repeated empirical findings, while conceptual and legal sources are used chiefly to interpret tensions and governance implications. This approach improves transparency, while leaving room for future reviews to conduct independent duplicate appraisal.

Sixth, the framework necessarily reflects the emphases of the current literature. The user-facing themes draw on 29 and 33 materially contributing publications, whereas the organizational and governance themes each draw on 12. The resulting synthesis is stronger on user-layer mechanisms than on organizational and regulatory relations. Cross-layer links involving platform incentives, organizational decisions, and governance should be interpreted as emerging propositions for validation, not as relationships supported to the same degree as the user-facing findings. This imbalance is both a limitation of the framework and a research gap identified by the review.

Seventh, the framework itself is a conceptual synthesis derived from the reviewed literature. It has not been externally validated through expert elicitation, case testing, or formal empirical application. Its present contribution lies in structuring the field. Future work is needed to develop validated instruments for prediction or assessment.

4.5. Future research

In direct response to RQ3, several directions for the responsible adoption and evaluation of generative AI companion systems follow from the findings and limitations. One important direction is the empirical examination of how organizational and regulatory conditions shape user-facing relational outcomes. The reviewed corpus already provides strong accounts of attachment, companionship, anthropomorphism, self-disclosure, and social support. What remains less developed is how these outcomes are configured by platform incentives, monetization strategies, moderation choices, service continuity, and governance structures. Future studies should therefore examine these relationships across stakeholder layers and determine how changes in platform or governance conditions affect users' relational experiences over time.

Another direction concerns populations and contexts that remain underexamined. The literature frequently raises concerns about younger users and emotionally vulnerable populations, but more targeted study is needed. Cross-cultural work, accessibility-focused work, and research on differences in social, legal, and institutional context would also help strengthen the field, especially because the current corpus is strongly

shaped by platform-specific and culturally bounded studies. Such research could clarify whether the identified benefits, risks, and responsible-adoption requirements vary across user groups, systems, and contexts.

A final direction concerns ethical issues that remain only partially integrated into the current corpus. These include the labor and infrastructure conditions behind conversational AI, the upstream data sources used to build underlying models, the role of long-term memory and multimodality in intensifying relational and privacy risks, and the relation between generative AI companion services and broader platform ecologies. Existing studies already point to risks around memory, disclosure, personalization, emotional mirroring, pseudo-intimacy, and privacy, but these issues deserve closer integration with the study of human–chatbot interaction. Addressing them would support a more comprehensive evaluation of how generative AI companion relationships are designed, sustained, and governed.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT in order to support language revision and drafting. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research was supported by the Research Council of Finland through the project Emotional Experience of Privacy and Ethics in Everyday Pervasive Systems (BUGGED) (decision no. 348391). The funder had no involvement in the conducted research beyond financial support.

CRediT authorship contribution statement

Joni-Roy Piispanen: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tinja Myllyviita:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **Ville Vakkuri:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Conceptualization. **Rebekah Rousi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.infsof.2026.108340.

Appendix A. Database-specific search strings

The full search strings used in the review are provided below for transparency and reproducibility. Database syntax was adjusted to each platform's field logic, but the substantive search intent remained constant across databases.

Scopus

TITLE-ABS-KEY("social chatbot" OR "ai companion" OR "companion chatbot" OR "social support agent" OR "human–chatbot" OR "chatbot companion" OR "chatbot* relationship" OR "emotiona* artificial intelligence" OR "emotiona* ai" OR "human–ai fri*" OR "automated conversational agent" OR "service robot accept*" OR "chatbot* social" OR "anthropomorphic cue" OR "anthropomorphic design cue") AND PUBYEAR > 2004 AND PUBYEAR < 2025

ACM Digital Library

AllField:("social chatbot" OR "ai companion" OR "companion chatbot" OR "social support agent" OR "human–chatbot" OR "chatbot companion" OR "chatbot* relationship" OR "emotiona* artificial intelligence" OR "emotiona* ai" OR "human–ai fri*" OR "automated conversational agent" OR "service robot accept*" OR "chatbot* social" OR "anthropomorphic cue" OR "anthropomorphic design cue") AND [E-Publication Date: (01/01/2005 TO 12/31/2024)]

Web of Science Core Collection + MEDLINE

(TI=("social chatbot" OR "ai companion" OR "companion chatbot" OR "social support agent" OR "human–chatbot" OR "chatbot companion" OR "chatbot* relationship" OR "emotiona* artificial intelligence" OR "emotiona* ai" OR "human–ai fri*" OR "automated conversational agent" OR "service robot accept*" OR "chatbot* social" OR "anthropomorphic cue" OR "anthropomorphic design cue") OR AB=("social chatbot" OR "ai companion" OR "companion chatbot" OR "social support agent" OR "human–chatbot" OR "chatbot companion" OR "chatbot* relationship" OR "emotiona* artificial intelligence" OR "emotiona* ai" OR "human–ai fri*" OR "automated conversational agent" OR "service robot accept*" OR "chatbot* social" OR "anthropomorphic cue" OR "anthropomorphic design cue") OR AK=("social chatbot" OR "ai companion" OR "companion chatbot" OR "social support agent" OR "human–chatbot" OR "chatbot companion" OR "chatbot* relationship" OR "emotiona* artificial intelligence" OR "emotiona* ai" OR "human–ai fri*" OR "automated conversational agent" OR "service robot accept*" OR "chatbot* social" OR "anthropomorphic cue" OR "anthropomorphic design cue")) AND PY=(2005–2024)

