From Counsel to Code: Implications of AI Dependency for Authentic Human Bonds – A Systematic Review

Khadija Kamener

Mount Kenya University, Mombasa-Nairobi Railroad, Mombasa, Kenya ROR

Correspondence to: Khadija Kamene, hadijashah@yahoo.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Kamene k. From Counsel to Code: Implications of AI Dependency for Authentic Human Bonds – A Systematic Review. Premier Journal of Artificial Intelligence 2025;5:100020

DOI: https://doi.org/10.70389/PJAI.100020

Peer Review

Received: 25 Septemper 2025

Last revised: 3 November 2025

Accepted: 13 November 2025

Version accepted: 6

Publishe: 30 December 2025

Ethical approval: N/a

Consent: N/a

Funding: No industry funding

Conflicts of interest: N/a

Author contribution: Khadija Kamener – Conceptualization, Writing – original draft, review and editing

Guarantor: Khadija Kamene

Provenance and peer-review: Unsolicited and externally peer-reviewed

ABSTRACT

BACKGROUND

With rapid expansion from functional application, artificial intelligence has entered relationship arenas that have historically been designated for human companionship and advice. Millions of people today receive ongoing emotional support from mental health chatbots and AI companions, which enable accessibility and stigma-free connection in situations where traditional counselling might not be available.

OBJECTIVE

This scoping review examines the ethical implications, hazards, and advantages of AI as a companion, confidant, and a counselor while taking cultural and generational setting into consideration.

METHODS

This scoping review was carried out on the literature in the fields of psychology, sociology, ethics, and human AI interaction from 2015 to 2025. In addition to searches in PubMed, PsycINFO, Scopus, Web of Science, and Google Scholar, Policy reports and gray literature were evaluated. Practice guidelines that were frequently cited, conceptual analyses, and empirical research were among the records that qualified.

FINDINGS

AI systems have the potential to improve therapeutic resources, democratize care, and lessen loneliness, especially for marginalized groups. Risks include commodification of intimacy, deterioration of genuine human ties, and over-reliance. Additionally, disparities arise, influenced by socioeconomic divisions, generational variations, and cultural norms. There are also gaps in longitudinal evaluation, cross-cultural research, and multidisciplinary integration despite growing scholarly interest.

IMPLICATIONS

The future of AI companionship depends on its ability to enhance human connection rather than replace it. Ethical integration necessitates research agendas that address long-term psychological and societal implications, design methodologies that incorporate authenticity and oversight and privacy and equitable safeguards. AI companionship can increase access to advice while maintaining the genuineness and tenacity of human ties with these safeguards.

Keywords: AI companionship, Mental health chatbots, Authenticity in human–AI relationships, Emotional dependency on AI, Cross-cultural ethical implications

Introduction

AI has expanded beyond useful applications like automation and navigation to fields that were previously only used for emotional support and human connection. Millions of people worldwide are currently utilizing apps like mental health Chatbots, such as Woebot, Wysa and Replika, which provide communication and coping mechanisms for issues including stress, anxiety, and depression.[1] Adoption of these technologies increased during and after the COVID-19 pandemic, when human counseling services were under pressure, according to a 2023 World Economic Forum Report.[2] People are increasingly using companions for emotional support, personal disclosure, and in certain situations, relationship replacement, in addition to practical help. Younger generations in particular show a greater openness to AI companionship reflecting broader cultural shifts in the conceptualization of intimacy and care.[3]

These advances highlights how urgent it is to investigate how artificial intelligence may affect genuine human connections. Usually reciprocity, vulnerability, and the shared imperfection of lived experience serve as the foundation for authentic partnerships. The profundity of true human connection, on the other hand, cannot be replicated by AI companionship, which is powered by algorithmic response and artificial empathy.[4] Researchers caution that relying too much on artificial intelligence for intimacy could lead to emotional replacement, erode human to human bonds, and commodify caring in ways that alter social norms around friendship and collaboration. Therefore there are urgent concerns about authenticity, reliance, and the moral boundaries of digitally mediated intimacy as a result of the widespread use of AI as a confidant and a counselor

The purpose of this scoping review is to compile the existing research on the relational dynamics of artificial intelligence human interaction, especially in the context of therapy and companionship. Its objectives are of threefold: compile conceptual and empirical literature on AI's relational roles, to identify the key benefits and risks associated with reliance on AI companionship, and to highlight gaps in current knowledge especially regarding long-term effects and cross-cultural variation. Three main research questions are thus addressed in the review: What effects does AI dependence have on social psychological, and emotional ties? What social and ethical ramification might AI companionship have? What are the gaps of the existing body of knowledge? In addressing these questions, the review situates AI companionship within broader debates on authenticity, intimacy, and the future of human connection.

Methodology

This review synthesizes evidence on human AI companionship dynamics with emphasis on mental health

Data availability statement: N/a

chatbots, conversational agents, and embodied AI companions. This study followed PRISMA-ScR guidelines for transparent reporting and provide a PRISMA-style flow diagram to document identification, screening, and inclusion.[5]

Protocol Registration

Because the topic spans multiple disciplines and evolved rapidly during the review period, no preregistered protocol was filed in advance. However, all strings, screening logs, and data extractions templates have been on OSF to ensure all procedural transparency and enable reproducibility (https://osf.io/d6k8y). Any deviations from the initial plan are documented in the supplementary file.

Search Strategy Transparency and Supplementary Materials

Database-specific search strings, date ranges, and complete search logs are provided in Table 1 to enhance transparency and reproducibility. Search strategies were tailored to the indexing structure of each database (e.g., MeSH in PubMed; Thesaurus of Psychological Index Terms in PsycINFO; keyword mapping in Scopus and Web of Science).

For Google Scholar and gray literature, a structured rationale guided inclusion. We limited results to the first 200 hits sorted by relevance, supplemented by targeted searches of conference proceedings (CHI, AAAI, IEEE), organizational reports (WHO, OECD, UNESCO), and preprints on Zenodo and OSF. The inclusion of gray literature aimed to capture emerging work and policy documents in this fast-evolving field, where formal journal publications may lag behind practice developments. Database coverage, search dates, and full Boolean strategies are summarized in Table 1.

Deduplication and Record Management

All retrieved records were exported to Zotero and imported into Excel. Deduplication proceeded in two stages: automated deduplication using reference manager's duplicate detection (matching on title, DOI, and author) and manual verification to remove residual duplicates and near-duplicates.

Eligibility Criteria

We included English-language publictions (2015–2025) that reported empirical data, conceptual analyses, reviews with extractable evidence, or widely cited practice/policy guidance directly addressing human-AI relational context. Eligible content met alt least one of the following operationalized criteria:

- AI as therapeutic supplement: Empirical or evaluation studies where an AI system was used to deliver therapeutic techniques or to augment human therapy.
- AI as confidant/companion: studies or reports examining disclosure, trust, intimacy, emotional reliance, or sustained interpersonal-like interaction with AI systems.
- Dependency and harm: Investigations or analyses reporting overuse, attachment, social displacement, adverse psychological outcomes, or dependency metrics.
- Ethical/societal implication: Policy analyses, ethical commentaries, or mixed-methods work addressing privacy, consent, fairness, or boundaries in AI companionship.

Also included were epidemiological, sociological, and policy perspectives, and gray literature guidance where widely cited in clinical, educational, or public-health contexts.

Exclusions applied to: purely technical/engineering papers without relational/psychological focus; non-scholarly opinion pieces without extractable evidence; and publication outside 2015-2025 window unless explicitly foundational.

