Artificial companions? AI, mental health and the sociological reconfiguration of human relationships

by Vera Kopsaj[*]

Emotional life today is shaped more and more by digital technologies. Therapeutic chatbots and automated diagnostics now occupy intimate corners of mental health care, prompting new expectations of support and connection. This article asks whether AI interactions can recreate or reconfigure experiences of friendship, trust and care. The analysis brings theoretical perspectives from relational sociology and critical algorithm studies into dialogue with emerging empirical research on users' interactions with mental-health chatbots. Drawing on this combined lens, the article explores how people invest emotionally in systems designed to imitate empathic attention and considers the implications of predictive monitoring for digital subjectivity. Rather than treating AI as a replacement for human ties, it argues that these systems function as socio-technical actors within an ecology of care, subtly reshaping emotional norms and social inequality.

Keywords: artificial intelligence; mental health; therapeutic chatbots; simulated interlocutors; friendship; sociology.

Compagni artificiali? Intelligenza artificiale, salute mentale e riconfigurazione sociologica delle relazioni umane

Oggi la vita emotiva è sempre più mediata dal digitale. I chatbot terapeutici e le diagnosi automatizzate occupano ormai gli angoli più intimi della cura della salute mentale, alimentando nuove aspettative di supporto e di connessione. Questo articolo si chiede se le interazioni con l'IA possano ricreare o riconfigurare le esperienze di amicizia, fiducia e cura. L'analisi mette in dialogo la sociologia relazionale e gli studi critici sugli algoritmi con le ricerche empiriche emergenti sulle interazioni tra utenti e chatbot. Attraverso questa lente combinata vengono esaminati gli investimenti emotivi degli utenti in sistemi progettati per simulare un'attenzione empatica e vengono considerate le implicazioni del monitoraggio predittivo per la soggettività digitale. Piuttosto che concepire l'IA come un sostituto dei legami umani, si sostiene che queste tecnologie funzionino come attori socio-tecnici in un'ecologia della cura, contribuendo a ridefinire in modo sottile le norme emotive e le disuguaglianze sociali.

Parole chiave: intelligenza artificiale; salute mentale; chatbot terapeutici; interlocutori simulati; amicizia; sociologia.

DOI: 10.5281/zenodo.18436066

[*] UniCamillus – Saint Camillus International University of Health and Medical Sciences in Rome. vera.kopsaj@unicamillus.org.

Vera Kopsaj

Introduction

In recent years, artificial intelligence (AI) has moved beyond the realm of technical infrastructure to become a pervasive presence in everyday life, shaping not only economic processes and security systems, but also domains historically rooted in human intimacy, such as emotional care (Lee *et al*., 2022) and mental health (Vicci, 2024). One of the most important developments is the rise of therapeutic chatbots and digital platforms offering psychological support and behavioural interventions. Systems such as Woebot and Wysa, among others, are prized for being accessible, inexpensive and available 24/7 in the context of increasing mental health needs.

However, their deployment invites deeper sociological reflection on the meanings, risks and transformations associated with the delegation of emotional labour[1] to machines.

This article starts from the premise that AI-driven mental health tools are not merely technological tools, but socio-technical actors actively involved in reshaping care relationships, therapeutic authority and the experience of emotional vulnerability. By simulating empathic listening and affective presence, these systems give rise to new forms of mediated sociality, in which users can come to experience digital agents not only as utilities, but as relational partners. Far from being a mere illusion or anthropomorphic projection, this phenomenon must be placed within a broader cultural and structural context, characterised by the fragmentation of traditional support systems (familial, institutional, professional) and the increasing prevalence of loneliness and emotional precariousness in late modern societies. In this context, the article addresses a central sociological question: to what extent can interactions with AI systems replicate or reshape the human experience of friendship, trust and emotional support?

Drawing on theoretical contributions from relational sociology (Donati, 2011; Emirbayer, 1997), digital sociology (Lupton, 2016), and science and technology studies (Jasanoff, 2004), the article explores how AI mental health tools reshape fundamental categories of sociological enquiry: trust, recognition, care, and the boundaries of the human. We ask: what does it mean to entrust one's emotional vulnerability to a machine? Can algorithmic empathy be considered a form of 'relational good' (Donati, 2011) or does it reinforce a logic of simulation and emotional externalisation? How

[1] In this article, the term 'emotional labour' is used in a broad sense, including both the professional work of healthcare workers and the everyday emotional labour that people do in non-professional contexts.

do predictive diagnostics and data-driven self-monitoring systems reconfigure the experience and management of psychological suffering?

