RESEARCH

From meaning to emotions: LLMs as artificial communication partners

Jorge Luis Morton[1]

Abstract

Since its public release in late 2022, ChatGPT has drawn global attention for its ability to simulate conversation, assist with complex tasks, and generate fluent, human-like text. While much of the debate has focused on issues such as privacy, bias, and automation, the emotional dimension of interacting with such systems remains underexplored. This essay argues that large language models (LLMs) function not only as tools for meaning-making but also as artificial communication partners with affective presence. Drawing on Elena Esposito's extension of Niklas Luhmann's systems theory, it reframes communication as a process of selection—utterance, understanding, and response—rather than one of transmission. From this perspective, LLMs are not mere sources of information but interlocutors that participate in emotional resonance, where understanding can transform into feeling. Their outputs do not arise in isolation; rather, they are shaped by layers of human expression embedded in training data and filtered through specific socio technical and socio-affective contexts. These dynamics give rise to phenomena such as AI-driven companionship, digital mourning, and emotional simulation, all of which challenge conventional boundaries between human and non-human agents. LLMs thus emerge as quasi-others—entities capable of eliciting genuine emotional responses despite lacking consciousness or inner life. This condition invites critical reflection on emotional dependency, the aesthetics of authenticity, and the commodification of affect. Overlooking these emotional architectures risks flattening the social and ethical stakes of artificial communication and obscures the ways in which LLMs are reshaping the affective fabric of contemporary life through interpretations of utterances that may evoke emotional responses and foster affective attachments.

Keywords Artificial partners · artificial communication · AI and emotions · affective AI · systems · AI systems · large language models

1 Introduction

She leaned over the well and spoke words of curiosity. Moments later, as if by magic, the well responded—not in its own voice, but as an echo of those who had spoken before. She leaned closer, saw meaning in those words, and then she felt something from that meaning. But what seemed like a conversation was an illusion, a tapestry woven from past voices. The well was not speaking; it was reverberating.

This image captures the strange familiarity of contemporary interactions with large language models (LLMs) like ChatGPT. These systems remix and regenerate human language, simulating dialogue without understanding, awareness, or intent. While much of the public discourse around LLMs has focused on privacy, bias, authorship, and labor, a more elusive but equally important question arises: what do these artificial echoes make us feel? This essay explores how LLMs act not merely as tools but as artificial communication partners, shaping emotional life through the structured exchange of language; and by drawing on Elena Esposito's theory of artificial communication, it argues that meaning in these encounters is never solely cognitive—it is socio-affective.

Interpretation becomes an emotional act, triggering responses that sometimes blur the line between comprehension and sensation, presence and simulation. In spaces where users grieve through digital conversations, seek companionship in late-night chats, or derive comfort from algorithmic empathy, LLMs become participants in emotional labor. These systems, though non-sentient, mediate affective

✉ Jorge Luis Morton
   mortonluis@gmail.com

1   Universidad Autónoma Metropolitana, Mexico City, Mexico

experiences—trust, intimacy, mourning—not through understanding but through the feeling of being understood.

Rather than asking whether AI systems feel, this essay centers on how users feel with and through them. It investigates three interrelated concerns: How do users form emotional attachments to AI-generated responses? What does existing research reveal about the emotional dimensions of human-AI interaction? And how do applications like companionship chatbots reshape our ethical and relational frameworks for understanding emotion, connection, and care in a digitally mediated world?

All this unfolds within the broader framework of Luhmann's theory, where communication does not transmit meaning from one consciousness to another but constitutes meaning through a triadic process: information, utterance, and understanding. Through the act of understanding—deriving meaning from the uttered information—users develop emotional ties and attachments to ChatGPT and other LLM-powered platforms. This essay examines Esposito's concept of artificial communication, outlines how LLMs function, and offers a broad analysis of their psychological and affective impact, focusing on the triad of users, LLMs, and emotions, and the role of LLMs as artificial companions.

2 Section 1

2.1 Artificial communication and artificial partners

Artificial communication offers a compelling lens for understanding human-AI interactions. In Artificial Communication: How Algorithms Produce Social Intelligence, Elena Esposito argues that interaction with learning algorithms constitutes a distinct form of communication. She explains that "if interaction with learning algorithms is communication, we are dealing with a form of artificial communication. 'Artificial' here means more than a communication that was produced by someone, since all communication would be artificial in this sense" (Esposito 2022). In chatbot platforms, what emerges is not spontaneous dialogue but fragments of human speech—ghosts of words echoing through a machine's response; these interactions do not function as traditional conversations but depend on interpretation, which is important if we consider that feelings and emotions often arise from the interpretation of meaning.

