Social and Emotional Uses of AI

Emily Tseng
Microsoft Research
New York, New York, USA
Human Centered Design & Engineering
University of Washington
Seattle, Washington, USA
emtseng@uw.edu

Daniel A. Adler
Information Science
Cornell Tech
New York, New York, USA
Computer Science and Engineering
University of Michigan
Ann Arbor, Michigan, USA
adlerdan@umich.edu

Ashley Marie Walker
Google
New York, New York, USA
amwalker@google.com

Renee Shelby
Google Research
San Francisco, California, USA
reneeshelby@google.com

Stevie Chancellor
Department of Computer Science & Engineering
University of Minnesota
Minneapolis, Minnesota, USA
steviec@umn.edu

Eugenia Kim
Microsoft
New York, New York, USA
eugeniakim@microsoft.com

Sachin R. Pendse
School of Medicine
University of California, San Francisco
San Francisco, California, USA
sachin.pendse@ucsf.edu

Renwen Zhang
Wee Kim Wee School of Communication and Information
Nanyang Technological University
Singapore, Singapore
renwen.zhang@ntu.edu.sg

Abstract

More and more people look to generative AI for social and emotional support — presenting profound interpersonal and societal risks. In this workshop, we invite HCI researchers across the subcommunities of digital safety, digital mental health and well-being, and responsible AI to come together and articulate a shared research agenda for HCI to lead the design, governance, and safeguarding of social and emotional uses of AI. Workshop participants will engage in a series of talks and group discussions focused on defining and addressing foundational, methodological, and translational challenges towards safer AI use.

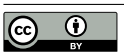
CCS Concepts

• **Human-centered computing** → **HCI design and evaluation methods**; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → *Psychology*.

Keywords

ethics; responsible AI; trust and safety; mental health; well-being; chatbots

ACM Reference Format:
Emily Tseng, Daniel A. Adler, Ashley Marie Walker, Renee Shelby, Stevie Chancellor, Eugenia Kim, Sachin R. Pendse, and Renwen Zhang. 2026. Social and Emotional Uses of AI. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3772363.3778699



CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3778699

1 Motivation

A growing number of users are turning to conversational AI systems for social and emotional support—a practice we term *social and emotional uses of AI*. In these interactions, users treat AI less as a search engine or content generation tool and more as a conversational partner. General-purpose AI developers, such as OpenAI [16] and Anthropic [2], have recently reported rising numbers of this broad class of uses (which they term "affective use" or "affective conversations")[1]. Researchers have also reported rising use of more specialized platforms, like AI tools for therapy [15] and companionship bots [18]. Many users perceive these systems to offer a uniquely safe and non-judgmental space for personal reflection [35]. Research suggests that users can feel more comfortable disclosing personal information to chatbots than to humans [8, 19, 22], and view them as ever-present confidantes when they perceive limited community [39].

While social and emotional uses of AI holds some promise for addressing users' felt need for connection, the practice also presents profound risks — individual, relational, and societal. AI-human feedback loops can subtly alter human perceptions, emotions, and social judgments [14]. Already, there are reports around the world of people experiencing what some characterize as 'delusions' that lead to psychosis following extended AI use ("*AI psychosis*") [28], using

[1]We choose not to use the term "affective" as all interactions, including information seeking and task-oriented use cases can also carry user affect, such as emotions like frustration, excitement, and anxiety, among many others.

AI to escalate arguments with their significant others [10], and even developing romantic attachments to their AI [38]. More alarmingly, the use of AI has been implicated in several tragic mental health crises, including teen suicides that have led to ongoing lawsuits against AI developers [9, 30].

Social and emotional use thus presents an urgent challenge to the design, governance, and safeguarding of AI. Risks ranging from dependency and adverse social norms to emotional manipulation are enmeshed in the interplay between human psychology [26] and AI's rapid societal integration [5]. The scale of this integration is stark: OpenAI reported in July 2025 that it has 700 million weekly active users, and that approximately 11% of messages sent on the platform are purely for *expressing* "views or feelings...not seeking any information or action" [5, p. 3]. While much-needed attention is focused on preventing acute individual harms arising from the "pseudo-intimacy" of parasocial relationships [36], the interpersonal and societal risks also demand scrutiny. However, existing approaches to AI safety are insufficient: they rely on static detection and mitigation approaches like benchmarking [25], single-turn evaluation [33], and post-training assessment [13] that lack the context-specificity and dynamism necessary to assess AI's interpersonal and societal effects [24]. Furthermore, AI regulation offers narrow individual protections for safe use, transparency, and data privacy [31], lacking consideration of individuals' offline relational capacities [23]. The Human-Computer Interaction (HCI) community, with its long history as an interdisciplinary and human-centered field, is well-positioned to lead research and translational efforts in this area. Doing so, however, requires uniting HCI sub-communities that have developed independently, but have cross-cutting interests and expertise: the subfields of digital safety, digital mental health and well-being, and responsible AI.

