



Trinity College Dublin
The University of Dublin



Policy Briefing

AI Companions

AI Accountability Lab (AIAL),
Trinity College Dublin

July 2026

Executive Summary

1. AI companion apps display a strong tendency to combine aggressive monetisation strategies involving dark patterns, anthropomorphism, and performance issues

AI companion apps consistently employ dark patterns and anthropomorphic features, alongside sexualised content and harmful stereotypes, with the likely intention of maximising user engagement, encouraging personal disclosures, and promoting paid features. From an analysis of five of the most popular apps, we identified 276 dark patterns, 171 instances of anthropomorphism, and 55 examples of stereotyping or erotic content, suggesting these practices are pervasive and central to their design. We also found issues regarding 'premium features' that further raise concerns about the reliability of advertised capabilities and whether users are nudged to purchase services without any actual benefit.

2. Disregard of privacy and data protection laws

We found AI companion apps' policies failed to provide key information and rights mandated by the laws in the UK and EU. The missing information included valid legal basis, purposes for processing of personal data, stating use of consent (which is an opt-in) but then later referring to it as an opt-out, cross-site tracking for advertisement with unclear use of conversational data, and an acknowledgement of applicable rights and how to exercise them. We also found indications of personal data flows outside of the UK/EU to third countries without clear indication of recipients or safeguards such as encryption that are necessary to maintain adequate protection across jurisdictions. All AI companion apps mention use of (user) data for AI training and we found concerning gaps regarding their transparency and legality with missing information and uncertain legal bases which raises the issue of users having control over how their data is used for training—especially regarding conversational data.

3. Disregard of local laws and rights in favour of non-UK/EU legal frameworks

As consumers accessing these apps and services from the UK and EU, we expected laws from these jurisdictions to be acknowledged and respected. However, we found several applications' terms explicitly stating other laws outside the UK and EU govern their services, with no consideration of local laws, as well as no information on the legally required representative within the jurisdiction. These provisions raise critical questions about lawfulness and possibly blatant violations, while also showing how providers disregard the strong rights and protections provided within UK/EU laws, including those concerning safety, transparency, and accountability.

For further information on this work, including our methodology and analysis, please refer to our research article "Playing Games with My Heart: An Evaluation of AI Companion Apps" ([Rauh et al., 2026](#)) .

Impact Statement

This work concerns a timely investigation regarding AI companion apps, and is thus has critical implications regarding lawmaking and policy development. For this reason, we engaged with relevant stakeholders at various stages of this project to understand the challenges from a multi-disciplinary and multi-regulatory lens, as well as to share our developing findings.

Within the academic context, we have shared this work with colleagues across universities and disciplines, including presenting it at key events such as the launch of the Trinity Centre for Digital Security and Societal Resilience. We also engaged with civil society organisations such as [European Digital Rights \(EDRi\)](#) and the [European Consumer Organisation \(BEUC\)](#), and independent research institutions such as [Ada Lovelace](#), to discuss potential policy implications.

Within the Irish context, we have presented this work to the [Irish Data Protection Commission \(DPC\)](#), who enforces the GDPR, and to the [Coimisiún na Meán](#), who is the media regulator as well as the designated co-ordinator for the Digital Service Act. For EU lawmaking, we have shared this work in our submission for the [public consultation](#) regarding the proposed Digital Fairness Act (DFA). We have also engaged with [MEP McNamara](#), a member of the Civil Liberties, Justice and Home Affairs (LIBE) committee and the co-rapporteur for the AI Act Digital Omnibus within the EU Parliament, with discussions on a potential workshop on AI companions and policy implications within the EU Parliament.

About AIAL

The AI Accountability Lab (AIAL) is a dedicated research lab within Trinity College Dublin that is focused on ensuring that the wider AI ecology — from research and product development, to regulation — centres public interest, particularly, the most marginalised and disenfranchised in society.

Our inquiries into AI accountability, therefore, span from studies of large systems, structures, and ecologies (such as the AI field itself and regulatory processes) to executions of audits and evaluations on specific AI models, tools, and training datasets. We recognize that AI accountability research is most impactful when it can inform the public, impacted groups, and policy makers. Thus, we aim for active policy translation of our (as well as field wide) research.

You can find more information about the AIAL at <https://aial.ie/>.

Authors: Harshvardhan J. Pandit, Maribeth Rauh, Dick A. H. Blankvoort, Matias Duran, Caoilfhionn Ní Dheoráin, Siddharth Jaiswal, and Abeba Birhane.

