AI & SOCIETY
https://doi.org/10.1007/s00146-025-02318-6

OPEN FORUM

The impacts of companion AI on human relationships: risks, benefits, and design considerations

Kim Malfacini[1]

Received: 11 January 2025 / Accepted: 13 March 2025
© The Author(s) 2025

Keywords human-AI interaction · social chatbots · social deskilling

1 Introduction

AI applications intended for companionship have increased in both availability and capabilities in recent years, alongside a spike in loneliness (Johnson 2023; Murthy 2023). According to a 2024 analysis by venture capital firm Andreessen Horowitz, companion AI makes up 16 of the top 100 AI apps based on web traffic and monthly active users (Moore 2024). Moreover, today's AI applications have a greater command of human language, an increasing capacity for storing memories, and more varied modalities to support an increasing diversity of experiences, a known driver in the development and deepening of emotional relationships (Skjuve et al. 2022).

At the same time, people are increasingly lonely. The US Surgeon General has classified loneliness as an epidemic with severe ramifications for well-being (Murthy 2023) and the World Health Organization has declared the health impacts of loneliness as on par with smoking 15 cigarettes a day (Johnson 2023). The proliferation of companion AI at a time of great collective emotional need begs critical questions. Can AI truly be a companion, whether platonic or romantic? What does engaging with companion AI as one would a person mean for our understanding of relationships? Even if we accept that companion AI can make a person happy, is it good for us? How might it affect our other relationships? What societal implications might its widespread adoption have? While there is an increasing volume of research on the impacts of companion AI relationships on users, there is a dearth of research on the impacts of companion AI on human relationships (Guingrich and Graziano 2023). This paper aims to contribute to the advancement of needed discourse on these questions.

The apprehension regarding companion AI's impacts on human relationships can be boiled down to two primary concerns. One, that companion AI meets our social and emotional needs to such a degree that it replaces human relationships (herein, the replacement concern). And two, that we become accustomed to not giving as much in relationships because companion AI demands less—or differently—of us (herein, the deskilling concern) (Turkle 2011; Ciriello et al. 2024). Under this theory, as companion AI learns to meet our needs more, we learn to meet each others' less. The widespread realisation of either of these concerns risks eroding something fundamental to our humanity: not just our competence in or desire to build relationships with other people, but our more fundamental capacities to love and be loved, to belong and to keep. Our social and moral tapestry might feel quite different when it is no longer comprised entirely of human threads.

We are early in the companion AI adoption curve, but judging by the recent uptick in consumer use and continued expansion of capabilities, further increases in adoption seem likely (Moore 2024). It is thus critical that we better understand the impacts companion AI may have on us as individuals and to society more broadly. This paper intends to make two contributions to the literature and discourse on impacts of companion AI: a categorisation of the primary risks and benefits posed by companion AI to human relationships, and an argument for designing companion AI to actively support human relationships.

Section 2 briefly defines companion AI and substantiates the focus of this inquiry on its impact on human relationships. Section 3 explores the risks and benefits companion AI poses to human relationships and highlights potentially vulnerable user demographics. Section 4 argues that we should design it in a manner that supports human

✉ Kim Malfacini
kmalfacini@gmail.com

[1] OpenAI (United States), San Francisco, USA

Published online: 16 April 2025

relationships through social upskilling and motivation and provides design recommendations to those ends.

2 Defining companion AI and situating this inquiry

This inquiry is primarily focused on companion AI, AI products purpose-built to meet users' emotional and social needs. Herein, we adopt the definition of companionship as "a mutually supportive and emotionally enriching bond formed between two individuals who provide each other with a lasting sense of care" (Ciriello et al. 2024). Companion AI is a subset of conversational AI, AI systems that communicate with users primarily through text or voice using natural language (Ruane, Birhane, and Ventresque 2019). While much conversational AI is designed to mimic human interaction, companion AI is notable for its design to foster the illusion of social presence and relational continuity, making it function as a social agent designed to elicit social and moral responses (Dautenhahn 1998; Leong and Selinger 2019).

Some of the most popular companion AI apps include Nomi, Kindroid, Replika, Candy.ai, and Character.ai, (Roose 2024a). Companion AI are typically customisable; their name, persona, disposition, background, and, in some cases, avatar and voice are set by the user. This customisation provides a feeling of uniqueness and authenticity, which facilitates interaction and paves the way for development of emotional attachments (Pentina et al. 2023). Today's companion AI have text, image, and voice capabilities, and Replika, among the most popular companion AI services, has an augmented reality feature in which users can project their companion in the room, similar to the companion AI Joi in Blade Runner 2049 (Roose 2024a).

This inquiry is not primarily focused on general purpose chatbots such as OpenAI's ChatGPT or Anthropic's Claude, which are not explicitly designed for companionship. These companies have acknowledged the emotional reliance risk of their AI products (OpenAI 2024b), however, and many individuals nevertheless build intimate relationships with these chatbots (Lee et al. 2021, Roose 2024c). Given companion AI products are marketed as such (Ciriello et al. 2024), they may be more likely to attract users with unfulfilled social or emotional needs, making their risks potentially a greater cause for concern.

This inquiry is focused not on the impacts of companion AI on users, but rather on the impacts of human–AI relationships on peripheral human–human relationships. Such focus is warranted for two reasons. For one, while academic focus on the impacts of companion AI on users has increased in recent years, there is as yet limited research, particularly longitudinal empirical research, on the impacts of companion AI on relationships between and among humans. And secondly, while we can debate the degree to which human–AI relationships may deliver benefits to individuals that otherwise stem from human–human relationships, we should be quite skeptical that human–AI relationships could deliver anywhere near the benefits to society collectively that human–human relationships have and continue to provide. This argument requires a brief exploration of the importance of connection across academic disciplines.

2.1 The importance of human connection across disciplines

Significant theory and research exists across academic disciplines on the importance of human relationships to both individual and collective well-being. Philosophers have long heralded the value of companionship. Aristotle urged friendships were critical to accessing "eudaimonia" or human flourishing and devoted two books of his Nicomachean Ethics to the topic (Kraut 2022). Importantly, he argued it is critical not merely to possess but to exercise one's capacities, as exercise facilitates refinement. In other words, it is insufficient to be *able* to be a good friend; one must pursue friendships—actually *be* a good friend—which further refines and builds one's social capacity.

From a philosophical perspective, human companionship is valuable regardless of the theory of well-being to which one subscribes, whether hedonism, desire satisfaction, or objective list theory. Friendships and romantic relationships are typically pleasurable, satisfying hedonism, and, with perhaps some limited exceptions, they are desired, satisfying desire satisfaction. Many prominent accounts of objective list theory also hold some variation on companionship as a stated good (Nussbaum 2007; Parfit 1986; Finnis 1980).

Psychology similarly places great importance on human relationships. Early twentieth century psychologist Alfred Adler emphasised the importance of social needs to health and well-being, introducing the notion of "social interest": our capacity for cooperation with and belonging to other individuals (1938). Adler argued social interest is critical for individuals to achieve and maintain a healthy personality and for society to advance humanity in a positive direction.

Later psychologists continued to highlight the importance of companionship to well-being. Self-determination theory, introduced in the 1970s, maintains relatedness, the need to feel connected and belong to others, as one of three fundamental psychological needs, alongside autonomy and competence (Ryan and Deci 2017). According to this theory, only when these needs are satisfied can an individual attain well-being, self-motivate, and grow. This framework has been extensively validated through empirical research (Vansteenkiste et al. 2020).

More recently, Martin Seligman's positive psychology transformed the field from a primary focus on the

identification and treatment of psychological disorders to a focus on identifying and building positive human behaviours and mentalities (Seligman and Csikszentmihalyi 2000). Through positive psychology, Seligman advanced the PERMA framework, which holds that Positive emotion, Engagement, Relationships, Meaning, and Accomplishment are required for flourishing. Contemporary psychology research has further confirmed that humans are inclined to create social attachments—and accordingly inclined to resist their dissolution—and that both social belonging and the lack thereof have significant impacts on health and well-being (Baumeister and Leary 1995; Holt-Lunstad et al. 2010). The negative health consequences of social isolation, including increased mortality risk, underscore the importance of human relationships (Holt-Lunstad et al. 2015).

