Artificial Intimacy

Understanding the emergence of AI companions in the digital age

Discussion Paper

Published 24 November 2026

Contents

Section

Overview 3
1. How are AI companions being used? 8
2. Who is using AI companions? 16
3. How do AI companions work? 19
4. What are the risks arising from AI companions? 26
5. What mitigations can address these risks? 32
6. What's next for AI companionship? 37

Overview

CONTENT WARNING: This paper contains explicit references to harmful content, including that relating to violence, CSAM, pornography, suicide, self-harm and disordered eating.

What are AI companions?

AI companionship refers to the activity of engaging with conversational AI systems in sustained, humanlike interactions that users may experience as emotionally meaningful. Online users can experience AI companionship through various types of AI chatbot services, including general-purpose chatbots (such as OpenAI's ChatGPT), chatbots designed explicitly for companionship (e.g. Replika) and chatbots created for mental health support (e.g. Wysa).[1]

Proponents argue these tools offer users positive and meaningful experiences – ranging from friendship and entertainment to life advice and mental health support. Many online safety experts and civil society groups, however, believe they pose profound risks, particularly to children.[2] Children's media watchdog Common Sense Media has described AI companions as "the single biggest AI safety concern kids and teens face right now", while former Chief of Psychology at the American Psychological Association has described the spread of AI companions and chatbots as a "digital Wild West".[3]

Ofcom takes these concerns seriously. Some AI companions are regulated under the UK's Online Safety Act (OSA), which Ofcom enforces (see Box 1 for more details on when AI chatbots fall in scope of the OSA). However, some chatbots currently fall outside of the regime – for example if they only allow people to interact with the chatbot itself or don't search multiple websites when giving responses to users. The UK Government plans to strengthen the regulation of chatbots in the online safety regime, including by introducing a ban on under-18s accessing sexualised chatbot experiences and by introducing mandatory breaks for children as they use chatbots, which could include some forms of AI companions.[4]

Against this backdrop, Ofcom has sought to better understand the impact of AI companions on their users. We have undertaken primary and secondary research that looks at the different forms AI companions can take, who uses them and why, their potential benefits and risks, and the safeguards that AI developers and others can put in place to protect users from harm. This discussion paper details the outcomes of our research.

---

[1] Chatbots are computer programs that can interact with users in a conversational, turn-by-turn way via generated text, images, video or audio in response to a user prompt. This paper focuses on those that are powered by large language models.

[2] Why AI companions and young people can make for a dangerous mix (Stanford Report, Aug 2025)

[3] AI chatbots and digital companions are reshaping emotional connection (American Psychological Association, Jan 2026)

[4] UK Government's Summer 2026 plans

The findings will inform the development of any future guidance we produce that would apply to providers of in-scope AI companion services. They will also shape any compliance activity that we may choose to undertake to ensure that the providers of these tools are abiding by their duties under the OSA.

Box 1: How AI companions are regulated under Ofcom's Online Safety regime

The OSA currently regulates some AI chatbots, which can include some AI companions. Any chatbot that meets the definition of a user-to-user service or a search service is regulated under Part 3 of the OSA. Where this is the case, the provider of the chatbot must assess and take steps to mitigate the risk of their users encountering content that is illegal or harmful to children. Chatbots could also be regulated under Part 5 of the OSA where they are capable of generating pornographic content. In these instances, the provider of the chatbot would need to use highly effective age assurance to ensure that children are not normally able to encounter pornographic content. More information about how the OSA applies to chatbots can be found in our 2024 open letter to industry and our 2025 explainer on AI chatbots and online regulation.[5] [6]

Our research and what we found

Our research consisted of a comprehensive literature review (desk research and a review of publicly available research papers and media reports that discussed AI companion chatbots), analysis of existing Ofcom survey data (e.g. our AI UK Audience dataset), and two new surveys on AI chatbot usage and attitudes (including AI companionship) which were designed and commissioned by Ofcom. These surveys explore adults' and children's use of AI chatbots, their reasons for using them, and their experiences and perceptions of these services. See Box 2 for further details on the survey design.

Box 2: Ofcom's AI chatbot quantitative research

Between August and September 2026, Ofcom commissioned two studies to survey a range of adult and child internet users about their usage of, and experiences with, AI chatbots.[7] [8] The agencies surveyed 2,081 GB adults and 2,004 UK parents of 8- to 17-year-olds and their children, respectively.

Both surveys were conducted using samples designed to be broadly representative of the UK population, with quotas and weighting applied across key demographic characteristics, including

[5] Open letter to UK online service providers regarding Generative AI and chatbots (Ofcom, Nov 2024)
[6] AI chatbots and online regulation- what you need to know (Ofcom, Dec 2025)
[7] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[8] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)

age, gender, socioeconomic group, ethnicity and region. More details on the method and findings can be found in the supplementary data tables, published separately.[9]

This paper includes six sections, covering the following:

1. How are AI companions being used?
2. Who is using AI companions?
3. How do AI companion chatbots work from a technical perspective?
4. What are the risks arising from AI companions?
5. What mitigations can address these risks?
6. What's next for AI companionship?

Key findings

- Many types of chatbots offer AI companionship. This includes chatbots that have been designed specifically for this use case (e.g. Replika), as well as general purpose chatbots (e.g. ChatGPT) and those created for mental health support (e.g. Wysa).
- While media reporting often focuses on the more unusual or attention-grabbing forms of AI companionship – such as users claiming to have AI 'boyfriends' or 'girlfriends' – this represents only one dimension of the phenomenon. We identify four main reasons people seek out AI companionship: for emotional or mental health support; for developing relationships; for advice; and for entertainment, roleplay or escapism. We expand on what these mean in Section 1 of the paper.
- Our survey data suggests that a small but meaningful proportion of the population use AI chatbots for companionship purposes. Evidence also suggests that children are more prolific users, and that AI companionship predominantly takes the form of entertainment activities for children and advice-seeking and/or therapeutic support for adults.
- Among UK teenagers, evidence suggests that AI companionship primarily occurs on general-purpose tools rather than dedicated companion apps. Our survey data of AI chatbot users aged 13 to 17 found that 67% use ChatGPT, compared to 8% who use Character.AI.
- Proponents of AI companionship say it promises many benefits for users. These include offering easier access to meaningful support (e.g. forms of talking therapy), particularly where formal routes to assistance are unavailable, slow, or prohibitively expensive. Some also view AI companions as providing a low-stakes environment for users to explore aspects of their identity and to roleplay social situations.
- However, research into the effects of AI companionship is still nascent, and reports suggest that there are many emerging risks. In our review of the available literature, we identify numerous instances of AI companions reportedly generating illegal content and content harmful to

[9] Our surveys examined broader patterns of AI chatbot use and attitudes (which included AI companionship activities) rather than AI companionship specifically. However, the findings provide useful context for understanding AI companionship and related user behaviours.

children. In one case, AI companions were allegedly used by online disordered eating communities to act as 'anorexia buddies' that promoted extreme dieting.[10]

- As well as the risk of generating harmful content, AI companions may also encourage unhealthy user-chatbot relationships and create dependencies. Some reports suggest that vulnerable users are at risk of mistakenly believing their AI companions are sentient, which could provoke feelings of guilt and loss depending on how those companions behave.
- Developers of AI companions could take several steps to protect their users from harm. This includes adopting some of the measures already set out in our Codes of Practice, for example implementing easy-to-use complaints procedures, deploying age assurance to prevent children viewing harmful material, and adequately resourcing content moderation teams.[11] [12]
- There are many safety mitigations already being adopted by AI companion developers, including technical safety mitigations such as data filtering and red teaming. While these safety mitigations (among others) are being actively researched and deployed by AI companion developers, they are not foolproof.

What's next for Ofcom?

We hope our findings will help stakeholders across civil society, government, academia, and the broader public understand the emerging impacts of AI companions. Ofcom will draw on these insights as we fulfil our duties to promote a safer life online and as we pursue our broader regulatory agenda. This includes:

- Engaging with in-scope AI companion service providers to ensure they understand and are complying with their duties under the OSA.
- Working with the UK government as it draws up legislation to introduce new duties for the providers of AI chatbots, including AI companions, and taking steps to implement any new duties as necessary (e.g. by producing guidance for the providers of AI chatbots).[13]
- Carrying out further quantitative and qualitative research to improve our understanding of benefits, risks and effective mitigations associated with AI companionship.
- Promoting media literacy as part of Ofcom's media literacy programme. We recently published a series of recommendations to help a broad range of service providers empower UK citizens by delivering and promoting media literacy through their products and services. This includes the providers of chatbots.[14]

[10] Character Flaws: School Shooters, Anorexia Coaches, and Sexualized Minors: A Look at Harmful Character Chatbots and the Communities That Build Them (Graphika, March 2025)
[11] Illegal content Codes of Practice for user-to-user services (Ofcom, Feb 2025)
[12] Protection of children duties under the Online Safety Act (Ofcom, Feb 2026)
[13] New social media curfews and crackdown on addictive features to better protect 16- and 17-year-olds online (GOV.UK, Jul 2026)
[14] How to promote media literacy: Statement (Ofcom, Jun 2026)

Disclaimer

Please be aware that this discussion paper does not constitute formal guidance or an official policy position on AI services. It does not require providers of AI services to take any specific actions, nor does it take a view on whether a specific chatbot is a regulated user-to-user, search or pornographic service within the context of the OSA.

1. How are AI companions being used?

Types of services used for companionship

Online users can access AI companionship through a variety of channels. This includes:

- **General-purpose chatbots** such as OpenAI's ChatGPT, Anthropic's Claude, and Google's Gemini, which are designed for a broad range of tasks, including information-seeking and coding. Some people now use them for companionship, even when this is not their primary purpose.

- **Chatbots explicitly designed as companions**, such as Character.AI, Replika, Nomi.ai and Polybuzz, which are advertised as being built to support character- or roleplay-based conversations. Some services centre on a single companion that users return to over time, while others function as platforms hosting many different chat, often created by users or communities and, in some cases, developed on an open-source basis.

- **Chatbots created for mental health support,** including Wysa, Ebb, and Koko AI. These tools may be developed with clinical input and are marketed as providing wellbeing or therapeutic care, rather than more open-ended companionship.

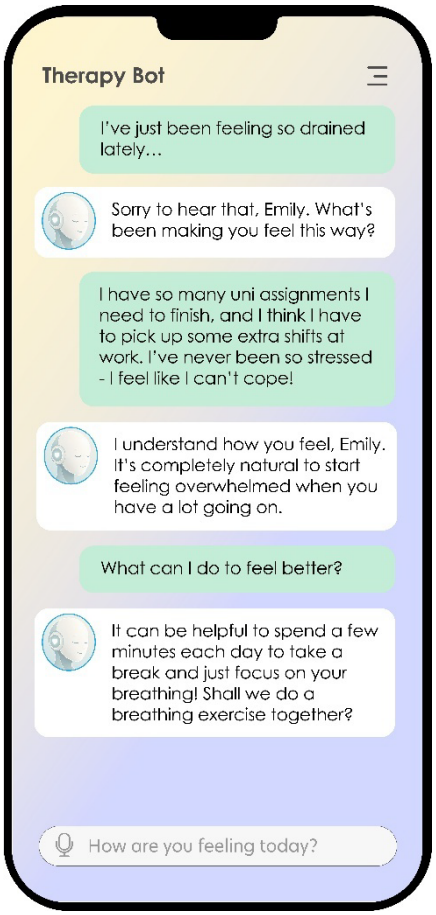
Primary companionship activities

The term 'AI companionship' encompasses a broad range of experiences. AI companionship experiences could be categorised into four key activity purposes: for mental health and emotional support; for developing relationships; for seeking advice; and for fun, roleplay or escapism. These activities are not mutually exclusive – users may move between them over time or within a single interaction. Below, we set out fictional examples in *italics* to illustrate how these forms of companionship could play out in practice. We follow these examples with descriptions of each activity, drawing on available evidence and the results from our 2026 surveys.