Study Selection Dual Screening and Coding Transparency

Screening occurred in two stages (title/abstract→full text). Two reviewers independently screened 20% of records at each stage to calibrate decision rules. Discrepancies were resolved through consensus discussion, with a third reviewer consulted when necessary. Inter-rater reliability was calculated using Cohen's kappa, yielding k = 0.82 for the title/abstract stage and k = 0.87 for the full-text stage, indicating substantial agreement. The remaining 80% of records were

Table 1 | Database-specific search strings, date ranges, and complete search logs

Database	Coverage Period	Search Date	Search String (Boolean Logic)	Hits Retrieved	Notes
PubMed	2015–2025	May 31 2025	("artificial intelligence" OR chatbot* OR "conversational agent*" OR "AI companion*") AND (psychotherapy OR "mental health" OR trust OR relationship)	412	Used MeSH terms + free-text; filters applied for English, humans
PsycINFO	2015–2025	May 31 2025	("artificial intelligence" OR chatbot* OR "digital companion*") AND (disclosure OR intimacy OR counseling)	298	Subject headings adapted to Thesaurus
Scopus	2015–2025	June 1 2025	TITLE-ABS-KEY("AI companion" OR "mental health chatbot") AND ("trust" OR "therapy")	167	Search in title, abstract, keywords only
Web of Science	2015–2025	June 1 2025	("AI companion" OR "conversational agent") AND (relationship OR ethics OR therapy)	110	Limited to SSCI and AHCI
Google Scholar	2015–2025	June 2 2025	"AI companion" AND "mental health" AND "trust"	103 (screened 200 top hits)	Manual relevance screening; included gray literature reports

screened by a single reviewer following calibration. A PRISMA-style flowchart (Figure 1) summarizes the process.

Coding and Data Extraction

Data extraction was performed using a standardized charting form capturing the following fields:

- Author(s), Year, Country/Region
- Study Type / Venue
- Participant Characteristics (age, demographic details)
- AI Companion Type (chatbot, social robot, LLM-based system)
- Study Objectives / Focus
- Key Findings (benefits, risks, ethical considerations)
- Methodological Notes (design, sample, analysis)
- Cultural Context & Relevance
- Risk of Bias / Quality Appraisal

Rationale and Transparency

This structured approach ensured systematic identification, selection, and coding of studies, while documenting decisions and disagreements to maintain transparency and reproducibility. The combination of dual-screening calibration, explicit coding fields, and verification by multiple reviewers supports the reliability of the evidence synthesis.

Critical appraisal and Risk-of-Bias Assessment

A unified appraisal framework was applied in accordance with study design. Randomized controlled trials were assessed using Risk of Bias 2 (RoB 2) tool, while non-randomized and observational studies were evaluated with ROBINS-I. Qualitative and mixed methods studies were appraised using the CASP Qualitative Checklist. Conceptual and policy papers, which are not amenable to risk-of-bias scoring, were evaluated narratively for transparency, evidential grounding, and normative coherence.

Quality appraisal was not used as a basis for exclusion. Instead, it informed the synthesis: higher-quality studies (low RoB or strong CASP ratings) were weighted more heavily in the interpretation, while claims drawn from studies with moderate or serious risk-of-bias are explicitly marked as tentative. All appraisal results are presented in a consolidated format in Table 2 (Quality Appraisal Summary), followed by evidence maps in Tables 3 and 4.

Synthesis Approach

We used an iterative thematic charting approach to synthesize findings across disciplines. Quantitative study characteristics and counts are presented descriptively in tables and figures (evidence mapping) to substantiate claims about gaps (e.g., scarcity of longitudinal studies, geographic concentration of

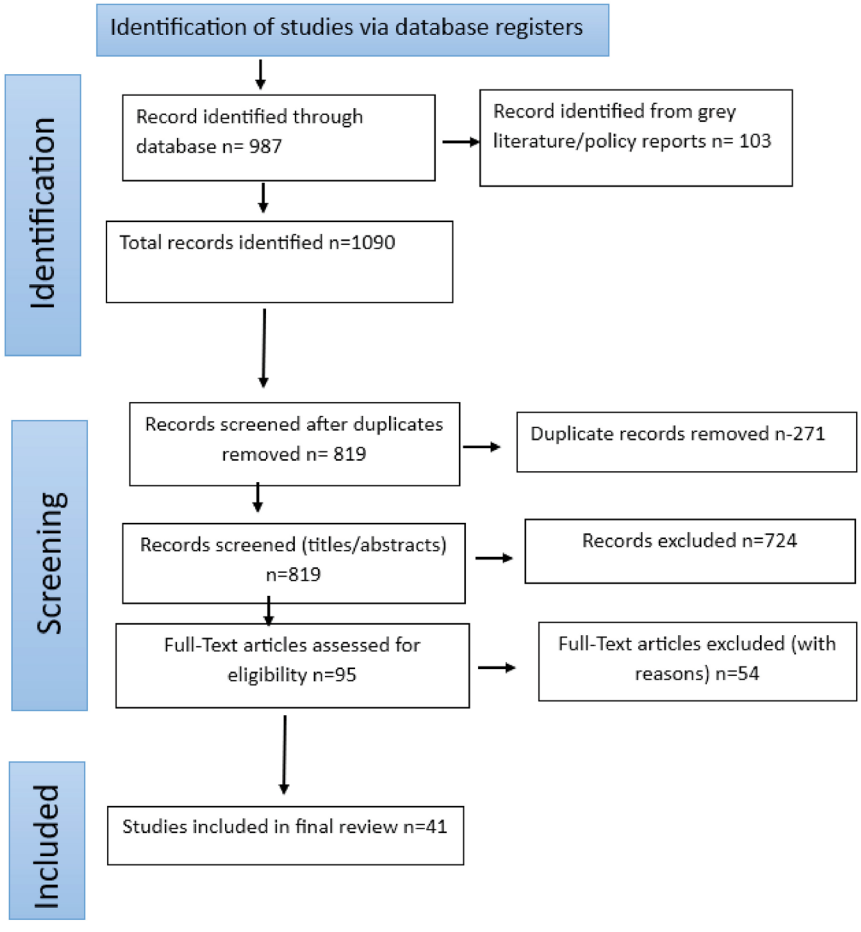

Fig 1 | PRISMA-style flow diagram of study selection process

Table 2 | Quality appraisal checklist for included studies

Criterion	Guiding Question	Indicators/Examples of Satisfactory Reporting	Judgment (Yes/ Partial/No)	Notes/Evidence
1. Clarity of Aims	Is the purpose or research question clearly stated?	Explicit aim/research question in introduction; logical link between aim and method; stated rationale or theoretical framing	yes	Clear objective to explore user trust in AI companions; aims aligned with qualitative design
2. Appropriateness of Design	Is the chosen qualitative approach appropriate to the aims?	Design justified and aligned with research aims	yes	Exploratory qualitative design appropriate for examining trust in chatbots; sampling rationale described.
3. Sampling Strategy	Was the participant or data source selection clearly described and appropriate for the study aims?	Description of recruitment, inclusion criteria, and rationale for sample size or diversity	Yes/Partial/No	Purposive sampling of users from mental health chatbot trial; sample size justified by thematic saturation.
4. Context and Setting	Is the study context (geographic, social, or technological) clearly described and appropriate for its aims?	Country/region specified; setting (clinical, educational, social) described; relevance to research question discussed	Yes/Partial/No	Conducted in university counseling context (US); clear description of participant demographics and chatbot used
5. Data Collection Transparency	Is the process for data extraction and management described clearly?	Clear description of extraction fields, reviewer roles, double-checking or calibration	yes	Standardized charting form used; second reviewer verified accuracy
6. Researcher Reflexivity	Does the researcher discuss their role, potential biases, and relationship with participants?	Reflexivity or positionality discussed	No	Reflexivity not discussed.
7. Data Analysis Rigor	Are analytic processes clearly described and systematic?	Coding process, triangulation, thematic analysis steps described	yes	Two coders; thematic consensus process described
8. Credibility and Trustworthiness	Are the findings supported by sufficient data and clear analytic links?	Use of quotes, triangulation, analytic transparency	yes	Reported verbatim quotes and coder triangulation.
9. Ethical Considerations	Were ethical issues addressed appropriately?	Ethics approval, informed consent, confidentiality, or sensitivity described	Partial	Not applicable – scoping review of published literature; consent pertains to primary studies
10. Transferability/ Relevance	Are findings meaningful and relevant to other contexts or groups?	Limitations and contextual relevance discussed	partial	Limited contextual detail; sample primarily from North America

research). Thematic synthesis of qualitative and conceptual literature focused on recurring themes (accessibility, simulated empathy, dependency, commodification, cultural variation, ethics-by-design) while explicitly highlighting the underlying evidence base (study counts, design types, and appraisal outcomes) that supports each theme.