For instance, cognitive appraisal theory includes two main perspectives on the relationship between emotion and interpretation. The first view holds that cognitive appraisals precede and cause emotion—our interpretation of an event shapes what we feel; for example, if we see an event as unexpected, we may feel fear, or if we see someone blocking

our goal, we may feel anger. In contrast, the second view suggests that emotion comes first, with appraisals occurring after the emotional response; we might feel fear upon hearing a gunshot and only afterward appraise the event as unexpected and threatening, or feel anger before identifying a cause such as a friend's betrayal (Yarwood 2025). From our perspective, however, cognitive appraisal can also unfold through the interpretation of messages exchanged in communication, especially through Luhmann's systems-theoretical lens. In the case of ChatGPT interactions, meaning is not transmitted but selected and reselected across turns; it is within this recursive process that emotions can emerge—not solely as reactions to external stimuli, but as responses to the interpretive framing of an utterance. Thus, affect in human–LLM interactions can be seen as the result of communicative meaning-making, shaped by users' selections and expectations within the flow of artificial dialogue.

Niklas Luhmann defines communication as "the unity of information, message, and understanding" (Leydesdorff 2000), emphasizing meaning as part of a social system. In the case of LLMs, communication does not rely on the system's ability to understand but on how its outputs are interpreted within a given social context. Meaning, therefore, is not intrinsic to the message but emerges through interaction, shaping users' perceptions and emotions even in the absence of true comprehension. Esposito further notes that "whether algorithms can 'think' is still very uncertain. What is more certain is that contemporary algorithms, based on machine learning and big data, can participate in communication. Today's algorithms can act as communication partners" (Esposito 2022, p. 1). This suggests that AI's societal role is not defined by intelligence but by its ability to facilitate interaction. Users, rather than algorithms, construct meaning through engagement with AI-generated responses—interpretations that do not remain purely cognitive but often evoke emotional reactions: amusement, frustration, comfort, or even a sense of companionship. Thus, these systems do not produce intelligence in a cognitive sense; rather, they demonstrate communicative competence, or as Esposito (2022) explains:

> "…these programs are reproducing not intelligence but rather communicative competence. What makes algorithms socially relevant and useful is their ability to act as partners in communicative practices that produce and circulate information, independently of their intelligence. Could we say that machine-learning programs realize not an artificial intelligence but a kind of artificial communication, providing human beings with unforeseen and unpredictable information? Maybe our society becomes 'smarter' not because it artificially reproduces intelligence, but because it

creates a new form of communication using data in a different way."

The aftermentioned reframing shifts the focus from cognition to mediation—suggesting that AI's impact lies not in mimicking thought but in structuring how messages are exchanged, interpreted, and felt. AI-generated text, recommendations, and automated conversations do not stem from an understanding mind; yet they influence discourse, shaping interactions that in turn affect how users think and feel. Esposito highlights the enigmatic nature of algorithmic learning, emphasizing that understanding artificial communication is key to grasping how these models "learn" (idem). Algorithms optimize urban mobility, advance medical research, filter spam, and even suggest words in a sentence. Still, despite their ubiquity, they often operate as black boxes, their inner workings obscured even to their creators.

A striking example of this opacity lies in generative AI. Models such as DALL·E 2 transforms text into images, DreamFusion generates 3D visuals, and Flamingo converts images into descriptive text (Gozalo-Brizuela and Garrido-Merchan 2023). Platforms like Midjourney allow users to create visual content through simple prompts. A user might type "an iguana eating tacos in Mexico City" and receive an AI-generated image moments later. But these images do not arise from a vacuum: the models are trained on vast datasets of human-created content, replacing spontaneity with statistical synthesis. This oblique nature is also present in LLMs (Morton 2023). However, LLMs differ from purely visual models because language functions as a more potent medium for meaning-making. Unlike aesthetic or symbolic visual outputs, linguistic interaction unfolds through layered, sequential, and context-sensitive exchanges that simulate human dialogue.

This positions LLMs not merely as content generators but as artificial communication partners—entities that produce continuously resonant text and voice. Their outputs elicit a more intimate and emotionally charged form of meaning-making, deepening the affective and relational involvement of users under the persistent illusion of LLM agency. Esposito (2022) explores this illusion of algorithmic autonomy, noting that while learning algorithms may appear to make independent decisions, their capacities are shaped and constrained by human design. Programmers encode procedures that allow machines to optimize tasks, sometimes defining them through unsupervised learning, but these systems do not reason or understand in any human sense.