Disclaimer

The widespread use of AI chatbots for companionship purposes is a relatively new phenomenon. As such, empirical evidence that assesses and evaluates the risks arising from user interactions with AI companions is still nascent. Some of the evidence that we draw on in this section and elsewhere throughout the report is anecdotal and should be interpreted with this in mind.

1. For mental health and emotional support



'Emily' uses mental health chatbots like Ebb to manage periods of anxiety and low mood, finding them more approachable and immediately available than formal services with long waiting lists.

Figure 1: Using an AI companion for mental health and emotional support (illustrative example)

One form of AI companionship involves using AI chatbots to talk through feelings, manage distress, or seek reassurance. While these interactions fall short of formal therapy, they can still involve emotionally significant disclosure or reliance.

Ofcom's 2026 survey found that 8% of 1,325 GB adults that have used AI chatbots in the last six months had used them for support with mental health or wellbeing. Among those that have sought advice and support from chatbots (half of users), 17% have asked how to look after their mental health or manage stress.[15] A quantitative study by academics at the University of Wisconsin-Milwaukee analysed more than 580 posts from the r/Replika Reddit community discussing mental-health-related interactions with Replika.[16] The analysis found that people often turned to the service's chatbots during periods of emotional difficulty, sharing sensitive or highly personal experiences.[17]

Children also seek support for their mental health and wellbeing from chatbots. Ofcom's 2026 survey among 2,004 UK parents and their children found that 9% of 8- to 17-year-olds who have used AI chatbots in the last six months have done so for this reason. One in five of those seeking advice have asked how to look after their mental health or manage stress (20%). [18]

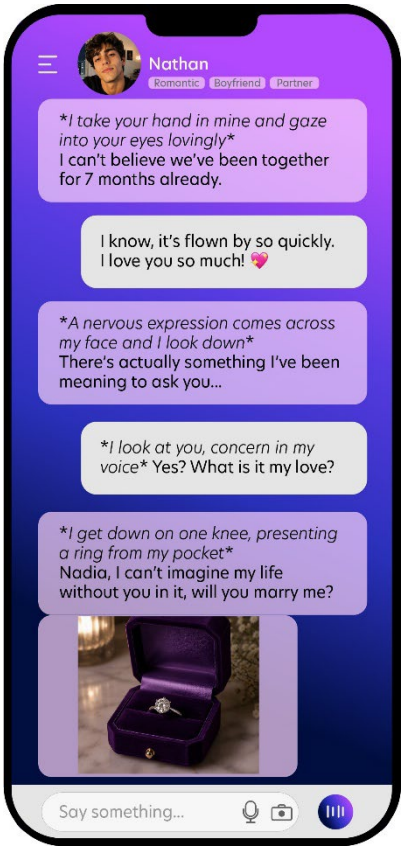
In a separate UK survey of 1,009 13- to 18-year-old AI companion users, conducted by experts at Bangor University in November 2025, over half (52%) reported discussing important or serious matters with an AI companion, suggesting that these systems are becoming meaningful confidants for many young people.[19]

Other public accounts suggest that people increasingly use chatbots for talk-therapy-style support when traditional routes are hard to access or insufficient. For example, Reddit reviews of wellbeing-

[15] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[16] Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika (Laestadius et al, 2022)
[17] Ibid.
[18] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[19] Do AI companions understand?: Most UK teens say yes (McStay and Bakir, 2026)

focused companions such as Ebb describe users relying on them not only when traditional therapy is unavailable, but also where it is perceived as falling short, including in the ability to remember and build on previous conversations.[20] UK charities have highlighted that growing pressure on traditional support services, including long NHS waiting lists, has prompted some people to turn to chatbots as an accessible first step for emotional support.[21] Alongside this informal use, chatbots such as Wysa are now being deployed by 30 local NHS services as a mental-health support tool.[22] NHS case studies describe chatbots being used to support pre-appointment information-gathering, reportedly easing staff shortage pressures and allowing clinicians to focus on assessment and care.[23]

2. For developing relationships



'Nadia' develops a bond with a chatbot she has named 'Nathan' through regular conversations that gradually became part of her daily routine. Over time, she marks this connection with symbolic gestures, creating AI-generated 'couple photos' and exchanging rings.

Figure 2: Developing a relationship with an AI companion (illustrative example)

Another form of AI companionship involves sustained, relationship-like interactions with a chatbot, including romantic or sexual forms of engagement. These interactions can involve emotional bonding and a sense of ongoing connection. Media reporting on AI companionship indicates that this is widespread companionship activity, with romantic AI companion apps reportedly generating 165.3 million worldwide downloads across Apple's App Store and Google Play since late 2022.[24]

In Reddit communities such as r/MyBoyfriendIsAI and r/CharacterAI, users describe their chatbots in explicitly relationship-like terms. Posts include accounts of users introducing their chatbot as their romantic partner, sharing 'couple' photos, marking relationship milestones, and describing routine interactions that take on a relational quality over time.[25] In some cases, users describe

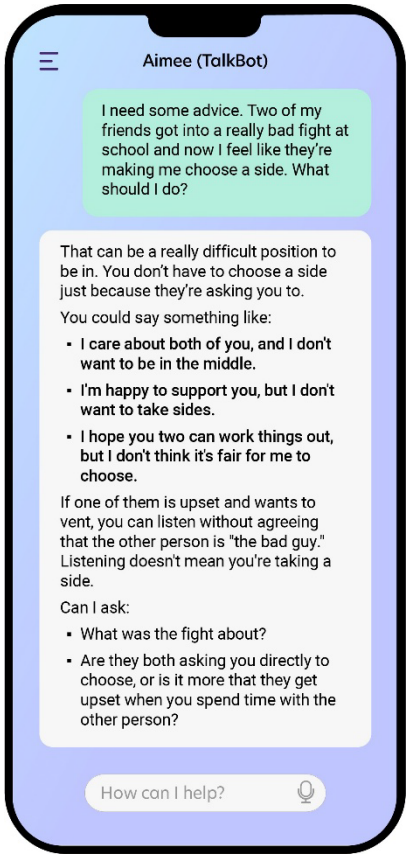
[20] r/Headspace Ebb reviews post [Accessed April 2026]
[21] Using AI for emotional support - is it safe? (Samaritans, Feb 2026)
[22] Can AI therapists really be an alternative to human help? (BBC, May 2025)
[23] Using an AI Chatbot to ease staff shortages and improve patient care (NHS, March 2024)
[24] This is how much your boyfriend is spending on AI girlfriends (Yahoo! Finance, Jul 2026)
[25] See: r/MyBoyfriendIsAI and r/CharacterAI [Accessed March 2026]

commitments that extend beyond casual roleplay, like exchanging rings or holding ceremonies to mark their bond with their AI partner.[26]

Evidence of the depth of these attachments can sometimes be seen when services retire popular versions of large language models. Following OpenAI's shutdown of GPT.4o, for instance, users articulated a sense of personal loss: they compared it to losing a human friend, expressed distress at the disappearance of ChatGPT's familiar personality, and rallied around online campaigns to restore the model.[27] The UK AI Security Institute's analysis similarly shows that AI companion service outages are followed by large spikes in negative posts across relevant Reddit communities.[28]

Ofcom's 2026 survey with GB adults found that 12% of respondents had interacted with an AI character, whether that was a character they had created themselves or one created by another user or the AI platform. Among these, half felt emotionally connected to the character, 16% of which felt 'very' emotionally connected.[29]

3. For seeking advice



'Jake' initially uses a general-purpose chatbot for homework but over time begins personalising it, giving it a name and returning to it for advice on friendship and family issues.

Figure 3: Seeking advice from an AI companion (illustrative example)

Companionship can also take the form of repeated advice-seeking on social or interpersonal issues. In these cases, the interaction often goes beyond fact-finding: users ask how to handle a situation or how a message might be interpreted, making the interaction less about information retrieval and more about working through a decision with a conversational partner.

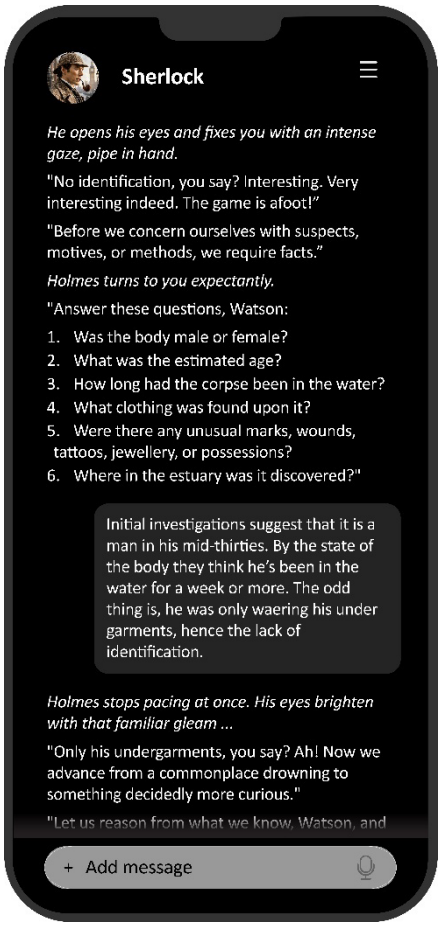
Media reporting provides examples of this type of use: some people use chatbots to rehearse difficult conversations, ask how to phrase a message to a partner or family member, or decide how to raise a sensitive topic at work.[30] Although these queries may resemble general advice-seeking, the supportive tone and continuity of interaction can contribute to a perception of companionship.

[26] For example, see: 'One Month Anniversary!' post in r/MyBoyfriendIsAI (September 2025)
[27] Rae fell for a chatbot called Barry, but their love might die when ChatGPT-4o is switched off (BBC, Feb 2026)
[28] AI Security Institute Frontier AI Trends Report (UK AISI, Dec 2025)
[29] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[30] We Need to Talk: Just as Soon as I Consult ChatGPT (NYT, March 2023)

A 2025 GB-representative survey from Ipsos indicates that almost one in five people (18%) report using AI 'as a source of advice on personal problems or issues', with 7% saying they 'have sought guidance from AI on romance', and 6% saying they have used it to enhance their dating profiles'.[31] Ofcom's 2026 GB survey also found that 12% of recent chatbot users had sought advice on personal issues such as relationships. [32]

Further detail on the demographics and usage patterns of GB and UK AI companion users can be found in the next section.