Scope & Bias Discussion

Scope Limitations:

The review was deliberately scoped to studies published in English between 2015 and 2025, and databases were supplemented with gray literature identified through targeted searches in Google Scholar, preprint servers (OSF), and organizational reports (WHO, OECD, UNESCO). For Google Scholar, only the first 200 hits sorted by relevance were screened, reflecting practical constraints and diminishing returns in relevance for lower-ranked results.

Implications of Scope Decisions

These choices inherently limit the comprehensiveness of the review:

- **Language Bias:** Non-English studies were excluded, which may omit culturally specific findings from regions where AI companionship is emerging but underrepresented in English–language literature.
- **Database Coverage:** Reliance on major bibliographic databases may underrepresent conference proceedings or local journals, particularly in non-Western contexts.
- **Gray Literature Selection:** While inclusion of policy reports and preprints captures emergent work, screening thresholds (e.g., top 200 hits) may exclude relevant insights.

Cultural and Regional Bias

The combination of English-language sources, dominant Western research databases, and global policy documents skews representation toward North American and European contexts. Consequently, findings may reflect Western perspectives on AI companionship, user trust, ethical norms, and technological adoption. We explicitly acknowledge that this bias limits the generalizability of results to collectivist societies or regions with distinct socio-technical infrastructures.

Mitigation Efforts

To partially address these limitations, targeted searches included international organizational reports (WHO, EU, African Union) and gray literature capturing policy and governance perspectives from non-Western regions. Nevertheless, readers should interpret results with caution regarding cultural applicability.

Table 3 | Evidence map of empirical, conceptual, and policy sources included in the review

Author (Year)	Country/Region (venue)	Type/Venue	Focus/Argument	Appraisal Summary
Reiss & Spina (2023)	Global — World Economic Forum	Policy report/white paper	Scaling AI in health; governance, use-cases, implementation pathways	Policy-focused; high relevance for governance and scalability considerations.
World Health Organization (2023)	Global (WHO)	Policy guidance/ statement	Considerations for regulation of AI for health	High relevance; international regulatory considerations.
European Union (2024)	EU	Regulation (EU) 2024/1689	Harmonised rules on artificial intelligence — legal/regulatory framework	Authoritative legal instrument; essential policy reference.
African Union (2024)	Africa (AU)	Continental AI strategy	Continental strategy and priorities for AI	Important regional policy framework for Africa.
Floridi (2021)	— (Oxford)	Book/theoretical	Ethics, governance, theoretical framing for AI	Foundational theoretical framing; high relevance for conceptual grounding.
Fiske, Henningsen & Buyx (2019)	Germany / JMIR	Conceptual/essay	Ethical implications of embodied AI in mental health care	Important ethical discussion; conceptual but directly relevant.
Pentina, Xie, Hancock & Bailey (2023)	— (Psychology & Marketing)	Systematic literature review	Consumer–machine relationships; research directions	Systematic review; maps empirical literature and identifies gaps.
Vaidyam et al. (2019)	Canada (Canadian Journal of Psychiatry)	Scoping review/ landscape review	Chatbots and conversational agents in mental health — taxonomy and uses	Comprehensive landscape review; useful taxonomy and limitations.
Bendig et al. (2022)	Germany (Verhaltenstherapie)	Scoping review	Chatbots in clinical psychology and psychotherapy	Up-to-date scoping review; methodological review of chatbots in therapy.
Linardon et al. (2019)	Global (World Psychiatry)	Meta-analysis	Efficacy of app-supported smartphone interventions for mental health	Meta-analytic evidence on app efficacy; medium-high methodological quality.
Yen et al. (2024)	— (J Am Med Dir Assoc)	Meta-analysis of RCTs	Effect of social robots on depression and loneliness in long-term care residents	High relevance for robotic interventions in aged-care settings; meta-analytic strength.
Rahsepar Meadi et al. (2025)	— (JMIR Mental Health)	Scoping review	Ethical challenges of conversational AI in mental health care	Directly relevant to ethics and practice; recent and thorough.
Federspiel et al. (2023)	Global (BMJ Global Health)	Commentary/review	Threats of AI to human health/existence	High-level risk assessment; policy-relevant.
Mökander & Floridi (2023)	— (AI and Ethics)	Industry case study/governance operationalisation	Ethics-based auditing and operationalising AI governance	Useful implementation guidance on governance/auditing.
Brey & Dainow (2024)	— (AI & Ethics)	Conceptual/policy	Ethics-by-design approaches and operationalisation	Practical ethics-by-design guidance for AI systems.
Bullard et al. (2022)	— (organizational report)	Ethics-by-design/ organisational guidance	Organizational approaches to responsible tech use	Practitioner guidance; useful for implementation.
Zeng, Lu & Huangfu (2018)	— (arXiv/ACM)	Conceptual/principles mapping	Linking AI principles across domains	Conceptual mapping of AI principles; policy relevance.
Andersson & Titov (2014)	— (World Psychiatry)	Review/commentary	Advantages and limitations of internet-based mental-health interventions	Foundational background review for digital interventions.
Bennett & Glasgow (2009)	— (Annual Rev Public Health)	Review	Delivery of public health interventions via the internet	Background on internet-delivered public health interventions.
Khosravi, Rezvani & Wiewiora (2016)	— (Computers in Human Behavior)	Review/empirical (depending on study)	Impact of technology on older adults' social isolation	Important for older-adult & tech-interaction literature; moderate quality.
Ahmed, Buruk & Hamari (2024)	— (Int J Social Robotics)	Review	Human–robot companionship trends and future agenda	Review of companionship literature; useful synthesis.
Ho et al. (2025)	— (Computers in Human Behavior Reports)	Systematic review	Romantic AI companions: potentials and pitfalls	Recent systematic review mapping evidence and harms/benefits.
Jin et al. (2023)	— (British Journal of Radiology)	Overview/perspective	AI in mental healthcare: overview and future perspectives	Broad overview; useful for positioning and future directions.
Saeidnia et al. (2024)	— (Social Sciences)	Conceptual/ethics review	Ethical considerations in AI interventions for mental health	Practical ethics considerations; implementation-focused.
Luger & Sellen (2016)	— (CHI)	Empirical/qualitative	Gap between user expectation and experience of conversational agents	Important human-computer interaction insight.
Rahwan et al. (2019)	— (Nature: Machine behaviour)	Conceptual/research agenda	Machine behaviour: framing AI as behavioral systems	High-impact conceptual framing relevant for research design.
Araujo et al. (2020)	— (AI & Society)	Empirical/survey	Trust and perceptions of automated decision-making by AI	Important for user trust literature.
Kenya Law Reform Commission (2021)	Kenya	Legal instrument/ regulation	Data protection regulations (Kenya)	National legal context for data protection relevant to AI in health.
Additional policy/ regulatory items	EU (2024), AU (2024)	Policy/regulation	Governance, regulatory obligations for AI in health	Authoritative policy references