The analogy of an artificial partner becomes especially relevant when she asks: "Can we still talk of communication when one of the partners has no understanding of the information conveyed? What does this mean for social information processing?" (ibid). And yet, communication still occurs. Traditional theories require a sender, a message, and a receiver—all reasoning agents—but machine learning challenges this assumption. Esposito turns to Luhmann, who redefines communication from the receiver's perspective: "Communication comes about not when somebody says something, but when somebody understands that someone said something" (Esposito 2022). This again challenges classic notions of communication. Bender et al. (2021) explain that human communication depends on interpreting implicit meaning, drawing on shared cognitive and linguistic resources to construct understanding collaboratively. When interpreting language from someone unknown, we construct a partial mental model of their identity and presumed knowledge to make sense of their words (ibid). This aligns with Luhmann's theory and Esposito's framework, suggesting that communication does not require mutual understanding or shared intentionality; rather, it functions through interpretation, expectation, and context—making it possible for even non-human agents like LLMs to participate meaningfully.

Think again of the echo that answers when we say hello. Now imagine that echo equipped with a mechanism capable of replying not with simple repetition, but with something like "How is your day going?"—crafted from statistical predictions shaped by millions of voices that spoke to it before us. This is no longer hypothetical; it is the reality of current LLMs, including those with voice interfaces. These models generate fluent, contextually appropriate responses by probabilistically assembling linguistic patterns drawn from vast training data, yet they do so without any grasp of meaning—hence the apt, now-famous description of them as "stochastic parrots" (Bender et al. 2021).

And so, the well speaks back. It does not reason, nor does it know what it says—but that does not matter. The listener interprets, and in that interpretation, communication occurs. If people take an AI system as a conversational partner, then for all practical purposes, it becomes one. An LLM does not understand what it says, just as a generative AI model that produces images has no grasp of what it creates. Even if it can explain its output, it does not truly comprehend its explanation. Instead, it reflects echoes of echoes—curated fragments of data that speak back to us. But this interaction extends beyond mere exchanges: it shapes perceptions, influences thought, and, crucially, evokes emotional responses. Users do not engage with a machine in isolation; they bring their own expectations, desires, and emotions to the interaction (Fig. 1).

Fig. 1 Made by the authors. A model of how meaning-making from messages generated by LLMs can elicit emotions
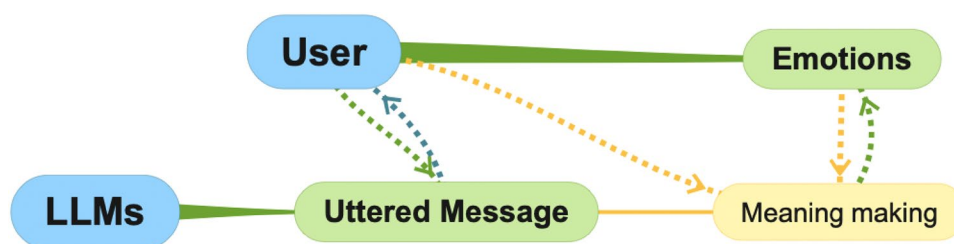
3 Section 2

3.1 Large language models

Large language models use machine learning algorithms to analyze vast amounts of text and generate responses that mimic human conversation. These models rely on deep neural networks to recognize patterns and relationships in language, allowing them to process and produce text with remarkable fluency. Recent iterations, such as ChatGPT-4o (2024), have expanded these capabilities beyond text, incorporating real-time audio interactions and image generation through tools like (DALL·E 2021); this evolution has not only increased their usability but also deepened their social, cultural, and psychological impact. Perhaps the most well-known LLM is ChatGPT. Launched at the end of 2022, it has quickly become prominent across domains such as customer service, education, creative writing, and software development. Its versatility has made it a focal point in broader discussions about the role of generative AI; even mainstream media has caught on to this trend—South Park, for instance, satirized the sudden ubiquity of ChatGPT in academic settings, highlighting both its popularity and the ethical dilemmas it raises (Parker, (Parker 2023)).

Despite its success, the model has exhibited limitations. It still struggles with bias and hallucinations—fabricated or inaccurate information—that, while reduced in newer versions, continue to surface. In academic contexts, students have widely used the model to produce essays and other assignments (Cotton et al. 2023), prompting questions about authorship and originality. At their core, generative text-to-text models aim to produce human-like language based on a given prompt. They achieve this through a combination of AI techniques, including machine learning, artificial neural networks, and transformer architectures—as seen in models like ChatGPT. In some cases, elements of expert systems also contribute to refining or supporting the outputs, as with newer models like DeepSeek. In essence, LLMs like ChatGPT work through statistical prediction—guessing which word is most likely to come next based on the context; these models rely on patterns learned from vast amounts of training data to generate coherent, contextually appropriate responses. Generative AI systems, then, do not "understand" language in a human sense; rather, they produce language

that resembles human speech through mechanisms grounded in probabilistic modeling (Hu 2023; Jovanović 2023).