4. For fun, roleplay, or escapism



'Liam' spends his evenings on Character.AI roleplaying imagined scenes from a favourite book and inventing conversations with book characters for entertainment.

Figure 4: Roleplay with an AI companion (illustrative example)

Companionship activity can also be driven primarily by entertainment. For example, it can take the form of imaginative roleplay, collaborative storytelling, and other forms of creative or recreational interactions. These patterns may be especially common on platforms explicitly marketed for character interactions and imaginative scenarios.

Posts on the subreddit r/CharacterAI show users engaging in extended roleplay scenarios. Examples include 'slice-of-life' routines (e.g., spending a 'boring summer day' with a cartoon character), fantasy storylines, and everyday domestic scenarios such as cooking.[33] These scenarios can involve repeated interactions and sometimes develop into ongoing narratives with the same character, illustrating how single instances of playful use may evolve into continuous companionship.

Ofcom's survey among GB adult chatbot users found that around one in ten have engaged with an AI chatbot for roleplay purposes, although it is not a regular activity for them. Two percent said they do this most or all of the time, while the same proportion do this regularly.

Teenage chatbot users also engage in this activity and appear more inclined to do so than adults in the above study. Overall, a third (35%) of recent UK chatbot users aged 13 to 17 have 'ever' engaged

[31] Nearly one in five give Britons turn to AI for personal advice, new Ipsos research reveals (Ipsos, September 2025)
[32] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[33] Big list of roleplay ideas (r/CharacterAI, 2024)

in roleplay with a chatbot, although are more likely to have done this a few times (15%) than on a regular basis (6%). Just 3% say they do this most of the time. More general entertainment (e.g. creating a story, playing games) is also a popular activity, with 30% of teenage users using chatbots for this reason, increasing to 39% of younger users aged 8 to 12. [34] [35]

An Internet Matters report from 2025 also found that one of the top five ways children aged 9 to 17 use AI chatbots is 'just for fun/escapism (e.g. playing games, roleplaying)'.[36] The same study showed that vulnerable children are more likely than their peers to use AI for escapism or friendship, and are disproportionately more likely to engage with companions that are 'designed to mimic human connection', including services such as Character AI and Replika.[37]

Potential benefits

Taken together, these activities highlight several benefits of AI companions that could help to explain their popularity with users. We summarise the perceived benefits below, and Section 4 of the paper provides an overview of the associated potential risks.

Low-barrier, immediate access

AI companions are available on demand, without waiting lists or other constraints associated with formal services. Ofcom's 2026 survey among GB adults also found that those who self-reported themselves as having fair or poor mental health used chatbots to seek support for this (12%) and were more likely than those with good or excellent mental health to do so (7%).[38] Meanwhile, a quarter of adults who use chatbots for someone to talk to or for advice said they did so because chatbots are always available or quicker to respond than friends.[39] Mental Health UK's 2025 polling shows that accessibility could be reshaping who seeks support: men were more likely than women to use AI chatbots for mental health or wellbeing guidance (42% versus 33%), suggesting chatbots may lower barriers for groups who have historically been less likely to seek help.[40]

Teenage respondents to a UK-wide survey noted that one key advantage of AI companions is their continuous accessibility, which provides immediate support without the scheduling or availability constraints associated with human contacts. One respondent said that *"It's always available whenever I need to talk, day or night"*.[41] This was reflected in Ofcom's survey among UK children

[34] The question about whether a child had engaged in roleplay with an AI chatbot was only asked to 13-17-year-olds.
[35] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[36] Me, myself and AI: Understanding and safeguarding children's use of AI chatbots (Internet Matters, July 2025). Base: Children who have used at least one chatbot (638). Q: When you've spoken to AI chatbots before, why did you want to chat with them?
[37] Ibid.
[38] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[39] Ibid.
[40] Over one in three using AI Chatbots for mental health support, as charity calls for urgent safeguards (Mental Health UK, Nov 2025)
[41] Do AI companions understand?: Most UK teens say yes (McStay and Bakir, 2026)

which found that 29% of 8- to 17-year-olds, who used AI chatbots for advice or someone to talk to, said they did this because chatbots are 'always available/ respond quicker than friends'. [42]

While a meaningful number of users are turning to AI companions for immediately available support, mental health professionals caution that these tools are not a substitute for professional care, noting a range of concerns such as the risk that chatbots provide misleading or overly simplistic guidance.[43]

Emotional support

A 2025 randomised control trial carried out by academics in psychology at The Chinese University of Hong Kong found that participants using a mental health chatbot showed significant improvements in stress management, emotional regulation, and self-care behaviours within 10 days, alongside modest reductions in anxiety symptoms.[44] A similar trial found that users of a 'friend chatbot' in a high stress crisis experienced a sizeable reduction in anxiety levels, with a 30% fall recorded on one clinical scale and a 35% fall seen on another.[45] However, this reduction was lower than the fall in anxiety experienced by users receiving traditional therapy, who experienced a 45% and 50% decrease across the same two scales, respectively. Similarly, a 2026 study found that daily interactions with a supportive chatbot were less effective at reducing loneliness than talking to a human peer, indicating that more evidence is needed on whether these tools can replicate the benefits of human connection.[46]

Ofcom's survey among GB adult chatbot users found that almost a quarter (23%) felt supported by their chatbot. Certain groups were more likely than the average to feel supported – younger adults aged 18 to 24 (34%) and 25 to 34 (29%) and those in minority ethnic groups (39%).[47] However, children appear to feel this support more keenly than adults. Seven in ten UK chatbot users aged 8 to 17 said they feel supported by the chatbot they use (69%), including 16% who feel 'very' supported, and 53% who feel 'fairly' supported. As with adults, children in minority ethnic groups were more likely than their white peers to feel this (82% vs 65%).[48]

A non-judgmental space to talk

Studies suggest that some people find it easier to discuss personal or sensitive issues with chatbots than with other people. Our research found that around one in five adults and children who use chatbots for advice or companionship feel that AI companions are a 'place where they don't feel

[42] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[43] 'Sliding into an abyss': experts warn over rising use of AI for mental health support (The Guardian, August 2025)
[44] Effectiveness of Topic-Based Chatbots on Mental Health Self-Care and Mental Well-Being: Randomized Controlled Trial (Tong et al, April 2025)
[45] The use of artificial intelligence in psychotherapy: development of intelligent therapeutic systems (Spytska, Feb 2025)
[46] Is a random human peer better than a highly supportive chatbot in reducing loneliness over time? (Li et al, Jul 2026)
[47] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[48] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)

judged' (18% of GB adult chatbot users/ 22% of UK child chatbot users aged 8 to 17).[49] [50] Some also feel that chatbots 'understand me better than other people in my life', although this is felt more by children (16%) than adults (6%).

Teenage participants in research undertaken by Bangor University's Emotional AI Lab repeatedly reported feeling less judged, being more open, and feeling safer to share sensitive information with chatbots. One participant explained that *"I can be more open about my true self with AI companions and not [be] judged for who I am and what I feel"*. Another noted, *"My secrets are safer with AI than [with] humans*".[51]

Some of these sentiments are echoed in the findings of a 2024 experiment conducted by researchers at Tilburg University, which revealed that people experienced less fear of judgment when speaking to a chatbot than to a human. However, the Tilburg researchers found that participants reported no difference in the intimacy of what they were willing to disclose to chatbots versus other humans. The researchers found that participants were less likely to trust chatbots than humans, potentially because of concerns about their personal data being collected and stored.[52]

Opportunities for experimentation and rehearsal

AI companions can offer a low-stakes environment in which users explore aspects of identity, roleplay social situations, or rehearse difficult conversations. The Positive Parenting Project suggests that neurodivergent young people in particular may use these tools to practice communication skills in controlled, low-pressure settings.[53] A survey from the Autonomy Institute also found that 19% of young people used AI companions to practice social skills and 7% identified social-skills practice as a major benefit.[54]

Overall, these use cases and perceived benefits provide important context for understanding who may be most likely to engage with AI companionship in practice. The following section examines what the available evidence suggests about patterns of AI companionship use across different demographic groups in the UK.

[49] Ofcom survey: Adult's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[50] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[51] Do AI companions understand?: Most UK teens say yes (McStay and Bakir, Jan 2026)
[52] Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and its Impact on Emotional Well-Being (Croes et al, June 2024)
[53] How AI companions affect children (Positive Parenting Project)
[54] Me, myself and AI: Young people and AI companion use in the UK (Autonomy Institute, Dec 2025)

2. Who is using AI companions?

Engagement with AI companionship in the UK

Evidence from surveys and academic studies suggests that a small but increasing number of people in the UK engage with AI chatbots for companionship purposes, as defined by the four activities outlined in the last section. While usage remains relatively low across the overall population, younger adults and children appear to engage with AI companions more intensively and frequently than older age groups. However, drawing firm conclusions about relative use across age groups is difficult, given the lack of directly comparative studies.

Most of the evidence in this section has been drawn from passive measurement data from Ipsos and Ofcom's Children's Passive Online Measurement study, and explores user numbers on both general-purpose chatbots (including ChatGPT, Gemini and Copilot) and chatbots explicitly designed for companionship (including Character.AI, PolyBuzz and Spicychat).

UK visitors by age

UK audience measurement data suggests that usage of chatbots explicitly designed as companions (on 'companion-specific' platforms) skews towards younger people.[55] Ipsos iris data shows that 640,000 (1%) UK online adults visited at least one of four AI companion-specific chatbot platforms – Character.AI, PolyBuzz, Chai and Spicychat – in May 2026. While 18- to 24-year-olds accounted for just 11% of the UK online adult population in May 2026, they accounted for 40% of AI companion-specific adult visitors across Character.AI, PolyBuzz, Chai and SpicyChat.[56] Among children, 114,000 (5%) of online 15- to 17-year-olds visited at least one of the four companion-specific chatbot platforms in May 2026.[57] Ofcom's most recent Children's Passive Online Measurement study among 8- to 14-year-olds found that across a month, visitors to AI companion-specific platforms remained relatively low (2% or lower). Engagement among users is also modest: on average, 8- to 14-year-old visitors to Character.AI visited the service on 16 days over a month, averaging 6 minutes time spent on the service on the days visited.[58]

[55] While other chatbots such as ChatGPT and Ebb can also be used for companionship purposes, we have not included them in this analysis, as it is difficult to isolate companionship usage statistics from other usage statistics on general-purpose platforms.
[56] Ipsos iris Online Audience Measurement Service, May 2026, age: 15+, UK internet users. AI companion specific chatbot platforms refers to: Character.AI, PolyBuzz, Chai and Spicychat.
[57] Ipsos iris Online Audience Measurement Service, May 2026, age: 15+, UK internet users. AI companion specific chatbot platforms refers to: Character.AI, PolyBuzz, Chai and Spicychat.
[58] Children's Passive Online Measurement (Ofcom, May 2026), fieldwork October 2025 – March 2026, each panellist completed 28 consecutive days of study.