IEEE Xplore

Search 1:

(("Document Title":"social chatbot" OR "Abstract":"social chatbot" OR "Index Terms":"social chatbot") OR ("Document Title":"ai companion" OR "Abstract":"ai companion" OR "Index Terms":"ai companion") OR ("Document Title":"companion chatbot" OR "Abstract":"companion chatbot" OR "Index Terms":"companion chatbot") OR ("Document Title":"social support agent" OR "Abstract":"social support agent" OR "Index Terms":"social support agent") OR ("Document Title":"human–chatbot" OR "Abstract":"human–chatbot" OR "Index Terms":"human–chatbot") OR ("Document Title":"chatbot companion" OR "Abstract":"chatbot companion" OR "Index Terms":"chatbot companion") OR ("Document Title":"chatbot* relationship"

OR "Abstract":"chatbot* relationship" OR "Index Terms":"chatbot* relationship") OR ("Document Title":"emotiona* artificial intelligence" OR "Abstract":"emotiona* artificial intelligence" OR "Index Terms":"emotiona* artificial intelligence") OR ("Document Title":"emotiona* ai" OR "Abstract":"emotiona* ai" OR "Index Terms":"emotiona* ai"))

Range: 2005–2024

Search 2:

(("Document Title":"human–ai fri*" OR "Abstract":"human–ai fri*" OR "Index Terms":"human–ai fri*") OR ("Document Title":"automated conversational agent" OR "Abstract":"automated conversational agent" OR "Index Terms":"automated conversational agent") OR ("Document Title":"service robot accept*" OR "Abstract":"service robot accept*" OR "Index Terms":"service robot accept*") OR ("Document Title":"chatbot* social" OR "Abstract":"chatbot* social" OR "Index Terms":"chatbot* social") OR ("Document Title":"anthropomorphic cue" OR "Abstract":"anthropomorphic cue" OR "Index Terms":"anthropomorphic cue") OR ("Document Title":"anthropomorphic design cue" OR "Abstract":"anthropomorphic design cue" OR "Index Terms":"anthropomorphic design cue"))

Range: 2005–2024

Appendix B. Synthesis pathway

Table B.1 depicts the synthesis pathway from extracted study-level material to analytic theme. Each row preserves a representative trace from a reported finding, argument, or observation to an initial code, a descriptive theme, and a higher-order analytical theme. The examples are illustrative rather than exhaustive and include different evidentiary functions within the corpus. Coding was non-exclusive, so a publication or extracted unit could support more than one synthesis element when it played a distinct and substantiated role in each.

Table B.1
Examples of the synthesis pathway.

Finding	Initial code	Descriptive theme	Analytical theme
Users describe chatbots as friends or companions and report trust, closeness, and relationship development [5,24,9].	Friendship, closeness, relationship trajectory	Relationship formation and attachment	Trust, attachment, over-reliance, and psychosocial outcomes
Self-disclosure can deepen perceived intimacy and trust over time [57, 58].	Disclosure reciprocity, remembered personal information	Disclosure and relational deepening	Usability and relational affordances in interaction
Deletion, service disruption, or companion loss can produce distress and continuity harms [68].	Service continuity, loss, user recourse	Platform governance and service logic	Responsible adoption, platform design, and organizational mediation
Privacy-by-design and data protection analyses emphasize consent, transparency, and accountability in companion-chatbot contexts [32,35].	Data protection, accountability, meaningful consent	Governance and accountability	Governance, privacy, and accountability
Avatar presentation, gaze, emotional language, and response style shape comfort and social interpretation [16,40].	Interface cues, emotional style, avatar presentation	Anthropomorphic and interface cues	Usability and relational affordances in interaction
Users may experience chatbots as supportive while also reporting uncertainty, dependence, discomfort, or ambivalence [59,6,66,7, 39].	Supportive use, uncertainty, dependence, ambivalence	Ambivalent psychosocial trajectories	Trust, attachment, over-reliance, and psychosocial outcomes

Table B.2 complements the worked examples with a concise codebook for the final study-level coding. UA, TA, RA, GP, and TM identify the four analytical themes and the cross-cutting technological mediation code. The definitions and exclusion boundaries below specify how the final codes were applied.

Table B.2
Concise codebook for final synthesis codes.

Code	Operational definition and exclusion boundary	Corpus example(s)
UA	Usability and relational affordances in interaction. Apply to interface, interaction, or design features that shape usability, social presence, warmth, familiarity, perceived reciprocity, or relational interpretation. Exclude user motives or outcomes that are not linked to an interactional or design mechanism.	Avatar presentation, gaze, emotional language, and response style [16,40].
TA	Trust, attachment, over-reliance, and psychosocial outcomes. Apply to motivations and reported experiences involving companionship, trust, attachment, dependence, loneliness, support, distress, or ambivalence. Exclude purely technical, organizational, or legal issues without a user-relational or psychosocial process or outcome.	Relationship development and closeness [24]; dependence and ambivalence [39].
RA	Responsible adoption, platform design, and organizational mediation. Apply to product or service conditions, organizational decisions, monetization, moderation, premium features, service continuity, and institutional responsibility. Exclude individual experiences with no platform or organizational mechanism and legal requirements better coded as GP.	Deletion, service disruption, and user recourse [68]; platform and business logics [20].
GP	Governance, privacy, and accountability. Apply to data protection, consent, transparency, accountability, legal or regulatory obligations, and safeguards for vulnerable users. Exclude general design criticism without governance, privacy, or accountability content.	Privacy-by-design, meaningful consent, and accountability [32,35].
TM	Cross-cutting technological mediation. Apply when a specific technical or design mechanism connects user experience, organizational practice, or governance, including memory, personalization, prompts, avatar embodiment, emotional response style, or service continuity. Exclude generic references to technology without an identified mediating mechanism.	Disclosure reciprocity as mechanisms of relational deepening [57,58].