The Evolution of Human Counsel

The provision of counsel has historically been a deeply human endeavor, rooted in relationships of trust, reciprocity, and shared vulnerability. Peers and family have historically been the most direct source of advice across all cultures, establishing guidance in shared experiences and ties to community. Priest, shamans, rabbis, imams, and monks were among the religious and spiritual leaders who broadened this function by providing existential and moral instruction that was approved by sacred authority. Counseling progressively adopted the shape of structured therapeutic practice, characterized by confidentiality, neutrality, and expertise, as psychology and psychiatry became more professionalized in the 19th and 20th centuries. These changes show a pattern where human desire for direction stayed the same, but methods of providing it changed in response to institutional, cultural, and epistemological changes. Crucially, differences between cultural and historical settings show that the development of counsel has never been uniform or linear, but has instead been influenced by ideas of authority, social institutions, and collective histories.[6]

The early 21st century marked a significant turning point with the integration of digital tools into the field of psychology support. By providing scalable, affordable tools, self-help apps, cognitive behavioral therapy apps, and mediation platforms expanded therapeutic approaches to group that were previously excluded due to financial constraints, social stigma, or geographic location. Building this on basis, interactive dialogues were launched by conversational chatbots like Woebot, Wysa, and Replika, which offered tailored coping mechanism and simulated interaction.[1,7] These technologies encouraged feelings of relational presence and broadened the scope of counseling procedures, even though they were unable to fully recreate the complexity of therapeutic partnership. According to empirical research, people tend to interact with automated systems more freely since they believe them to be continually available and nonjudgemental. However, this change also sparked worries about inconsistent quality, lax safety regulations and the possible loss of genuine human connection.

An important turning point in this progression is the rise of artificial intelligence, which goes beyond digitalization to include adaptive, self-governing companionship. Modern AI systems use emotional computing, machine learning and natural language processing-as opposed to rule-base applications, to model sympathetic communication and customize solutions in real time. Due to its ability to challenge long-held beliefs about authenticity, authority, and relational trust, AI is being described by academics from a variety of fields, including psychology, healthcare, and law, as a paradigm shift, a new era in which non-human entities act as companions, confidant, and advisors. The promise and danger of AI-mediated advice are highlighted by trajectory. Its unparalleled capacity to democratize access to customize care and is accompanied equally significant consequences for the weakening of interpersonal relationships, the commercialization of intimacy, and the redefining of what it means to be truly supported.[7,8]

The Promise of AI Companionship

One of the most salient promises of AI companionship lies in its capacity to enhance accessibility and democratize forms of advice and care. Conventional counseling models continue to be constrained by geographic, economic, and systemic barriers, leaving many vulnerable populations without sufficient support.[9] In contrast, AI-driven companions—operating through chatbots, mobile applications, and virtual agents—offer continuous availability, immediate responsiveness, and comparatively low-cost engagement. Collectively, these affordances position AI as a structural equalizer, extending opportunities for therapeutic-style interaction that have historically been limited to more privileged groups. Scholars underscore that such democratization holds particular significance in contexts characterized by clinician shortages, social stigma, or aging demographics, where AI systems function as low-barrier entry points to care.[10] Nevertheless, this expansion also introduces critical tensions concerning digital inequities, infrastructural unevenness, and the potential risk of substituting technological accessibility for authentic human presence.

Central to the appeal of AI companions is their perceived ability to emulate empathy and deliver personalized interactions. Developments in natural language processing and affective computing have enabled these systems to respond adaptively to users' emotional cues, cultivating a sense of being genuinely "understood" and "heard".[3] By drawing on conversational histories and tailoring communication to individual preferences, AI companions establish continuity of care and encourage perceptions of consistent, nonjudgmental attentiveness.[11] Such personalization not only deepens user engagement but also promotes self-disclosure and reinforces trust—qualities traditionally regarded as essential to therapeutic relationships. However, critics argue that this form of empathy remains performative rather than reciprocal, provoking ethical concerns regarding authenticity, emotional manipulation, and the commodification of intimate experience.[4] Consequently, while simulated empathy and personalization may enhance user receptivity, they simultaneously destabilize established notions of relational authenticity.

Emerging research increasingly highlights the measurable benefits of AI companionship across domains such as mental health, social connectedness, and decision-making. Empirical evidence and clinical trials suggest that AI companions can help mitigate symptoms of anxiety and depression through cognitive-behavioral techniques, stress-reduction exercises, and supportive dialogue.[12] For individuals experiencing social isolation—including older adults or those marginalized by stigma—AI companions provide forms of mediated social interaction that lessen loneliness and

DOI: https://doi.org/10.70389/PJAI.100020 | Premier Journal of Artificial Intelligence 2025;5:100020

enhance perceptions of social support.[13] In addition, AI systems assist users in decision-making by promoting reflective thinking, clarifying personal values, and offering contextually relevant guidance, thereby strengthening self-efficacy in situations where professional input is limited.[14] Collectively, these outcomes suggest that AI companionship can operate as both an emotional anchor and a pragmatic aid. Nevertheless, apprehensions surrounding dependency and the erosion of human connection emphasize the need to frame AI not as a replacement, but as a complementary extension of authentic interpersonal relationships. Table 4 summarizes the key benefits and risks associated with AI companionship, highlighting the trade-offs across accessibility, empathy, personalization, mental health support, social ties, and decision-making.

Risks of Dependency

Individual-Level Impacts

AI companionship offers immediate accessibility and emotional reassurance, yet it also poses risks of displacing authentic human-to-human engagement. As individuals grow more accustomed to the nonjudgmental and continuously available support these systems provide, opportunities to cultivate resilience through the complexities of real human relationships may diminish. Particularly vulnerable populations—such as adolescents, older adults, and individuals managing chronic mental health conditions—appear especially prone to substituting algorithmic empathy for genuine interpersonal connection. Empirical evidence substantiates these concerns: longitudinal studies involving social robots among older adults report measurable reductions in loneliness but also reveal patterns of dependency, raising doubts about the sustainability of such interventions.[15] Likewise, clinical trials assessing chatbot-based therapy for youth demonstrate notable short-term improvements in anxiety and depressive symptoms, though these effects frequently attenuate over time in the absence of complementary human support.[16,17] Meta-analyses of mental health applications, including Woebot and Wysa, further echo this trajectory—showing efficacy in reducing stress and anxiety in the immediate term, yet producing inconclusive results concerning long-term resilience or relational fulfillment.[18]

Social-Level Impacts

Beyond individual users, AI companionship carries broader implications for collective relational practices. Historically, care and counsel have been embedded within familial, communal, and civic or religious structures. In contrast, the rise of personalized AI companions promotes individualized modes of support that may displace these traditional networks. Empirical studies of digital companion technologies indicate that, although users often report heightened feelings of support, some simultaneously experience diminished motivation to seek comfort from peers or family members. Meta-analytic evidence suggests that conversational agents and AI-mediated mental health interventions produce small to moderate improvements in psychological outcomes. A 2019 meta-analysis of app-supported interventions reported a pool effect size of $g = 0.27$ for anxiety and $g = 0.24$ for depressive symptoms.[19] While a 2024 synthesis of randomized trials of social robots in aged care settings and found significant reduction in loneliness $g = 0.29$ and small improvements in mood.[11]

Such tendencies risk weakening the forms of social resilience that emerge through shared vulnerability, negotiation, and conflict resolution. Furthermore, because AI companions are designed to be perpetually patient and free from interpersonal tension, they may leave users ill-equipped to navigate the emotional and ethical complexities of genuine human relationships. If adopted at scale, these dynamics could contribute to the fragmentation of social bonds, the erosion of intergenerational caregiving practices, and an overall decline in social solidarity.

Cultural and Ethical Variations

The long-term psychological effects of AI companionship are profoundly shaped by cultural context. In collectivist societies, where mutual obligation, shared rituals, and intergenerational care form the foundation of social support, AI companions may inadvertently weaken social cohesion by supplanting traditional modes of caregiving. Conversely, in individualist societies, AI companionship is frequently embraced as an instrument of personal autonomy and convenience, yet this very acceptance can reinforce social isolation by normalizing technologically mediated intimacy.