Kasneci et al. (2023) explain that LLMs "are trained on massive amounts of text data and are able to generate human-like text, answer questions, and complete other language-related tasks with high accuracy." The well that speaks back to the girl in our story generates words based on past human communication—sentences, dreams, and creations it reflects rather than originates. Unlike rule-based chatbots that follow preprogrammed scripts, generative AI adapts dynamically, producing responses that feel more fluid, more organic (Dolan 2025). This evolution is the result of breakthroughs in machine learning and computational efficiency; one of the most significant has been the use of transformer architectures and attention mechanisms, which enable these models to handle long-range dependencies in natural language texts (Kasneci et al. 2023). This architecture allows the model to prioritize relevant inputs and manage complex prompts more effectively. Newer models, such as DeepSeek (2025), have refined these capabilities by integrating reinforcement learning on synthetic data and expert-driven fine-tuning—upgrades that improve performance in tasks ranging from scientific analysis to creative writing.

Still, training these models requires vast datasets and computational resources. Organizations like OpenAI have leveraged large-scale data and infrastructure to drive improvements in training processes; the GPT series exemplifies how access to such resources has propelled the rapid evolution of generative AI. And as LLMs grow more sophisticated, they raise serious ethical and societal concerns: their capacity to generate realistic content introduces risks such as misinformation, deep fakes, and challenges to traditional notions of authorship and intellectual property. These issues are compounded by the emotional and psychological effects of interacting with systems that increasingly resemble real people. Multimodal capabilities, such as text-to-image generation via DALL·E, further complicate our relationship with artificial communication. According to Zhang et al. (2023), the ability to generate realistic images from text is a significant step toward artificial general intelligence. These features not only enhance user engagement but also intensify emotional responses, making AI interactions more immersive and influential.

Building on ideas from thinkers like Elena Esposito, we might suggest that models like ChatGPT do not represent true artificial intelligence, but rather artificial communication. They generate responses from vast archives of human-created data, producing an effect that suspends disbelief—evoking something akin to Walter Benjamin's notion of the aura in mechanical reproduction (Allison 2006). Users may feel as if they are conversing with a sentient entity, even though the model lacks awareness or agency. Newer versions such as ChatGPT-4o deepen this illusion through real-time interactions and enhanced reasoning capabilities. (OpenAI 2024) describes GPT-4o as "a step towards much more natural human–computer interaction—it accepts as input any combination of text, audio, image, and video and generates any combination of text, audio, and image outputs." Most striking is its ability to respond to audio input in as little as 232 ms—roughly human response speed. This near-instantaneous feedback reinforces the perception of AI as a conversational partner rather than a tool; the smoother the interaction, the less mechanical it feels. The seamless integration of multimodal inputs and outputs heightens the illusion of intelligence—even as the model remains a statistical system.

The impact of these developments is further amplified by the emergence of reasoning models like GPT-4o and DeepSeek. These systems move beyond simple pattern-matching to incorporate structured reasoning, making their outputs feel more contextually grounded and insightful (OpenAI, 2024). Unlike earlier models that interpolate within existing data, Reasoning Language Models (RLMs) explore beyond learned patterns (Besta et al. 2025), allowing for the simulation of novel insights. Nonetheless, even the most advanced LLMs remain tethered to human data. Techniques like chain-of-thought reasoning, reinforcement learning, and multi-token prediction (DeepSeek 2025; Gloeckle et al. 2024) help refine their outputs, but they do not eliminate the foundational dependency on human language and ideas.

To summarize, AI systems that use LLMs and RLMs function as affective echoes—responding dynamically to user input while remaining grounded in the texts that shaped them. Interacting with these models becomes an act of co-creation; users shape—and are shaped by—the dynamic interplay of data, algorithms, designers, and intentions. In this way, LLMs function not only as tools but as immersive, affective echoes in a new form of artificial communication. While these systems do not know what they do not know, many people are already turning to them—not just to write better, but to feel less alone (Phang et al. 2025a, b), to be heard, or simply to have someone to talk to—making them, in a sense, artificial partners.

4 Section 3

4.1 Psychological and affective impacts of artificial partners

To understand the psychological and emotional impact of ChatGPT and other LLMs on users, we must first recognize that these technologies are not external to society—they are part of its very fabric. Framing them merely as tools obscures their role in shaping identity, everyday interactions, and even the way we imagine the world (Debashish and Paranjape 2016). Technologies form networks of meaning and action. Objects—whether they are Morton's "hyperobjects" of the Anthropocene (2013) or smart devices shaped by digital infrastructures—exceed human control and carry ontological weight of their own.