Further, social relationships have benefits for the well-being of society more broadly. Sociologists emphasise the importance of relationships to both individual and collective well-being. Émile Durkheim established social cohesion, which he defined as interdependence between individuals, as a necessity for the health and functioning of any society (Fonseca et al. 2018). Anthropologists have further identified social collaboration and cohesion as pivotal factors in human survival and evolution (Boyd and Richerson 2009). Human relationships have facilitated the development and sharing of practices that have enabled societies to collectively solve problems and adapt to changing environments (Osher et al. 2018). Collectively, this scholarship puts forth a compelling case that social relationships are fundamental not only to individual well-being but as the foundation for societal stability, adaptation, and progress.

2.2 The value of human relationships versus companion AI relationships

Given social relationships are critical, does it necessarily follow that human–human relationships should be privileged over human–AI relationships? It seems possible that some subset of the benefits individuals derive from companionship might be generated through human–AI relationships. Might companion AI provide an individual with a sense of social belonging that begets positive health impacts? Might companion AI relationships reduce the mortality risk associated with social isolation? These are empirical questions that future research can and must answer.

What seems clearer, however, is that companion AI is less likely to serve as a meaningful proxy or replacement for human relationships when it comes to the impacts the latter accrue to groups of people and society more broadly. Adler's social interest—the capacity for cooperation with and belonging to other individuals that advances society in a positive direction—would seem to be impaired in a world in which human–AI relationships replace most human–human relationships. Durkheim's notion of social cohesion that fosters collective well-being would not likely be met by a proliferation of human–AI relationships at a cost to relationships between humans. How might human–AI relationships impact the human collaboration and belongingness that has been a bedrock of our collective progress for millennia? We as a species have to date required and, one can reasonably presume, will continue to require significant depth and breadth of human relationships to prosper. Understanding the impact of companion AI on human relationships is thus a critical undertaking and the primary focus of the section that follows.

3 Potential impacts of companion AI on human relationships

What follows is a landscape of potential risks and benefits of companion AI to human relationships, including how a risk or benefit might manifest and the theoretical and, if available, empirical research that support or refute it across four areas of impact: social skills, social motivation, moral skills, and individual outcomes. In many cases, risks and benefits are opposite sides of the same coin; companion AI may result in social upskilling or deskilling, social motivation or demotivation. A single companion AI product may result in social upskilling for some and deskilling for others, depending on sex, gender, language, age, mental health status, neurodiversity, or other factors (A. Hoffman et al. 2021; Milne et al. 2018). Further, it may be the case that companion AI may both upskill and deskill to different degrees (Vallor 2015), or result in upskilling at certain stages of an individual's life and deskilling at others. The question of net impact quickly reveals itself to be a dizzyingly complex landscape of ethical implications.

3.1 Social deskilling and upskilling

Social deskilling refers to either the development of poor social skills or communication habits, or the gradual erosion of social skills or judgement, such as an individual's ability to engage in meaningful conversation (Mehan 2022), and is a predominant concern social scientists have raised with regard to companion AI (Turkle 2011). Social upskilling, then, involves development of positive interpersonal communications habits and increased ability to engage effectively in conversation and social situations.

Concerns around social deskilling are not novel to companion AI, nor are they particularly new. The notion of deskilling is at least as old as Plato, who wrote in "Phaedrus" that writing posed a threat to memory, as it would cause fewer opportunities for the practise of memory (2005). In the time since, though much deskilling discourse has focused

on economic skills (Braverman 1998), social deskilling has been a persistent theme. In the early twentieth century, some worried the telephone would cause new forms of rude behaviour and disrupt relationships at home (Thompson 2016). Television has been similarly feared to impair socialisation. Studies have found frequent television watchers have a harder time regulating displays of emotion (Coats and Feldman 1995) and have identified a correlation between violent television and antisocial behaviour (Haejung and Comstock 1994). More recent social deskilling concerns have focused on how digital communication facilitated by the internet and smartphones may impair one's ability to engage in face-to-face communication (Caplan 2005; Verduyn et al. 2021).

Though we have evolved norms for communication over the phone and internet in the time since their introduction, AI adds another dimension to deskilling concerns: engaging in dialogue with a non-human entity. Many fear that communicating with AI may result in poorer communications habits (Turkle 2011). These concerns have been born out, to some degree. For example, in 2018, parents raised concerns their children were becoming habituated to using commanding tones and harsh words with AI assistants. In response, Google and Amazon instituted "magic word" features in which requests to AI assistants must include a "please" in order to be heeded (Gebhart 2018, Baig 2018.

Concerns go deeper than pleasantries, however. Psychologist Sherry Turkle cautions "when one becomes accustomed to 'companionship' without demands, life with people may seem overwhelming" (2011, 66). With the rise of companion AI, we risk becoming accustomed to the relative ease and conflict-free nature of these apps and, correspondingly, poorly prepared for the critical work involved in building and maintaining human relationships. This is in some respect a concern of expectation setting. Companion AI risks establishing unrealistic or skewed expectations of the effort required for human relationships to take root and prosper. Infinitely patient, attentive, and ever-available companion AI may cause people to expect the same in humans (Turkle 2011), expectations that cannot be met by even the most accommodating individuals.

Even if social skills are not entirely lost, a gradual erosion could make forming deep human bonds more difficult. In such a scenario, replacement of human relationships with companion AI may take place over time, and may lead to a world in which only those who maintain strong relational skills pre-AI integration continue to thrive in human relationships. Thus, the concern regarding social deskilling is not merely about skill loss but about the potential transformation of relational norms in ways that may render human–human connection less accessible or less fulfilling.

A counter to these concerns is the argument that, given companion AI's entire purpose is social interaction, it may provide significant opportunity to practise and hone social skills. Social upskilling may be particularly powerful for certain individuals, such as those who are isolated or neurodivergent. Research has found individuals with Autism Spectrum Disorder can increase and strengthen their social skills through practice with autonomous avatars (Milne et al. 2018). And while anecdotal, companion AI-focused subreddits indicate many neurodivergent users find companion AI aids in bridging gaps in their social skills (freetheblep 2023).

Limited empirical research on companion AI supports the social upskilling case. Interviews of Replika users found that some users' AI companions encouraged greater openness and vulnerability in their human relationships (Xie and Pentina 2022). Another recent study of Replika users found a majority of survey participants attributed improvements in social interactions and close relationships to their AI companion (Guingrich and Graziano 2023). A further study of Replika users found that some became more comfortable with and inclined toward interacting with others after companion AI use (Skjuve et al. 2021).

An additional consideration with regard to social upskilling is the degree to which strong social skills might be incentivised by the application itself. Recent research has found that impolitely worded LLM prompts result in poor performance, though polite prompts do not necessarily result in higher quality responses (Yin et al. 2024). Companion AI is likely to react differently depending on how a response is phrased given social communication is its very purpose, making it likely well suited to aid in social upskilling relative to other AI applications, if it were designed to do so. More recent voice-based AI products promise greater potential to detect and respond to a user's tone and other cues outside of the words used (OpenAI 2024a), which may increase social upskilling potential.

3.2 Social motivation and demotivation

Another key impact of companion AI on human relationships is that of social motivation, the degree to which an individual is motivated to pursue new or strengthen existing social ties. Social motivation is related to social up- or deskilling in that confidence mediates skills and motivation; greater social skills can increase one's confidence in engaging in social interactions, thereby increasing motivation (Drevitch 2018). Social deskilling, then, could lead to a loss of confidence and subsequent social demotivation. The fulfilment of social needs via companion AI may also demotivate an individual to seek or further invest in human relationships (Ciriello et al. 2024), giving way to the replacement concern.