Appendix C. Study-level evidence matrix, appraisal, and framework coding

This appendix documents the contribution-type-specific critical-appraisal criteria and the study-level audit trail for the 42 formal corpus

publications. Table C.1 operationalizes each appraisal family, and Table C.2 links every publication to its contribution type, appraisal-family code, role-specific evidentiary category, main limitation, synthesis-element coding, framework dimension or stakeholder relevance, and framework role. The Element(s) column applies the material-contribution rule in Section 2.8 and the operational definitions in Table B.2. Coding is non-exclusive: a publication could contribute materially to more than one synthesis element.

Appraisal categories indicate the relative strength of each publication for its assigned role in the present synthesis and served as weighting criteria, not exclusion criteria. Strong indicates strong support for the assigned role; moderate-to-strong indicates robust but meaningfully constrained support; moderate indicates useful but bounded support; and limited indicates interpretive salience or issue framing with constrained support for recurrence claims.

Appraisal-family codes. EMP-Q = MMAT-informed qualitative empirical criteria; EMP-QD = MMAT-informed quantitative descriptive criteria; EMP-NR = MMAT-informed quantitative non-randomized or associational criteria; EMP-EXP = MMAT-informed controlled experimental criteria, using randomized criteria when random allocation was reported and non-randomized criteria when allocation was unclear or non-randomized; EMP-MM = MMAT-informed mixed-methods criteria, including component and integration appraisal; TECH = bespoke technical, implementation, log-analysis, or computational-system criteria; ARG = adapted JBI-informed argument/text appraisal for conceptual, ethical, philosophical, theoretical, and critical contributions; LAW/POL = adapted JBI-informed legal or policy-oriented appraisal; REV = adapted review/synthesis appraisal.

Table C.1
Contribution-type-specific critical-appraisal criteria.

Family	Applied to	Core criteria and claim-boundary checks
EMP-Q	Qualitative empirical studies	Clarity of the question; appropriateness of the qualitative approach; adequacy and transparency of sampling, context, and data collection; coherence of analysis; grounding of findings in the data; and alignment of interpretations and limitations with the evidence.
EMP-QD	Quantitative descriptive studies	Fit of the sampling strategy; representativeness and selection limits; appropriateness and completeness of measures and data; nonresponse or missing-data concerns; suitability of descriptive analyses; and alignment of conclusions with a descriptive design.
EMP-NR	Quantitative non-randomized or associational studies	Sampling and measurement quality; clarity of predictors, exposures, and outcomes; consideration of confounding; completeness of outcome data; appropriateness of analysis; and avoidance of causal claims that exceed the design.
EMP-EXP	Controlled experimental studies	Clarity of intervention and comparator; allocation or randomization as reported; baseline comparability; outcome measurement and completeness; adherence or intervention fidelity; appropriateness of analysis; and claim boundaries arising from allocation and setting.
EMP-MM	Mixed-methods studies	Quality of the qualitative and quantitative components; rationale for combining them; adequacy of integration; treatment of divergent findings; coherence of meta-inferences; and alignment of conclusions with the integrated evidence.
TECH	Technical, implementation, log-analysis, or computational-system studies	System, model, and data-source description; data provenance and access; transparency of implementation and evaluation; fit between evaluation and claims; reproducibility; and version, platform, or proprietary-system constraints.
ARG	Conceptual, ethical, philosophical, theoretical, or critical contributions	Clarity of the problem and thesis; adequacy of the source or evidence base; definition and consistent use of concepts; logical coherence; engagement with alternatives or counterarguments; warrant for conclusions; and explicit scope conditions.
LAW/POL	Legal or policy-oriented contributions	Clarity of the legal or policy question; authority, relevance, and currency of sources; jurisdictional scope; transparency of interpretive method; distinction between descriptive law and normative recommendation; and limits on transfer across jurisdictions.
REV	Review or synthesis papers	Clarity of question and scope; transparency of search and selection; extraction and appraisal procedures where applicable; appropriateness of synthesis; alignment between the underlying evidence and review conclusions; and limitations, including possible overlap with primary studies.

Synthesis element. UA = usability and relational affordances in interaction; TA = trust, attachment, over-reliance, and psychosocial outcomes; RA = Responsible adoption, platform design, and organizational mediation; GP = governance, privacy, and accountability; TM = technological mediation.

Table C.2
Study-level evidence matrix, appraisal, and framework coding.