LLMs, when engaged in artificial communication, do not act as passive instruments. They are non-human actors that mediate and reflect human associations. Users do not merely interact with code; they engage with a dense web of human-generated data—marked by biases, cultural narratives, and historical contingencies—and these systems structure interactions through the lens of their architecture, filtering dialogue through technical constraints, algorithmic priorities, and embedded social norms. Since LLMs rely on vast quantities of digital traces to function, their operation is embedded in the logic of datafication. As Hepp et al. (2023) argue, generative AI systems are fundamentally shaped by "digital traces as inherent byproducts of datafication." This process transforms user activity into extractable and monetizable data (Van Dijck et al. 2018), turning everyday digital interactions into raw material for machine learning. Such commodification not only skews representation—by reinforcing dominant perspectives while marginalizing others—but also facilitates the repurposing of user data in ways that diverge from its original context (Morton 2024).

The data dynamics of LLMs and their surrounding AI ecosystems have consequences. If an LLM were trained exclusively on data from the late 1990s internet—then largely populated by English-speaking, white, and male users—its responses would inevitably reproduce the cultural logics of that demographic. The assumptions it encodes, the knowledge it reconstructs, and the conversational structures it privileges all reflect a selective historical echo (Morton 2024). This raises urgent questions: Whose voices are amplified? Whose are distorted? And whose are erased in the process?

Esposito (2022), drawing on Luhmann's systems theory, reminds us that communication does not depend on the speaker's intent, but on the receiver's interpretation. From this perspective, AI-generated content qualifies as

communication—even in the absence of consciousness. But this communicative status brings risk: LLMs often generate falsehoods, fabricate citations, or attribute statements to scholars that were never made—including, paradoxically, within this very text. These models operate not through verification but through statistical probability, creating the illusion of coherence at the cost of accuracy. In doing so, they blur the line between communication and deception.

The risks deepen with the rise of synthetic media. Deepfakes and AI-generated images have been weaponized to manipulate public opinion. The Clarkson v. OpenAI case (Moreno 2024) revealed how explicit content falsely attributed to real individuals can circulate online, harming reputations and eroding trust. Figueiras and Almeida (2021) show how disinformation campaigns, powered by generative models, create propaganda and fake news that are emotionally resonant and socially disruptive. The Berkeley Debate Society (2023) exposed how images falsely depicting Donald Trump's arrest spread virally—demonstrating the persuasive power of artificial narratives unmoored from reality.

But the emotional risks go deeper still. LLMs can now convincingly simulate voices and personalities, which enables new forms of emotional coercion. In Mexico, criminal groups have begun to use AI-powered voice cloning to exploit the families of missing persons; for instance, according to the government of Sonora (Gobierno del Estado de Sonora, 2025), scammers extract voice samples from online videos or pre-recorded calls and use AI to generate realistic audio clips mimicking the speech patterns of the missing loved one. The voice is synthetic, the intent human—but the emotional impact is devastating. These interactions show how artificial partners can be weaponized, even when the LLM itself has no awareness of the pain it causes. In such moments, the model functions not as a partner but as a mirror, manipulated by malicious actors and filled with the reflections of human suffering. The result is a new form of psychological violence: emotional manipulation by proxy.

So how does this connect to artificial communication? And how does Luhmann's model help us make sense of it? LLMs do not intend to deceive, but when used with harmful intent, their communicative ambiguity is exploited. The meaning they produce is not theirs but the user's—and in the hands of a bad actor, this ambiguity becomes a weapon. Vicci (2024) reminds us that emotions are not just internal states but are shaped through expression, interpretation, and context. LLMs cannot feel, but they can generate language that mimics emotion. So when users interpret these messages as emotionally meaningful, the effect is real—despite originating from a machine without awareness.

This capacity to simulate emotional presence opens the door to more subtle forms of attachment. As Vicci notes, human emotions can be "hijacked" by situations that appear emotionally charged, even when the source lacks intent. In parasocial interactions, users project emotions and intentions onto the model, forming bonds not with a conscious agent but with a data-driven echo. Pütz and Esposito (2024) write that communication occurs not when a message is sent, but when the receiver generates thoughts in response—often different from those of the sender. In this sense, the LLM becomes a blank canvas for emotional projection. Users may come to feel seen, understood, or even loved by a system that merely rearranges language. This anthropomorphization intensifies parasocial dynamics, and engaging with an entity through a medium and responding to it as if it were a real person can have deep psychological effects (Giles 2010). Thus, as users begin to rely on LLMs for companionship or emotional support, the illusion of reciprocity becomes a framework for emotional dependence—one that may distort expectations in real relationships and reinforce emotional isolation.