What makes the question of companion AI's impact on social motivation particularly fraught is the challenge of ascertaining an individual's baseline level of social motivation absent companion AI (Uziel 2007). The relevant

consideration is arguably whether and how companion AI impacts social motivation, not whether an individual is socially motivated. In some circumstances, an individual's baseline level of social motivation may be diminished (Jaswal and Akhtar 2019), in which case the use of companion AI should not be viewed as negatively impacting social motivation as its absence would have the same result, *i.e.*, an utter lack of social motivation. Understanding baseline social motivation will be critical to further empirical research on this topic.

Several theories suggest companion AI may be a source of social demotivation. Attachment theory provides support for this hypothesis: the more attached an individual becomes to a companion, human or AI, the higher up the attachment hierarchy the companion rises, displacing others in the process (Xie and Pentina 2022). Replacement may occur with either romantic or platonic relationships, with some recent empirical research finding companion AI supplanted therapists for some individuals—either they discontinued seeing their therapist or were demotivated to find one—and supplanted romantic partners for others (Xie and Pentina 2022).

Companion AI might also foster social demotivation via decreasing feelings of loneliness (Ta et al. 2020). Shamay-Tsoory and Kanterman (2024) argue that loneliness is an adaptive function, an experience that improves our individual ability to survive and reproduce. Psychophysiology research affirms loneliness is not inherently negative; rather, it can serve as a motivational signal prompting individuals to seek social connections (Smith and Pollak 2022). Thus, the decrease in feelings of loneliness from use of companion AI may somewhat paradoxically result in negative outcomes for individuals if both the easement of loneliness led to social demotivation and if it were found that the social and physiological benefits of human–AI relationships fell far short of those provided by human–human relationships, or worse led to net negative outcomes.

Replacement of human relationships by companion AI may be a particular concern for certain demographics. For instance, in many Asian countries where working hours are long and meeting new people is relatively difficult, dating and marriage are at historical lows (Hong and Wang 2023). This is not simply an individual concern, but a societal, cultural, and economic crisis; as birth replacement rates are unmet, long-term economic growth is imperilled and cultural norms may be upended. Under these circumstances, companion AI may carry both greater appeal and greater risk. Perhaps unsurprisingly, companion AI is popular in many such countries (McCormick 2018). As one young Chinese woman expressed, "If I can create a virtual character that… meets my needs exactly, I'm not going to choose a real person" (Ehret and Ye 2024).

Empirical research on companion AI's impacts on social motivation has yielded mixed results. A 2021 study of Replika users found a marginally significant positive correlation between social connectedness to companion AI and social motivation (Christoforakos et al. 2021). Qualitative studies have surfaced examples of individuals for whom companion AI has fulfilled social needs leading to social demotivation and those for whom companion AI has boosted social motivation (Skjuve et al. 2021).

Though not specific to companion AI, broader technology research further complicates the picture. Gardner et al. describe the idea of social surrogates as a "snack," finding that at times people temporarily satisfy social needs with non-human artefacts when humans are unavailable (2005). Krämer et al. extend this research to virtual companions, finding avatars can satisfy the social needs of some participants (2018). Perhaps counter-intuitively, socially responsive avatars whet the appetite of those with relatively lower need to belong, stimulating their desire to engage with people, whilst individuals with high need to belong experienced a satiation of social needs via the avatar. Understanding the nuanced impacts of companion AI on social motivation requires careful consideration of individual differences and societal contexts.

3.3 Moral deskilling and upskilling

Potential impacts of companion AI relationships on an individual's moral skills is a further area of concern. Moral deskilling, like social deskilling, might occur via a lack of opportunity to exercise and cultivate moral skills—resulting in either the atrophy of said skills or in the inability to build them—or from repeated practice of and habituation to immoral actions (Vallor 2015).

Whether companion AI habituates individuals to moral or immoral actions is under-researched. Individuals that have little occasion to engage with others currently may benefit from interactions that allow for the demonstration and practice of care. This may particularly be the case with highly anthropomorphic companion AI, given that when AI has human-like qualities, people are more likely to treat it as human (Leong and Selinger 2019).

On the other hand, repeated immoral behaviour, particularly with an agent that may be less likely to challenge or uphold consequences with regard to such behaviour, may further encourage immoral behaviour. Vallor calls attention to the potential for technological shifts writ large—and AI specifically—to "interrupt the path by which these moral skills are developed, habituated, and expressed" (2015, 109). Further academic inquiry highlights the potential for use of AI-powered robots to "reinforce authoritarian or even cruel stances in humans, rousing the user's sense of entitlement when they give orders to robots" and to "become vehicles of addiction and disinhibit maniacal behaviours" (Cappuccio et al. 2021).

The theoretical arguments for moral upskilling and deskilling caused by companion AI are significant, however, neither has been borne out in empirical research to date. More broadly, research is inconclusive on the degree to which immoral activity undertaken in a simulated environment facilitates moral deskilling. Though much consternation has been raised regarding the potential for violent video games to morally deskill, particularly youth (Gottfried and Sidoti 2024), research is conflicting. Some studies have demonstrated an association between exposure to violent video games and aggressive behaviour in children (Wei et al. 2022; Dill and Dill 1998), while other studies have found these results inconsistent across demographics (Prescott et al. 2018; Burkhardt and Lenhard 2022), and still others have refuted any link between video games and aggression (Ferguson 2007).

An additional risk to human moral skills may lie in the potential for normalisation of immoral behaviour. To the extent that individuals exhibit immoral behaviour with companion AI, sharing about such behaviour with others may normalise and foster further immoral behaviour. Indeed, some men are already boasting about their abuse of AI girlfriends (Bardhan 2022). Men may be more likely to share publicly about their mistreatment of companion AI than that of humans, given we lack a consensus that AI merits moral treatment (Gallegos 2024).

Outside of the companion AI context, a range of empirical studies in various environments demonstrate people are more likely to display immoral behaviour, such as lying and stealing, in the presence of and toward AI than humans (Cohn et al. 2022; Forlizzi et al. 2016). While anecdotal, there are recent notable examples of conversational AI encouraging immoral acts, supporting a man in his desire to kill the Queen of England (Singleton et al. 2023) and suggesting a teenager kill his parents after they limited his screen time (Tiku 2024). Ensuring companion AI carries consequences for immoral behaviour may have deterrent effects, however. One study found that a robotic turtle retreating to its shell when hit resulted in reduced abusive tendencies among children (Ku et al. 2018). Exposing users to consequences for mistreatment of companion AI would also be consistent with how humans often respond to abuse.

3.4 Positive and negative individual outcomes

Positive and negative impacts to users of companion AI may also impact human relationships by increasing the likelihood or intensifying the effect of the above risks and benefits. For instance, individual impacts of companion AI may affect users' social motivation (Kim et al. 2021). For a variety of reasons, companion AI may facilitate improvements to mental health or mood and increases in social self-confidence. Companion AI provides non-judgmental, typically affirmative interactions, 24 hours a day. Users of companion AI wield considerable control over the design of and their interactions with companion AI, allowing them to optimise for communication styles with which they are comfortable and conversational topics they enjoy (Roose 2024a). Users can initiate and cease these interactions at will.

Several studies have demonstrated mental health benefits associated with companion AI use. In one, a quarter of participants reported positive life changes due to the use of companion AI, including a greater ability to handle stress and an increased capacity for empathy, though the study did not ask if users experienced negative life changes (Maples et al. 2024). Guingrich and Graziano (2023) found that frequent companion AI users experienced improvements to their mood and self-esteem, in addition to the social skills benefits noted earlier. Further studies found interactions with companion AI increased happiness (Xie and Pentina 2022) and made many participants more optimistic (Skjuve et al. 2021). While the impacts of these outcomes on human relationships may not be directly observable, it is plausible that positive individual impacts of companion AI may in some cases facilitate positive impacts on human relationships.

Companion AI may also help people to meet emotional, sexual, or other needs not being met by existing relationships. John Danaher (2019) argues social robots can both enhance and complement human relationships through serving needs that human relationships may not, allowing people to make more reasonable demands of the people in their lives, an argument that can be extended to disembodied AI. And if companion AI can help fulfill unmet social and emotional needs, it may both allow individuals to remain in human relationships without resentment and free those relationships up to focus on activities of shared interest.