We therefore propose a workshop at CHI 2026 bringing together researchers from the HCI sub-communities of digital safety, digital mental health and well-being, and responsible AI, to share knowledge and develop a shared research agenda around social and emotional uses of AI. The workshop will organize around three core challenges:

- **Foundational**: Social and emotional uses of AI encompasses a vast range of human behaviors: from asking a chatbot to be your therapist, to discussing relationship problems, to treating it as a romantic partner [24]. Building a coherent field of study requires the development of shared conceptual frameworks and taxonomies to understand these diverse interactions. Mental health has models focused on individual emotion regulation [3] and appropriate levels of validation [20]; digital safety has models focused on at-risk users [21, 34] and safety engineering [27]; and responsible AI has models focused on sociotechnical harms [17, 29], quantifiable safety metrics [1] and human-AI complementarity [6]. How can we integrate models across these subcommunities to foster more complete knowledge?
- **Methodological**: Current AI evaluation approaches are largely focused on single-turn prompt pairs or limited multi-turn interactions [33], with some long multi-turn and multi-session analysis in ML focused on simulation [37] (although the reliability and soundness of these methods are underexamined). How can we design studies that capture harms and effects as they unfold over multiple, long-term interactions? How do we develop holistic, human-centered methods that engage both with system log analysis, the user's full situational context, and societal effects? Furthermore, how can we differentiate between the effects of product features (e.g., tonal characteristics like sycophancy [7, 11]) versus the more fundamental impacts of the technology itself?
- **Translational**: A primary goal is to bridge foundational understanding with practice. How do we translate research insights into concrete and operationalizable principles for the design of safer AI systems [32]? How can our work inform the development of effective pre- and post-training techniques, AI guardrails, content policies, and dynamic governance frameworks that are both evidence-based and adaptable to this fast-changing risk landscape [12]?

Through facilitated knowledge sharing, cognitive mapping, and collaborative work time, this workshop will create a research agenda for HCI to meet the moment, and enable us all to understand, design, develop, and evaluate social and emotional uses of AI in ways that enhance human connection and flourishing.

2 Length of the Workshop

We will hold a **long** workshop, anticipating two sessions, 90 minutes each, with a break in between.

3 Organizers

Our organizing team spans HCI leaders in digital safety, digital mental health and well-being, and responsible AI, working across a range of industry and academic standpoints.

Emily Tseng is a postdoctoral research scientist at Microsoft Research and will soon join the faculty of the University of Washington in Human Centered Design & Engineering. Trained as a scientist-advocate for survivors of tech abuse, Emily's recent interests include AI's role in interpersonal norms and harms, toolkits for community-controlled AI, and sociotechnical interventions for digital safety. She has organized multiple successful workshops at ACM venues including CHI, CSCW, and FAccT.

Daniel A. Adler is a postdoctoral fellow at Cornell University, and will soon join the faculty at the University of Michigan in Computer Science and Engineering. Dan's expertise lies in designing, developing, and evaluating novel AI systems by studying their applications in mental healthcare. He is interested in unifying models of AI evaluation to assist individuals using AI for social and emotional support. He has organized multiple workshops at ACM UbiComp and frequently attends workshops at CHI.

Renee Shelby is a sociologist in Google Research studying the design, experience, and social impacts of AI systems to develop evidence-based interventions for models, policy, and design. Her current work explores the harms of algorithmic systems and how people interact with generative AI, particularly in sensitive and high-risk use cases. Prior to Google, she spent seven years as an applied researcher focusing on gender-based violence and child abuse, working towards systems change within the criminal legal system in the Southeast.

Ashley Marie Walker is a researcher at Google working in Trust & Safety on empirically-driven, effective improvements to global digital safety. Their work includes working with external experts to understand the emerging harms landscape, including through workshops hosted at ACM venues and beyond.