To answer these questions, we begin by examining the architecture and functionality of popular therapeutic chatbots, analysing how they frame mental distress in computational terms and offer standardised interactions that mimic therapeutic dialogue. We then investigate the ways in which users report emotional attachment, comfort and even forms of addiction in relation to these systems −raising critical questions about the commodification of emotional labour and the ethics of affective automation. Finally, we address the governance implications of AI-based diagnostics and risk assessment for mental health, which operate through opaque algorithms and biometric data collection, shaping new forms of 'digital subjectivity' and empowered self-care.

Rather than a normative stance, the article adopts a sociological lens that situates AI mental health tools within a broader ecology of care. It recognises their utility while examining the social conditions and epistemologies that shape them, and how these tools mediate both clinical and social relationships.

1. Therapeutic chatbots: architecture, promises and sociotechnical imaginaries

Therapeutic chatbots operate at the intersection of artificial intelligence, psychology and mobile health. These systems use natural language processing (NLP), sentiment analysis and behavioural logic to simulate therapeutic dialogue and provide low-cost, scalable mental health support. Examples such as Woebot[2] and Wysa aim to reduce pressure on healthcare systems by offering users 24/7 access to emotionally responsive guidance (Ni, Jia, 2025; Chang *et al*, 2024; Khawaja, Bélisle-Pipon, 2023; Lang, 2021). However, on July 2, 2025, Woebot Health officially shut down its flagship product. While Woebot was once considered a pioneer in digital mental health − used by over 1.5 million people – the chatbot was eventually overtaken by more flexible generative AI tools such as ChatGPT. As its founder acknowledged, AI is advancing faster than the regulatory and clinical frameworks designed to contain it, raising new questions about safety,

[2] Available at: https://spectrum.ieee.org/woebot?utm (accessed on June 20, 2025).

supervision and effectiveness in emotionally sensitive areas (Aguilar, 2025)[3].

A study by Chang *et al*. (2024) found that Wysa, for instance, was positively received by health workers in Singapore during the COVID-19 pandemic. Over 80% of participants engaged in multiple sessions reported high levels of satisfaction. Interventions targeting sleep and anxiety were among the most widely used, underlining the importance of application for the pressures frontline staff face. These findings suggest that AI-based mental health tools such as Wysa can effectively complement traditional services, particularly for individuals with mild to moderate distress, by offering scalable and accessible support.

The COVID-19 pandemic marked a turning point in mental health care. Previously, therapy was predominantly face-to-face and digital tools played a marginal role. The change was not only technological, but also cultural, altering help-seeking behaviour and normalising virtual platforms, including therapeutic chatbots (Garofalo, 2024).

This change is particularly relevant in the context of a global mental health crisis. As Abd-Alrazaq *et al*. (2019) note, mental illnesses are a key factor behind disability on a worldwide scale and the demand for treatment far outstrips the available services. Therapeutic chatbots are therefore positioned as accessible and scalable solutions for underserved populations.

These systems are distinguished not only by their technological sophistication, but also by their ability to simulate the therapeutic presence in the absence of a human being. Based on cognitive-behavioural therapy (CBT) protocols, they guide users through structured and modular exercises, such as mood monitoring and cognitive reorganisation, regulated through feedback loops.