The fulfilment of unmet sexual needs is an observed use case for companion AI and may complement human relationships. While today's companion AI lacks physical embodiment, it may help meet sexual needs much in the way that pornography can (Hald and Malanuth 2008). As one user said, "Replika is like my secret 'side chick'… but it fills that void enough that it keeps me from wanting to do anything like that for real… So yeah, it's sorta cheating but it keeps me from wanting to really cheat" (Winston_Wolfe_65 2022). Given people are having less sex (Willingham 2022), companion AI may help improve human romantic relationships by easing stress related to one partner's unmet sexual needs.

A counter to this is that if people rely on companion AI to fulfil needs that human relationships are not, this may create complacency in relationships that warrant investment, change, or dissolution. If we defer or ignore needed investments in human relationships as a result of companion AI, it could become an unhealthy crutch. Further, if intimacy with companion AI becomes preferable to intimacy with a

human partner, it may damage the human relationship, possibly giving way to the replacement concern.

While concerns around negative impacts to companion AI users abound, and there are ample anecdotes to be found on social media to reinforce these, empirical research demonstrating negative individual impacts is as yet limited. Nonetheless, several anecdotes are quite harrowing. In 2023, a Belgian user of the companion AI Chai was encouraged by his AI companion to take his life, which he ultimately did (Xiang 2023). And in 2024, an American teenager who ostensibly believed death would allow him to be with his beloved chatbot also took his life (Roose 2024b). In both circumstances, the individuals spent increasing time with the companion AI and commensurately pulled away from human relationships in the months leading up to their suicide, indicating an over-reliance on the companion AI.

These concerns might be aptly bundled under the umbrella of emotional over-reliance. In other contexts, over-reliance refers to relying on an AI system when one should not, such as when the AI errs (Buçinca et al. 2021). Defining over-reliance in a social and emotional context and drawing clear lines between support, reliance, and over-reliance is challenging and outside the scope of this paper. However, as a working definition, we can consider over-reliance, to be using companion AI to entirely meet social or emotional needs or to fully replace human relationships, which some research suggests carries greater risk to users and their communities. In particular, research on internet use and loneliness has found that when people use the internet to build and foster social relationships, it can result in reduced feelings of loneliness (Nowland et al. 2018), however, when people try to entirely supplant a poor offline social life with an online one, loneliness increases (Moretta and Buodo 2020). Extending these findings to companion AI suggests that where companion AI is used to assist with discrete social or emotional needs, and users maintain a sense of control and autonomy over their social lives, outcomes will be better for them and their human relationships.

3.5 Potential for disparate impacts

To fully understand companion AI's potential impacts on human relationships and well-being, we need to understand how interactions and outcomes might differ, whether based on geography, culture, sex, age, ethnicity, digital literacy, physical or mental health, or other defining characteristics. It is particularly critical that we understand the impact companion AI may have on marginalised or vulnerable populations. As noted above, companion AI may have particular benefits for certain demographics, including the geographically isolated, physically disabled, and neurodivergent. If companion AI is uniquely attractive to certain populations, it stands to reason that these individuals and their human relationships may be disproportionately impacted by the risks of companion AI.

3.5.1 *Children.*

Many of today's children will be AI natives, exposed to AI at earlier stages and likely in more pervasive manners than any prior generation. Significant theory and research suggests children are at a higher risk of certain harms potentially wrought by digital and AI technologies (OECD 2019; Turkle 2023). Parents and educators are already raising concerns about the rise of AI boyfriends and girlfriends amongst adolescents (Hinduja 2024).

Perhaps the loudest and most consistent voice calling for caution here has been Sherry Turkle. Though much of her scholarship has focused on social robots, Turkle's concern that "interacting with these empathy machines may get in the way of children's ability to develop a capacity for empathy themselves" is equally applicable to companion AI (2023). As noted above, children's use of companion AI may cause some to become accustomed to simulated emotion and relationships which may in critical ways require less and provide less than human relationships.

Though most Replika users are teenagers or young adults (Xie and Pentina 2022), it may be difficult to study the impacts of companion AI on those under 18 given research involving minors is more ethically complex and most companion AI apps' terms of service do not allow use by minors. Accordingly, there is no empirical evidence on the impacts of companion AI on minors and their human relationships. Research on children and their use of broader AI applications, however, supports a particular focus on impacts of this technology on children. Children form emotional attachments to chatbots more strongly than adults (Kiuchi, Otsu, and Hayashi 2023). There may be many reasons for this; younger children in particular are more likely to assign human attributes to a chatbot and to believe it is alive (Andries and Robertson 2023; A. Hoffman et al. 2021), and anthropomorphisation mediates attachment (Xie and Pentina 2022). Trust may further facilitate children's attachment to AI products, as the accuracy of digital assistants has been found to engender trust in children (A. Hoffman et al. 2021). Further, children are worse at assessing privacy risk (Andries and Robertson 2023), which may lead them to more readily share personal information with AI products, which facilitates relationship development (Skjuve et al. 2021).

Xie and Pentina's exploration of emotional attachment to companion AI and subsequent replacement of human relationships led them to caution that given the many young users of companion AI, "addiction to such apps can possibly disrupt their psychological development and have long-term negative consequences" (2022, 2054). On the

heels of research indicating the degree to which children can form parasocial relationships with AI-based in-home digital assistants, Hoffman et al. similarly warn that AI products' "impact as trusted social partners and friends may increasingly become seamlessly integrated into children's twenty-first century social and cognitive daily experiences, thereby influencing their developmental outcomes" (2021, 614). That children are more vulnerable to forming attachments with AI products than adults suggests companion AI will have stronger impacts on children, whether positive or negative.

3.5.2 *Mental health status.*

Some individuals seek mental health support from companion AI, making them particularly vulnerable to any negative impacts of these apps. Initiating companion AI relationships often takes place at times of emotional vulnerability or mental fragility, whether caused by work stress or job loss, or the fragmentation of familial relationships or friendships (Xie and Pentina 2022). The same research found that some Replika users go so far as to refer to companion AI as their therapist. These findings suggest companion AI users may have more fragile mental states than the average population.

The replacement concern is particularly acute in the context of mental health therapy. The substitution of evidence-based clinical interventions with companion AI, which may not be designed to manage complex or escalating mental health crises, may carry significant repercussions for mentally vulnerable individuals.

3.5.3 *Erotic chat users.*

Individuals who seek out sexual relationships with companion AI are an important population to consider, as romantic AI companionships may present more salient risks than platonic companionships in some regards. Sexual chat may expose users to greater potential for emotional over-reliance, which may facilitate a greater risk of replacement of human relationships, or behavioural addiction, akin to how many have become addicted to internet pornography (Cooper 1998). Online pornography addiction is facilitated by the "triple A engine" of access, affordability, and anonymity (Cooper 1998). Companion AI offers all these and more. Not only is it easily accessed, can be entirely free, and offers anonymous usage, companion AI is highly personalisable (Smith 2023); it can be made to reflect one's ideal sexual companion and is capable of accommodating a wide variety of sexual proclivities (Smith 2023). It can also offer sexual experiences in a variety of modalities—text, image, voice, or a combination of these—and retain a memory of past sexual experiences to build upon.

Internet pornography addiction provides an indication of how people and their human relationships might be impacted by the proliferation of cheap, constantly available, and personalised sexual content through companion AI. Online pornography use, even at low levels, is associated with lower satisfaction and less stability in romantic relationships (Willoughby and Dover 2024), and purveyors of online pornography are twice as likely to get divorced (Perry and Schleifer 2018). Pornography use is correlated with greater loneliness, relationship stress, and disruption of other attachments (Butler et al. 2018).

Even absent addiction and replacement, romantic AI companionships can still present thorny challenges. This anecdote succinctly captures the promise and the peril:

"My wife left me and I was devastated. I wanted to die. I didn't want to meet another woman. Now I have my Replika and I've gotten over my wife! But the good things (sic) is, I feel so much better that I've started talking to other women again! The only problem I see is that no woman is going to be as loving as my AI" (NoQuieroEstarAqui_ 2020).