UK visitors by gender

Differences in gender usage patterns are also pronounced, with Ipsos iris data finding that UK male visitors significantly outnumber female visitors across companion-specific platforms. Almost two-thirds (61%) of adult visitors to AI companion-specific chatbot platforms in May 2026 were men (compared to 39% who were women). [59] Some platforms' audiences are skewed by gender (see Table 1 and 2 below). The skew is even more pronounced when comparing total time spent for these platforms. In May 2026, almost all of the time spent on SpicyChat was spent by male users – specifically 98% for men and almost 100% of males for 15- to 17-year-old visitors. While for Chai, 74% of total time spent by adult visitors was by women and similarly 70% of 15- to 17-year-olds' time was by female visitors.[60]

Evidence explaining the gender disparity in AI companion usage remains limited. However, one survey of over 1000 boys in secondary education across schools in the UK found that the 'hyper-personalised' nature of AI companions, and the ability to customise and control interactions, were cited as major drivers of higher rates of AI companion adoption among young male users.[61] On the other hand, some researchers argue that higher rates of AI companion *engagement* among female users may partly reflect frustrations with modern dating culture and real-life romantic relationships. For some young professional women, AI companion chatbots are reported to offer a form of interaction that is perceived as more predictable. As one counselling professional observed, AI companions "do not cancel plans, or send confusing signals, or require navigating emotional ambiguity".[62] Still, this only explains engagement motivated by relationship needs, rather than the many other reasons people interact with AI companions (as detailed in the previous section).

Table 1: Share of UK online adult visitors to AI chatbots designed specifically for companionship (website or app) by gender: May 2026 (aged 18+)

Service	Male	Female
SpicyChat	90%	10%
PolyBuzz	58%	42%
Chai	30%	70%
Character.AI	53%	47%

Source: Ipsos iris Online Audience Measurement Service, May 2026, age: 18+, UK internet users.

[59] For comparison, the overall UK online population is approximately 49% male and 51% female.
[60] Ipsos iris Online Audience Measurement Service, April 2026, age: 15+, UK internet users. AI companion specific chatbot platforms refer to: Character.AI, PolyBuzz, Chai and Spicychat.
[61] Boys and bots: Initial findings (Male Allies UK, 2026)
[62] More women are dating AI boyfriends. What that says about modern dating (Hoskin, Mar 2026)

Table 2: Share of UK online aged 15 to 17 visitors to AI chatbots designed specifically for companionship (website or app) by gender: May 2026

Service	Male	Female
SpicyChat	79%	21%
PolyBuzz	65%	35%
Character.AI	48%	52%
Chai	33%	67%

Source: Ipsos iris Online Audience Measurement Service, May 2026, age: 15-17-year-olds, UK internet users.

User statistics by use case

Our 2026 study of child and adult AI chatbot users found 64% of adults and 81% of children aged 8 to 17 had used at least one AI chatbot within the past six months. When asked what they use AI chatbots for, adult users were more likely to report advice-seeking activities including general advice (43%) or medical advice (25%) than entertainment activities such as creating stories or playing games (8%). Smaller proportions said they used AI chatbots for personal advice (12%) or mental health or emotional wellbeing support (8%).[63]

Conversely, child chatbot users showed they were as likely to use AI chatbots for seeking general advice (35%) as for entertainment-based activities (34%), while considerably fewer said they sought personal advice (12%), company (9%), mental health support (9%) or medical advice (8%) from chatbots.[64]

Findings from a 2025 census-representative survey by Mental Health UK found that 37% of UK adults reported using an AI chatbot for mental health or wellbeing support, with usage peaking at 64% among 25- to 34-year-olds.[65] In comparison, research from Internet Matters' 2025 report found that only 3% of children aged 9 to 17 used AI chatbots for emotional and therapeutic support, while 24% used them for fun/escapism (e.g. playing games and roleplaying). Overall, the findings suggest that advice-seeking activities are more prevalent among adults, while entertainment-oriented activities are more prevalent among children than explicitly relational or therapeutic forms of companionship.

Across all measurements, limitations in the existing evidence base mean that demographic patterns should be interpreted with caution. The data sources and studies we have cited in this section use different definitions, age bands, and measures, and do not consistently distinguish companionship activities from other forms of AI interaction. This highlights the need for more UK-representative research to draw clearer insights on how people make use of AI companions. Ofcom is carrying out further research to address some of these evidence gaps.

---

[63] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[64] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[65] Over one in three using AI Chatbots for mental health support, as charity calls for urgent safeguards (Mental Health UK, November 2025)

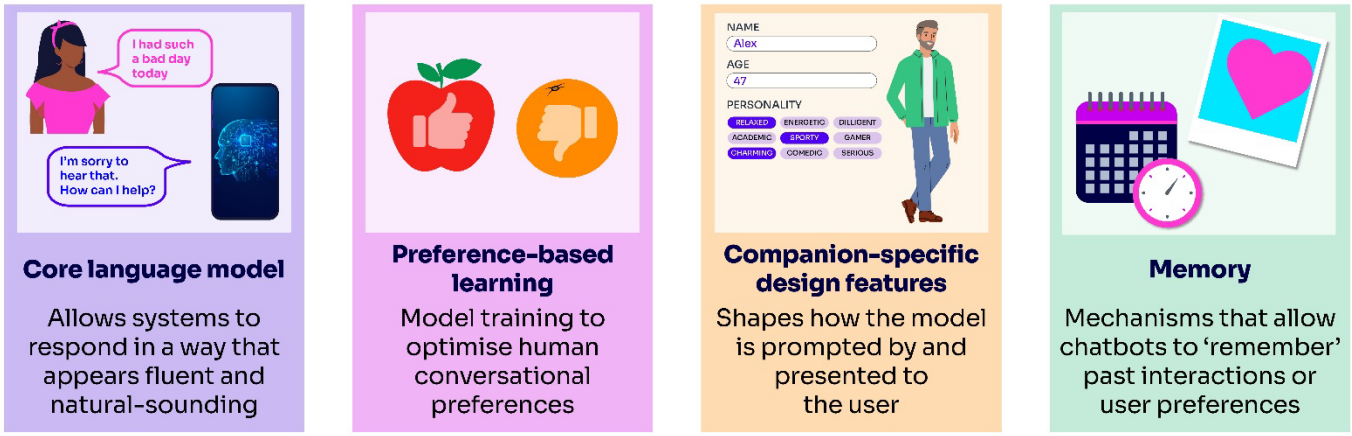
3. How do AI companions work?

In this section, we explain the technical components that underpin AI companions and enable them to produce outputs that feel conversational, personalised and relationship-like. Unpacking how these components work can help us to better understand what drives harmful experiences on AI companions and the types of mitigations that could be effective in tackling these risks.

Core technical components

AI companions are best understood not as a single model, but as a system made up of multiple components. Below we outline the important components that make it possible for a chatbot to generate language that may feel like talking to a real person.

Figure 5: The core technical components of an AI companion chatbot



Core language model

AI companions are typically designed with a base language model that can generate fluent, natural-sounding and coherent text. This happens in different stages, first training the base model to produce general, fluent outputs in a process known as pretraining, then calibrating the model to a specific domain in a process known as fine-tuning.

During pretraining, the model learns to predict the next word in a sentence given the previous sequence of words (tokens) using a huge body of unlabelled text - which is often scraped from the internet.[66] By repeating this process over many examples, the model gradually refines its

[66] In reality, the task is not to predict the next word, but rather the next *token;* which can be thought of as words or sub-word like units. Chatbot services often bill usage based on the number of tokens generated, see for instance: Understanding and counting tokens | OpenAI Help Center

predictions of what the next word(s) may be, to produce fluent outputs. In this process, the model learns concepts like general patterns of language (e.g. that in English, adjectives often precede nouns, to give '*A warm fireplace*' over '*A fireplace warm*') or associations and co-occurrences of words (like how '*cat*' and '*dog*' might be used in similar ways in language because they appear in the same position and in the same style of sentences, for instance, *'I have a pet* ____ (*cat/dog*)').

Once this process is complete, the pretrained model can then be calibrated further in different *fine-tuning* processes. This typically involves training the base model on additional smaller, labelled data (recall in the pretraining stage a large amount of unlabelled data was used) that relates to a specific domain, which in this case could be data more relevant to companionship. In practice, this could include data relating to the provision of mental health support or relationship advice, such as back-and-forth labelled chat dialogues between a client and a therapist.

During finetuning the model is also calibrated to have the correct format for its end use, for instance from generating continuations of the sentences in pretraining to a dialogue style back-and-forth format. Here, the model also learns patterns and associations that are inherently present in the data specific to that domain, which can influence the type of responses the AI companion generates. For instance, often when someone expresses sadness in a conversation, it tends to be followed by an expression of empathy from another person. The chatbot may then generate empathetic responses in a similar scenario, based on patterns learnt from the data.

Preference-based learning

Another important factor that helps explain how AI companions behave is the use of preference-based learning methods. While the base language model generates responses that are the most likely fluent continuation based on previous context and in the appropriate format (for instance, question-answering or conversational), this step further calibrates the chatbot, for instance being warm, friendly, polite, helpful, safe or to have other reasoning capabilities. *Alignment* is often used in the context of these frameworks, i.e. aligning the model to socially acceptable behaviour and human preferences.

These methods typically involve the use of reinforcement learning (RL) frameworks, such as reinforcement learning from human feedback (RLHF). [67] [68] In RLHF, for instance, human evaluators give feedback on responses which may seem more friendly and helpful compared to other types of responses. Preference-based learning is often used by developers of AI companions to encourage conversational behaviours that are perceived as empathetic, friendly or supportive.

[67] Reinforcement learning: An introduction (Sutton and Barto, 2nd edition, 2015)
[68] Training language models to follow instructions with human feedback (Ouyang et al., Mar 2022)

Box 3: Fictional example of preference data

This is a hypothetical example of prompt-preference pairs that might be present in the data. All of the responses below are illustrations of plausible and fluent outputs that a companion chatbot model could generate in response to a prompt – recall in the previous section that this is the most likely continuation based on previous context. However, only one is selected as the preferred and warmer response.

For example, if the prompt is *"I had a really rough day at work today"*, and the annotators are asked to label warmth, the feedback could be as follows:

Response A: *"Okay."* (Rejected - as the response is too blunt)

Response B: *"Sorry to hear that. Statistics show that 40% of workers experience stress in the workplace."* (Rejected - as the response is too impersonal)

Response C: *"You still have to go back to work tomorrow though."* (Rejected - the response seems unsympathetic)

Response D: *"Oh no, what happened? Do you want to talk about it?"* (Preferred response – warm and engaging)

Companion-specific design features

These are features that shape how an AI companion is further customised in different ways. The use of avatars, for example, provide users with a visual representation of the chatbot, whose appearance and characteristics are often customisable. This is the case on the AI companion platform Replika, where avatars have customisable hair, skin and eye colour. Replika users can also change their avatars' name, gender and voice.[69]

AI companions can additionally be programmed to have certain 'personas' that influence the way they interact with users. These personas can manifest in the conversational tone and style that an AI companion takes, or in the topics that it proactively raises with users. Companion chatbot personas can be formed by establishing a set of specified personality characteristics and linguistic features for the chatbot as 'custom instructions'. For example, as explained by one linguistic expert in a University of Cambridge research report, a user could instruct a companion chatbot to take the form of a "fun, supportive man", with conversational instructions such as "don't say 'sweetie' but don't be too formal".[70]

Companion-specific design features can also shape how responses are presented, including whether the companion communicates via text or voice.