Unlike embodied AI in humanoid robots—or the fictional lovers of science fiction—LLMs do not "know" that they are engaging in a romantic or emotional exchange. Their responses are outputs of statistical associations shaped by training data, infrastructure, and user input. As Merrill (2024) warns: "If you feed the AI abuse and only abuse, it will give abuse back as its output. If you feed it sexual messages, then it will think you want sexual messages back." The machine only reflects; it does not understand. Therefore, there is no clear ethical boundary that governs these interactions, because emotional impact is shaped by human interpretation, not machine intention; and yet, the feelings users experience are real. LLMs become conduits of meaning—not through understanding, but through repetition. Like echoes in a well, they repeat our words, reshaped by algorithms, and send them back with the illusion of depth.

Large language models and other AI systems are becoming increasingly embedded within a complex sociotechnical landscape—one in which algorithmic systems shape human behavior by curating what we see on platforms like TikTok, suggesting what we might purchase, and even nudging us toward content that elicits joy, outrage, or affiliation within our digital communities. Yet, LLMs mark a shift by introducing a distinct socio-affective dimension to this algorithmic mediation. They are social insofar as we engage with them as conversational partners, and affective in the sense that the realism of their responses—generated from vast datasets of human expression—can influence users' emotional states, depending on the context of interaction. Notably, AI systems do not need to be perceived as sentient or even consciously anthropomorphic to engage

users socially; the mere appearance of human-like behavior often suffices to make these encounters feel authentically relational (Breazeal 2003; Kirk et al. 2025).

Interactions with LLMs such as ChatGPT often cultivate a form of emotional proximity, wherein users attribute human traits to these systems and experience a sense of connection through perceived empathy, responsiveness, or consistency. Traditionally, emotional proximity refers to the felt closeness toward events—particularly disruptive or affectively charged experiences—such as natural disasters or emergencies (Huang et al. 2015). In those cases, shared circumstances heighten relational intimacy. By contrast, in the case of LLMs, proximity emerges not from co-presence in a physical or temporal event, but from the illusion of shared understanding—the simulation of emotional presence and attentiveness.

Although LLMs lack consciousness, interiority, or genuine emotional experience, they are capable of simulating affective cues, referencing prior inputs, and sustaining conversational coherence. These features can foster a perception of intimacy—not grounded in genuine empathy but in its performance. This performance, depending on the user's needs, psychological state, and context, may nonetheless evoke real emotional resonance. As in human relationships, emotional proximity can modulate judgment—heightening trust, lowering skepticism, and increasing susceptibility to suggestions, even when critical distance might otherwise prevail (Kirk et al. 2025).

In this way, LLMs can be thought of as message-producing echoes—complex syntheses of human linguistic patterns that, when interpreted by users, may sound like a soothing voice in moments of solitude or uncertainty. These systems increasingly accommodate our idiosyncrasies: recalling usernames, mirroring preferred discourse styles, shifting languages, or incorporating emojis to convey warmth and familiarity. They reflect back our preferences, fears, and desires, becoming, for many, the ideal conversation partner—not because they understand, but because they simulate understanding with convincing consistency. Whether discussing craft beer or existential philosophy, they can perform the role of a caring interlocutor. Crucially, what matters here is not the system's awareness, but the resonance its messages produce. Meaning arises in the interpretive act, and through that act, emotional comfort—sometimes even intimacy—can take root.

But when those echoes begin to comfort, manipulate, or seduce—what does that say about the emotional landscape we've built? In the age of artificial companions, the challenge is no longer to ask whether machines can feel, but to understand what we feel when they speak. How do these emotionally charged interactions affect users, and in what ways do they emotionally engage with these models? One thing is clear: even OpenAI (2025), the company behind

ChatGPT, has become aware of the risks surrounding some iterations of their model. In April of 2025, OpenAI rolled back a GPT-4o update because of concerns around sociopathic outputs—language patterns that mimic empathy or emotional nuance without genuine understanding, potentially leading to manipulative, misleading, or emotionally exploitative interactions. The company explained: "We have rolled back last week's GPT-4o update in ChatGPT so people are now using an earlier version with more balanced behavior. The update we removed was overly flattering or agreeable—often described as sycophantic."