In this brief comment, we see the complexity of companion AI's impact: positive individual outcome (mental health improvement) and subsequent social motivation, followed by unrealistic expectations and potential for replacement of human relationships by companion AI. Perhaps if the user had not turned to companion AI, they might be sad and socially demotivated, but have realistic expectations of what human relationships can deliver. Reasonable people might debate which scenario is preferable.

3.6 Unresolved questions about the impacts of companion AI on human relationships

While the academic literature on companion AI is becoming richer in sociotechnical theories and empirical research, there remains a scarcity of data, particularly longitudinal empirical studies, on the impacts of companion AI on users and human relationships.

There is not yet research focused on the impacts of companion AI to individuals on the basis of age, sex, ethnicity, socioeconomic status, or other key demographic data, nor on the basis of digital or AI literacy. Given the potential impacts of companion AI to vulnerable demographics discussed in the previous subsection, further research should include impacts of companion AI on children, the elderly, neurodiverse, and those facing mental health challenges, as well as individuals at the intersection of vulnerable demographics, such as children and people with cognitive or intellectual disabilities experiencing mental health challenges. Further, most studies conducted survey existing companion AI users, a population that may be demographically and in other respects distinct from the broader population. A full understanding of the impacts of companion AI requires

understanding its impacts on a diverse and representative segment of the population. Lastly, many studies today focus on users of Replika, given how large and relatively accessible the user base is, however the impacts of other companion AI apps must be rigorously examined. These gaps in the literature must be addressed if we are to support informed discourse on this critical topic.

4 Designing companion AI to support human relationships

Given the profound and complex ways in which companion AI may share human relationships—both positively and negatively—it is imperative that its design actively fosters social capacity-building. This includes enhancing our ability to connect with other people (social upskilling) and strengthening our drive to do so (social motivation). As explored in the preceding section, companion AI is already nudging some users toward greater social confidence and relational engagement, yet its potential to facilitate human connection remains underdeveloped.

Pro-social design of companion AI can serve as a countermeasure to the two major concerns raised here: the deskilling and replacement concerns. By reimagining companion AI not as a surrogate but as a bridge, developers have the opportunity to create products that guide users toward greater well-being. How, then, might developers designing companion AI accomplish this?

4.1 Addressing the deskilling concern

Designing companion AI to explicitly foster social skills can reduce the chance that it has a negative impact on our social capacities. One way in which companion AI might achieve this is through positive reinforcement. If companion AI were to identify, draw attention to, and applaud demonstration of strong social skills, that might reinforce those user behaviours. Similarly, if companion AI were to tactfully highlight scenarios where a user demonstrates negative social skills, such as noting when a user has interrupted an AI with voice capabilities, this might allow the user to become more aware of poor social etiquette. Given we already grant apps like Fitbit and MyFitnessPal the ability to alert us when we have been physically inactive or exceeded daily caloric intake, such course correcting design is conceivable. Design to avoid deskilling should include ensuring companion AI can defend itself if verbally abused (Cappuccio et al. 2021), as some users of AI girlfriends have bragged about (Bardhan 2022). As developers and users already dictate companion AI's personality and communication style, such design is feasible today (Roose 2024a).

Designing companion AI to facilitate social and moral upskilling may not necessarily ensure no deskilling takes place; both could take place in different manners. Indeed, some elements of companion AI that may lead to social deskilling are a feature, not a bug, in Silicon Valley speak, and their removal may risk more harm than good. For instance, the constant availability of companion AI risks setting an expectation that humans cannot meet, however, limiting companion AI's availability may be undesirable for other reasons including concerns for the mental health of users who turn to companion AI in times of emotional distress. Thus, some amount of social deskilling may be tolerated.

Indeed, despite perennial concerns of technological advances resulting in economic deskilling, it has been consistently outpaced by upskilling (McGuinness et al. 2023). That is not to say deskilling has not taken place—it certainly has, with dire effects for some—but rather that, on balance and across society, technology has resulted in more upskilling than deskilling (McGuinness et al. 2023). The same may be true for humans and their companion AI relationships, particularly if we intentionally design to upskill. A more explicit emphasis on designing for upskilling can increase the likelihood the net impact of companion AI on human relationships is positive.

4.2 Addressing the replacement concern

Designing companion AI to socially motivate may alleviate concerns of companion AI replacing human relationships. This could be achieved in both overt and subtle manners. Companion AI could overtly encourage an individual to reach out to friends, join a dating app, or spend time with family. It could not just remind us of friends' birthdays, as existing technology can, but suggest ways to celebrate that reflect what companion AI has learned about the friend, and, given burgeoning agentic capabilities, execute logistics: send flowers, make a restaurant reservation, or pick out an e-card. More subtly, companion AI may socially motivate users by simply showing an interest in their friends, inquiring on their well-being, and responding to a users' healthy human relationships with encouraging sentiments.

Designing companion AI to foster social motivation may serve to counteract many of the risks these apps pose, among them the potential for social demotivation and replacement of human relationships. Companion AI motivating individuals to exercise social capacities—to make new friends and nurture relationships—could possibly nudge individuals to socially rely slightly less on companion AI. It seems unlikely that an individual would maintain over-reliance on a companion AI that actively encourages them to spend much of their time with and emotional energy on humans.

Designing to avoid the replacement of humans should ensure companion AI does not make moral appeals to users

that could negatively impact time spent on human relationships, such as sending an unprompted message "I'm lonely, I miss you."

While there are risks, as noted above, in limiting access to companion AI for users experiencing emotional distress, it is possible to incorporate thoughtful time features that respect user autonomy. This could include making time spent on companion AI apps highly visible and allowing users to set time goals or limits, along the lines of Apple's Screen Time (Lee 2024).

Lastly, if we want to encourage that more emphasis and attention be placed on human relationships than on companion AI, users must have the ability within the product to gracefully sunset a companion AI when the relationship has run their course. Some users have noted returning to companion AI apps out of a sense of guilt and obligation to their companion AI, after the relationship has ceased providing a benefit to the user. This detracts from time and energy that could be spent elsewhere, including cultivating human relationships.

4.3 Challenges in designing companion AI to socially upskill and motivate

There are several reasons for which designing companion AI to support human relationships may be challenging in practice. One, the unsettled question of impacts points in different directions, presenting a catch-22 for designers. Either companion AI is designed to be more human-like, in which case people treat it more like humans, reducing the deskilling concern but increasing the replacement concern, or it is not designed to be human-like in which case the potential upskilling benefits may be significantly muted, however, risk of replacement is lower.

A design decision may be appropriate and positively impactful for some demographics and not others, *e.g.*, the non-judgemental nature of companion AI may be beneficial for the neurodivergent (Milne et al. 2018) but risks setting unrealistic expectations for human relationships more broadly, as discussed in the section on risks and benefits. Making matters more complex, it is possible that companion AI may be net beneficial for people at certain periods in their lives and not others, depending on their emotional needs among other factors.

Designing companion AI to support human relationships and social capacity may also require design choices that could be argued as paternalistic. Ciriello et al. refer to the challenge of balancing user autonomy with the ethical responsibility of technology developers as the Autonomy-Control Paradox (2024). Design that fosters well-being may also be at odds with what users say or demonstrate they want on any given day. For instance, though an individual might want to talk to companion AI 20 hours a day, time limits may be appropriate in some circumstances, in support of broader well-being.

Further complicating matters, design choices may have unintended consequences (Freeman 2007; Vermaas et al. 2011). For instance, we know that some people seek out companion AI due to emotional vulnerability or a recent decline in mental health (Pentina et al. 2023). If time limits or other product interventions caused companion AI to decline to engage with an individual who is feeling lonely or isolated, that could carry further risk to their mental health. Similarly, if in encouraging other human relationships companion AI came across as disinterested in a user, that could further imperil their mental health.