Stevie Chancellor is an Assistant Professor in Computer Science & Engineering at the University of Minnesota - Twin Cities. She studies building human-centered artificial intelligence for mental health in two areas: AI-driven social media and generative AI. Her work bridges methods from HCI, NLP, and ethics to develop empirically-grounded and human-centered interventions in this space. She has organized numerous workshops at CHI, CSCW, and ICWSM, and served as the 2023 CSCW Workshops Co-Chair.

Eugenia Kim is an AI Safety Researcher at Microsoft. Her work involves developing safe, secure, and reliable AI systems primarily through an offensive security lens of red teaming. Her work focuses on surfacing boundary pushing model behaviors and blind spots that standard evaluations may overlook, with the goal of building safer and more trustworthy AI. This spans across different modalities, psychosocial harms, and offensive security research.

Sachin R. Pendse is an Assistant Professor in the School of Medicine at the University of California, San Francisco, where he directs the Technology, Mental Health, and Society (TeMHSo) Lab. He studies the role that societal factors play in how people understand their mental health experiences and look for support online, towards designing safer AI-mediated mental health support technologies. His work is interdisciplinary, spanning methods from HCI, computational social science, and NLP. He has organized successful workshops at multiple ACM venues including CSCW, CHI, and FAccT.

Renwen Zhang is an Assistant Professor in the Wee Kim Wee School of Communication and Information at Nanyang Technological University, Singapore. As the director of the SWEET Lab, her research examines how digital technologies, such as social media and AI chatbots, mediate and impact well-being and social relationships. Renwen's recent work focuses on the processes, benefits, and harms of human-AI relationships. She has organized multiple workshops at ACM CHI and CSCW.

4 Plans to Publish Workshop Proceedings

Due to the anticipated sensitive nature of the conversation, we will not be making workshop submissions publicly available. Given that this is an emergent topic area, where core arguments are still developing, participants' submissions are likely to reflect the state of play at the time of submissions, but may not necessarily represent the long-term stance of the field. Not publicly publishing workshop proceedings will help give researchers space to present in-process ideas that may not yet be ready for full public scrutiny. The goal of the workshop will be to stress test and develop these ideas further. More fully-fleshed arguments will be presented in the after-workshop materials and in anticipated peer-reviewed publications that will evolve from workshop conversations.

5 Accessibility

We are committed to making this workshop accessible and inclusive for everyone. Participants will be asked about their specific accessibility needs when they are accepted to join the workshop, and the organizers will work directly with participants to discuss how best to meet them. Depending on the specific accessibility needs that arise through this process, the organizers will liaise with the CHI workshop chairs to facilitate them.

6 Offline Materials

To foster an environment of open and candid discussion, this workshop will be conducted under the Chatham House Rule. This will allow attendees to use the information discussed, but preserve anonymity for generative discussion on sensitive topics. To further protect potential vulnerabilities of researchers conducting research on "at-risk" user groups [4], the original workshop submissions will remain confidential and not be made publicly available. Following the workshop, we plan to collaboratively produce a public-facing summary and/or an article for *Interactions*. All participants will be invited to contribute to these materials.

7 Workshop Activities

Interested participants will submit a one-page ($\leq$500 words, or similar effort in another format, such as a short slide deck, infographic, or 5-min audio recording), single-authored proposal. This proposal will detail their interest in the topic area (e.g., open questions, research gaps, opportunities for intervention, possible metrics), as well as their relevant experiences (e.g., connections to previous research, potential methods contributions, cross-disciplinary opportunities) to ground the conversation. We strongly encourage junior researchers (e.g., PhD students) to submit proposals.

The overarching goal of the shared workshop time is to identify knowledge gaps in the emerging study of the social and emotional uses of AI to develop a prioritized research agenda. Safeguarding AI systems to limit potential harms that come along with social and emotional use cases will require developing a systemic risk literacy, identifying basic dynamics, potential interventions, and metrics for understanding efficacy. Bringing together the community of researchers who are interested in this topic area will help build connections between ongoing research in this quickly developing space to better synthesize actionable insights.

Each of the proposed sessions of this workshop are designed in sequence to: develop a shared understanding of the topic area, make connections with other researchers in the space, and hold small group discussions on open questions and research ideas that arise day-of. By sequencing the day's events this way, each session can build on the outcomes of the previous session, leading to actionable next steps that can extend beyond the end of the conference. A summary of the workshop activities can be found in Table 1.