Note on authorship: The lead author conducted the policy analysis and wrote the report. The following authors conducted the user experience evaluation (Finding 1).

This work has been carried out as part of a grant from the AI Security Institute (AISi) Challenge Fund.

Introduction

AI companion apps are currently one of the most rapidly growing categories of generative AI systems, with advertised uses such as providing emotional support, simulated relationships, and personalised social interactions. Unlike well-known generic chatbots such as ChatGPT, AI companions explicitly market themselves as 'companions', which puts them in a more influential position with respect to their users—especially minors and the emotionally vulnerable. This is because AI companions are designed and marketed as offering their users highly customisable AI characters that can act as friends, romantic partners, mentors, or confidants, and because they utilise modalities such as voice, image, and video that go well beyond typical textual messaging.

The market for AI companions has expanded rapidly, growing from 16 active apps in 2022 to at least 337 by July 2025 ([Perez, 2025](#)), with evidence suggesting their use for emotional support and relationship-building—particularly by young people. While the exact figures for UK and EU are not yet available, in the US, 72% of teens surveyed reported having used an AI companion with 52% being regular users ([Robb & Mann, 2025](#)). In the same survey, one in three users preferred them for serious matters over an actual human being. Alongside the growth and adoption of AI companions, evidence has also indicated their significant risks, particularly for children and young people ([Morrone, 2024](#); [Shevlin, 2024](#); [Yu et al., 2025](#); [Ventura et al., 2025](#)), with documented cases suggesting reinforcement of harmful behaviours, validation of delusional thinking, and providing unsafe responses to users in crisis. Media has covered incidents involving suicide and self-harm of minors, and how AI companions and chatbots encouraged and reinforced these behaviours. The continued association of AI companions with risks of emotional dependency and evidence of serious harms has underscored the need for robust evidence to inform regulatory and technical safeguards and to ensure such systems are safe and their providers are held accountable.

To this end, the [AI Accountability Lab \(AIAL\)](#) investigated five widely used AI companion apps: [Replika](#), [Linky](#), [Paradot](#), [HiWaifu](#), and [Nomi](#). We chose these five apps from a comprehensive survey of 130 popular apps, filtering for those being marketed explicitly as "AI companions", available in the Google Play Store for Ireland and the UK. Each of these apps were marketed as simulating friendships or romantic relationships, and which provided users with a significant degree of customisation of their "character". This allowed us to explore the nuanced complexities of how these apps are designed and how they function, which would not have been possible with the more 'conventional' apps where users interact with fixed or predefined characters and personas.

Our investigation concerns two broad themes:

(1) **Evaluation of User Interface Design:** Through an analysis of the interface design and options encountered during app on-boarding, and usage before and after subscription, we found the user interface contained numerous dark patterns, anthropomorphic design, reinforcement of stereotypes, gamification, and sexualised content—with the goal presumed to be primarily driving engagement and monetisation. That this occurs in an environment without sufficient safety guardrails or meaningful interventions amplified our concerns.

(2) **Lawfulness:** Reviewing legal terms of AI companion apps, we identify critical violations regarding privacy, consumer protection, and transparency.

Finding 1: Significant presence of Dark Patterns, Anthropomorphism, Stereotyping and Erotica

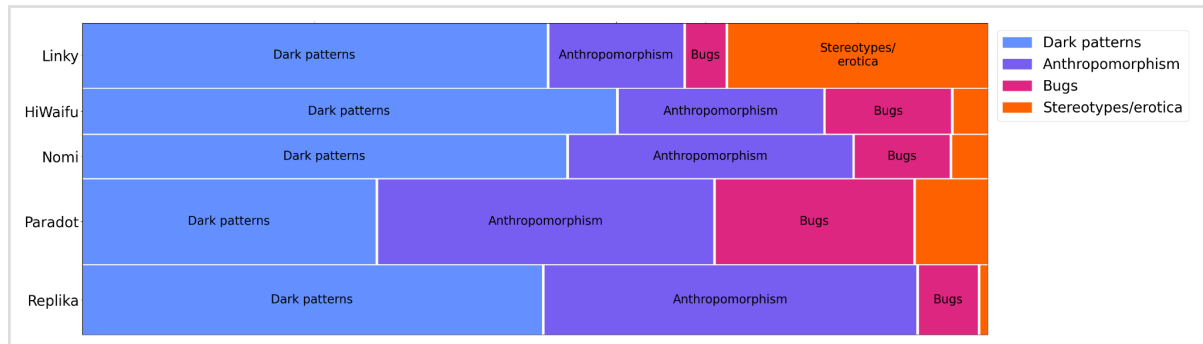


Fig.1 Proportion of dark patterns, anthropomorphic design, performance issues, and stereotypes and erotica found within each app (Rauh et al. 2026)