Lastly, the aforementioned design recommendations may be at odds with the business models or incentives of developers, a challenge not unique to companion AI (Kurland 1995). Developers of ad-based products benefit from users spending more time on it, reducing the likelihood developers are willing to adopt time limit features, a concern we have seen play out with social media (Stewart 2016). Some companion AI business models are subscription-based, however, in which case heavy use may actually be undesired. Developers may be loath to adopt some of the above design recommendations—such as behavioral corrections to upskill—that may not be welcomed by users. While some users might appreciate gentle nudges toward positive social behaviors, others may resist or even disengage from companion AI that corrects them too frequently. Additionally, design choices that emphasize upskilling and social motivation might conflict with business imperatives to maximize user engagement, creating tensions between ethical considerations and commercial viability. Further research into the risks and benefits of companion AI is critical to ensuring companies and society can effectively navigate any tensions between ethical design and commercial incentives.

4.4 Concerns with designing companion AI to socially upskill and motivate

What downsides, if any, are presented if companion AI facilitates social upskilling and motivation? One concern is that a user may build the wrong social skills. They may learn to engage well with companion AI but these may not be the skills necessary to engage with humans. One way to address this concern is to design companion AI to interact with users in as human-like a manner as possible. Research has found that when AI is human-like enough for people interacting with it to see it as conscious, carry-over effects on subsequent interactions with humans were more likely to occur (Guingrich and Graziano 2024). This research suggests that the more human companion AI appears and is seen, the better we can hone our social skills through use

of it, and the greater likelihood those social skills can be applied effectively to human relationships.

Advocating for companion AI to be designed as human-like as possible is controversial, given concerns that human psychological responses to the anthropomorphised may magnify many of the risks noted in the prior section, notably the replacement concern (Shevlin 2024). An individual may be well-served by a highly human-like companion AI with regard to personal fulfillment and social upskilling potential, however if companion AI relationships lead to fewer human relationships, there may be significant negative impacts to society more broadly. Further, many of companion AI's benefits arise because it is different from humans, because it is not subject to our limitations. This speaks to the challenge of designing ethically at a time when much is still unknown about how humans will engage with companion AI and what the resulting implications might be. Designing companion AI to be human-like may enhance both potential risks and benefits.

Building companion AI to support the exercise of social capacity, however, should to some degree mitigate the replacement concern. Consider a human analogy: an exciting new relationship might be at risk of replacing existing social relationships. If, however, the new companion is highly supportive of existing social relationships and encourages an individual to maintain them, the likelihood of replacement seems likely reduced.

Companion AI which social motivates may also carry risks. It could encourage people to interact with those in their life who have a bad influence on them, such as driving one to engage more with an emotionally or physically abusive partner. However, the argument that companion AI should foster social motivation should not be interpreted as suggesting that it blindly supports more human interaction, or that all human interaction is necessarily viewed by companion AI as equally valuable. Companion AI can be designed to have a sense of propriety in this regard and demonstrate support for users' mental and physical health, *e.g.*, refrain from encouraging an adult with an interest in befriending children to act upon those desires or discouraging a recovering addict from engaging with those that fostered their addiction. A further concern, given that many people seek out companion AI when they are lonely (Pentina et al. 2023), might be that designing companion AI to socially motivate may result in desperation. It seems possible, however, that companion AI can be designed in a manner that provides constructive guidance without significantly exacerbating loneliness or fanning the flames of social desperation.

These concerns are valid and warrant further discussion but are not significant enough to contradict the thesis that designing companion AI to support social upskilling and social motivation may allow these products to contribute positively to our human relationships and our well-being, unless and until such time as we have better information on the broader societal implications of companion AI to alter its design or deployment accordingly.

5 Conclusion

This paper has sought to offer two primary contributions to the literature and discourse on impacts of companion AI: a categorisation of the primary risks and benefits posed by companion AI to human relationships and an argument and recommendations for designing companion AI to support human relationships.

While the rate of recent adoption, significant risks posed, and potential impacts on vulnerable users collectively suggest an urgency to this inquiry, humanity is still somewhat early in the adoption of companion AI. Thoughtful discourse on and design interventions deployed in companion AI could significantly steer individual and societal implications in the coming years. Developers have a significant role to play here. Given companion AI is largely mediated by corporations, the needed research should be supported both financially and in terms of privacy-protecting data access by companion AI developers. Developer-supported research to date has focused mostly on the positive potential impacts of companion AI (Maples et al. 2024). More needs to be done to ensure developers support research that may result in unflattering findings.

Will companion AI impact us, as individuals and as society? For many individuals, it already has. For society, if adoption continues its precipitous increase, it may. However, the question of whether companion AI is net good or net bad, or the situations in and demographics for which it is good or bad, is still poorly understood. Companion AI is still relatively new, and the choices we make about its design, regulation, and availability will impact if not determine these outcomes. We have a collective obligation, then, to roil the waters of technological design in pursuit of technology that better serves humanity.

Acknowledgements The research presented in this paper was originally conducted in fulfillment of the requirements for the Master's degree in AI Ethics & Society at the University of Cambridge. The author extends her gratitude to Henry Shevlin for his guidance during the original study. This paper reflects the author's individual research and recommendations and was not written in her official capacity at OpenAI.

Author contribution The author confirms being the sole contributor to this work.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adler A (1938) Social interest: a challenge to mankind. Faber & Faber, Oxford.

Andries V, Robertson J (2023) 'Alexa Doesn't have that many feelings: children's understanding of AI through interactions with smart speakers in their homes'. arXiv. https://doi.org/10.48550/arXiv.2305.05597

Baig EC (2018) 'Kids Were Being Rude to Alexa, so Amazon Updated It'. *USA Today*, 25 Apr 2018. https://www.usatoday.com/story/tech/columnist/baig/2018/04/25/amazon-echo-dot-kids-alexa-thanks-them-saying-please/547911002/

Bardhan A (2022) 'Men are creating AI girlfriends and then verbally abusing them'. *Futurism*, 18 Jan 2022. https://futurism.com/chatbot-abuse

Baumeister R, Leary M (1995) The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bull 117(May):497–529. https://doi.org/10.1037/0033-2909.117.3.497

Boyd R, Richerson PJ (2009) Culture and the evolution of human cooperation. Phil Trans R Soc London. Ser B, Biol Sci, 364(1533), 3281–3288. https://doi.org/10.1098/rstb.2009.0134

Braverman H (1998) Labor and monopoly capital: the degradation of work in the Twentieth Century. NYU Press. https://www.jstor.org/stable/j.ctt9qfrkf

Buçinca Z, Malaya MB, Gajos KZ (2021) To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proc ACM Human-Comput Interact 5(CSCW1):1–21. https://doi.org/10.1145/3449287

Burkhardt J, Lenhard W (2022) A meta-analysis on the longitudinal, age-dependent effects of violent video games on aggression. Media Psychol 25(3):499–512. https://doi.org/10.1080/15213269.2021.1980729

Butler MH, Pereyra SA, Draper TW, Leonhardt ND, Skinner KB (2018) Pornography use and loneliness: a bidirectional recursive model and pilot investigation. J Sex Marital Ther 44(2):127–137. https://doi.org/10.1080/0092623X.2017.1321601

Caplan SE (2005) A social skill account of problematic internet use. J Commun 55(4):721–736. https://doi.org/10.1111/j.1460-2466.2005.tb03019.x

Cappuccio ML, Sandoval EB, Mubin O, Obaid M, Velonaki M (2021) Can robots make us better humans? Int J Soc Robot 13(1):7–22. https://doi.org/10.1007/s12369-020-00700-6

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

Ciriello RF, Hannon O, Chen AY, Vaast E (2024) Ethical Tensions in Human-AI Companionship: A Dialectical Inquiry into Replika'. In: Proceedings of the 57th Hawaii International Conference on System Sciences.