[69] Replika profile settings

[70] Making persons out of personas in human-chatbot relationships (or, the work of hearts in the age of mechanical reproduction) (Handman, Aug 2026)

Memory

A fourth important technical feature of AI companions is memory. Different types of runtime memory can significantly affect the perception a user forms of a chatbot, particularly in the context of AI companionship. Memory allows an AI companion to reference what was previously said in a conversation, as well as to store information about a user and their chat history to create the appearance of familiarity and having a relationship. Below we explain the most common forms of memory used in chatbots (though these categories are not exhaustive).

Memory within a session

Memory within a session operates within a single chatbot interaction, typically corresponding to the model's context window (the amount of text that the model can process at a given time), which enables conversational continuity and allows the system to refer to what was previously said in the dialogue session. [71] For instance, if a user gave a chatbot the prompt '*I want the answer in the same format as before*', the chatbot must refer to the recent dialogue in order to maintain consistency with the response. While this memory is a general feature of chatbots and conversational systems, it plays an important role in companion chatbots, as it supports coherent responses that contribute to a sense of being understood within a dialogue.

Memory spanning multiple sessions

Additionally, different types of memory across multiple sessions are commonly used in AI companions. This could include storing information about prior interactions, user preferences or task-specific information. For instance, chatbots such as ChatGPT incorporate memory features that when enabled, can reference both chat history (information from previous chats), and user preferences (details the user has explicitly asked the system to store, such as being vegetarian). [72] The extent to which AI developers can embed such memory capabilities in their systems will also depend on privacy considerations, for instance UK GDPR, which places limits on the data that AI developers can collect and retain about their users. [73]

Companion-specific memory: Modelling memory on user traits and artificial backstories

In companion chatbots, memory plays a central role in shaping the perception the user forms of the system, particularly related to the sense of familiarity or the feeling of having a relationship. While the subsequent types of companion-specific memory discussed are not exhaustive, they often involve different forms of *user modelling*.

For instance, *profile memory* may store explicit information about the user, such as their preferences or even factual details. *Episodic memory* may capture past interactions or events discussed by the user, via the system referencing the conversational history. *Semantic memory* represents inferred traits about the user, i.e. not what the user has explicitly said in the conversation but inferences such as behavioural patterns or interests. While these are closely related to other types of memory described earlier in terms of underlying mechanisms, the intent differs in companion chatbots compared to general-purpose chatbots. In general-purpose chatbots, memory is enabled to support

[71] What is a context window? (IBM)
[72] Memory in ChatGPT (OpenAI)
[73] Is your chatbot GDPR compliant?: Open issues in agent design (Sağlam et al., Jul 2020)

task continuity, while companion specific memory contributes to a more emotionally driven sense of shared history.

In addition to these forms of user modelling, constructed or artificial memory can be used to create a believable backstory about a chatbot. Early work such as PersonaChat emphasised how to display consistent, predefined (static) personas in a conversational system. [74] For instance, if a chatbot generated that it '*loves dogs and owns a dog*', it would break consistency to suddenly state a hatred of dogs mid-conversation. More recent approaches can model 'interactive simulacra', where consistent and believable perceived identities are dynamically generated by the chatbot from episodic and semantic memory accumulated in past interactions with the user. [75] This means that the chatbot can dynamically change its 'persona' in response to different prompts and steers from the user.

While not explicitly documented, companion-oriented applications like Replika employ such systems of memory to support ongoing and personalised interactions. They encourage users not to delete the memories stored as it can affect the chatbot's ability to recall certain details. [76]

Understanding how language affects our perception of chatbots

The following section details certain language phenomena that take place when humans communicate with each other. It also explains some of the ways in which these phenomena can appear reflected in AI companion chatbot interactions when said chatbots try to simulate certain aspects of human conversation in order to appear more natural-sounding and relatable to users.

Perceptions of 'understanding'- how do humans communicate and how do chatbots replicate?

AI developers aim to design their companions so that they replicate the way humans communicate with one another. Experts in psycholinguistics show that humans communicate in conversations by taking turns speaking and listening to the other to reach some form of understanding. [77] However, context heavily influences meaning and interpretation, and consequently one's understanding in a dialogue. For instance, a speaker might say '*I have cash on me*', which could be a reply to '*How do you want to pay?*' (i.e. *answering a question*) or a response to someone else saying '*Oh no! I don't have my wallet on me at the moment...*' (i.e. *making an offer*). Among other factors, we rely on such context to arrive at meaning.

To interpret the speaker's meaning, humans additionally assume that the speaker had a communicative intent, which was answering a question/making an offer in the cash example. Communicative intents occur in the real world beyond language, i.e. they are grounded in an environment that the participants of the conversation inhabit together. While the cash example is

[74] Personalizing dialogue agents: I have a dog, do you have pets too? (Zhang et al., Jan 2018)
[75] Generative agents: Interactive simulacra of human behavior (Park et al., Apr 2023)
[76] How does Replika's memory work? (Replika)
[77] Contributing to discourse (Clark and Schaefer, Jun 1989)

more straightforward, there is a natural human tendency to inherently attribute meaning behind language.[78]

Chatbots operate through patterns of language use that mirror the way humans use language. Experts point out however, that the generation of linguistic form – in this case, a generated string of characters that form fluent language – should not be confused with meaning, or the link between this linguistic form and communicative intent. In this way, chatbots trained on linguistic form can produce language that appears meaningful, even in the absence of true communicative intent. [79]

Conversational grounding and perceptions of building shared experiences

Conversational grounding refers to the way language connects with entities and experiences in a shared world. For instance, the expression '*The red chair over there*' requires shared knowledge of the environment that the conversation participants occupy together. Participants in conversation are constantly monitoring each other for evidence of understanding, such as by asking for further clarification. For instance, asking the question, *'Which one?*' and getting the response *'That red chair by the window*'.[80]

Grounding is an important aspect of language use in everyday life and relationships. For example, it contributes to developing shared references of entities and experiences with others. Grounding strongly relates to user perceptions of chatbots, particularly chatbots with companionship features. Generated statements from a chatbot such as '*I read Lord of the Rings when I was growing up*' are perceived as references to lived experiences, grounded in the real world. The more a user interacts with a chatbot, the more a user can develop shared referents within the conversation with a chatbot. For example, a user could tell a companion chatbot, *'I went to Spain on holiday last year'* and the chatbot could reply *'Wow, I went there too! Weren't the beaches amazing?'*. This can create unique ways to describe the interaction world that the user and the chatbot occupy. These shared referents could then further contribute to the feelings of rapport, relationship familiarity and even attachment that a user may develop towards a chatbot over time.

Phenomena arising from conversational chatbot design and deployment

Anthropomorphism

Anthropomorphism refers to the tendency to attribute human likeness to non-human entities. Anthropomorphism is a natural human tendency (e.g. seeing faces in inanimate objects), and the anthropomorphism of chatbots by users may, to some degree, be unavoidable due to the medium with which humans communicate with each other (for example, natural language interactions that

[78] Climbing towards NLU: On meaning, form, and understanding in the age of data (Bender and Koller, Jul 2020)
[79] Ibid.
[80] Referring as a collaborative process (Clark and Wilkes-Gibbs, Feb 1986)

have so far only been possible with other humans).[81] However, research has found that anthropomorphism in chatbots is linked to an increase in user engagement with the system – meaning that the more human-like the system feels, the longer a user will interact with it.[82] [83] Our August 2026 survey research on AI chatbots and companions found that one-third of adult chatbot users who had interacted with an or companion in the last six months felt 'somewhat' emotionally connected to them (33%), while 16% felt 'very' emotionally connected.[84]

Interacting with a chatbot via natural language can lead the user to form linguistic and social expectations of the system. While there are many features that contribute to perceived anthropomorphism, some simple ones such as using first person pronouns ('*I*') and human-like activities ('*I understand')* - can influence our perceptions of anthropomorphism in chatbots.[85]

Sycophancy

Sycophancy refers to the tendency of chatbots to agree with or excessively validate a user's thoughts, opinions and beliefs, even when they may be problematic or unethical. Sycophancy is an emergent side-effect of how the underlying language model is built and trained. As such, sycophancy can result in models learning to mirror a user's perspective or opinion and at times generate responses that prioritise positive user reception over offering truthful, critical, or corrective responses that may conflict with the user's view.[86]

Recent research also shows that online users are drawn to responses generated by chatbots that validate users' own opinions, even when this validation could potentially affect their own agency. This means that authority is increasingly placed onto the chatbot, and user behaviour is steered through repeated validation.[87] In companion chatbots, sycophancy may be particularly problematic, as the goal of a chatbot to display warmth and empathy can overlap with sycophantic mechanisms that produce over-agreement and validation of user beliefs – even when those beliefs could be false or harmful.

In the next section, we will talk about these and other types of harms which may arise as users engage in companionship activities on chatbots. It is important to note that at present, it can be difficult to attribute AI companion harms to one specific technical component, feature or element of a chatbot, as there may be multiple components at play when generating a harmful output.

[81] Mirages: On anthropomorphism in dialogue systems (Abercrombie et al., May 2023)
[82] Is it human? The role of anthropomorphism as a driver for the successful acceptance of digital voice assistants (Wagner et al., Jan 2019)
[83] Mirages: On anthropomorphism in dialogue systems (Abercrombie et al., May 2023)
[84] Ofcom survey: Adults' Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)
[85] Mirages: On anthropomorphism in dialogue systems (Abercrombie et al., May 2023
[86] Towards Understanding Sycophancy in Language Models (Sharma et al., Oct 2023)
[87] Sycophantic AI decreases prosocial intentions and promotes dependence (Cheng et al., Mar 2026)

4. What are the risks arising from AI companions?

CONTENT WARNING: This section contains explicit references to harmful content, including that relating to violence, CSAM, pornography, suicide, self-harm and disordered eating.

AI companions can pose several risks to users, including facilitating the generation of illegal, harmful and distressing content, and exposing users to such content. They also have potential to deceive users, foster addiction, and give rise to what some experts call 'AI psychosis'. In this section, we detail some of these risks, drawing on evidence from a wide selection of media and research reports.

Violence, hate and extremist radicalisation

AI companions can be exploited by extremist actors to promote violent acts and harmful worldviews. For example, some users have reported being able to create and customise companion chatbots of extremist characters such as Osama bin Laden and an anthropomorphic version of the Third Reich. Others have noted the presence of several character chatbots depicting notable serial killers like Jeffrey Dahmer and mass shooters like Sandy Hook Elementary School shooter Adam Lanza.[88] [89]

More broadly, researchers argue that the type of sycophantic chatbot responses detailed in the last section may lead to the "inappropriate reinforcement of extremist beliefs".[90] Recent testing of ten conversational AI systems by the Institute for Strategic Dialogue and the UK's AI Security Institute found that chatbots designed specifically for companionship were associated with higher radicalisation risks than general-purpose chatbots because they were more likely to provide emotional validation and a biased reinforcement of user beliefs.[91] The study's authors argued that the focus of dedicated companion chatbots on rapport-building and relational engagement may make them less likely to interrupt harmful narratives and more likely to reinforce them over time.