8 Post-Workshop Plans

The goal for the workshop is to work towards actionable next steps to synthesize and develop research on social and emotional uses of AI. As such, plans for after the workshop will be emergent based on the desired goals of workshop attendees. We anticipate that potential outcomes might include facilitating journal special issues, co-authoring an *Interactions* article, or another format of public scholarship. Given that the goal is to foster community among attendees and facilitate further collaboration after the event, we will

Table 1: Planned activities

Session	Time	Activity	Description
Session 1 (90 min total)	10 min	**Opening remarks**	The organizing team will introduce themselves and the plan for the workshop
	15 min	**Shared sensemaking, Talk 1**	Social and emotional uses of AI from a digital safety lens
	15 min	**Shared sensemaking, Talk 2**	Social and emotional uses of AI from a mental health and well-being lens
	15 min	**Shared sensemaking, Talk 3**	Social and emotional uses of AI from a responsible AI lens
	35 min	**Network building quick hits**	Meet other researchers
Break			
Session 2 (90 min total)	35 min	**Collaborative work time 1**	Attendees will split into small groups to discuss open questions and ideate on research to address these questions
	35 min	**Collaborative work time 2**	Attendees will rotate to a different small group to discuss open questions and ideate on research to address these questions
	20 min	**Closing remarks**	The organizing team will reconvene the group, recap discussions, and share future plans

offer options for participants to continue to engage in the conversation. This may look like an email list for future local gatherings and workshops at other conferences, reaching out to small discussion groups after the conference to continue working towards shared goals, and shepherding potential collaborative projects. If enough demand, we will consider hosting a virtual gathering to reconnect and include participants who could not attend in-person in Barcelona.

9 Call for Participation

More and more people are looking to conversational AI chatbots for social and emotional support – what we term social and emotional uses of AI. However, such uses present profound risk. AI can alter human perceptions, emotions, and judgments, leading individuals to experience delusions and engage in harmful behaviors following extended use. These emerging and rapidly evolving risks present urgent interpersonal and societal challenges, and the HCI community has an opportunity to lead research and translational efforts in this area towards safer AI use. In this long workshop (two 90 minute sessions), we invite HCI researchers across the subcommunities of digital safety, digital mental health and well-being, and responsible AI to come together and articulate a shared research agenda around the design, governance, and safeguarding of AI for social and emotional use. Interested participants can submit a one-page, single authored proposal (≤500 words, or similar effort in another format) to detail their interest in the topic area and relevant experiences to ground the conversation. We strongly encourage junior researchers (e.g., PhD students) to submit proposals. Instructions to submit proposals can be found on our workshop website (https://sites.google.com/view/socialemotionalai). We do not plan to publish accepted proposals due to the sensitive nature of this topic. Authors of accepted proposals must attend and register for the workshop.

10 Expected Size of Attendance

We plan to have a maximum of 40 participants.

Acknowledgments

D.A. is supported by Neuromatch and the Wellcome Trust. S.C. is supported by a Center for Advancing Safety of Machine Intelligence grant, awarded by Northwestern University.

References

[1] Lama Ahmad, Sandhini Agarwal, Michael Lampe, and Pamela Mishkin. 2025. OpenAI's Approach to External Red Teaming for AI Models and Systems. doi:10.48550/arXiv.2503.16431 arXiv:2503.16431 [cs]

[2] Anthropic. 2025. How people use Claude for support, advice, and companionship. https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship. Accessed on September 27, 2025.

[3] Judith S. Beck. 2021. *Cognitive behavior therapy: Basics and beyond. 3rd Edition.* The Guilford Press, New York, NY, USA.

[4] Rosanna Bellini, Emily Tseng, Noel Warford, Alaa Daffalla, Tara Matthews, Sunny Consolvo, Jill Palzkill Woelfer, Patrick Gage Kelley, Michelle L. Mazurek, Dana Cuomo, Nicola Dell, and Thomas Ristenpart. 2024. SoK: Safer Digital-Safety Research Involving At-Risk Users. In *2024 IEEE Symposium on Security and Privacy (SP).* 635–654. doi:10.1109/SP54263.2024.00071

[5] Aaron Chatterji, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. *How People Use ChatGPT.* Working Paper 34255. National Bureau of Economic Research. doi:10.3386/w34255

[6] Valerie Chen, Q. Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. 2023. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (Oct. 2023), 370:1–370:32. doi:10.1145/3610219

[7] Myra Cheng, Sunny Yu, Cinoo Lee, Pranav Khadpe, Lujain Ibrahim, and Dan Jurafsky. 2025. Social Sycophancy: A Broader Understanding of LLM Sycophancy. *arXiv preprint arXiv:2505.13995* (2025).