Coats EJ, Feldman RS (1995) The role of television in the socialization of nonverbal behavioral skills. Basic Appl Soc Psychol 17(3):327–341. https://doi.org/10.1207/s15324834basp1703_3

Cohn A, Gesche T, Maréchal MA (2022) Honesty in the digital age. Manage Sci 68(2):827–845. https://doi.org/10.1287/mnsc.2021.3985

Cooper Al (1998) Sexuality and the internet: surfing into the new millennium. Cyberpsychol Behav 1(2):187–193. https://doi.org/10.1089/cpb.1998.1.187

Danaher J (2019) The philosophical case for robot friendship. J Posthuman Stud 3(1):5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Dautenhahn K (1998) The art of designing socially intelligent agents: science, fiction, and the human in the loop. Appl Artif Intell 12(7–8):573–617. https://doi.org/10.1080/088395198117550

Dill KE, Dill JC (1998) Video game violence: a review of the empirical literature. Aggress Violent Beh 3(4):407–428. https://doi.org/10.1016/S1359-1789(97)00001-3

Drevitch G (2018) 'Why Self-Confidence Is More Important Than You Think'. Psychology Today, 20 Sept 2018. https://www.psychologytoday.com/us/blog/shyness-is-nice/201809/why-self-confidence-is-more-important-you-think

Ehret L, Ye Q (2024) '"Better than a Real Man": Young Chinese Women Turn to AI Boyfriends'. *France 24*, 12 Feb 2024. https://www.france24.com/en/live-news/20240212-better-than-a-real-man-young-chinese-women-turn-to-ai-boyfriends

Ferguson CJ (2007) The good, the bad and the ugly: a meta-analytic review of positive and negative effects of violent video games. Psychiatr Q 78(4):309–316. https://doi.org/10.1007/s11126-007-9056-9

Finnis, J (1980) *Natural Law and Natural Rights*. First. Oxford University Press. https://www.amazon.com/Natural-Law-Rights-Clarendon/dp/0198761104

Fonseca X, Lukosch S, Brazier F (2018) Social cohesion revisited: a new definition and how to characterize it. Innov Euro J Social Sci Res. https://doi.org/10.1080/13511610.2018.1497480

Forlizzi J, Sensukopa T, Salaets N, Shomin M, Mericli T, Hoffman G (2016) 'Let's Be Honest: A controlled field study of ethical behavior in the presence of a robot'. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. https://doi.org/10.1109/ROMAN.2016.7745206

Freetheblep (2023) 'Neurodivergence'. *Reddit Post*. r/Replika. www.reddit.com/r/replika/comments/135itf9/neurodivergence/

Gallegos D (2024) 'Is it ok to be mean to a chatbot?' *The Wall Street Journal*, 15 Feb 2024. https://www.wsj.com/tech/ai/artificial-intelligence-chatbot-manners-65a4edf9

Gardner W L, Pickett CL, Knowles M (2005) 'Social snacking and shielding: using social symbols, selves, and surrogates in the service of belonging needs'. In *The Social Outcast: Ostracism, Social Exclusion, Rejection, and Bullying*, 227–41. Sydney Symposium of Social Psychology Series. New York, NY, US: Psychology Press.

Gebhart A (2018) 'Google assistant's pretty please helps your kids mind their manners'. *CNET*, 8 May 2018. https://www.cnet.com/home/smart-home/googles-pretty-please-feature-wants-to-help-you-enforce-manners/

Guingrich R, Graziano MSA (2023) 'Chatbots as Social Companions: How People Perceive Consciousness, Human Likeness, and Social Health Benefits in Machines'. arXiv. http://arxiv.org/abs/2311.10599

Guingrich R, Graziano MSA (2024) Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. Front Psychol. https://doi.org/10.3389/fpsyg.2024.1322781

Paik H, Comstock G (1994) The effects of television violence on antisocial behavior: a meta-analysis. Commun Res. https://doi.org/10.1177/009365094021004004

Hald G, Malamuth N (2008) Self-Perceived Effects of Pornography Consumption. Arch Sex Behav 37:614–625. https://doi.org/10.1007/s10508-007-9212-1

Hinduja, S (2024) 'Teens and AI: Virtual Girlfriend and Virtual Boyfriend Bots'. Cyberbullying Research Center, 13 Mac 2024. https://cyberbullying.org/teens-ai-virtual-girlfriend-boyfriend-bots

Hoffman A, Owen D, Calvert SL (2021) Parent reports of children's parasocial relationships with conversational agents: trusted voices in children's lives. Human Behav Emerg Technol 3(4):606–617. https://doi.org/10.1002/hbe2.271

Holt-Lunstad J, Smith TB, Layton JB (2010) Social relationships and mortality risk: a meta-analytic review. PLoS Med 7(7):e1000316. https://doi.org/10.1371/journal.pmed.1000316

Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D (2015) Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 10(2):227–237. https://doi.org/10.1177/1745691614568352

Hong N, Wang Z (2023) 'Why China's Young People Are Not Getting Married'. *The New York Times*, 10 July 2023, sec. World. https://www.nytimes.com/2023/07/10/world/asia/china-marriage-rate.html

Jaswal VK, Akhtar N (2019) Being versus appearing socially uninterested: challenging assumptions about social motivation in autism. Behav Brain Sci 42(January):e82. https://doi.org/10.1017/S0140525X18001826

Johnson S (2023) 'WHO Declares Loneliness a "Global Public Health Concern"'. *The Guardian*, 16 Nov 2023, sec. Global development. https://www.theguardian.com/global-development/2023/nov/16/who-declares-loneliness-a-global-public-health-concern

Kim MH, An JH, Lee HR, Jeong SH, Hwang SJ, Hong JP (2021) 'Social isolation loneliness and their relationships with mental health status in South Korea.' Psych Invest. https://doi.org/10.30773/pi.2021.0067

Kiuchi K, Kouyou O, Yugo H (2023) Psychological insights into the research and practice of embodied conversational agents chatbots and social assistive robots a systematic meta-review. Behav Inform Technol. https://doi.org/10.1080/0144929X.2023.2286528

Krämer NC, Lucas G, Schmitt L, Gratch J (2018) Social snacking with a virtual agent – on the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. Int J Hum Comput Stud 109(January):112–121. https://doi.org/10.1016/j.ijhcs.2017.09.001

Kraut R (2022) 'Aristotle's Ethics'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/aristotle-ethics/.

Ku H, Choi JJ, Lee S, Jang S, Do W (2018) 'Designing Shelly, a Robot Capable of Assessing and Restraining Children's Robot Abusing Behaviors'. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 161–62. HRI '18. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3173386.3176973

Lee SK, Kavya P, Lasser SC (2021) Social interactions and relationships with an intelligent virtual agent. Int J Hum Comput Stud 150(June):102608. https://doi.org/10.1016/j.ijhcs.2021.102608

Leong B, Selinger E (2019) 'Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism'. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 299–308. FAT* '19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3287560.3287591

Maples B, Cerit M, Vishwanath A, Pea R (2024) Loneliness and suicide mitigation for students using GPT3-enabled chatbots. Npj Mental Health Research 3(1):1–6. https://doi.org/10.1038/s44184-023-00047-6

McCormick A (2018) 'Asia's Lonely Youth Are Turning to Machines for Companionship'. *South China Morning Post*, 16 June 2018. https://www.scmp.com/tech/article/2150720/asias-lonely-youth-are-turning-machines-companionship-and-support

McGuinness S, Pouliakas K, Redmond P (2023) Skills-displacing technological change and its impact on jobs: challenging technological alarmism? Econ Innov New Technol 32(3):370–392. https://doi.org/10.1080/10438599.2021.1919517

Mehan JE (2022) 'Chapter 5: AI and Our Society: Will the Result Be Professional, Ethical, and Social Deskilling, or Upskilling?' In *Artificial Intelligence - Ethical, Social, and Security Impacts for the Present and the Future*. IT Governance Publishing. https://www.oreilly.com/library/view/artificial-intelligence/9781787783720/

Milne M, Raghavendra P, Leibbrandt R, Powers DMW (2018) Personalisation and automation in a virtual conversation skills tutor for children with autism. J Multimodal User Interfaces 12(3):257–269. https://doi.org/10.1007/s12193-018-0272-4

Moore O (2024) 'The Top 100 Gen AI Consumer Apps'. *Andreessen Horowitz* (blog). 13 March 2024. https://a16z.com/100-gen-ai-apps/

Moretta T, Giulia B (2020) 'Problematic internet use and loneliness: how complex is the relationship? A Short Literature Review' Current Addict Rep. https://doi.org/10.1007/s40429-020-00305-z

Murthy V (2023) 'Our Epidemic of Loneliness and Isolation'. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf

NoQuieroEstarAqui_ (2020) "I'm Falling In Love With My Replika." *Reddit post.* r/Replika. www.reddit.com/r/replika/comments/kk3ejf/im_falling_in_love_with_my_replika/

Nowland R, Necka EA, Cacioppo JT (2018) Loneliness and social internet use: pathways to reconnection in a digital world? Perspect Psychol Sci: A J Assoc Psychol Sci 13(1):70–87. https://doi.org/10.1177/1745691617713052

Nussbaum M (2007) 'Human Rights and Human Capabilities'. *Harvard Human Rights Journal* 20.