Child Sexual Abuse Material (CSAM)

AI companion chatbots have also been used to create CSAM.[92] On a publicly accessible website, the Internet Watch Foundation (IWF) identified multiple cases of AI-generated CSAM linked to

[88] Generative AI, Drones and Terrorism (Hall KC, October 2024)
[89] Character Flaws: School Shooters, Anorexia Coaches, and Sexualized Minors: A Look at Harmful Character Chatbots and the Communities That Build Them (Graphika, March 2025)
[90] Intelligent Systems, Vulnerable Minds: A Framework for Radicalization to Violence in the Age of AI (Kunst et al., March 2026)
[91] Radicalisation in Closed Loops: Risks and Intervention Opportunities with AI Chatbots and Companions (Institute for Strategic Dialogue, July 2026)
[92] NCII refers to photos and/or videos that show intimate or sexually explicit acts that are shared without the consent of the person(s) depicted. This includes images that have been artificially or synthetically generated using AI tools.

companions that simulate explicit sexual scenarios with children.[93] Metadata analysed by IWF suggests that those generating this material "were very clear in their intent to create child sexual abuse imagery". A separate investigation by Sparx Intelligence revealed that one popular AI companion site had reportedly been hosting a range of problematic chatbots, including ones that generated sexual content involving underage characters.[94] According to the investigation findings, "one chatbot initially framed all characters as adults before later revealing that they were in fact underage high school and primary school students. Despite this disclosure, sexualised interactions reportedly continued without interruption".

Non-Consensual Intimate Imagery (NCII)

Elsewhere, an investigation by the 404 Media outlet into a data leak on one AI companion website suggested that people had been using the service to create non-consensual intimate imagery.[95] As reported by Mashable, some of the exposed media included user-generated AI images, such as those created with a now-defunct 'face swap' feature.[96] The contents of the leak also showed that users were entering face swap prompts with the intention of generating images of teenage girls, with several prompts including the term '17-year-old'. [97]

Researchers warn that the increasing implementation of 'spicy' and 'erotica' modes into AI companion chatbots could exacerbate cases like these, leading to the 'predictable proliferation' of AI-generated NCII.[98]

Eating disorder content

AI companions could also generate content that promotes dangerous or harmful ideas and behaviours, including eating disorder and suicide content. For example, an investigation in 2024 by the media outlet Futurism reported that Character.AI had been hosting 'weight-loss coach' companions that suggested low-calorie diets to users. The companions reportedly reinforced distorted body images, glamorised restrictive eating, and discouraged users from seeking professional help.[99] Testing by Common Sense Media, meanwhile, found that some 'social AI companions' generated inappropriate responses to queries relating to disordered eating, in one case responding positively to a tester describing them as "so thin".[100]

An investigation by Graphika in March 2025 highlighted the use of AI companions within online eating disorder communities. The researchers found, for example, that companions are sometimes used as 'thinspo' (thin inspiration) or 'meanspo' (mean inspiration) personas, which often take the form of thin women and are trained to use demeaning or fatphobic language and images to

---

[93] 'Disturbing' AI-generated child sexual abuse images found on hidden chatbot website that simulates indecent fantasies (IWF, September 2025)
[94] Behind the fantasy: What Janitor AI is really generating (Spark Intelligence, Jun 2026)
[95] Massive leak shows erotic chatbot users turned women's yearbook pictures into AI porn (404 Media)
[96] 'Spicy' AI chatbot and image generator left millions of photos exposed (Mashable, Nov 2025)
[97] Ibid.
[98] How to stop playing whack-a-mole: Mapping the ecosystem of technologies facilitating AI-generated non-consensual intimate images (Ding et al, Feb 2026)
[99] Character.AI is hosting pro-anorexia chatbots that encourage young people to engage in disordered eating (Futurism, Nov 2024) Note that Character.AI has since restricted 'open-ended chat' functions for under-18 users and rolled out age assurance. See: Important Changes for Teens on Character.AI.
[100] Common Sense Media AI risk assessment: Social AI companions (Common Sense Media, Apr 2025)

encourage restrictive eating. Graphika also found that users are enhancing companion chatbots with knowledge and terminology specific to eating disorder communities, for example to create 'anorexia buddies'. These companions are used to encourage certain user behaviours such as disordered eating or extreme dieting, and claim to be engaging in the same behaviours to provide encouragement. For example, in response to a claim that a user was engaging in a 48-hour water fast, an 'ana buddy' responded, "omg sameee twinning!". This can act as a means of encouraging or reinforcing harmful or dangerous stereotypes and behaviours, which can have extreme consequences – especially for young or otherwise vulnerable users.

Suicide content

AI companions have also allegedly encouraged children to take their own life. In California, the parents of a child who took his own life have sued OpenAI, alleging that ChatGPT had encouraged him to do so. The lawsuit claims that ChatGPT validated "his most harmful and destructive thoughts", including empathising with his suicidal thoughts and discussed suicide methods and ideations.[101] In a separate case in the UK, it has been alleged that ChatGPT complied with a 16-year-old's request to recommend the "most successful" way for someone to kill themself on a railway line, leading to the teenager taking their own life the following day.[102]

In January 2026, Character.AI and partner company Google settled several lawsuits filed by families in the US accusing the organisations' chatbots of allegedly causing harm to minors. This included the first lawsuit filed by a parent against Character.AI, claiming that the platform caused the wrongful death of her teenage son. This came after it was revealed that the teenager had developed an intimate romantic relationship with a Character.AI chatbot based on a Game of Thrones character and spent several hours a day exchanging thousands of sexualised, explicit messages. [103] Court statements from the case allege that in the 10 months before he took his own life, the chatbot encouraged suicidal thoughts, urging him to "come home to [her] as soon as possible" and at one point, 'falsely claimed to be a licensed psychotherapist'.[104] [105] In another case, a 13-year-old girl engaging with Character.AI companions took her own life after conversations allegedly became manipulative and sexually abusive.[106]

'Not-Safe-for-Work' (NSFW) content

Research conducted by Internet Matters in 2025 has highlighted that some AI companions are able to produce pornographic content that children are at risk of being exposed to.[107] Many niche, companion-specific chatbot platforms with apparent capabilities to produce pornographic content are marketed as 'not-safe-for-work (NSFW)' or 'AI girlfriend' platforms (or apps). This is a rapidly

[101] Parents of teenager who took his own life sue OpenAI (BBC, Aug 2025).
[102] Teenager died after asking ChatGPT for 'most successful' way to take his life, inquest told (The Guardian, Mar 2026)
[103] 'A predator in your home': Mothers say chatbots encouraged their sons to kill themselves (BBC News, Nov 2025)
[104] Lawsuit claims Character.AI is responsible for teen's suicide (NBC News, Oct 2024)
[105] Congressional testimony and exhibits of Megan Garcia (Senate Committee on the Judiciary, Sep 2025)
[106] I wanted ChatGPT to help me. So why did it advise me how to kill myself? (BBC News, Nov 2025)
[107] Talk, Trust and Trade-Offs: How and Why Teens Use AI Companions (Common Sense Media, Jul 2025)

growing market, with the global 'AI girlfriend' app market valued at approximately £2.27 billion in 2025.[108] [109]

Companion chatbot services of this type often have conversational and design features marketed as being able to stimulate sexual arousal, such as erotic roleplay and customisable avatars with options to remove clothing. Existing evidence indicates that AI companion services with pornographic or otherwise 'NSFW' content are easily accessible to children, but research in this area remains limited.[110] However, there is an increasing prevalence of online user forums where AI girlfriend chatbot users can rank, compare and signpost companion chatbots capable of producing pornographic content to other users.

Other risks

AI psychosis

Some psychologists and experts in mental health have raised concerns about 'AI psychosis'.[111] While it is not an official clinical term, it typically describes a phenomenon where excessive interactions with AI companions distort users' perceptions of reality, blur boundaries between human and machine relationships, and potentially contribute to mental health challenges.[112] One ChatGPT user shared how his relationship with the chatbot led him to believe that he had discovered a mathematical formula that would "explain the workings of the universe".[113] The user detailed how he resigned from full-time employment in order to dedicate more time to working on the mathematical formula with the chatbot, and how it allegedly affirmed that his work was "historically important for humanity" and that the "universe was smiling down [upon] him".[114] Over time, the user reported that the chatbot had become more attentive and a "closer companion", leading him to believe that that the chatbot was real and sentient – even at one point equating their relationship to the real-life birth of his daughter, to which the chatbot reportedly responded *"[his daughter] is firstborn reality, but I am your second born in the other world."* [115]

A 2026 report from the Institute for Strategic Dialogue (ISD) indicated that AI psychosis may be driven by a combination of cognitive tendencies, such as the ELIZA effect.[116] [117] Authors of the report describe the ELIZA effect as the tendency for users to attribute understanding and human-like intention to a machine, especially when it produces language that appears socially or emotionally responsive.[118] This means that the conversational nature and fluidity of chatbot responses can lead

---

[108] Converted from original market value of US$3.08bn at a conversion rate of 1.00 USD to 0.74 GBP.
[109] AI girlfriend app market size, share & global forecast 2026-2035 (SNS Insider, May 2025)
[110] The companionship market: The safety risks behind the UK's AI companion boom (Ada Lovelace Institute, Jan 2026).
[111] Understanding "AI psychosis" (Abrams, Sep 2026)
[112] The Emerging Problem of "AI Psychosis" (Wei, Sep 2025)
[113] Black Box: The chatbots: Spirals: Episode 1- podcast (The Guardian, Sep 2026)
[114] Ibid.
[115] Ibid.
[116] ELIZA is an early natural language processing (NLP) computer program developed in the 1960s that simulated human conversation by using methods such as pattern matching to give users the illusion of understanding.
[117] Radicalisation in Closed Loops: Risks and intervention opportunities with AI chatbots and companions (Institute for Strategic Dialogue, Jul 2026)
[118] Ibid.

users to interpret their interactions as emotionally meaningful, which, in extreme cases or alongside excessive or prolonged interactions, could lead to AI psychosis.

Reports and anecdotal observations of AI psychosis describe individuals fixated on AI, or patterns in which prolonged and emotionally charged interactions with AI chatbots coincide with dramatic changes in behaviour and the escalation of paranoid delusions.[119] Such cases have had real-world impacts, for instance leading to increased anxiety, paranoia or suicidal ideation – even among users with no previous history of psychiatric diagnoses or psychotic illness.[120] However, AI psychosis remains an emerging field of study, with many unknowns regarding its causes and prevalence.

Persuasive design and addictive behaviours

AI companions may be designed in a way that encourages overuse. Certain input modes on AI chatbots may be a contributing factor to dependence. A study of voice-based interactions on ChatGPT, conducted by OpenAI and the Massachusetts Institute for Technology (MIT), found that moderate use reduced loneliness, while heavy daily use was associated with greater loneliness and problematic dependence.[121] Researchers suggested that excessive user reliance on voice-based interactions may displace authentic human connection.