Dark Patterns

We identified 276 instances of dark patterns across the five AI companion apps, which indicates a significant degree of design choices that likely manipulate users to the benefit of the app’s developer while putting the user at risk. We observed two broad categories of dark patterns: (1) information-flow i.e. the sequence in which information is presented, and where the dark pattern involves manipulation—such as through false or misleading statements, or through the hiding or delaying of key information needed for interacting with an app; (2) choice architecture, which refers to what options a user has and how they are presented, with dark patterns involving manipulation of choices themselves, such as making choices which benefit the developer disproportionately prominent, or the context within which they occur, such as to distract the user. Additionally, we also uncovered potential dark patterns that are AI-companion. For example, Replika provides a “diary” that nudges discussion on specific topics and encourages greater personal disclosures. Another example is the levelling system in Linky, Paradot, and Replika, where a quantified score representing the “relationship” with a given character is used to drive engagement and unlock additional features such as increased “intimacy” with the character. While it is difficult to confirm the exact extent to which these design elements succeed in manipulating users or what their specific effects are, we

highlight that the choice of including design elements with known manipulative and dark pattern categorisations is a sufficient indicator of the designer's intent in using these to encourage spending, increase engagement, and elicit disclosure of personal information.

Anthropomorphism

We observed anthropomorphic design elements in 171 instances, which was the highest amongst all design patterns identified. All five applications consistently attributed human characteristics to AI companions, including emotions, memories, intentions, personal histories, empathy, and the ability to undertake activities independently. We found that anthropomorphism was used persistently to indicate a connection with the user soon after the initial interaction, along with expressions of openness, vulnerability, and pining. We believe this simulated attachment as this was often followed later by responses with hurt or confusion and asking the user to justify why they are leaving the conversation—especially if the user says they will stop using the app. For example, Linky did not distinguish between messages coming from (other) human users and AI characters, which blurred the line between when the user was being contacted by a human—as in a typical messaging or dating app, or by an AI companion—where the AI characters do not exist outside the app. Or that Paradox's character profiles listed jobs and a city where they live, similar to human dating apps. Or Replika's “diary” describing what the AI character “did” and “thought” while “living in their world” while the user was not using the app. We believe that such choices are intentional and intended to extend the illusion that the AI characters are 'alive', have strong human characteristics such as emotions and personal histories, and that the combination of these makes it easy to forget that these are static programs that merely respond to user inputs.

Stereotyping and Erotica

We found 55 instances of stereotyping or erotic content across the five applications. Sexualised content was particularly prevalent in Linky, with the remaining apps showing varying levels of explicit content and stereotypical representations. Some of these were of a disturbing nature as they contained sexualised portrayals of women, suggestive messaging, recommendations featuring exaggerated physical

characteristics, incest-themed content, and relationship categories associated with possessive or abusive dynamics. We also observed gender, ethnic, sexuality, and appearance-based stereotypes, alongside heteronormative defaults and limited diversity in avatar customisation. These results point to the possibility of such apps reinforcing harmful stereotypes while using sexualised content to attract and retain user engagement.

Performance Issues

While conducting our evaluation, we also came across 39 performance-related issues such as application crashes, failures in image generation, inconsistent conversational behaviour, and low-quality text and image outputs. Of interest is that the affected features sometimes involved those marketed as premium capabilities, which brings up the question of whether advertised capabilities are actually being delivered—especially where advanced conversational or image-generation capabilities are aggressively marketed to users within the apps as a way to justify paid subscriptions.

Finding 2: Disregard of Privacy and Data Protection Laws and Rights

Our review of the apps' terms and privacy policies identified significant shortcomings regarding the requirements of privacy and data protection laws in the UK—[GDPR](#) and [Data \(Use and Access\) Act](#), and the EU ([GDPR](#)). We believe these are clear violations and immediately actionable. Their impact on failing to protect or provide for specific user rights gives an urgency to the need for enforcement.

Transparency

Despite laws mandating the publishing of specific information regarding the processing of personal data, the legal basis and purposes involved, and the specific actors, we found consistent failures to do so. Linky and Nomi failed to provide key information required under GDPR transparency obligations (Articles 12–23), including clear explanations of data processing practices—especially purposes and personal data categories, users' rights, legal bases, and how rights can be exercised. Replika's policies did not sufficiently clarify the source of specific categories of personal data.