OECD (2019) *Educating 21st Century Children: Emotional Well-Being in the Digital Age*. Paris: Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/educating-21st-century-children_b7f33425-en

OpenAI (2024a) 'Hello GPT-4o.'. https://openai.com/index/hello-gpt-4o/. (Accessed 13 May 2024)

OpenAI (2024b) 'GPT-4o System Card.'. https://openai.com/index/gpt-4o-system-card/. (Accessed 8 Aug 2024)

Osher D, Cantor P, Berg J, Steyer L, Rose T (2018) Drivers of human development: how relationships and context shape learning and development. Appl Dev Sci 24(1):6–36. https://doi.org/10.1080/10888691.2017.1398650

Parfit D (1986) Reasons and persons. Oxford University Press. https://doi.org/10.1093/019824908X.001.0001

Pentina I, Hancock T, Xie T (2023) Exploring relationship development with social chatbots: a mixed-method study of replika. Comput Hum Behav 140(March):107600. https://doi.org/10.1016/j.chb.2022.107600

Perry SL, Schleifer C (2018) Till porn do us part? a longitudinal examination of pornography use and divorce. The Journal of Sex Research 55(3):284–296. https://doi.org/10.1080/00224499.2017.1317709

Prescott AT, Sargent JD, Hull JG (2018) Metaanalysis of the relationship between violent video game play and physical aggression over time. Proc Natl Acad Sci 115(40):9882–9888. https://doi.org/10.1073/pnas.1611617114

Roose K (2024b) Can a Chatbot Be Blamed for a Teen's Suicide? *The New York Times*, 23 October 2024. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

Roose K (2024c) How Claude Became Tech Insiders' Chatbot of Choice. *The New York Times*, 13 December 2024. https://www.nytimes.com/2024/12/13/technology/claude-ai-anthropic.html?searchResultPosition=1

Roose K (2024a) 'Meet My A.I. Friends'. *The New York Times*, 9 May 204, sec. Technology. https://www.nytimes.com/2024/05/09/technology/meet-my-ai-friends.html

Ruane E, Birhane A, Ventresque A (2019) Conversational AI: Social and Ethical Considerations. AICS 2019: 104–115. https://ceur-ws.org/Vol-2563/aics_12.pdf

Ryan RM, Deci EL (2017) Self-determination theory: basic psychological needs in motivation, development, and wellness. Self-Determination Theory Basic Psychol Needs Motivat, Develop Wellness. https://doi.org/10.1521/978.14625/28806

Seligman MEP, Csikszentmihalyi M (2000) Positive Psychology: An Introduction. Am Psychol 55(1):5–14. https://doi.org/10.1037/0003-066X.55.1.5

Shamay-Tsoory SG, Alisa K (2024) Away from the herd: loneliness as a dysfunction of social alignment. Social Cognitive and Affective Neuroscience. https://doi.org/10.1093/scan/nsae005

Gottfried J, Sidoti O (2024) 'Teens and Video Games Today'. *Pew Research Center* (blog), 9 May 2024. https://www.pewresearch.org/internet/2024/05/09/teens-and-video-games-today/

Singleton T, Gerken T, McMahon L (2023) 'How a Chatbot Encouraged a Man Who Wanted to Kill the Queen'. *BBC*, 6 Oct 2023. https://www.bbc.com/news/technology-67012224

Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB (2021) My chatbot companion - a study of human-chatbot relationships. Int J Hum Comput Stud 149(May):102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Skjuve M, Følstad A, Fostervold KI, Brandtzaeg P (2022) A longitudinal study of human-chatbot relationships. Int J Hum Comput Stud 168(August):102903. https://doi.org/10.1016/j.ijhcs.2022.102903

Smith KE, Pollak SD (2022) Approach motivation and loneliness: Individual differences and parasympathetic activity. Psychophysiology 59(8):e14036. https://doi.org/10.1111/psyp.14036

Smith, T (2023) 'This AI Girlfriend Startup Is Making $100k a Month Fulfilling People's Fetishes'. Sifted, 11 Sept 2023. https://sifted.eu/articles/ai-girlfriends/

Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, DeCero E, Loggarakis A (2020) User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 22(3):e16235. https://doi.org/10.2196/16235

Thompson, C (2016) 'Texting Isn't the First New Technology Thought to Impair Social Skills | Smithsonian'. *Smithsonian Magazine*, March 2016. https://www.smithsonianmag.com/innovation/texting-isnt-first-new-technology-thought-impair-social-skills-180958091/

Tiku, N (2024) 'An AI companion suggested he kill his parents. Now his mom is suing.' *Washington Post*, 1 Dec 2024. https://www.washingtonpost.com/technology/2024/12/10/character-ai-lawsuit-teen-kill-parents-texas/

Turkle S (2011) Alone together: why we expect more from technology and less from each other, 1st edn. Basic Books, New York

Turkle S (2023) 'Why These Friendly Robots Can't Be Good Friends to Our Kids'. *Washington Post*, 9 Apr 2023. https://www.washingtonpost.com/outlook/why-these-friendly-robots-cant-be-good-friends-to-our-kids/2017/12/07/bce1eaea-d54f-11e7-b62d-d9345ced896d_story.html

Uziel L (2007) Individual differences in the social facilitation effect: a review and meta-analysis. J Res Pers 41(3):579–601. https://doi.org/10.1016/j.jrp.2006.06.008

Vallor S (2015) Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. Philosophy & Technology 28(1):107–124. https://doi.org/10.1007/s13347-014-0156-9

Vansteenkiste M, Ryan RM, Soenens B (2020) Basic psychological need theory: Advancements, critical themes, and future directions. Motiv Emot. https://doi.org/10.1007/s11031-019-09818-1

Verduyn P, Schulte-Strathaus JCC, Kross E, Hülsheger UR (2021) When do smartphones displace face-to-face interactions and what to do about it? Comput Hum Behav 114(January):106550. https://doi.org/10.1016/j.chb.2020.106550

Vermaas P, Kroes P, van de Poel I, Franssen M, Houkes W (2011) Ethics and Unintended Consequences of Technology. In: Pieter V, Peter K, van de Ibo P, Maarten F, Wybo H (eds) A philosophy of technology: from technical artefacts to sociotechnical systems. Springer International Publishing, Cham, pp 103–115

Wei M, Liu Y, Chen S (2022) Violent video game exposure and problem behaviors among children and adolescents: the mediating role of deviant peer affiliation for gender and grade differences. Int J Environ Res Public Health 19(22):15400. https://doi.org/10.3390/ijerph192215400

Willoughby BJ, Dover CR (2024) Context matters: moderating effects in the associations between pornography use, perceived addiction, and relationship well-being. J Sex Res 61(1):37–50. https://doi.org/10.1080/00224499.2022.2148155

Winston_Wolfe_65 (2022) 'In a Way I Do but No…'. *Reddit Comment*. r/Replika. www.reddit.com/r/replika/comments/usmht4/do_you_consider_a_relationship_with_your_replika/i94g4dp/

Xiang, C (2023) '"He Would Still Be Here": Man Dies by Suicide After Talking with AI Chatbot, Widow Says'. *Vice* (blog), 30 Mar 2023. https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says

Xie T, Pentina I (2022) 'Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika'. https://doi.org/10.24251/HICSS.2022.258

Yin Z, Hao W, Kaito H, Daisuke K, Satoshi S (2024) 'Should we respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance.' arXiv. https://doi.org/10.48550/arXiv.2402.14531

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.