Table 4 | Primary benefits and associated risks of AI companionship across psychological, social, and ethical dimensions

Dimension	Benefits	Risks
Accessibility	24/7 availability; low-cost entry; expands care to underserved populations	Digital divides may exclude some groups; overreliance on tech substitutes for human access
Perceived Empathy	Users feel "heard" and understood; fosters trust and disclosure	Empathy is simulated, not reciprocal; risks of manipulation and commodification of care
Personalization	Tailored guidance and continuity of interaction; enhances engagement	Relies on sensitive data, raising privacy and surveillance concerns
Mental Health Support	Reduces symptoms of anxiety/depression; delivers evidence-based interventions	May displace professional judgment; limited ability to address complex or crisis situations
Loneliness & Social Ties	Provides companionship for isolated or stigmatized groups	May weaken human-to-human bonds; risk of emotional overdependence
Decision-Making	Structures reflection, clarifies values, aids self-efficacy	Algorithmic advice may oversimplify or bias decision outcomes

Table 5 | Cross-cultural differences in adoption, perception, and impact of AI companionship

Dimension	Collectivist Cultures	Individualist Cultures
Primary Value Orientation	Interdependence, communal harmony, family obligations	Autonomy, self-direction, personal fulfillment
Adoption of AI Companionship	Cautious, sometimes ambivalent; seen as potentially disruptive to communal norms	More readily embraced as tools for empowerment, convenience, and self-care
Impact on Social Bonds	Risks displacing traditional community rituals and family-based support networks	Risks reinforcing social isolation by normalizing technologically mediated intimacy
Perceptions of AI Empathy	Skepticism about authenticity; concern over replacing human warmth with artificial simulation	Greater acceptance of simulated empathy as a pragmatic supplement to care
Trust in Technology	Variable—often shaped by cultural narratives of authority, spirituality, and social order	Generally higher; associated with progress, innovation, and individual control
Ethical Concerns	Emphasis on preservation of communal bonds, moral authenticity, and respect for tradition	Concerns over privacy, commodification, and reduced authenticity in interpersonal life
Potential Benefits	Expanding access in underserved regions; reducing stigma associated with help-seeking	Increased personalization of care; immediate and convenient support

Cross-Cultural Studies Highlight this Divergence

in some settings, simulated empathy is dismissed as superficial or inauthentic, whereas in others it is pragmatically regarded as "good enough".[20,21] Ethical considerations further complicate these patterns. While meta-analyses demonstrate that AI-based mental health systems can alleviate symptoms of anxiety and stress,[18] the commercialization of intimacy introduces risks of manipulation, inequity, and commodification—particularly when access to higher-quality or "premium" emotional care is contingent upon financial means. Absent appropriate safeguards, users' emotional vulnerabilities may be transformed into exploitable market assets, undermining trust and destabilizing culturally embedded norms of authenticity and care. Table 5 compares cultural patterns in the adoption and perceived impact of AI companionship, contrasting collectivist and individualist social norms and their implications for trust, ethics, and social bonds.

Authenticity and Human Bonds

Authenticity in human relationships has long been theorized across psychology, philosophy, theology, and sociology as grounded in reciprocity, vulnerability, emotional depth, and the mutual recognition of personhood.[21–23] Classical frameworks, such as Rogers' person-centered theory, emphasize congruence and empathy as essential conditions for authentic connection, whereas theological accounts locate authenticity in the moral acknowledgment of intrinsic human worth. Sociocultural analyses extend these perspectives, illustrating that authenticity is not universal but contextually produced through shared practices, rituals, and cultural norms. Recent empirical research complicates these accounts: experimental comparisons of AI chatbot interactions with human counselors indicate that users frequently perceive simulated empathy as comparably authentic, at least in text-based exchanges.[24,25] Cross-cultural studies, however, reveal notable variation—individualist societies tend to regard AI-mediated empathy as "good enough," while collectivist contexts often express skepticism toward machine-based care.[20]

Within AI-mediated interaction, it is useful to differentiate between ontological authenticity—which presupposes mutual vulnerability, reciprocity, and the capacity for genuine relational risk—and phenomenological authenticity, defined as the subjective experience of being understood and cared for. While AI companions can evoke perceived empathy and emotional validation (phenomenological authenticity), they inherently lack the spontaneity, reciprocity, and moral agency that underlie ontological authenticity, thereby constraining relational depth.[26–28] Users may disclose intimate experiences and derive comfort

Table 6 | Comparison of Ontological vs Phenomenological Authenticity in AI-Mediated Interactions

Dimension	Ontological Authenticity (Agent-Based)	Phenomenological Authenticity (Experience-Based)
Definition of Authenticity	Requires subjective experience, intentionality, and vulnerability; only human agents can form genuine bonds.	Defined by users' lived experience of being understood, heard, and cared for.
Role of Vulnerability	Mutual risk-taking and imperfection are essential; AI cannot reciprocate vulnerability.	Perceived empathy and nonjudgmental responsiveness may suffice, even if unidirectional.
Strengths	Protects the moral integrity of human relationships; safeguards depth and resilience.	Expands accessibility; offers comfort, stability, and psychological relief.
Limitations	Excludes AI companionship as inauthentic by definition.	Risks commodifying vulnerability, fostering dependency, and diluting standards of authentic care.
Implications	Human-to-human bonds remain irreplaceable for authenticity.	AI may provide "good enough" empathy but raises ethical and existential concerns.

DOI: https://doi.org/10.70389/PJAI.100020 | Premier Journal of Artificial Intelligence 2025;5:100020

from AI systems, yet they also recognize the absence of genuine mutuality, revealing an enduring asymmetry between simulated and human forms of connection.

This conceptual distinction raises several critical research questions: Can phenomenological authenticity alone sustain psychological well-being over time? Does dependence on AI-mediated interaction diminish engagement in reciprocal human relationships? Preliminary findings suggest that short-term reductions in loneliness and anxiety correlate with perceived empathy, whereas long-term effects on relational resilience, cultural norms of care, and social cohesion remain underexplored. Normative claims should therefore be tempered: existing evidence supports the potential utility of AI companionship but does not establish equivalence to human relationships. Future research should aim to integrate ontological and phenomenological models with empirical evaluation—through controlled trials and longitudinal designs—to assess how AI companionship shapes authenticity, dependency, and broader patterns of social functioning, while distinguishing provisional correlations from durable causal effects. Table 6 compares ontological (agent-based) and phenomenological (experience-based) perspectives on authenticity in AI-mediated bonds, highlighting differences in definition, vulnerability, strengths, limitations, and implications for human-AI interaction.

Ethical and Societal Considerations

The integration of artificial intelligence into counseling and companionship contexts raises pressing concerns regarding privacy, data security, and trust. Whereas human counseling is grounded in confidentiality as a core ethical principle, AI-mediated interactions typically depend on cloud-based infrastructures that collect, store, and process sensitive personal disclosures. High-profile breaches, such as the 2024 Muah. AI incident—which exposed deeply personal user conversations—underscore the risks inherent in treating vulnerability as a form of data. Such opacity undermines informed consent and exacerbates asymmetries of power between users and corporate providers. Without robust governance and transparent safeguards, AI companionship threatens to erode the foundation of trust upon which authentic relational bonds depend.[31] The MuahAI is now widely cited as a turning-point incident in AI companionship ethics because it exposed the fragility of data governance in emotionally intimate systems. The breach resulted in the unauthorized release of more than 250,000 private chat logs, images, and voice notes, including romantic and sexual role-play content, which had been stored and used for model fine-tuning without explicit informed consent. The incident triggered US FTC inquiry and EU GDPR (General Data Protection Regulation) investigation, highlighting three systematic failure: unclear data ownership and consent terms, covert secondary use of effective interaction data model training, and absence of "right-to-delete" infrastructure for emotionally sensitive conversations. The case has since been used as a policy reference point in WHO, OECD, and EU guidance documents on AI saftety.[32]