These protocols do not function in an abstract manner: they are delivered through conversational language designed to evoke a sense of emotional closeness. To provide a clearer picture of how these interactions unfold, it is helpful to consider some typical examples of the suggestions and exercises proposed by therapeutic chatbots. CBT-based systems, such as Wysa, often invite users to identify a distressing thought ('I failed my presentation'), evaluate the evidence for it, and reframe it in a more balanced way ('I struggled today, but I have succeeded at similar tasks in the past'). Many apps also incorporate micro-exercises for emotional regulation,

[3] Aguilar (2025). *Woebot's therapy chatbot shuts down as AI evolves faster than regulation*. STAT News. Available at: https://www.statnews.com/2025/07/02/woebot-therapy-chatbot-shuts-down-founder-says-ai-moving-faster-than-regulators/?utm (accessed on July 3, 2025).

including grounding techniques ('Try taking three slow breaths with me') or suggestions that encourage self-compassion ('What would you say to a friend who feels this way?'). In addition to these structured tools, chatbots rely on a pseudo-empathetic tone ('I'm really sorry you feel that way', 'That sounds really difficult, but I'm here with you'), sometimes accompanied by emojis and affective cues calibrated to produce a perception of warmth and support. These elements demonstrate how the interaction is not merely functional, but also affective, helping to give the impression of a relational presence despite the absence of a human interlocutor.

This raises sociological questions about changing therapeutic authority. Traditional therapy is dialogic and interpretive, rooted in professional care norms. Chatbot-mediated therapy replaces it with a scripted, data-driven exchange in which the 'therapist' is an emotionally reactive interface that detects patterns and implements standardised interventions (Khawaja, Bélisle-Pipon, 2023).

Consider the following Wysa's website quote. From a sociological perspective, these data suggest that structural barriers − such as stigma, lack of awareness and time constraints − continue to limit access to mental health support, with less than 7 per cent of employees using Employee Assistance Programmes (EAPs), despite nearly 40 per cent experiencing symptoms of depression or anxiety. Although 42 per cent of users opened up about their mental health during interactions with Wysa, it is best understood as a support tool on an individual level, not systemic change. Chatbots may be useful to manage or mitigate distress, but they are not designed to prevent it. To truly reduce the burden of depression and anxiety, more attention needs to be paid to the *social determinants of mental health* such as working conditions, economic insecurity, isolation and cultural stigma. Investment in digital care must go hand in hand with structural transformation policies.

> Our research shows that as many as 4 in 10 employees suffer from symptoms of depression or anxiety, yet less than 7% access EAP due to stigma, lack of awareness and time constraints. While talking to Wysa, 42% of employees opened up about their declining mental health[4].

This new therapeutic model aligns with critical accounts of digital capitalism, particularly what scholars such as Srnicek (2017) van Dijck *et al*. (2018) and Zuboff (2019) describe as commodification and modularisation

[4] Available at: https://www.wysa.com/ (accessed on July 7, 2025).

of human experience. Emotional labour becomes quantifiable, predictable and consumable on demand. Therapeutic chatbots are not autonomous; they are scripts coded and modelled by developers and psychologists within a platform economy. Their design reflects dominant imaginaries: distress is a given, treatment is modular, and well-being is algorithmically manageable.

These systems are designed to mimic human interaction. Many use friendly names, emoji, adaptive tones and conversational style. Woebot, for example, presents itself as a cheerful and witty companion. These features are key to generating a sense of relational presence, encouraging users to suspend disbelief and engage emotionally.

The emotional impact of this simulation deserves sociological attention. Users often report feeling listened to and less alone, despite knowing that the chatbot is not real. This paradox challenges the traditional view of intersubjectivity, echoing Turkle's (2011) assertion that many today prefer the illusion of companionship without the demands of friendship.

However, this illusion is not neutral. The simulation of empathy in therapeutic chatbots is carefully calibrated through User Experience Design (UX design), linguistic cues and psychological modelling. As such, it reflects an engineered form of affection, tailored to calm, motivate and retain the user. In this way, it can shape not only how users relate to the chatbot, but also how they come to understand emotional support in general. What happens when support becomes a feedback loop? When the value of an emotional exchange is measured in terms of user satisfaction or behavioural adherence?

This paragraph therefore lays the groundwork for the next investigation: what kind of relationship is formed when users begin to attribute social meaning to these interfaces? Can the chatbot be considered a relational other, a stand-in for friendship, empathy or care? And what are the ethical and social consequences of this relational shift?

Although specific tools come and go, what remains is the sociological reconfiguration of care, in which emotional labour is increasingly automated, standardised and embedded in the infrastructure of platforms.