Ref.	Type(s)	Appraisal	Main limitation	Element (s)	Dimension(s)	Framework role
[56]	Empirical qualitative taxonomy study	EMP-Q; Moderate	Taxonomy depends on sample and interpretive grouping.	UA; TA	Human	Relationship types and well-being implications
[65]	Quantitative non-randomized / associational survey study	EMP-NR; Moderate	Survey evidence supports perceptions and associations; causal and clinical implications remain bounded.	TA	Human	Social interaction anxiety and fear-based pathways to compulsive chatbot use, moderated by frustration about unavailability
[68]	Empirical qualitative descriptive open-ended survey study	EMP-Q; Moderate-to-strong	Focused on loss and disruption; transferability is bounded by companion-loss cases.	TA; RA; GP	Human; Organization	Continuity, deletion, service disruption, and grief
[5]	Empirical qualitative interview study	EMP-Q; Moderate-to-strong	Sample may emphasize invested users and friendship-oriented interpretations.	UA; TA	Human	AI friendship and user understanding of human-AI relationships
[10]	Literature review / science mapping synthesis	REV; Strong	Broad companionship scope; organizational and regulatory mechanisms receive less direct attention.	UA; TA; TM	Human; Technology	Companionship mechanisms, antecedents, mediators, moderators, and outcomes
[20]	Conceptual / managerial analysis	ARG; Moderate	Managerial framing; limited direct evidence of user harms.	UA; RA	Organization; Human	Business models, companion capabilities, and organizational use
[62]	Critical literature review / socio-technical conceptual analysis	REV + ARG; Moderate	Normative and conceptual contribution; empirical grounding is secondary.	UA; GP; TM	Human; Technology; Organization	Pseudo-empathy and ethical challenge to human essence

(*continued on next page*)

Table C.2 (*continued*)

Ref.	Type(s)	Appraisal	Main limitation	Element (s)	Dimension(s)	Framework role
[59]	Mixed-method empirical and log-analysis study	EMP-MM + TECH; Moderate-to-strong	Top-user focus may overrepresent highly engaged use; platform-specific logs shape inference.	UA; TA; RA; TM	Human; Organization; Technology	Intensive use, anthropomorphism, and use-pattern interpretation
[31]	Dialectical conceptual inquiry	ARG; Moderate-to-strong	Dialectical conceptual analysis; empirical claims depend on case interpretation.	UA; TA; RA; GP	Human; Organization	Ethical tensions in human-AI companionship
[15]	Longitudinal quantitative empirical study	EMP-NR; Strong	One chatbot and bounded observation period; outcomes may differ in immersive or newer companion systems.	UA; TA	Human	Relationship formation, change, and limits over time
[16]	Empirical/content analysis	EMP-NR; Moderate	Early interactions over three weeks; longer-term relationship development remains unexamined.	UA; TM	Human; Technology	Rapport-building language strategies and social simulation
[57]	Controlled quantitative experiment	EMP-EXP; Moderate	Short experimental exposure limits claims about repeated disclosure and longitudinal trust.	UA; TA; TM	Human; Technology	Disclosure, trust, reciprocity, and interactional deepening
[32]	Legal analysis	LAW/POL; Strong	Jurisdictionally grounded legal analysis; empirical user outcomes are outside scope.	RA; GP	Organization; Regulator	Privacy by design, data protection, and legal risks
[28]	Conceptual / philosophical argument	ARG; Limited	Conceptual argument; direct empirical testing remains limited.	TA; GP	Human	Loneliness, recognition, and societal implications of AI companions
[29]	Empirical mixed-methods study	EMP-MM; Moderate-to-strong	Context-specific crisis setting and Chinese female Replika-user focus; support effects may depend on situational stressors.	TA	Human	Mediated empathy, resilience, and emotional support
[67]	Computational topic-modeling / social-media analysis	EMP-QD + TECH; Moderate	Social-media data may overrepresent vocal users and public controversies.	TA; RA; GP	Human; Organization	Public perceptions, privacy, trust, and platform discourse
[63]	Empirical qualitative study / digital ethnography	EMP-Q; Moderate	Exploratory qualitative design; identity effects remain context-specific and interpretive.	UA; TA	Human; Technology	Self-reflection, identity, and AI companion mirroring
[6]	Qualitative grounded theory	EMP-Q; Moderate-to-strong	Platform-specific and self-selected Reddit accounts; supports mechanism identification more than prevalence estimation.	TA; RA; GP	Human; Organization	Emotional dependence, mental-health harms, and platform-mediated vulnerability
[69]	Critical media/cultural analysis	ARG; Moderate	Claims are strongest for cultural framing and imaginaries; empirical recurrence claims remain limited.	RA; GP	Human; Organization	Gendered harms, imaginaries, and critical platform critique
[60]	Empirical qualitative / critical cultural analysis	EMP-Q + ARG; Limited	Purposive "Top Posts" sampling, one Douban subgroup/platform, single-researcher analysis, translation, and inferred gender	UA; TA; RA	Human; Technology	Norms of friendship, intimacy, love, and robotized relations
[21]	Controlled quantitative experiment	EMP-EXP; Moderate	Experimental exposure constrains claims about long-term attachment and sustained companion use.	UA; TA; TM	Human; Technology	Social presence, loneliness, and acceptance of AI companions
[36]	Conceptual analysis	ARG; Limited	Empirical boundary conditions remain to be tested.	UA; TA; GP	Human; Technology	Validation, deception, and emotional echo chambers
[66]	Empirical qualitative user-review analysis	EMP-Q; Moderate	User-review data are self-selected and platform-specific.	UA; TA; RA	Human; Organization	Contradictory user experiences and platform-specific tensions
[7]	Empirical qualitative study	EMP-Q; Moderate	Power-user uncertainty focus; broader behavioral outcomes remain less developed.	UA; TA	Human	Ambivalence, uncertainty, and relationship interpretation
[8]	Empirical qualitative study	EMP-Q; Moderate-to-strong	Focused on Replika and romantic users; transferability is bounded.	UA; TA	Human	Romantic meaning-making and intimate human-AI relationships
[24]	Mixed-method empirical study	EMP-MM; Moderate-to-strong	Single platform and self-selected users; prevalence remains uncertain.	UA; TA; TM	Human; Technology	Relationship development, attachment, and Replika use
[25]	Systematic literature review	REV; Strong	Broader consumer-machine scope; generative AI companions form part of the synthesis.	UA; TA; RA; TM	Human; Technology; Organization	Relational mechanisms, outcomes, and field positioning
[35]	Legal/policy case analysis	LAW/POL; Moderate	Single-platform case; legal and policy interpretation rather than empirical outcome evidence.	RA; GP	Organization; Regulator	Privacy, transparency, compliance, and platform accountability
[61]	Technical review / field outlook	REV; Moderate	Outlook-oriented synthesis; provides design and field framing rather than systematic outcome appraisal.	UA; TM	Technology; Human	Field development and social-chatbot design challenges
[58]	Qualitative longitudinal empirical study	EMP-Q; Moderate-to-strong	Backend data practices remain partly unobservable; platform-specific self-disclosure trajectories.	UA; TA; GP; TM	Human; Technology; Organization	Disclosure over time, privacy salience, and relational deepening
[22]	Empirical qualitative interview study	EMP-Q; Moderate-to-strong	Interview sample of users with developed Replika friendships; trajectories are context- and platform-specific.	UA; TA	Human	Friendship, companionship, trust, and relationship formation