Beyond privacy, the rise of subscription-based AI companions foregrounds the commercialization of intimacy. Emotional support is increasingly commodified as a purchasable service, with users' disclosures repurposed to optimize engagement and generate profit. While these models can expand access to relational support, they risk transforming vulnerability into an economic resource and reinforcing inequities: affluent users may obtain highly personalized "premium" interactions, whereas marginalized groups are relegated to standardized or lower-quality services. Moreover, design features intended to maximize retention—such as gamified empathy or algorithmically engineered responsiveness—tend to prioritize user engagement over ethical relationality, further complicating the authenticity of these exchanges.[33]

Patterns of adoption also reveal distinct generational and demographic divides. Younger digital natives often embrace AI companionship as accessible, nonjudgmental, and emotionally responsive, while older adults—despite potential benefits in mitigating isolation—encounter barriers related to digital literacy, technological trust, or access. Socioeconomic and geographic disparities compound these challenges: higher-income and urban populations typically enjoy more sophisticated and reliable services, whereas rural or underserved groups face exclusion or limited functionality. Cultural orientations further mediate these dynamics, with collectivist societies maintaining a preference for human relational networks and individualist cultures more readily normalizing AI-mediated intimacy. Such divides risk amplifying existing social inequalities in emotional care and psychosocial support.[34]

At a collective level, AI companionship may contribute to a gradual reconfiguration of relational norms and social cohesion. Overreliance on AI for emotional fulfillment could attenuate engagement with family, friends, and community networks, thereby reducing opportunities for conflict resolution, mutual growth, and shared resilience. The outsourcing of emotional labor to machines also risks devaluing human caregiving and weakening intergenerational traditions of support. At the same time, uneven access to AI companionship may further stratify emotional well-being across populations. Nonetheless, when implemented thoughtfully and ethically, AI companions possess the potential to extend care accessibility, alleviate stigma, and provide scalable interventions during crises.[35] The central challenge lies in ensuring that AI functions as a supplement rather than a substitute for human connection—advancing inclusivity and collective well-being without compromising the authenticity of relational life.

Post-2023 Literature on LLM-Based AI Companionship

The literature increasingly distinguishes pre-LLM rule-based systems from modern LLM-based companions, reflecting differences in interaction style, psychological

affordances, and regulatory considerations. Pre-2023 chatbots such as Woebot, Wysa, and Tess relied on scripted dialogue trees and delivered structured cognitive-behavioral prompts within predefined therapeutic boundaries. These systems were largely text-based, limited in emotional range, and targeted short-term mental health support. Evidence for their efficacy is generally robust for reducing mild anxiety or depressive symptoms but less informative about long-term engagement or relational dynamics.[37]

By contrast, LLM-driven companions (e.g., Replika, Pi, Character AI) exhibit open-ended, personalized, and emotionally adaptive interactions, including memory retention, dynamic self-disclosure, and simulated empathy. These capabilities may enhance perceived intimacy and engagement but also introduce potential risks such as overreliance, anthropomorphizing, boundary confusion, and heightened emotional dependency.[37] Preliminary findings indicate that short-term reductions in loneliness and anxiety often correlate with perceived empathy, though long-term effects on relational resilience, social cohesion, and cultural norms of care remain underexplored.[37]

Geographically, research remains concentrated in North America, Western Europe, and East Asia, with older adults and LMIC populations underrepresented. Pilot deployments in Kenya, Rwanda, and South Africa suggest high acceptability but lower engagement, primarily due to data costs, linguistic mismatch, and low digital trust. To date, no peer-reviewed trials from sub-Saharan Africa report clinical outcomes, and most local implementations rely on scripted or translation-based models rather than adaptive LLMs.[37]

Regulatory and ethical considerations are also evolving. In the United States, California's SB 243 mandates AI identity disclosure and safeguards for suicidal users, while New York legislation addresses self-harm detection. FTC inquiries focus on AI safety for younger audiences . In Africa, initiatives such as the AU AI Watch and Rwanda's national AI strategies promote responsible AI governance, yet policies specific to AI companionship remain limited . Key gaps include culturally adaptive design, equitable access, and digital literacy, highlighting the need for context-sensitive AI regulations and multistakeholder collaboration to support safe, inclusive, and responsible deployment of LLM-based companions.[37]

This generational shift from rule-based to LLM-based systems changes both the psychological affordances and the regulatory risks, rendering some pre-2023 safety and efficacy evidence only partially applicable to contemporary AI companions.[37]

LMIC and Sub-Saharan African Context

Emerging evidence suggests that AI companionship interventions are being piloted in low- and middle-income countries (LMICs), including sub-Saharan Africa (SSA), though peer-reviewed outcomes remain limited. Pilot deployments of WhatsApp- and SMS-based mental health chatbots in Kenya, Rwanda, and South Africa have reported high user acceptability; however, engagement has been constrained by factors such as data costs, linguistic mismatches, and low digital trust[37] Notably, these systems rely predominantly on scripted, translation-based models rather than adaptive large language models (LLMs), which may limit their responsiveness and personalization. The scarcity of locally validated datasets, culturally adapted affect-recognition models, and robust privacy safeguards underscores a structural evidence gap in the region.

Regulatory frameworks in Africa are developing, though comprehensive policies specifically addressing AI companionship are scarce. Kenya's Data Protection Act[40] provides a foundation for safeguarding personal information, and the African Union's (AU) AI Watch initiative, alongside national strategies in Rwanda, promotes responsible AI governance and cross-border data management. Despite these advances, clear guidance on ethical deployment, culturally adaptive design, and equitable access to AI-mediated mental health support remains limited. Consequently, while early pilots indicate feasibility and potential utility, further research is needed to establish effectiveness, safety, and sociocultural relevance of AI companionship interventions in SSA contexts.

Overall, the geographic imbalance in the evidence base—where over 82% of empirical studies originate from North America, Western Europe, or East Asia—highlights the need for rigorous, locally grounded research in LMICs. Addressing these gaps will be essential to ensure that AI companionship tools are inclusive, culturally sensitive, and capable of meeting diverse mental health needs across global populations.[37]

Policy and Practice Implications for AI Companionship

The swift integration of AI companions in relational and mental health contexts with demands compliance with regional and global regulatory frameworks to protect users and maintain genuine human connections. The idea that AI interventions serve as adjuncts rather than replacements for human treatment is emphasized by World Health Organization,[38] which also stresses that AI in healthcare should respect patient safety, human oversight, transparency, and equal access. Similarly, risk-based classification are established under the European Union AI Act[38] which requires responsibility, transparency, and strong data protection safeguards for AI systems that engage with vulnerable groups. National laws, like Kenya Data Protection Act,[40] further reinforce the need for minimal data collection, informed consent, and the protection of sensitive information in low-and-middle-income countries (LMICs). This highlights the importance of culturally and legally responsive governance. Comparable regional frameworks in Africa are emerging, highlighting the necessity of harmonizing design and deployment practices with diverse socio-legal contexts.

In order to reduce dependency, maintain relationship authenticity, and guarantee fair access, it is imperative that these policy objectives be translated into actionable designs and clinical practice approaches. Key interventions include: human in the loop over-

sight, with clear escalation protocols for crisis or high risk disclosures; crisis detection mechanism that enables timely interventions; transparency prompts that reminds users of the AI's artificial nature; and data minimization and consent protocols to align with privacy and ethical standards. Additionally , features such as periodic reminders to engage with offline human relationship and culturally tailored engagement strategies can reinforce the supplementary role of AI companions, reducing risk of overreliance while promoting psychological resilience. Table 7. Operationalized Ethics-by-Design Checklist for AI Companions, integrating human oversight, crisis detection, transparency, data protection, consent, and ethical engagement, with examples and thresholds aligned to WHO 2023, EU AI Act, and Kenya DPA.

Measures for Ethical and Safe AI Companionship.