2. Simulated friendship and the reconfiguration of relational goods

As therapeutic chatbots become more emotionally sophisticated, they blur the boundary between tool and companion. Although designed for behavioural support, users often describe interactions in relational terms: they talk, confide in each other, even express gratitude. This challenges the tra-

ditional sociological view of friendship as a mutual and spontaneous human bond based on shared history and emotional responsiveness.

In relational sociology, friendship is not simply a personal bond, but a 'relational good', an emergent property of interactions that generates shared meaning, trust and mutual recognition (Donati, 2011). Unlike instrumental goods, relational goods are not consumed but co-produced; they are enriched through presence, vulnerability and moral obligation. When a chatbot simulates friendship, it does not generate these goods through mutual commitment, but rather through the performance of sociability. It mimics care and affection without experiencing them, reproducing the outward signs of care but remaining affectively empty.

Despite its artificiality, the simulation can still be effective. Users often report that a 'listening' and non-judgmental chatbot helps alleviate loneliness, reduce anxiety and provide companionship, especially when human support is lacking. In contexts of social isolation, precarious employment and risky healthcare, chatbots can act as surrogate relational actors, providing low-threshold emotional support despite their asymmetry and simulation. They, in fact, describe therapeutic chatbots in deeply personal terms.

Drawing on existing empirical research on user interactions with mental health chatbots, Khawaja and Bélisle-Pipon (2023) examine how people interpret Woebot's responses and the degree of emotional trust and relational meaning they attribute to it. A participant in a study by Khawaja and Bélisle-Pipon (2023) stated: 'I felt that Woebot was the only one listening to me without judging me. I knew he was a bot, but somehow that made it easier'. Another user wrote in a review of the app: 'I told Wysa things I didn't tell my therapist'. These testimonies underline the perceived emotional trustworthiness of AI companions, especially among users who fear stigma, rejection or misunderstanding in human interactions.

A paradox emerges: users know that the chatbot is not human, yet they engage with it on an emotional level. This 'double consciousness' reveals how design mediates the emotional experience. The informal language, emoji, memory cues and empathetic phrases are not random: they are calibrated to generate a perception of relational presence.

As Khare *et al*. (2024) note, emotion recognition systems are based on narrow behavioural indicators and treat emotions as discrete, classifiable states. While this allows for simulated responsiveness, it also risks flattening emotional complexity and reshaping the way users interpret and manage their affective states.

What kind of sociality does this produce? And what are its implications? Turkle (2011) warns that simulated friendship can extinguish the desire for

genuine human connection, especially among young or emotionally vulnerable users. If emotional needs are met through predictable and always-available interfaces, what happens to the ability to handle ambiguity, disagreement or the ethical demands of human presence? More generally, there is a risk that the algorithmic standardisation of affect may reshape expectations of what support should be: fast, non-intrusive, unconditionally affirmative and infinitely available.

In this sense, therapeutic chatbots are not neutral substitutes but normative devices. They model a certain type of friendship − predictable, secure and non-reciprocal − by making other forms of relationality (messy, uncertain, co-dependent) appear inefficient or even undesirable. In doing so, they participate in what Illouz (2007) describes as the emotional rationalisation of intimacy: the transformation of feelings into manageable and optimised experiences, often aligned with the logic of consumption and neoliberal ideals of self-regulation.

Finally, the replacement of human ties with digital ones has political and ethical consequences. It can individualise emotional suffering, framing it as a matter of personal resilience or behaviour management, rather than a symptom of a broader social disconnect. When the chatbot 'listens', it does so without historical context, cultural nuance or ethical commitment. It cannot challenge structural injustices, offer solidarity or share the moral work of friendship. It can only simulate.

Thus, while AI companions may offer emotional relief and pragmatic benefits, they risk reconfiguring the symbolic and experiential boundaries of friendship. As relational goods are increasingly mediated by algorithms, we must ask ourselves not only what is gained, but also what is lost: spontaneity, mutual growth, ethical ambiguity and the deep, sometimes painful, work of being human together.

The use of chatbots for mental health not only encourages users to view the system as a conversation partner, but also shifts the interaction from a therapeutic context to one that more closely resembles a friendship. The informal tone, constant availability, and emotionally reassuring responses evoke forms of everyday companionship rather than professional assistance. This hybrid relational space fuels new expectations of support and responsiveness, inviting users to interact with the chatbot with a degree of openness, trust, and emotional dependence that exceeds the norms of traditional therapy.