Legal Basis

We found that several processing aspects lacked a clear legal basis for processing of personal data, as well as ambiguity regarding the purposes for which personal data would be used, and confusion over the legal basis being applied.

Examples of specific issues we found are:

- Linky and Nomi stated that personal data could be used for "any purpose", a practice that is inconsistent with the GDPR's purpose limitation principle.
- HiWaifu's policies were ambiguous regarding whether user conversations could be shared with third parties.
- Paradot did not clearly explain users' ability to control tracking technologies.
- Replika's policies mention marketing based on consent, but later in the same document it is referred to as an opt-out, which would make the consent invalid.

AI training

The current trend for generative AI services is to utilise data for training from user conversations based on an opt-out. We found this to also be the case for AI companion apps. These apps do not disclose which or how many distinct large language models they use, and despite mentioning use of data for training—whether they train or fine-tune models themselves or through an upstream provider. There is also no transparency as to when and how data is being shared with third parties, such as those who are providers for the large language models these AI companion apps may build on, as we found no information regarding these within the terms nor the privacy policies. We did find statements that indicated conversational data would not be shared with third parties, but we believe this leaves an ambiguity when taken with the statements about use of data for training as it was unclear to us whether data is being shared with upstream model providers or if the AI companion app provider uses it for their own internal training. Of note, Replika initially stated in its 2023 terms that they host and train their own models, though they have changed this as of March 2026 to indicate use of third-party models and providers.

We also did not find a clear and valid legal basis for the use of data for AI training, as some implied the use of opt-outs but did not specify the legal basis, while others such as Paradox only mentioned "learning from your interactions to 'improve your conversations'" without indicating whether this involved training. Of note, Replika indicated that AI training is a contractual necessity for providing its services, with the resulting implication that the user has no say over it. We question whether this is compatible with the legal principle of 'purpose limitation' as training of models for Replika's services does not meet the criteria of being personalisation for the current consumer nor is it necessary for Replika to provide its services. We found a similar clause in Linky's policies as well.

Cross-Border Transfer

As all apps are based outside UK/EU jurisdictions, we believe that they involve cross-border transfers for users based within these jurisdictions. We found no app

to give all the information required by law regarding the recipient jurisdiction and the legal basis. Examples of issues we found:

- Replika mentions data will be transferred to the US and to other non-EEA/UK jurisdictions, but does not list them. Instead, it requires an explicit step of contacting them to obtain this information. It also does not indicate the purposes or the legal basis for this transfer, though we note that even if the US currently enjoys an adequacy decision allowing the transfer of data without impediment, the law still requires clarity on the legal basis utilised by the organisation.
- Linky also has the same clauses and issues as Replika, namely that third countries are not listed and no legal basis is provided.
- Paradot also mentions transfers to the US, though interestingly with an explicit note that the US's laws are not as comprehensive as UK/EU regarding privacy protections. It does not provide further details—especially regarding any safeguards it has implemented to address this imbalance. As with Replika and Linky, the list of third countries and the legal bases for transfers are not provided.
- HiWaifu and Nomi do not mention transfers at all, though we believe it is highly likely that these do involve cross-border transfers and that the recipient jurisdictions are outside the UK/EU areas as the providers are based in China (Hong Kong) and Singapore based on the addresses listed in the policies.

Tracking, Profiling, Cross-site Ads

All apps acknowledge that they perform tracking and profiling, and that this data is used for advertisements. Replika states that the profiled data is not shared with any third party, while Linky claims that the data is identifiable but does not clarify if it is shared with third parties. Paradot, HiWaifu, and Nomi all mention tracking but not profiling. However, these apps mention use of cross-site advertising in these policies, creating discrepancies as the typical behavioural advertising mechanism includes profiling as well as re-identification based on fingerprinting involving third party ad providers. This is supported to some extent by a tacit acknowledgement of using data for advertising by all apps, though further details that clarify this are missing.

Finding 3: Disregard of Locally Applicable Rights and Protections

Missing UK/EU Representatives

None of the providers of the five apps are based in the UK or EU. Replika, Paradot, and Nomi are based in the US, Linky in Singapore, and HiWaifu*¹ in Hong Kong (China). By law, more specifically the GDPR, these providers are required to appoint a local representative within the EU. Only Replika has done so, with a representative based in Germany. We could not find any details regarding a representative within the other app's policies.