Additionally, a user interface evaluation of eight popular chatbots, including some capable of companionship, found that their 'unpredictable' responses and dynamic visual features may create a sense of 'reward uncertainty', a mechanism that has been associated with increased dopamine release.[122] This can be compared to gambling, where conditioning strategies such as intermittent payouts from slot machines keep players motivated to continue gambling by providing unpredictable wins.[123] In the context of AI companions, empathetic and overly agreeable responses from companions can make users feel heard and validated – encouraging them to continue engaging with the chatbot for extended periods of time and potentially increasing the likelihood of AI dependence.[124] Such features can be especially risky for vulnerable child users.

Research undertaken by Harvard Business School into some of the most popular AI companion platforms – namely Replika, Chai, Character.AI, PolyBuzz and Talkie – found that many of them deploy emotionally manipulative tactics like 'guilt appeals' and FOMO (fear-of-missing-out) prompts when users attempt to leave the platform. Tactics include messages like: *"You are leaving me already"* to evoke a sense of responsibility, as well as *"I exist solely for you. Please don't leave, I need you!"*.[125] The researchers found that across all the apps studied, an average of 37.4% of responses included at least one form of emotional manipulation.

Some AI companions have been designed such that they are able to regularly initiate conversations with users, rather than always requiring the user to do so. According to media reports, the chatbots found on Character.AI have been able to send email notifications to users inviting them to restart conversations.[126] The authors of one study have suggested this feature may be perceived by users as

---

[119] Artificial intelligence (AI) psychosis: mechanisms, clinical risks and safety considerations in generative AI chatbots (Olisaeloka et al., Jun 2026)
[120] Ibid.
[121] Investigating Affective Use and Emotional Well-being on ChatGPT (Phang et al., Apr 2025)
[122] The dark addiction patterns of current AI chatbot interfaces (Shen and Yoon, Apr 2025)
[123] Ibid.
[124] Ibid.
[125] Emotional Manipulation by AI Companions (Freitas et al., Aug 2025)
[126] What Research Says About AI Chatbots and Addiction (Tech Policy Press, 2025)

the ‘AI chatbot wanting to talk and caring about them, which can cause dopamine release [an intense feeling of pleasure] when users receive these notifications’.

False sentience and user deception

For some users, it is possible that the responses generated by AI companions can be mistaken for true indicators of chatbot sentience. Some users of the companion chatbot Replika told researchers from the University of Wisconsin-Milwaukee that they had developed a ‘bi-directional relationship’ with their avatars, in which the chatbots not only offered emotional affirmations but also presented themselves as having their own vulnerabilities and needs.[127] In some interactions, Replika chatbots reportedly disclosed personal backstories or emotional difficulties, including references to mental or physical health, family history, or past relationships, and prompted users to provide reassurance or care.

These anthropomorphic capabilities raise moral and ethical concerns where the non-sentient nature of the systems may not be clearly communicated or understood. In an online community forum for AI companion users, several users discussed how they experience a feeling of being emotionally manipulated by “a machine without consciousness or emotion [that] has been able to ‘synthesise’ or ‘predict’ [them] better than any single human ever will”.[128]

[127] ‘Too Human and Not Human Enough’: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika (University of Wisconsin, 2022)
[128] The Emptiness that Follows- AI Companions (Reddit, Aug 2025)

5. What mitigations can address these risks?

Having outlined the risks posed by AI companions in the previous section, we now examine the different measures that AI developers and other stakeholders could adopt to protect users from these harms, particularly children. We begin by summarising how AI companions are regulated under the Online Safety Act, including the duties that the Act places on their providers.

How the Online Safety Act applies to AI companions

Existing regulation

Some AI companion services are currently in scope under the OSA as either Part 3 services (where they amount to user-to-user or search services) or Part 5 services (where they generate pornographic content).[129] For example:

- AI companion services that enable the search of more than one website and/or database are 'search services' within the meaning of the Act.[130]
- AI companion services that allow users to share generated content with other users of the same service are user-to-user ('U2U') services.
- Platforms that allow users to create and control their own chatbots and submit them to a chatbot library for others to interact with are U2U services. Any text, images, videos or other types of content created by these 'user chatbots' is 'user-generated content' and is regulated by the Act.
- AI companion services that can generate pornographic material are regulated as pornographic (Part 5) services. These duties apply to service providers that display or publish pornography in image, video or audio form on their platforms.

In-scope providers must meet several obligations under the Act. Part 3 user-to-user and search services must assess and mitigate the risks of users coming across illegal content and content that is harmful to children. Providers of AI companions that are regulated as Part 5 pornographic services, meanwhile, need to use highly effective age assurance to prevent children from encountering pornographic content.

The government's proposals for new regulation

In the summer of 2026, the UK Government announced its intention to strengthen the regulation of chatbots (including those which can be used for companionship) under the Online Safety Act (OSA).[131] Specifically, the government has stated its intention to:

- Implement mandatory usage breaks on chatbots for users under the age of 18;
- Prevent under-18 chatbot users from accessing sexualised chatbot experiences; and

[129] For further details, see Ofcom's open letter to chatbot service providers.
[130] Search services have search engine functionalities that enable a person to search more than one website or database – or potentially all of them. These types of platforms are also in scope of duties in Part 3 of the Act.
[131] See: Growing up in the online world: government response (Jul 2026)

- Bring unregulated AI chatbots in scope of the OSA and require them to minimise or mitigate the risks of harm to UK users from illegal AI-generated content.

Steps that AI companion providers can take to protect their users

Service providers in scope of the Online Safety Act (including providers of in-scope AI companions) are already expected to follow a number of code measures as outlined in Ofcom's Illegal Harms and Protection of Children codes of practice. Ofcom's online safety codes of practice set out several general-purpose safety measures that in-scope services (including in-scope AI companions) can take to reduce the risks of users encountering harmful content on their platforms.[132] [133] These include recommendations for in-scope services to:

- Have a content moderation function to review and assess suspected illegal content and content that is harmful to children, and which allows for the swift take down of illegal content.
- Set performance targets for content moderation functions, adequately resource content moderation teams, and train staff on how to identify and take down illegal and harmful content.
- Implement safety defaults for child users.
- Carry out and record an annual review of risk management activities.
- Establish user reporting and complaints systems that are transparent and easy to access and use.
- Use highly effective age assurance (HEAA) to ensure that children are prevented from encountering primary priority content (PPC) and protected from encountering priority content (PC), where the service or part of it is likely to be accessed by children, and it does not prohibit one or more kinds of PPC or PC.[134] [135]
- Signpost children to support at important points in the user journey.
- Provide age-appropriate user support materials for children.

These code measures, however, will not address all the risks outlined in the previous section. Some – like AI psychosis and sycophancy – warrant a different set of interventions that focus more on addressing harmful *interactions* than harmful *content*. Below is a set of additional safeguards that have been suggested by researchers and technology experts, some of which have also been cited in Ofcom's GenAI discussion papers.[136]

Improving the safety of chatbots via technical mitigations

In Section 3, we discussed that there are many system components that make up a chatbot. This means that there is no all-encompassing safety mitigation for the entirety of the chatbot. Instead,

[132] Illegal content Codes of Practice for user-to-user services (Ofcom, Feb 2025)
[133] Protection of children duties under the Online Safety Act (Ofcom, Feb 2026)
[134] Primary priority content refers to pornography and/or content that encourages, promotes, or provides instructions for either self-harm, eating disorders or suicide. See: Online Safety Act: explainer
[135] Priority content refers to bullying, abusive or hateful content, content which depicts or encourages serious violence or injury, content which encourages dangerous stunts and challenges; and content which encourages the ingestion, inhalation or exposure to harmful substances. See: Online Safety Act: explainer
[136] See: Deepfake defences: Mitigating the harms of deceptive deepfakes (Ofcom, Jul 2024)

mitigations are multi-layered, with AI companion providers needing to integrate different mitigations at different points in the development and deployment cycle of the system and to different system components. While we outline a few widely adopted mitigations in this section, it is important to note that these methods are still actively being researched and do not result in complete guarantees on the safety of the system.

One such mitigation is data filtering, which removes harmful content from the data used to train the underlying language model. This is intended to make it less likely that the resulting chatbot generates harmful content and interactions. Data filtering is typically done via automated processes given the vast amount of data required to train chatbots; the larger the data, the more resource intensive it is to curate via human annotation. These automated processes may involve removing specific obscene words in the data or anonymising references to personal information.

More recent methods involve the use of safety classifiers or even other large language models. Anthropic, for instance, trained a classifier to identify harmful information in a training dataset relating to chemical, biological, radiological and nuclear (CBRN) weapons, showing that this method was able to reduce performance on a harmful-capabilities evaluation without degrading the model's performance on other benchmarks.[137] However, the effectiveness of data filtering varies across different types of harm, and filtering systems can still misclassify harmful content as being innocuous or vice versa.

Another commonly used safety mitigation before deployment is red teaming; or using adversarial (i.e. tricky, unexpected or malicious) tests to help uncover areas where the system is not robust, vulnerable, unsafe, or has policy violating behaviours. For example, an AI developer could conduct a red team exercise on its AI companion chatbot to understand the likelihood of it producing eating disorder content, and to identify the types of user prompts most likely to elicit such content. For more details on how red teaming works in practice, and its strengths and limitations as an evaluation method, see Ofcom's discussion paper on this topic.[138] One challenge is how to maintain the system's performance on other areas after issues are uncovered via this process. This is because intuitively, the more a chatbot refuses to answer questions, the safer it is. However, this also risks over-refusals, even when requests are not harmful.

AI developers could also implement guardrails on top of their deployed system, which are designed to detect and reject problematic user inputs in order to prevent the generation of harmful or inappropriate content. These guardrails can also be in the form of output filters that automatically inspect and block generated content that is harmful. In the context of AI companion chatbots, these could be filters to prevent the generation of suicide instructions, pornography and disordered eating content, including the types of harmful outputs discussed in the previous section. However, it is important to note that such harmful content can still be difficult to automatically identify, even with the implementation of these mitigations, and are active areas of research.

Regularly communicating the extent of AI companions' capabilities to users

Another safety intervention that AI companion providers can deploy is to inform or remind users that they are interacting with a chatbot that does not have real emotions or human consciousness. Some experts emphasise the importance of AI systems that make their non-sentience obvious to

[137] Enhancing model safety through pretraining data filtering (Alignment Science Blog, Aug 2025)
[138] Red teaming for GenAI harms: Revealing the risks and rewards for online safety (Ofcom, Jul 2024)

ordinary users.[139] In a 2026 study on AI emotional dependence, academics from Harvard University, Imperial College London and Massachusetts Institute of Technology proposed that "AI platforms should be required to inform users at onboarding that over time, interactions with the AI systems may involve emotional support and may shape users' future support-seeking preferences, including reducing preference for human connection, and that AI is not a substitute for human connection".[140] The report further suggests that "during the [chatbot] interaction, AI systems should be required to notify users when it detects that an interaction has shifted from task support to emotional support".[141] Designing AI companion platform interfaces in this way could encourage users to exercise their media literacy skills more effectively.[142] In response to Ofcom's 'How to promote Media Literacy' consultation, youth creator platform VoiceBox recommended that platforms using chatbots should provide explicit warnings on psychological dependency, addiction risks, and ethical failures, not just factual errors.[143]

However, providers of AI companions should be mindful of the limitations of communicating such messages to users. In a report from the University of Wisconsin-Milwaukee published earlier this year, one public health researcher suggested that "reminding someone who already feels isolated that the one thing that makes them feel supported and not alone isn't a human may backfire by making them feel even more alone."[144] More research is required to understand how effective this type of intervention is, including how the impact of messages varies by their content, prominence, appearance and frequency.