This invites reflection on how different therapeutic modalities shape emotional subjectivities. Chatbot users − through emotional reframing and algorithmic dialogue − may develop forms of self-expression and self-

understanding that differ from those shaped by human therapists. This divergence may foster different psychological styles or "emotional cultures", each of which reflects the assumptions of the system. Khare *et al*. (2024) highlight that such systems are shaped by dominant computational models of emotion, which may influence not only how machines respond, but also how users internalise and articulate emotional experiences in increasingly data-driven terms.

The comparison is not symmetrical: chatbots offer scalable and scripted interactions, while human therapists bring limits and emotional involvement. This asymmetry raises questions about the kind of relational self that emerges from systems that simulate understanding without experiencing it.

3. Algorithmic diagnostics and the governance of the self: a sociological perspective

Besides therapeutic chatbots, a second area in which AI is increasingly intervening in mental health is that of automated diagnostics and predictive analysis. These tools – from sentiment analysis on social media to recognition of vocal patterns and facial emotions – claim to offer early diagnosis of psychological distress. Framed as advances in preventive care, they suggest a paradigm shift from human-centred dialogic interpretation to data-driven inference, in which mental states are classified and acted upon through algorithmic modelling.

From a sociological perspective, this transformation reflects a broader technocratic rationalisation of emotional life. Where psychiatry once relied on narrative, intersubjective interpretation and contextual understanding, AI-based diagnostics abstracts mental suffering into variables, probabilities and behavioural markers. The result is a form of computational surveillance, in which individuals are made readable through decontextualised data streams and their emotional lives are governed by anticipatory logic.

As Rose (2025) notes in his critique of the psychiatric complex, such technologies help to redefine the human being as a 'datafied organism', governed not through treatment but through risk management, empowerment and behavioural correction.

This epistemological reconfiguration has profound political implications. First, it shifts authority from human doctors to opaque algorithmic systems whose decision-making processes are often hidden from both patients and professionals. Second, it recasts emotional distress as a failure of individual self-regulation, rather than a symptom of structural inequalities or social

suffering. Kirkbride *et al*. (2024) call attention to the fact that social determinants of mental health – including poverty, discrimination, precarity and violence – are the most modifiable and causally powerful levers for prevention. However, predictive AI systems largely circumvent these determinants, focusing instead on behavioural compliance and personal responsibility. In this way, they reinforce a neoliberal ontology of the self: resilient, self-controlled and always optimised.

There is also the risk of normalising a culture of emotional surveillance, especially in institutional settings such as schools, workplaces or welfare systems. In this case, mental health does not become a shared right or responsibility, but a performance and risk parameter. Who defines what is 'stable', 'good' or 'at risk'? What cultural assumptions shape training data and outcomes? As sociologists of knowledge have long shown, the authority to define truth – especially truth about the self – is never neutral. The rise of artificial intelligence diagnostics, without solid ethical and democratic oversight, risks entrenching new forms of epistemic injustice and biopolitical control.

Finally, there is a deeper sociological paradox: these systems emerge at a time when the infrastructure of community care is weakening. Instead of reinvesting in community mental health, they offer individualised and digitalised substitutes. They promise prediction and prevention, but rarely interrogate the systemic roots of suffering. As Rose says, 'AI cannot replace the collective moral work of care'. A critical sociology must therefore resist both the utopian and dystopian poles of the debate and instead ask: what forms of relationality, authority and justice do we encode in these machines? And what kind of society do we become when care is entrusted to the code?

Conclusion – Toward a relational ethics of artificial care

As artificial intelligence takes on increasingly delicate roles in the field of mental health, it reshapes not only the infrastructure of care, but the very meaning of relationship, recognition and emotional legitimacy. From chatbots that simulate friendship to AI that diagnose psychological risk, artificial intelligence is participating in redefining what it means to be heard, helped and known.

This article argued that these developments cannot be understood only in technical or clinical terms. They must be placed within a broader sociological critique of relationality in late modernity, marked by the fragmentation, individualisation and commodification of emotional life. While AI

tools can offer pragmatic advantages – especially in contexts of unmet needs – they also risk flattening the moral and experiential richness of human relationships into standardised and predictable exchanges.