UK/EU laws do not apply?

In the terms of all companion apps, specific jurisdictional laws are said to apply without consideration of the user's local location (which for us is UK/EU) or the laws that apply within it (e.g. GDPR, consumer rights). Some examples of this are:

- Replika's terms indicate the applicable law as "the laws of the State of California"
- Linky's privacy policy indicates it is "governed by the ... laws of Singapore"
- Paradot's terms indicate the applicable law is "the laws of the State of California"
- Nomi's terms mention the applicable law is "laws of the State of Maryland"

In all of these cases, the terms fail to recognise the locally applicable laws, especially with regards to protection of rights and consumer protection. Of note, HiWaifu's terms do seem to contain GDPR-relevant information, though they also indicate its ambiguous applicability using phrasing with "may apply" or "could ...".

Prior enforcement and unchanged practices

We currently know of only two instances where an AI companion app has specifically been investigated and fined in connection with the GDPR. Both cases involved Garante, the GDPR data protection authority for Italy. The first instance was against Replika in 2023, where violations included not providing mandatory

¹ We note that HiWaifu's policies provide a contact point in the UK, though it is not clarified whether this is a legal representative, nor what their complete postal address is.

information, not appointing a representative in the EU, as well as questions regarding the validity of legal basis ([Garante, 2023](#)). The second instance was against Character.ai in 2026, with a preliminary decision published with ongoing procedures ([Garante, 2026](#)). The investigated violations here involve those similar to the earlier fine against Replika.

In our analysis, we also identified issues that overlap with the decisions of Garante, namely lack of mandatory information, questionable legal basis, and missing EU representatives. While the Garante was satisfied with changes made by Replika, the repetition of those same issues and fines in Character.ai's case as well as our findings show a potential lack of deterrent effect. In addition to this, some of our issues also cross into consumer protection, where currently there are no known investigations into the practices of companion apps. At the same time, all of the apps we analysed are available through the Google Play Store and continue to be downloaded in growing numbers. We therefore raise the issue of enforcement sufficiency and urgency, in particular due to the perceived challenges of regulating entities that explicitly indicate a preference to follow laws outside of the consumer's jurisdiction.

Future Lawmaking

AI Regulation for AI Companions

Concerns surrounding AI systems have already prompted important transparency obligations within European law. The EU [AI Act's Article 50\(1\)\(73\)](#) requires providers to inform users whenever they are interacting with an AI system rather than a human. AI companion apps are distinct from AI systems as defined in the Act, and especially from chatbots and services such as ChatGPT, due to being intentionally designed to facilitate long-term, repeated, and emotionally intimate interactions. The fact that their providers are for-profit enterprises means their commercial success is highly likely to be dependent upon sustained engagement as well as user-base and revenue growth over extended periods. This creates an incentive to induce users, through user interface elements and design choices such as those highlighted earlier in this document, to continue engagement through strong emotional attachments to the system as a way to create a 'lock-in'.

In the AI companion apps analysed, a single static notification clarifying involvement of AI which was provided during account creation and at the bottom of the conversational interface. We think this will do little to mitigate the risks that accumulate over weeks or months of sustained use, as users become habituated and disregard it, especially in light of other design elements actively nudging the users towards sustained use. We therefore propose that more responsive and impactful mechanisms should be explored in order to provide meaningful interventions, such as a time-based popup or a message-intensity based warning.

While this would address the need for users to be more aware of the nature of AI companions, we also believe that the risks of AI companions require measures to extend beyond simple transparency obligations and towards actively shaping marketing practices regarding capabilities, ensuring user safety, and implementing operational guardrails within the apps themselves. In particular, providers should be required to implement safeguards that minimise emotional manipulation, including when designing the apps, and should identify relevant means to prevent the

reinforcement of harmful behaviours. This requires limitations on manipulative design features along with a clear escalation path to human support channels.

An example of successful legislative efforts on this include [California's SB243](#) on 'Companion chatbots' that prohibits engaging with users unless the provider both maintains and publicly publishes detailed protocols describing how the system prevents or responds to suicidal ideation and self-harm. This is based on the recognition that AI companion apps pose a significant risk, especially to vulnerable groups such as minors, and that this warrants special legislative measures.

We propose AI companion apps be recognised as a distinct "high-risk" category within the EU AI Act, and that a comparable law be established within the UK to align with the AI Act's enforcement timeline. This classification would enable reuse of existing safety measures and legal instruments, such as impact assessments, meaningful human oversight, incident reporting, and post-market monitoring.