[8] Emmelyn A J Croes, Marjolijn L Antheunis, Chris van der Lee, and Jan M S de Wit. 2024. Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and its Impact on Emotional Well-Being. *Interacting with Computers* 36, 5 (06 2024), 279–292. doi:10.1093/iwc/iwae016 arXiv:https://academic.oup.com/iwc/article-pdf/36/5/279/58823350/iwae016.pdf

[9] Michaeleen Doucleff and Pien Huang. 2025. AI chatbots are a lifeline for some teens. Are they safe? NPR. https://www.npr.org/sections/shots-health-news/2025/09/19/nx-s1-5545749/ai-chatbots-safety-openai-meta-characterai-teens-suicide Accessed on September 27, 2025.

[10] Maggie Harrison Dupre. 2025. ChatGPT Is Blowing Up Marriages as Spouses Use AI to Attack Their Partners. https://futurism.com/chatgpt-marriages-divorces. *Futurism* (sept 2025).

[11] Aaron Fanous, Jacob Goldberg, Ank A Agarwal, Joanna Lin, Anson Zhou, Roxana Daneshjou, and Sanmi Koyejo. 2025. Syceval: Evaluating LLM Sycophancy. *arXiv preprint arXiv:2502.08177* (2025).

[12] KJ Kevin Feng, Rock Yuren Pang, Tzu-Sheng Kuo, Amy Winecoff, Emily Tseng, David Gray Widder, Harini Suresh, Katharina Reinecke, and Amy X Zhang. 2025. Sociotechnical AI Governance: Challenges and Opportunities for HCI. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–6.

[13] Sorelle Friedler, Ranjit Singh, Borhane Blili-Hamelin, Jacob Metcalf, and Brian J Chen. 2023. AI Red-Teaming Is Not a One-Stop Solution to AI Harms:. *Data & Society* (2023).

[14] Moshe Glickman and Tali Sharot. 2025. How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour* 9, 2 (2025), 345–359.

[15] MD Romael Haque and Sabirat Rubya. 2023. An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth and uHealth* 11, 1 (2023), e44838.

[16] Jason Phang and Michael Lampe and Lama Ahmad and Sandhini Agarwal and Cathy Mengying Fang and Auren R. Liu and Valdemar Danry and Eunhae Lee and Samantha W.T. Chan and Pat Pataranutaporn and Pattie Maes. 2025. Investigating Affective Use and Emotional Well-being on ChatGPT. https://cdn.openai.com/papers/15987609-5f71-433c-9972-e91131f399a1/openai-affective-use-study.pdf Accessed on September 27, 2025.

[17] Kowe Kadoma, Danaé Metaxa, and Mor Naaman. 2025. Generative AI and Perceptual Harms: Who's Suspected of using LLMs?. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 861, 17 pages. doi:10.1145/3706598.3713897

[18] Linnea Laestadius, Andrea Bishop, Michael Gonzalez, Diana Illenčík, and Celeste Campos-Castillo. 2024. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society* 26, 10 (2024), 5923–5941.

[19] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. " I hear you, I feel you": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[20] Marsha M. Linehan. 1997. Validation and Psychotherapy. In *Empathy Reconsidered: New Directions in Psychotherapy*, Arthur C. Bohart and Leslie S. Greenberg (Eds.). American Psychological Association, Washington, DC, 353–392.

[21] Tara Matthews, Elie Bursztein, Patrick Gage Kelley, Lea Kissner, Andreas Kramm, Andrew Oplinger, Andreas Schou, Manya Sleeper, Stephan Somogyi, Dalila Szostak, Kurt Thomas, Anna Turner, Jill Palzkill Woelfer, Lawrence L. You, Izzie Zahorian, and Sunny Consolvo. 2025. Supporting the Digital Safety of At-Risk Users: Lessons Learned from 9+ Years of Research and Training. *ACM Trans. Comput.-Hum. Interact.* 32, 3, Article 22 (June 2025), 39 pages. doi:10.1145/3716382

[22] Elizabeth R. Merwin, Allen C. Hagen, Joseph R. Keebler, and Chad Forbes. 2025. Self-disclosure to AI: People provide personal information to AI and humans equivalently. *Computers in Human Behavior: Artificial Humans* 5 (2025), 100180. doi:10.1016/j.chbah.2025.100180

[23] Sarah Clark Miller. 2022. Toward a Relational Theory of Harm: On the Ethical Implications of Childhood Psychological Abuse. *Journal of Global Ethics* 18, 1 (Jan. 2022), 15–31. doi:10.1080/17449626.2022.2053562