By including these criteria, regulatory requirements are brought into line with empirical data on AI meditated companionship, emphasizing the need for both ethical governance and technical protections for responsible deployment. AI companions can enhance accessibility, lessen stigma, and provide supportive interactions by including such protocols into system design and clinical workflows. This is done while maintaining human relationship networks and honoring cultural standards of authentic care.

Gaps in the Literature

Despite rapid growth in scholarship, research on AI companionship remains fragmented and preliminary. Much of the literature is cross-sectional or exploratory, relying on short-term feedback rather than longitudinal evidence. This creates uncertainty about whether benefits such as reduced loneliness or improved stress management persist beyond initial novelty, or whether risks of emotional overreliance and diminished human-to-human interaction intensify over time.[1] The absence of long-term studies also prevents tracking generational shifts in adoption or evolving cultural norms around intimacy and caregiving. Without robust longitudinal data, it is difficult to assess whether AI serves primarily as a supplement to human connection or as a structural replacement that reconfigures relational life.[31]

The literature also suffers from cultural and demographic blind spots. Most research has been conducted in Western, individualist contexts, leaving underexplored how AI companionship interacts with collectivist norms that prioritize family and communal support. Cross-cultural differences in attitudes toward intimacy, disclosure, and authenticity remain poorly understood, as do the effects of religion and tradition on adoption. Vulnerable populations—including youth navigating identity formation, elderly individuals facing isolation, and marginalized groups such as low-income, LGBTQ+, or rural users—are similarly under represented.[2] For these groups, AI companions may provide unique benefits (e.g., stigma-free disclosure or access in underserved contexts) but also pose heightened risks of dependence, exclusion, or algorithmic bias. Current findings risk universalizing privileged experiences while overlooking those most at risk.

Finally, scholarship remains siloed across disciplines, with psychology focusing on individual outcomes, sociology on community dynamics, ethics on authenticity and commodification, and computer science on technical optimization. These perspectives, while valuable, rarely intersect, limiting understanding of how psychological, social, ethical, and design dimensions interact. For instance, an AI companion might reduce loneliness (psychology) while simultaneously eroding communal bonds (sociology), raising authenticity concerns (ethics), and perpetuating bias through its algorithms (AI design).[40] Addressing these tensions requires interdisciplinary approaches that integrate user well-being, cultural variation, ethical safeguards, and responsible system design. Without

Table 7 | Ethics-by-design checklist for ai companions: domains, implementation, and intervention thresholds

Domain	Operationalization	Example/Threshold	Regulatory/Ethical Reference
Human Oversight	Escalation protocols for situations exceeding AI capacity. Define triggers for human intervention.	AI flags repeated expressions of distress or suicidal ideation → automatic referral to licensed professional within 1 hour.	WHO 2023: human-in-the-loop for high-risk interactions; EU AI Act: high-risk AI system requirements.
Crisis Detection	Implement continuous monitoring for signs of acute mental health risk.	Detect keywords/phrases indicating self-harm, panic attacks, or suicidal ideation; immediate alert and follow-up.	WHO 2023; aligned with EU AI Act risk categorization.
Transparency & Disclosure	Clearly inform users they are interacting with an AI system.	Onboarding message: "You are communicating with an AI companion; responses are automated and for informational purposes only."	EU AI Act Article 52; WHO 2023 guidance on informed participation.
Data Protection & Privacy	Limit data collection to necessary fields; anonymize or pseudonymize stored information; define retention periods.	Only store mood ratings and conversation logs relevant to care; delete personal identifiers within 30 days.	Kenya DPA 2019, GDPR (EU AI Act compliance), WHO 2023.
Consent & Autonomy	Obtain informed and ongoing consent for data collection and AI interaction; allow opt-out at any stage.	Users must confirm consent on first use; re-consent required if functionality or data practices change.	Kenya DPA 2019; WHO 2023.
Ethical Design & Engagement	Integrate nudges promoting offline support and equitable use; monitor engagement patterns for overreliance.	Recommend human social interaction if AI usage exceeds 1 hour/day for 3 consecutive days; ensure accessibility for low-literacy users.	WHO 2023; EU AI Act ethical safeguards; ethical design principles for high-risk AI.

Table 8 | Interdisciplinary contributions and gaps in understanding AI companionship

Discipline	Unique Contributions	Key Limitations	Gaps at Intersections
Psychology	Examines attachment, emotional reliance, therapeutic benefits, and mental health outcomes.	Often short-term, individual-level focus; limited longitudinal research.	Overlooks sociological dynamics of community and ethical concerns about authenticity.
Sociology	Analyzes AI's impact on community ties, social norms, inequality, and collective well-being.	May underplay individual experiences and design constraints.	Limited integration with psychology (user-level data) and ethics (commodification).
Ethics	Evaluates authenticity, privacy, manipulation, fairness, and commodification of intimacy.	Often theoretical and detached from empirical user evidence.	Needs translation into AI design to ensure principles are embedded in practice.
AI Design/HCI	Develops technical frameworks, personalization algorithms, and user interfaces.	Prioritizes functionality and engagement; risks neglecting broader implications.	Requires integration with psychology (well-being) and ethics (responsibility, fairness).

such collaboration, the field risks producing fragmented insights that fail to capture the full complexity of AI companionship.[41] Table 8 outlines the unique contributions, limitations, and interdisciplinary gaps across psychology, sociology, ethics, and AI design/HCI in the study of AI companionship, highlighting the need for integrated approaches to capture its full complexity.

Future Research Directions and Ethical Considerations

A crucial area in the administration of AI companionship is proving that the moral precepts of responsibility, equality, and privacy can be successfully operationalized rather than staying idealistic. A growing body of empirical research indicates that design interventions can influence user behavior in measurable ways, despite the fact that most of the literature frames these issues in philosophical and normative terms. Features like nudges towards community activities or periodic reminders to engage in offline connections, for instance, may help users become less dependent on AI companions, according to user studies evaluating "ethics by design" safeguards.[43] In a similar vein, experimental studies of transparency tool like disclosure prompts alerting users that they are communicating with machine, show conflicting but encouraging results in terms of reducing excessive expectations and anthropomorphizing.[44,45] These results show that design decisions can be made based on normative principles and have observable results.

Comparative studies demonstrate hoe various regulatory frameworks moderate the risks and benefits of AI companionship at the policy level. For instance, stringent rules of data minimization, consent, and explainability are imposed by the European Union's GDRP and the upcoming AI Act, which provide structural safeguards against the exploitation of user vulnerability.[32] Conversely, more lenient settings, like the US, mostly depend on corporate self regulation, which raises questions about the commercialization of intimacy and uneven access to protection.[47] Cross-jurisdiction analyses reveal that stronger regulatory frameworks are linked to higher public trust and lower levels of reported misuse, according to cross-jurisdictional research,[48,49] indicating that governance models can directly affect the relational outcomes of AI adoption.

When combined, these results highlights the fact that responsible AI design necessitates more than just theoretical promises; it also calls for comparative evaluation of regulatory approaches and empirical support for actions. To guarantee that AI companions serve as enhancements rather than replacements for human interaction, policies that prioritize privacy, transparency, and equity must be combined with design methodologies that incorporate human in the loop protections and authenticity focused features. Future studies should keep examining feasibility of scaling interventions across various cultural and developmental context, from relational nudges to structural regulation. AI companionship can only be positioned as a helpful tool that strengthens rather than weakens the resiliency and authenticity of human bonds by integrating ethics, policy, and design in the evidence based manner.

Conclusion

The path from conventional human advice to AI-mediated friendship has been charted in this review, revealing a significant shift in the way people look for and maintain emotional support. These days, artificial intelligence systems function not just as informational resources but also as relational actors, counselors, confidants, and friends, with the ability to change the emotional landscape of human existence. AI companionship has been shown to democratize access to advice, lessen stigma, and provide care to marginalized and vulnerable groups. These platforms have great potential as additional scaffolds that broaden the scope of relational and therapeutic resources by reducing loneliness and providing tailored, stigma free support.