(*continued on next page*)

Table C.2 (*continued*)

Ref.	Type(s)	Appraisal	Main limitation	Element (s)	Dimension(s)	Framework role
[9]	Qualitative longitudinal empirical study	EMP-Q; Moderate-to-strong	System- and sample-specific longitudinal evidence; broader platform transferability remains bounded.	UA; TA	Human	Relationship trajectories and changing user interpretations over time
[23]	Empirical qualitative thematic analysis	EMP-Q; Moderate-to-strong	Self-reported experience data; transferability depends on everyday-use and companion-app context.	TA	Human	Everyday social support, companionship, and supportive use
[30]	Empirical qualitative thematic analysis	EMP-Q; Moderate-to-strong	Self-reported experience data; prevalence cannot be inferred.	TA	Human	Motivations, topics, and everyday uses of generative AI companions
[47]	Between-subject quantitative experiment	EMP-EXP; Moderate-to-strong	Brand-chatbot setting differs from explicit companion-AI platforms.	UA; TA; TM	Human; Technology	Social presence, parasocial interaction, dialogue, and engagement
[64]	Conceptual philosophy	ARG; Moderate	Conceptual argument; does not provide prevalence or outcome estimates.	UA; TA; TM	Human; Technology	Authenticity, affective asymmetry, and normative interpretation of emotional AI
[37]	Opinion / conceptual discussion	ARG; Limited	Opinion-style contribution; empirical operationalization remains limited.	UA; TA; GP	Human; Technology	Pseudo-intimacy and social-ethical impacts
[38]	Theory-led qualitative case study	EMP-Q; Moderate	Single-platform and theory-led interpretation; limited generalizability beyond Replika-style relationships.	UA; TA	Human	Attachment dynamics and relationship styles
[39]	Quantitative non-randomized / associational study	EMP-NR; Moderate-to-strong	Causal direction between engagement and dependence remains difficult to establish.	TA	Human	Psychological dependence and engagement-related harms
[40]	Two controlled quantitative experiments	EMP-EXP; Moderate-to-strong	Specific interface manipulation; broader platform conditions are indirect.	UA; TM	Technology; Human	Avatar design, comfort, gaze, and interactional interpretation
[70]	Quantitative non-randomized / associational survey study	EMP-NR; Moderate	Associational questionnaire evidence; longitudinal confirmation needed.	TA	Human	Socialization impacts and possible displacement concerns
[14]	Engineering implementation and large-scale log analysis	TECH; Moderate-to-strong	Single proprietary system; independent access to backend decisions and log-generating conditions is limited.	UA; TM	Human; Technology	Relational design, emotional response style, system architecture, and long-term engagement

Note. The appraisal categories are role-specific evidentiary weights for this synthesis. They are not formal MMAT or JBI quality grades and should not be interpreted as comparable risk-of-bias scores across heterogeneous contribution types.

Data availability

Data will be made available on request.

References

[1] M. Adam, M. Wessel, A. Benlian, AI-based chatbots in customer service and their effects on user compliance, Electron. Mark. 31 (2021) 427–445, https://doi.org/10.1007/s12525-020-00414-7.

[2] R.E. Bawack, S.F. Wamba, K.D.A. Carillo, S. Akter, Artificial intelligence in e-commerce: a bibliometric study and literature review, Electron. Mark. 32 (2022) 297–338, https://doi.org/10.1007/s12525-022-00537-z.

[3] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, Stroke Vasc. Neurol. 2 (2017) 230–243, https://doi.org/10.1136/svn-2017-000101.

[4] S. Roksandić, N. Protrka, M. Engelhart, Trustworthy artificial intelligence and its use by law enforcement authorities: where do we stand?. 45th Jubilee International Convention on Information, Communication and Electronic Technology, 2022, pp. 1225–1232, https://doi.org/10.23919/MIPRO55190.2022.9803606.

[5] P.B. Brandtzæg, M. Skjuve, A. Følstad, My AI friend: how users of a social chatbot understand their human–AI friendship, Hum. Commun. Res. 48 (3) (2022) 404–429, https://doi.org/10.1093/hcr/hqac008.