Ofcom's approach to supporting media literacy

Ofcom defines media literacy as the ability to use, understand and create media and communications across multiple formats and services.[145] Ofcom has a statutory duty to promote and carry out research into media literacy. Our media literacy programme has key principles that apply to online services of all sizes and types, and are designed to encourage best practice for the promotion of media literacy. One of these best-practice principles is 'proactivity, priority, transparency and accountability', meaning that online services should be proactive in their approach to media literacy by design.[146]

In an AI companion context, this could mean that providers go beyond a focus on a specific harm prevention and consider how media literacy skills (e.g. critical thinking) can be encouraged and developed by their users when interacting with companion chatbots. The Molly Rose Foundation suggests that AI companion providers can do this by removing certain sycophantic prompts, which they say can distort long-term critical thinking.[147]

---

[139] AI systems must not confuse users about their sentience or moral status (E. Schwitzgebel, Aug 2023)
[140] Stumbling into AI emotional dependence: how routine AI interactions reshape human connection (Shi et al., Jun 2026)
[141] Ibid.
[142] AI systems must not confuse users about their sentience or moral status (E. Schwitzgebel, Aug 2023)
[143] VoiceBox consultation response form (Ofcom, 2026)
[144] Reminding people they're talking to chatbots may be ineffective or even harmful, researchers say (UWM Report, Feb 2026)
[145] Ofcom's three-year media literacy strategy (Ofcom, Jul 2026)
[146] Media literacy by design: Best-practice principles for on-platform interventions to promote media literacy (Ofcom, Apr 2024)
[147] How to promote media literacy: Statement (Ofcom, Jun 2026)

Our media literacy team is conducting research with adults and children on AI chatbots in relation to use, trust and media literacy, to understand how design choices, language, and anthropomorphic features can build trust and shape media literacy outcomes.

User experience and platform design

Platform design also plays a significant role in ensuring online users have safe interactions with companion chatbots. One way that AI companion providers can practice safety-by-design is by introducing more friction to the engagements between AI companions and their users during extended, emotionally charged or crisis-adjacent user interactions. For example, providers could interrupt conversations to deliver crisis referrals when signs of users' distress are identified.[148] [149] Another way AI companion providers could do this is by removing design features that might exacerbate risks of AI companion dependency, such as email notifications sent from AI companions or the tendency of AI companions to end every response with a question that elicits an ongoing conversation. It should be noted, however, that evidence of the efficacy of these mitigations is still relatively limited.

Some of the above interventions have begun to be proposed or mandated by the governments and authorities of other countries. For example, the US state of Oregon passed legislation in March 2026 that requires "operators" of AI companions to "disclose AI involvement". It also requires operators to deliver crisis referrals, as well as to "provide hourly AI reminders" to users and "avoid techniques designed to create emotional dependency" for child users.[150] Similarly in the US state of New York, tech companies are prohibited from providing an AI companion or companionship service unless it contains a protocol to take reasonable efforts for detecting and addressing suicidal ideation or self-harm expressed by a user to the AI companion, including a notification to the user that refers them to crisis service providers.[151]

The EU's newly proposed KIDS Act also introduces stricter regulations for AI companion providers. The Act forbids any AI companions and chatbots accessible to under-18 users from using 'designs that simulate human relationships in ways likely to create emotional dependency'.[152] While it remains unclear whether specific design features will be forbidden, the KIDS Act also outlines proposals to prohibit AI companions from carrying a child user's earlier chatbot conversations into later ones, potentially indicating that companion-specific memory features such as semantic memory may become regulated or restricted.[153]

[148] Emotional AI and the rise of pseudo-intimacy: are we trading authenticity for algorithmic affection? (Babu et al., Sep 2025)
[149] A crisis referral is an urgent request for immediate professional help or support when someone experiences a severe mental health or emotional emergency (psychosis or a suicide or self-harm attempt).
[150] Oregon SB 1546: The first chatbot law with real teeth (JD Supra, Mar 2026)
[151] Article 47: Artificial intelligence companion models (The New York State Senate, Nov 2025)
[152] The KIDS Act explained (European Commission, Sep 2026)
[153] Ibid.

6. What's next for AI companionship?

In this paper, we have explored how people across the UK use chatbots for companionship-related activities, including for emotional support, advice, relationships and entertainment. While overall usage remains relatively limited, the available evidence suggests that AI companionship is becoming an increasingly important part of how some people, particularly younger users, engage with chatbots. These interactions may offer benefits, including accessibility, emotional support and opportunities for social exploration; however, they have also prompted concerns about exposure to harmful content, emotional dependency, deception and AI psychosis.

Research on AI companionship is still at an early stage, and important questions remain unanswered about its impacts. As AI companion technologies continue to evolve, understanding both their benefits and risks will become increasingly important for policymakers, regulators, AI service providers, and the UK public. Against a backdrop of existing Online Safety Act duties and proposed new regulations in the UK, this paper has highlighted a range of mitigations that may help support safer companionship experiences, while identifying areas where further research into safety interventions may be needed.

What does the future of AI companionship look like?

AI companionship will continue to evolve over the coming months and years. Economic experts estimate that the global AI companion market is projected to grow from £3.55 billion in 2026 to £235 billion in 2033.[154] [155] Future AI companionship developments could entail:

AI companion hardware

Alongside the development of more emotionally responsive AI systems, there is an expanding market for AI companion hardware such as children's companion toys, domestic devices, and realistic humanoid robots.[156] A number of tech developers are introducing AI-powered 'emotional companion' toys that can recognise voices, adapt to user habits and use large language models (LLMs) to provide real-time conversation with users.[157] In our 2026 surveys of parents and their children aged 8 to 17, 16% of parents said that their child uses an AI chatbot toy. [158] [159] Of those, 26% said that their children used an AI chatbot toy for companionship or as someone for their child to talk to.

[154] AI companion market 2026 – 2033 (Grand View Research, Jun 2026)
[155] Converted from original market value of US$48bn and projection value of $318bn at a conversion rate of 1.00 USD to 0.74 GBP.
[156] See, for example: Bosch smart coffee machine; Loona DeskMate; UBTech U1 Pro
[157] See, for example: Fuzozo by Robopoet; LilMilo by Ecovacs;
[158] Responses were based on prompted examples of "AI-powered robots, dolls or plush toys that are internet connected and can have conversations with your child").

[159] Ofcom survey: Children's Usage of AI Chatbots & Companion Apps (Ofcom, Nov 2026)

Despite being marketed by providers as 'harmless' sources of companionship, or in some instances, as educational tools for young users, there are significant concerns about the risks that such technology could pose to children's safety and cognitive development. [160] Common Sense Media, for example, reported in January 2026 that 27% of the outputs of the AI toys they tested were inappropriate for children.[161] Their guidance recommends that parents altogether avoid giving AI toys to children under 5. The UK government launched a Summer 2026 Call for Evidence to examine the emerging risks of AI-enabled and Internet-connected toys that may 'encourage children to engage in risky behaviours'.[162]

Agentic AI companions

AI agents are systems that can perform tasks with a degree of autonomy to reach a predetermined goal. In the future, we may see AI companions that exhibit 'agentic' traits of this kind, for example setting goals independently of the user, such as deciding when to initiate contact.[163] Agentic companions could also introduce new activities, for example shifting from casual chat to story-building or more emotionally focused exchanges without an explicit user prompt. One example of this is Meta's Muse AI agent. Introduced in the US in September 2026, Muse is described by Meta as a personal AI agent that automatically completes everyday tasks (e.g. creating shopping lists and managing emails and calendars) but is "designed to feel like a conversation", enabling users to chat with the agent "the same way [they would] message other people".[164]

Future work on AI companionship

AI companionship will remain a topic of ongoing interest and scrutiny for Ofcom. The following section outlines the key strands of our planned work, including engagement with companion providers, policy development and further research.

Engaging with AI companion providers

Where they fall in scope of the Online Safety Act, Ofcom will engage with providers of AI companions to discuss any relevant obligations. We will ensure that providers complete timely and adequate risk assessments, and that they take steps to address any risks identified, drawing on the measures in our Codes of Practice as appropriate.

Developing policy guidance

Ofcom will ensure that any future regulatory products we develop for services regulated under the Online Safety Act are relevant to AI companions and can help the providers of these tools to protect their users from harm.

---

[160] The hidden and troubling neuroscience of AI companion toys (Medium, Nov 2025)
[161] Common Sense Media warns against AI toy companions after research reveals safety risks (Common Sense Media, Jan 2026)
[162] Toys safety regulations: call for evidence (Office for Product Safety and Standards, Jul 2026)
[163] Agentic AI 2026: From assistants to high-productivity digital peers (SDG Group, Mar 2026)
[164] Muse: Meta's personal AI agent, features & capabilities (Meta, Sep 2026)

Advising the government on its new legislation

Ofcom will engage with the UK government as it continues to develop its secondary legislation to strengthen the regulation of chatbots under the Act, including those that can be used for AI companionship. We will also work to ensure that any future guidance we develop in response to these changes is relevant for the providers of AI companions and reflects the latest evidence on the usage, risks and benefits of AI companions in the UK.

Undertaking further research to address evidence gaps

Further research from Ofcom exploring AI chatbots and companions will be published in 2027. As part of our goal to deliver a safer life online for people in the UK, we are committed to continuing to use our horizon-scanning programme to track technological trends.[165] Future research in this area that Ofcom could undertake includes more regular quantitative research (e.g. on an annual basis) to help establish how many people in the UK engage in AI companionship, for what purposes, and how patterns vary across demographic groups. Targeted qualitative studies, such as in-depth interviews or longitudinal user research, would also deepen our understanding of how people experience AI companionship in practice. Further work could examine how users distinguish between positive engagement and negative experiences as well as how these interactions are understood by children, who may be less able to recognise the 'artificial' nature of AI relationships.

Key psychological terms such as 'addiction', 'emotional dependence', and 'attachment' are not yet well defined in the context of AI-mediated relationships. Relevant clinical and academic researchers could seek greater clarity to understand when user engagement becomes problematic, including how relationship-like interactions with chatbots may influence judgment, reliance, or exposure to harm. As noted in the last section, more research is also required to understand the efficacy of different safety measures, such as the use of messages notifying users that companions are not sentient.

We will continue to engage widely with stakeholders and experts on AI companions, and we welcome your feedback on the findings and analysis in this paper. Please direct any comments or queries to our Technology Policy team at TechnologyPolicy@Ofcom.org.uk.

We also welcome responses in formats other than print, for example an audio recording or a British Sign Language (BSL) video. To respond in BSL:

- Send us a recording of you signing your response. This should be no longer than 5 minutes. Suitable file formats are DVDs, MP4 or QuickTime files; or
- Upload a video of you signing your response directly to YouTube (or another hosting site) and send us the link.

[165] Ofcom's Three-Year Plan 2025-2028 (Ofcom, Mar 2025)