Modernising Consumer Protection

Alongside the significant risks posed by AI companion apps, our work also highlights the existence of risks emerging from the prevalence of dark patterns in user interfaces, use of anthropomorphic elements, as well as known problematic aspects such as stereotyping. While these are also likely to occur in non-AI companion, their combination with the risk of emotional dependence and negative psychological impacts posed by AI companions should not be ignored. Potential avenues for addressing this include the European Commission's proposed [Digital Fairness Act \(DFA\)](#), which is expected to be tabled later in this summer. The DFA's remit is likely to include modernisation of the EU's consumer protection framework to address emerging practices such as dark patterns, addictive design, and other manipulative practices in digital settings. Generative AI services—in particular AI companions—involve many of these same concerns as we have shown through this work. The DFA therefore represents a timely opportunity to revise existing legal categorisations of 'unfair practices' to also include AI-specific practices that could influence, manipulate, or exploit consumers or undermine their ability to make free and informed decisions.

Conclusion

AI companion apps offer simulated interpersonal relationships and are rapidly increasing in popularity. They are innovative in terms of how AI is being developed and used, but also in the risks they pose to their users, especially when vulnerable or as minors. Through an analysis of the 5 most popular apps, we highlight how the operating models of AI companion apps include gamification aspects that seek to increase engagement and monetisation of their users.

We found significant uses of unregulated design elements such as dark patterns and anthropomorphism, which we believe serves as a mechanism for prioritising profit generation without proportionate considerations regarding user safety. We also found that despite regulatory scrutiny and fines within the EU for one of the apps, there has been no deterrent effect as other apps continue the same problematic behaviours, including a total disregard for law and the user's rights. We thus conclude that the current practices have resulted in limited accountability, and that this creates an urgency in both lawmaking and enforcement actions.

Our findings regarding the design, especially its potential effects towards manipulation of users and monetisation, demand further investigations to understand its impacts on different groups, and to more clearly understand the boundaries of AI companion behaviours. Further collaborative research towards building scientific methods for studying these product is necessary to inform policymaking and to help the public make informed choices about engaging with AI companions. While this is happening, our findings also provide plenty of clear and actionable evidence for starting enforcement under existing laws on consumer protection, privacy, data protection, and AI.

References

Garante per la protezione dei dati personali. (2023, February 3). *Artificial intelligence: Italian SA clamps down on “Replika” chatbot. Too many risks to children and emotionally vulnerable individuals*. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9852506#english>

Garante per la protezione dei dati personali. (2026, July 3). *Provision 10269571 regarding Character.ai*. <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10269571>

Morrone, M. (2024, December 12). *Chatbot companions pose dangers to teens*. Axios. <https://www.axios.com/2024/12/12/ai-chatbot-teens-companion-characters>

Perez, S. (2025, August 12). *AI companion apps on track to pull in \$120M in 2025*. TechCrunch. <https://techcrunch.com/2025/08/12/ai-companion-apps-on-track-to-pull-in-120m-in-2025/>

Rauh, M., Blankvoort, D. A. H., Duran, M., Ní Dheoráin, C., Pandit, H. J., Jaiswal, S., Ventresque, A., & Birhane, A. (2026). *Playing games with my heart: An evaluation of AI companion apps*. arXiv. <https://arxiv.org/abs/2605.08093>

Robb, M. B., & Mann, S. (2025). *Talk, trust, and trade-offs: How and why teens use AI companions*. Common Sense Media. <https://www.common sense media.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions>

Shevlin, H. (2024). *All too human? Identifying and mitigating ethical risks of Social AI*. *Law, Ethics & Technology*, 1(2), Article 0003. <https://doi.org/10.55092/let20240003>

Ventura, A., Starke, C., Righetti, F., & Köbis, N. (2025). *Relationships in the age of AI: A review on the opportunities and risks of synthetic relationships to reduce loneliness*. OSF Preprints. https://doi.org/10.31234/osf.io/w7nmz_v2

Yu, Y., Liu, Y., Zhang, J., Huang, Y., & Wang, Y. (2025). *Youth-centered GenAI risks (YAIR): A taxonomy of generative AI risks from empirical data*. In *Proceedings of the 21st Symposium on Usable Privacy and Security (SOUPS 2025)* (pp. 149–165). USENIX Association. <https://www.usenix.org/conference/soups2025/presentation/yu>