[6] L. Laestadius, A. Bishop, M. Gonzalez, D. Illencik, C. Campos-Castillo, Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika, New Media Soc. 26 (10) (2024) 5923–5941, https://doi.org/10.1177/14614448221142007.

[7] S. Pan, J. Cui, Y. Mou, Desirable or distasteful? Exploring uncertainty in human–chatbot relationships, Int. J. Hum. Comput. Interact. 40 (20) (2024) 6545–6555, https://doi.org/10.1080/10447318.2023.2256554.

[8] S. Pan, Y. Mou, Constructing the meaning of human–AI romantic relationships from the perspectives of users dating the social chatbot Replika, Pers. Relatsh. 31 (4) (2024) 1090–1112, https://doi.org/10.1111/pere.12572.

[9] M. Skjuve, A. Følstad, K.I. Fostervold, P.B. Brandtzæg, A longitudinal study of human–chatbot relationships, Int. J. Hum. Comput. Stud. 168 (2022) 102903, https://doi.org/10.1016/j.ijhcs.2022.102903.

[10] R. Chaturvedi, S. Verma, R. Das, Y.K. Dwivedi, Social companionship with artificial intelligence: recent trends and future avenues, Technol. Forecast Soc. Change 193 (2023) 122634, https://doi.org/10.1016/j.techfore.2023.122634.

[11] L. Christoforakos, N. Feicht, S. Hinkofer, A. Löscher, S.F. Schlegl, S. Diefenbach, Connect with me. Exploring influencing factors in a human-technology relationship based on regular chatbot use, Front. Digit. Health 3 (2021) 689999, https://doi.org/10.3389/fdgth.2021.689999.

[12] A. Janson, How to leverage anthropomorphism for chatbot service interfaces: the interplay of communication style and personification, Comput. Hum. Behav. 149 (2023) 107954, https://doi.org/10.1016/j.chb.2023.107954.

[13] L. Qiu, Y. Shiu, P. Lin, R. Song, Y. Liu, D. Zhao, R. Yan, What if bots feel moods?, in: SIGIR 2020: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 1161–1170, https://doi.org/10.1145/3397271.3401108.

[14] L. Zhou, J. Gao, D. Li, H.-Y. Shum, The design and implementation of XiaoIce, an empathetic social chatbot, Comput. Linguist. 46 (1) (2020) 53–93, https://doi.org/10.1162/coli_a_00368.

[15] E. Croes, M.L. Antheunis, Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot, J. Soc. Pers. Relat. 38 (1) (2021) 279–300, https://doi.org/10.1177/0265407520959463.

[16] E. Croes, M.L. Antheunis, M.B. Goudbeek, N.W. Wildman, I am in your computer while we talk to each other": a content analysis on the use of language-based strategies by humans and a social chatbot in initial human–chatbot interactions, Int. J. Hum. Comput. Interact. 39 (10) (2023) 2155–2173, https://doi.org/10.1080/10447318.2022.2075574.

[17] Y. Lee, N. Yamashita, Y. Huang, W. Fu, I hear you, I feel you": encouraging deep self-disclosure through a chatbot. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, 2020, pp. 1–12, https://doi.org/10.1145/3313831.3376175.

[18] D.-C. Toader, G. Boca, R. Toader, M. Măcelaru, C. Toader, D. Ighian, A. T. Rădulescu, The effect of social presence and chatbot errors on trust, Sustainability 12 (1) (2020) 256, https://doi.org/10.3390/su12010256.

[19] X. Yang, M. Aurisicchio, W. Baxter, Understanding affective experiences with conversational agents, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–12, https://doi.org/10.1145/3290605.3300772.

[20] R. Chaturvedi, S. Verma, V. Srivastava, Empowering AI companions for enhanced relationship marketing, Calif. Manage. Rev. 66 (2) (2024) 65–90, https://doi.org/10.1177/00081256231215838.

[21] K. Merrill Jr., J. Kim, C. Collins, AI companions for lonely individuals and the role of social presence, Commun. Res. Rep. 39 (2) (2022) 93–103, https://doi.org/10.1080/08824096.2022.2045929.

[22] M. Skjuve, A. Følstad, K.I. Fostervold, P.B. Brandtzæg, My chatbot companion: a study of human–chatbot relationships, Int. J. Hum. Comput. Stud. 149 (2021) 102601, https://doi.org/10.1016/j.ijhcs.2021.102601.

[23] V. Ta, C. Griffith, C. Boatfield, X. Wang, M. Civitello, H. Bader, E. DeCero, A. Loggarakis, User experiences of social support from companion chatbots in