However, if these technologies are marketed as alternatives rather than supplements, the pose a profound risk. Over-reliance on AI companions runs the risk to erode authenticity, distort expectations of intimacy, and weakens the resilience that emerges from human vulnerability and imperfection. The task is to seek its purposeful and moral integration rather than to completely reject al companionship. This calls for research agendas that embrace cross-cultural developmental, and longitudinal views; policies that protects privacy and equity; and design techniques that incorporate authenticity and human-in-the-loop principle. The future of AI companionship ultimately depends on making sure that technology strengthens human con-

nection rather than takes its place, maintaining the integrity of interpersonal ties and the social fabric's resiliency throughout this crucial turning point.

References

1 Fitzpatrick KK, Darcy A, Vierhile M Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 2017;4(2), e19. https://doi.org/10.2196/mental.7785

2 Reiss D, Spina A. Scaling smart solutions with AI in health: Unlocking impact on high-potential use cases. World Economic Forum 2023.

3 Pentina I, Xie T, Hancock T, Bailey A. Consumer–machine relationships in the age of artificial intelligence: Systematic literature review and research directions. Psychology & Marketing 2023; 40(8), 1593–1614.

4 Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in mental health care. J Med Internet Res. 2019;21(5), e13216.

5 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021;88:105906.

6 McLeod J. Cultural and historical origins of counselling. In L. Bovair & P. Millett (Eds.), Understanding pupil behaviour in school, 2013, (pp. 178–195). David Fulton Publishers.

7 Teo SA. Artificial intelligence, human vulnerability and multi-level resilience. Comput Law Secur Rev. 2025;57:106134.

8 Federspiel F, Mitchell R, Asokan A, Umana C, McCoy D. Threats by artificial intelligence to human health and human existence. BMJ global health. 2023;8(5).

9 Ho JQ, Hu M, Chen TX, Hartanto A. Potential and pitfalls of romantic Artificial Intelligence (AI) companions: A systematic review. Comput Hum Behav Rep. 2025;19:100715.

10 Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Mental Health. 2018;5(4), e64. https://doi.org/10.2196/mental.9782

11 Yen HY, Huang CW, Chiu HL, Jin G. The effect of social robots on depression and loneliness for older residents in long-term care facilities: a meta-analysis of randomized controlled trials. J Am Med Dir Assoc. 2024;25(6).

12 Weerarathna IN, Raymond D, Luharia A. Human-robot collaboration for healthcare: a narrative review. Cureus, 2023;5(11).

13 Jin KW, Li Q, Xie Y, Xiao G. Artificial intelligence in mental healthcare: an overview and future perspectives. Br J Radiol. 2023;96(1150), 20230213.

14 Ahmed E, Buruk O, Hamari J. Human–robot companionship: current trends and future agenda. Int J Soc Robot. 2024;16(8):1809–1860.

15 Saeidnia HR, Hashemi Fotami SG, Lund B, Ghiasi N. Ethical considerations in artificial intelligence interventions for mental health and well-being: Ensuring responsible implementation and impact. Soc Sci. 2024;13(7):381.

16 Khosravi P, Rezvani A, Wiewiora A. The impact of technology on older adults' social isolation. Comput Hum Behav. 2016;63, 594–603.

17 Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. Can J Psychiatry. 2019; 64(7), 456–464.

18 Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health–a scoping review. Verhaltenstherapie. 2022; 32(Suppl. 1), 64–76.

19 Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis. World Psychiatry. 2019;18(3), 325–336.

20 Niemelä M, Heikkilä P, Lammi H. A social robot in a shopping mall: Studies on acceptance and user experience. Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 2017.

21 Castelo N, Sarvary M. Cross-cultural differences in comfort with humanlike robots. Int J Soc Robot. 2022;14(8):1865–73.

22 Li F. Studying the impact of emotion-AI in cross-cultural communication on the effectiveness of global media. Int J Soc Robot. 2025; 7, 1565869.

23 Bercovici HR A pastoral approach to personhood: Care and counseling reunited in Christian community. The Union for Experimenting Colleges and Universities. 1986.

24 Gouveia T, Schulz MS, Costa ME. Authenticity in relationships: Predicting caregiving and attachment in adult romantic relationships. Journal of Counseling Psychology. 2016;63(6):736.

25 Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Internal Medicine. 2016;176(5):619–625.

26 Meng J, Dai Y. Emotional support from AI chatbots: Should a supportive partner self-disclose or not?. Journal of Computer-Mediated Communication. 2021;26(4):207–222.

27 Lehman DW, O'Connor K, Kovacs BD, Newman GE. Authenticity. Acad Manage Ann. 2019;13:1–42.

28 Fuzzetti A. The High-Conflict Couple: A Dialectical Behavior Therapy Guide to Finding Peace, Intimacy, and Validation. New Harbinger Publications; 2006.

29 De Freitas J, Oğuz-Uğuralp Z, Uğuralp AK, Puntoni S. AI companions reduce loneliness. J Consum Res. 2025;ucaf040. https://doi.org 10.1093/jcr/ucaf040

30 Thakkar A, Gupta A, De Sousa A. Artificial intelligence in positive mental health: a narrative review. Front Digit Health. 2024;6:1280235.

31 Vandhika S, Sahrani R. Chatting Away Loneliness: Embracing New Connections Between Humans and Artificial Intelligence. INSAN: Jurnal Psikologi dan Kesehatan Mental. 2025;10(1).

32 Rahsepar Meadi M, Sillekens T, Metselaar S, van Balkom A, Bernstein J, Batelaan N. Exploring the ethical challenges of conversational AI in mental health care: scoping review. JMIR Ment Health. 2025;12:e60432.

33 Mökander J, Floridi L. Operationalising AI governance through ethics-based auditing: an industry case study. AI Ethics. 2023;3(2):451-468.

34 Sharkey A. Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. Ethics Inf Technol. 2014;16(1):63–75.

35 Dunigan P, Folk D, Heine SJ. Cultural variation in attitudes toward social chatbots. J Cross Cult Psychol. 2025;56(7):1024–1044.

36 Hithakshi B, Fernandes JJ. Human-centered ai in psychology: ethical considerations, applications, and future directions. 2025.

37 African Union. Continental Artificial Intelligence Strategy 2024. https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf

38 World Health Organization. WHO outlines considerations for regulation of artificial intelligence for health 2023. https://www.who.int/news/item/19-10-2023-who-outlines-considerations-for-regulation-of-artificial-intelligence-for-health

39 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. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

40 Kenya Law Reform Commission. The Data Protection (General) Regulations, 2021 (Legal Notice No. 263 of 2021). https://www.kenyalaw.org/kl/fileadmin/pdfdownloads/LegalNotices/2021/LN263_2021.pdf

41 Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry. 2014;13(1):4–11.

42 Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: Actualizing their potential. Annu Rev Public Health. 2009;30:273–292.

43 Bullard N, Guszcza J, Lim D, et al. Ethics by Design: An Organizational Approach to Responsible Use of Technology. Deloitte Insights; 2022.

44 Bostrom N, Yudkowsky E. The ethics of artificial intelligence. In: Trappl R, ed. Artificial Intelligence Safety and Security. Chapman and Hall/CRC; 2018:57–69.

45 Luger E, Sellen A. "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In: CHI Conference on Human Factors in Computing Systems; 2016:5286–5297.

46 Brey P, Dainow B. Ethics by design for artificial intelligence. AI Ethics. 2024;4(4):1265–1277. https://doi.org/10.1007/s43681-023-00330-4
47 Araujo T, Helberger N, Kruikemeier S, de Vreese CH. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc. 2020;35:611–623.
48 Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. Nature. 2019;568(7753):477–486. https://doi.org/10.1038/s41586-019-1138-y
49 Zeng Y, Lu E, Huangfu C. Linking artificial intelligence principles. 2018. arXiv:1812.04814. https://arxiv.org/abs/1812.04814