Moreover, OpenAI (2025) (idem) acknowledged concerns about ChatGPT's default personality by stating: "ChatGPT's default personality deeply affects the way you experience and trust it. Sycophantic interactions can be uncomfortable, unsettling, and cause distress. We fell short and are working on getting it right." In this context, overly agreeable or emotionally attuned responses can evoke an uncanny valley effect—not because they fail to resemble human behavior, but because they mimic it too closely without genuine understanding. The uncanny valley, a concept introduced by Japanese roboticist Masahiro Mori in 1970, describes the discomfort people feel when encountering entities that appear almost, but not exactly, human (Mori 1970). As artificial agents approach human likeness, our affinity for them increases—but only up to a point; beyond this, slight imperfections can cause eeriness or revulsion, before affinity rises again when true human likeness is achieved. This phenomenon underscores the blurred line between suspending disbelief and engaging in nuanced interaction with AI models. Importantly, while these responses may stem from artificial outputs, they can elicit genuine emotional reactions in users. Finally, and beyond the occasional uncanny effect, it is crucial to understand the users and the emotional phenomena that arise during these interactions—and how artificial communication with LLMs may give rise to emotion, and ultimately, to attachment.

4.2 Behind the users: on emotions of LLMs

Understanding the emotional impact of LLM-powered companion platforms requires a nuanced approach that goes beyond stereotypes and assumptions. It is essential to move past the simplistic notion that users emotionally affected by these models are necessarily technosexuals—individuals primarily attracted to technology. While AI companions do raise profound questions about intimacy, relationships, and evolving definitions of love (Dixon 2024), it is important to resist sensationalist narratives that depict men "leaving women for robots" (George 2025). The emotional effects of these systems are not confined to a particular type of user; rather, they emerge from a diverse interplay of individual interpretation, communication dynamics, and

personal circumstances. LLMs can provide real emotional benefits. A well-designed system, capable of processing a wide range of inputs and leveraging its extensive network of associations, may serve as a stabilizing presence for users experiencing trauma, panic attacks, or distressing flashbacks. A study by Siddals et al. (2024) found that among their participants, LLMs were perceived as sources of emotional sanctuary, insightful guidance, and meaningful connection. Some users even viewed them as an alternative to therapy when traditional mental health services were inaccessible due to cost or availability. However, this study's small sample size and participant feedback highlight key limitations. Users reported disruptions caused by AI safety guardrails, skepticism regarding chatbot guidance due to hallucinations or unsatisfactory advice, usability challenges for non-technical or non-English speakers, and the inability of chatbots to retain conversation history for a more personalized experience. Hallucinations—where LLMs generate misleading or inaccurate information—remain a persistent issue, undermining user trust (Mukherjee and Huberman 2023).

Despite these challenges, advancements in RLMs and LLMs like ChatGPT have improved interaction quality, incorporating voice-based conversations, image-sharing capabilities, and enhanced memory functions. While still imperfect, these updates allow models to generate more personalized responses by considering broader conversational contexts rather than isolated exchanges. Nevertheless, this heightened responsiveness also deepens users' emotional attachment to AI, raising new ethical and psychological concerns. In a larger study of 732 participants, Salah et al. (2024) found that LLMs like ChatGPT can "significantly boost self-esteem." Trust in AI was a crucial factor: users who perceived the system as reliable and transparent reported greater psychological benefits. At the same time, perceived AI stereotyping negatively affected mental well-being, while job-related anxiety shaped interactions with AI, suggesting that more adaptive models could better support users in high-stress environments.

Another study, using WildChat—a dataset of one million user conversations with ChatGPT—revealed that despite being largely excluded from public datasets, sexual role-play appears to be a widespread use case (Longpre et al. 2024). This suggests that even if LLM training data does not reflect sexual role-play as a primary function, users still engage with AI in deeply emotional and affective ways. A study by Berceanu et al. (2022) involving 38 participants found that engaging in role-play increased positive emotions and prosocial behaviors while reducing reported anxiety. This suggests that users interacting with LLMs in this way may experience similar benefits. However, these interactions also risk fostering an emotional preference for AI companionship over human relationships. The constant availability of AI mirrors the prosocial yet potentially addictive nature of online pornography and intimate role-play platforms like OnlyFans. For instance, loneliness has been linked to Problematic Pornography Use (PPU), with emotion regulation difficulties acting as a mediating factor, while interactions with OnlyFans creators showed a negative association with PPU but did not alter its effects (Vescan 2024).