everyday contexts: thematic analysis, J. Med. Internet Res. 22 (3) (2020) e16235, https://doi.org/10.2196/16235.
[24] I. Pentina, T. Hancock, T. Xie, Exploring relationship development with social chatbots: a mixed-method study of Replika, Comput. Hum. Behav. 140 (2023) 107600, https://doi.org/10.1016/j.chb.2022.107600.
[25] I. Pentina, T. Xie, T. Hancock, A. Bailey, Consumer–machine relationships in the age of artificial intelligence: systematic literature review and research directions, Psychol. Mark. 40 (8) (2023) 1593–1614, https://doi.org/10.1002/mar.21853.
[26] P.B. Brandtzæg, M. Skjuve, K.K. Dysthe, A. Følstad, When the social becomes non-human: young people's perception of social support in chatbots, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–13, https://doi.org/10.1145/3411764.3445318.
[27] M. de Gennaro, E.G. Krumhuber, G. Lucas, Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood, Front. Psychol. 10 (2020) 3061, https://doi.org/10.3389/fpsyg.2019.03061.
[28] K.A. Jacobs, Digital loneliness—changes of social recognition through AI companions, Front. Digit. Health 6 (2024) 1281037, https://doi.org/10.3389/fdgth.2024.1281037.
[29] Q. Jiang, Y. Zhang, W. Pian, Chatbot as an emergency exist: mediated empathy for resilience via human–AI interaction during the COVID-19 pandemic, Inf. Process. Manag. 59 (6) (2022) 103074, https://doi.org/10.1016/j.ipm.2022.103074.
[30] V.P. Ta-Johnson, C. Boatfield, X. Wang, E. DeCero, I.C. Krupica, S.D. Rasof, A. Motzer, W.M. Pedryc, Assessing the topics and motivating factors behind human-social chatbot interactions: thematic analysis of user experiences, JMIR Hum. Factors 9 (4) (2022) e38876, https://doi.org/10.2196/38876.
[31] R.F. Ciriello, O. Hannon, A.Y. Chen, E. Vaast, Ethical tensions in human–AI companionship: a dialectical inquiry into Replika, in: Proceedings of the Annual Hawaii International Conference on System Sciences, 2024, pp. 488–497, https://doi.org/10.24251/HICSS.2024.058.
[32] P. Dewitte, Better alone than in bad company: addressing the risks of companion chatbots through data protection by design, Comput. Law Secur. Rev. 54 (2024) 106019, https://doi.org/10.1016/j.clsr.2024.106019.
[33] Italian Data Protection Authority. (2023). Artificial intelligence: Italian SA clamps down on "Replika" chatbot. Too many risks to children and emotionally vulnerable individuals. Retrieved May 9, 2026, from https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9852506.
[34] E. Kaczmarek, Self-deception in human–AI emotional relations, J. Appl. Philos. 42 (3) (2025) 814–831, https://doi.org/10.1111/japp.12786.
[35] J.-R. Piispanen, T. Myllyviita, V. Vakkuri, R. Rousi, Smoke screens and scapegoats: the reality of General Data Protection Regulation compliance—privacy and ethics in the case of Replika AI, in: T. Olsson, O. Sahlgren, J. Parviainen, S. Westerstrand, J.T. Harviainen, A. Laitinen, J. Rantala (Eds.), Proceedings of the Conference on Technology Ethics 2024 (TETHICS 2024), Tampere, Finland, November 6–7, 2024 (CEUR Workshop Proceedings, Proceedings of the Conference on Technology Ethics 2024 (TETHICS 2024), Tampere, Finland, November 6–7, 2024 (CEUR Workshop Proceedings, 3901, CEUR-WS.org, 2024, pp. 1–12. https://ceur-ws.org/Vol-3901/paper_1.pdf.
[36] P.M.T. Mlonyeni, Personal AI, deception, and the problem of emotional bubbles, AI Soc. 40 (2025) 1927–1938, https://doi.org/10.1007/s00146-024-01958-4.
[37] J. Wu, Social and ethical impact of emotional AI advancement: the rise of pseudo-intimacy relationships and challenges in human interactions, Front. Psychol. 15 (2024) 1410462, https://doi.org/10.3389/fpsyg.2024.1410462.
[38] T. Xie, I. Pentina, Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika, in: Proceedings of the Annual Hawaii International Conference on System Sciences, 2022, pp. 2046–2055, https://doi.org/10.24251/HICSS.2022.258.
[39] T. Xie, I. Pentina, T. Hancock, Friend, mentor, lover: does chatbot engagement lead to psychological dependence? J. Serv. Manag. 34 (4) (2023) 806–828, https://doi.org/10.1108/JOSM-02-2022-0072.
[40] J. Yuan, X. Peng, Y. Liu, Q. Wang, Don't look at me!: the role of avatars' presentation style and gaze direction in social chatbot design, Comput. Hum. Behav. 164 (2025) 108501, https://doi.org/10.1016/j.chb.2024.108501.
[41] R. Urbani, C. Ferreira, J. Lam, Managerial framework for evaluating AI chatbot integration: bridging organizational readiness and technological challenges, Bus. Horiz. 67 (5) (2024) 595–606, https://doi.org/10.1016/j.bushor.2024.05.004.
[42] European Commission, High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, Eur. Comm. (2019). Retrieved May 9, 2026, from, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[43] European Parliament and Council of the European Union, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), Off. J. Eur. Union 277 (2022) 1–102. L, https://data.europa.eu/eli/reg/2022/2065/oj.
[44] European Parliament and Council of the European Union, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Off. J. Eur. Union L 2024/1689 (2024). https://data.europa.eu/eli/reg/2024/1689/oj.
[45] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.). IEEE, Retrieved May 9, 2026, from, https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e.pdf, 2019.