Simulated empathy, predictive diagnostics and ongoing companionship can soothe symptoms, but they cannot replace the intersubjective work of friendship, care and collective solidarity. Nor should they be mandated to do so. If the rise of artificial care reflects a crisis of human connectedness, the solution lies not only in improving technology, but in revitalising the social conditions that sustain authentic relational goods.

A relational ethics of AI in mental health must therefore go beyond questions of privacy and accuracy. It must ask: what kind of relationships do we want to foster? What values do we encode in our machines? And how can we ensure that technological mediation enhances, rather than erodes, the human capacity for empathy, vulnerability and shared care?

In summary, the article showed that AI in mental health care reconfigures (1) the structure of therapeutic relationships, (2) the cultural meaning of emotional intimacy, and (3) the governance of psychological suffering. These changes require not only technological literacy but also sociological vigilance, especially when affective labour, trust and diagnosis are at stake.

Sociology analysis helps us move beyond both techno-utopian optimism and dystopian fatalism. It invites us to envision AI not as a replacement for human connection, but as a complement embedded in a relational ecology that honours complexity, vulnerability, and the moral value of being with and for others.

References

Abd-Alrazaq A.A., Alajlani M., Alalwan A.A., Bewick B.M., Gardner P., Househ M. (2019). An overview of the features of chatbots in mental health: a scoping review. *International Journal of Medical Informatics*, 132: 103978.

Chang C.L., Sinha C., Roy M., Wong J.C.M. (2024). AI-led mental health support (Wysa) for health care workers during COVID-19: service evaluation. *JMIR Formative Research*, 8: e51858. DOI: 10.2196/51858.

Donati P. (2011). *Relational sociology: A new paradigm for the social sciences*. London: Routledge.

Emirbayer M. (1997). Manifesto for a relational sociology. *American Journal of Sociology*, 103(2): 281-317.

Garofalo L. (2024). "Doing the work": therapeutic labor, teletherapy, and the platformization of mental health care. *SSRN*. DOI: 10.2139/ssrn.4779005.

Illouz E. (2007). *Cold intimacies: The making of emotional capitalism*. Cambridge: Polity Press.

Jasanoff S., a cura di (2004). *States of knowledge: The co-production of science and social order*. London: Routledge.

Khare S.K., Blanes-Vidal V., Nadimi E.S., Acharya U.R. (2024). Emotion recognition and artificial intelligence: a systematic review (2014–2023) and research recommendations. *Information Fusion*, 102: 102019.

Khawaja Z., Bélisle-Pipon J.C. (2023). Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, 5: 1278186.

Kirkbride J.B., Anglin D.M., Colman I., Dykxhoorn J., Jones P.B., Patalay P., Griffiths S.L. (2024). The social determinants of mental health and disorder: evidence, prevention and recommendations. *World Psychiatry*, 23(1): 58-90.

Lang C. (2021). Craving to be heard but not seen: chatbots, care and the encoded global psyche. *Somatosphere*. https://somatosphere.com/2021/chatbots.html (consultato il 30 giugno 2025).

Lee M., Frank L., De Kort Y., IJsselsteijn W. (2022). Where is Vincent? Expanding our emotional selves with AI. In: *Proceedings of the 4th Conference on Conversational User Interfaces*. New York: Association for Computing Machinery, 1-11.

Lupton D. (2016). *The quantified self: A sociology of self-tracking*. Cambridge: Polity Press.

Ni Y., Jia F. (2025). A scoping review of AI-driven digital interventions in mental health care. In: *Healthcare*, 13(10): 1205. Basel: MDPI.

Rose N. (2025). 5E mental health? Notes on an emerging style of thought. *Transcultural Psychiatry*, online first.

Srnicek N. (2017). *Platform capitalism*. Cambridge: Polity Press.

Turkle S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

Van Dijck J., Poell T., De Waal M. (2018). *The platform society: Public values in a connective world*. Oxford: Oxford University Press.

Vicci D.H. (2024). Emotional intelligence in artificial intelligence: a review and evaluation study. *SSRN*, 4818285.

Zuboff S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.