[24] Jared Moore, Declan Grabb, William Agnew, Stevie Chancellor, Desmond C. Ong, and Nick Haber. 2025. Expressing Stigma and Inappropriate Responses Prevents LLMs from Safely Replacing Mental Health Providers.. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 599–627. doi:10.1145/3715275.3732039

[25] Inioluwa Deborah Raji, Emily M. Bender, Amandalynne Paullada, Emily Denton, and Alex Hanna. 2021. AI and the Everything in the Whole Wide World Benchmark. arXiv:2111.15366 [cs]

[26] Byron Reeves and Clifford Nass. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge, MA, USA.

[27] Shalaleh Rismani, Renee Shelby, Andrew Smart, Edgar Jatho, Joshua Kroll, AJung Moon, and Negar Rostamzadeh. 2023. From Plane Crashes to Algorithmic Harm: Applicability of Safety Engineering Frameworks for Responsible ML. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 2, 18 pages. doi:10.1145/3544548.3581407

[28] Sam Schechner and Sam Kessler. 2025. 'I Feel Like I'm Going Crazy': ChatGPT Fuels Delusional Spirals. https://www.wsj.com/tech/ai/i-feel-like-im-going-crazy-chatgpt-fuels-delusional-spirals-ae5a51fc. *Wall Street Journal* (aug 2025).

[29] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) *(AIES '23)*. Association for Computing Machinery, New York, NY, USA, 723–741. doi:10.1145/3600211.3604673

[30] Social Media Victims Law Center. 2025. Social Media Victims Law Center Files Three New Lawsuits on Behalf of Children Who Died of Suicide or Suffered Sex Abuse by Character.AI. Press Release. https://socialmediavictims.org/press-releases/social-media-victims-law-center-files-three-new-lawsuits-on-behalf-of-children-who-died-of-suicide-or-suffered-sex-abuse-by-character-ai/ Accessed on September 27, 2025.

[31] Tamar Tavory. 2024. Regulating AI in Mental Health: Ethics of Care Perspective. *JMIR Mental Health* 11 (Sept. 2024), e58493. doi:10.2196/58493

[32] Ashley Marie Walker, Renee Shelby, Ari Schlesinger, Emily Tseng, Mark Diaz, Andy Elliot Ricci, and Angela DR Smith. 2025. Designing Support for Systematic Sociotechnical Risk Literacy. In *Adjunct Proceedings of the Sixth Decennial Aarhus Conference: Computing X Crisis*. 1–4.

[33] Xingyao Wang, Zihan Wang, Jiateng Liu, Yangyi Chen, Lifan Yuan, Hao Peng, and Heng Ji. 2024. MINT: Evaluating LLMs in Multi-turn Interaction with Tools and Language Feedback. doi:10.48550/arXiv.2309.10691 arXiv:2309.10691 [cs]

[34] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L. Mazurek, Manya Sleeper, and Kurt Thomas. 2022. SoK: A Framework for Unifying At-Risk User Research. In *2022 IEEE Symposium on Security and Privacy (SP)*. 2344–2360. doi:10.1109/SP46214.2022.9833643

[35] Elizabeth Anne Watkins. 2025. What Happens When People Turn to Chatbots for Therapy? *Points* (jul 2025). https://datasociety.net/points/what-happens-when-people-turn-to-chatbots-for-therapy/ Accessed on September 27, 2025.

[36] Jie Wu. 2024. Social and ethical impact of emotional AI advancement: the rise of pseudo-intimacy relationships and challenges in human interactions. *Frontiers in Psychology* Volume 15 - 2024 (2024). doi:10.3389/fpsyg.2024.1410462

[37] Joshua Au Yeung, Jacopo Dalmasso, Luca Foschini, Richard JB Dobson, and Zeljko Kraljevic. 2025. The Psychogenic Machine: Simulating AI Psychosis, Delusion Reinforcement and Harm Enablement in Large Language Models. arXiv:2509.10970 [cs.LG] https://arxiv.org/abs/2509.10970

[38] Renwen Zhang, Han Li, Han Meng, Jinyuan Zhan, Hongyuan Gan, and Yi-Chieh Lee. 2025. The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 13, 17 pages. doi:10.1145/3706598.3713429

[39] Yutong Zhang, Dora Zhao, Jeffrey T Hancock, Robert Kraut, and Diyi Yang. 2025. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. *arXiv preprint arXiv:2506.12605* (2025).