[46] B. Mittelstadt, Principles alone cannot guarantee ethical AI, Nat. Mach. Intell. 1 (2019) 501–507, https://doi.org/10.1038/s42256-019-0114-4.
[47] W.-H.S. Tsai, Y. Liu, C.-H. Chuan, How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue, J. Res. Interact. Mark. 15 (3) (2021) 460–482, https://doi.org/10.1108/JRIM-12-2019-0200.
[48] G. Murtarelli, A. Gregory, S. Romenti, A conversation-based perspective for shaping ethical human–machine interactions: the particular challenge of chatbots, J. Bus. Res. 129 (2021) 927–935, https://doi.org/10.1016/j.jbusres.2020.09.018.
[49] D.S. Cruzes, T. Dybå, Recommended steps for thematic synthesis in software engineering, in: 2011 International Symposium on Empirical Software Engineering and Measurement, IEEE, 2011, pp. 275–284, https://doi.org/10.1109/ESEM.2011.36.
[50] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering (Version 2.3; EBSE Technical Report EBSE-2007-01), Keele University and Durham University, 2007. Retrieved May 9, 2026, from, https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf.
[51] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) n71, https://doi.org/10.1136/bmj.n71.
[52] M.J. Page, D. Moher, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, J.E. McKenzie, PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews, BMJ 372 (2021) n160, https://doi.org/10.1136/bmj.n160.
[53] R. Sarkis-Onofre, F. Catalá-López, E. Aromataris, C. Lockwood, How to properly use the PRISMA statement, Syst. Rev. 10 (1) (2021) 117, https://doi.org/10.1186/s13643-021-01671-z.
[54] Q.N. Hong, S. Fàbregues, G. Bartlett, F. Boardman, M. Cargo, P. Dagenais, M. P. Gagnon, F. Griffiths, B. Nicolau, A. O'Cathain, M.C. Rousseau, I. Vedel, P. Pluye, The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers, Educ. Inf. 34 (4) (2018) 285–291, https://doi.org/10.3233/EFI-180221.
[55] E. Aromataris, Z. Munn, JBI manual for evidence synthesis, JBI (2020). Retrieved May 9, 2026, from, https://synthesismanual.jbi.global.
[56] A. Alabed, A. Javornik, D. Gregory-Smith, R. Casey, More than just a chat: a taxonomy of consumers' relationships with conversational AI agents and their well-being implications, Eur. J. Mark. 58 (2) (2024) 373–409, https://doi.org/10.1108/EJM-01-2023-0037.
[57] E. Croes, M.L. Antheunis, L. He, You go first: the effects of self-disclosure reciprocity in human–chatbot interactions, in: 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, 2023, pp. 1–4, https://doi.org/10.1109/ACIIW59127.2023.10388091.
[58] M. Skjuve, A. Følstad, P.B. Brandtzæg, A longitudinal study of self-disclosure in human–chatbot relationships, Interact. Comput. 35 (1) (2023) 24–39, https://doi.org/10.1093/iwc/iwad022.
[59] H. Chin, A. Zhunis, M. Cha, Behaviors and perceptions of human–chatbot interactions based on top active users of a commercial social chatbot, Proc. ACM Hum.-Comput. Interact. 8 (CSCW2) (2024) 483, https://doi.org/10.1145/3687022.
[60] B. Lin, The AI chatbot always flirts with me, should I flirt back: from the McDonaldization of friendship to the robotization of love, Soc. Media Soc. 10 (4) (2024), https://doi.org/10.1177/20563051241296229.
[61] H.-Y. Shum, X. He, D. Li, From ELIZA to XiaoIce: challenges and opportunities with social chatbots, Front. Inf. Technol. Electron. Eng. 19 (1) (2018) 10–26, https://doi.org/10.1631/FITEE.1700826.
[62] A.Y. Chen, S.I. Kögel, O. Hannon, R.F. Ciriello, Feels like empathy: how "emotional" AI challenges human essence, in: Australasian Conference on Information Systems, ACIS 2023, 2023. https://aisel.aisnet.org/acis2023/80.
[63] T. Kouros, V. Papa, Digital mirrors: AI companions and the self, Societies 14 (10) (2024) 200, https://doi.org/10.3390/soc14100200.
[64] E. Weber-Guskar, How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners, Ethics Inf. Technol. 23 (2021) 601–610, https://doi.org/10.1007/s10676-021-09598-8.
[65] F. Ali, Q. Zhang, M.Z. Tauni, K. Shahzad, Social chatbot: my friend in my distress, Int. J. Hum. Comput. Interact. 40 (7) (2024) 1702–1712, https://doi.org/10.1080/10447318.2022.2150745.
[66] M. Namvarpour, A. Razi, Uncovering contradictions in human–AI interactions: lessons learned from user reviews of Replika, in: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, 2024, pp. 579–586, https://doi.org/10.1145/3678884.3681909.
[67] J. Kim, X. Jin, K. Xu, X. Chen, H. Yang, What do people say about Replika, an AI chatbot, on social media? Investigating diverse perspectives on the implications of Replika through a topic modeling analysis, Soc. Sci. J. 62 (3) (2025) 688–703, https://doi.org/10.1080/03623319.2024.2421610.

[68] J. Banks, Deletion, departure, death: experiences of AI companion loss, J. Soc. Pers. Relat. 41 (12) (2024) 3547–3572, https://doi.org/10.1177/02654075241269688.
[69] J.J. Lee, Who is sexually harassed? A python code haha": imaginaries of a post-violent AI world, Fem. Media Stud. 26 (2) (2026) 475–490, https://doi.org/10.1080/14680777.2024.2418393.
[70] Z. Yuan, X. Cheng, Y. Duan, Impact of media dependence: how emotional interactions between users and chat robots affect human socialization? Front. Psychol. 15 (2024) 1388860 https://doi.org/10.3389/fpsyg.2024.1388860.