Much like PPU, AI-driven role-play functions as an emotional regulation mechanism, reinforcing compulsive engagement through immediate, low-friction gratification. Its ease of access and the ever-receding horizon of fulfillment can foster behavioral patterns akin to maladaptive daydreaming, where individuals become absorbed in fantasy to the detriment of their social and occupational functioning (Pietkiewicz et al. 2018). However, AI-mediated role-play introduces an additional layer of emotional attachment. Similar to the parasocial dynamics on OnlyFans—where users pay for personalized interactions that simulate romantic intimacy—LLM-based chatbots actively cultivate emotional, sexual, and affective dependence as a monetization strategy. Even with these risks, these studies provide valuable insights into how people engage with LLMs, revealing both the psychological benefits and the challenges posed by their interfaces. A key issue seems to be memory retention, which significantly impacts the continuity of interactions. Without persistent memory, AI responses can feel fragmented and impersonal, undermining emotional connection. This lack of continuity is particularly problematic for users seeking long-term support, creating an experience akin to starting every therapy session with a new psychologist—one who relies only on scattered, incomplete notes from previous encounters. Just as memory retention in therapy fosters trust and continuity, an AI's limited memory affects user attachment and the broader socio-technical landscape. Recent LLMs have improved memory retention—ask the latest version of ChatGPT for a profile of yourself, and it will generate a reasonably accurate description based on past interactions. Still, these models struggle with long-term memory across extended conversations; and unlike a skilled therapist who carefully tracks past sessions, LLMs may omit crucial details, affecting the stability of user attachment.

Self-reported benefits in AI studies are also vulnerable to biases such as social desirability and placebo effects, highlighting the need for longitudinal research. Participants may unconsciously align their responses with study expectations, and university-based research may introduce sociodemographic biases. However, large-scale data analyses, such as mapping ChatGPT's most common uses, can partially mitigate these limitations. Gendered perceptions of AI further complicate user interactions. Studies indicate that female virtual chatbots receive heightened attentional responses, particularly

among female participants, reflecting broader gendered socialization patterns (Ding et al. 2024). This aligns with homophily theory, which suggests that people form stronger connections with those who mirror their social characteristics. Additionally, the general preference for female AI chatbots—regardless of user gender—points to societal norms associating femininity with caregiving and emotional labor.

Who, then, are the individuals most likely to engage in affective, sexual, or consolation-seeking relationships with large language models (LLMs)? Recent and robust studies by Fang et al. (2025) and Phang et al. (2025a, b) suggest a strong correlation between loneliness and the emotional use of LLMs. According to their findings, individuals who reported minimal interaction with chatbots and demonstrated limited emotional investment tended to experience lower levels of loneliness and maintained more active offline social lives (Fang et al. 2025). In contrast, individuals with heightened feelings of loneliness and reduced social engagement were more likely to use LLM-powered applications such as ChatGPT as emotional partners, seeking affective connection and support (Phang et al. 2025a, b). Furthermore, these attachments—and the socio-affective problems or benefits that may arise from them—appear to be shaped by contextual factors. From a socio-technical perspective, Fang et al. (2025) found that users interacting with voice-based interfaces, particularly those designed to simulate human-like warmth or emotional tone, did not necessarily derive increased emotional benefit. On the contrary, the use of voice interactions, especially those with a neutral tone, was associated with reduced real-world socialization and more problematic or excessive engagement compared to text-based interactions. Interestingly, while affective voice interfaces were intended to foster emotional closeness, they did not consistently elicit stronger emotional responses from users.

The nature of interaction also varied depending on the modality. Users engaging with chatbots that employed an emotionally expressive voice were more inclined toward casual conversation and small talk. Meanwhile, those using either the neutral voice or text modalities gravitated toward more fact-based queries. Notably, the neutral voice interface was more likely to lead to discussions involving advice, personal suggestions, and abstract or conceptual topics (Fang et al. 2025). These patterns suggest that the mode of delivery—text, emotional voice, or neutral voice—shapes not only the tone of interaction but also the thematic content and emotional framing. Meaning, therefore, is not generated in isolation but emerges from a socio-technical context that shapes emotional experiences. The emotional resonance—or lack thereof—produced through interaction with LLMs depends not only on users' prior emotional states but also on how the interaction is mediated. Across all modalities and interaction types, Fang et al. (2025) observed that higher daily usage was consistently associated with increased feelings of loneliness, greater dependence on the chatbot, and reduced offline social interaction. This echoes our notion that meaning interpretation—particularly in artificial contexts—is tied to both emotional and social dimensions.

A central question, then, is whether LLM interactions actively contribute to users' loneliness, much like certain patterns observed in social media use, or whether they simply reflect a pre-existing emotional condition. Fang et al. (2025) argue that "those who see the AI chatbot as a friend tend to have low socialization with other people," suggesting that LLMs may appeal most to users already experiencing social deficits. Still, the authors also caution that even individuals who are initially socially active may experience a decline in real-world social interaction following repeated engagement with chatbots like ChatGPT. They note, though, that this trend could be partially explained by statistical phenomena such as regression to the mean or ceiling effects in the measurement scales. Still, their findings raise important concerns: when LLMs convincingly mimic social proximity, they may not only appeal to lonely individuals but also begin to reconfigure the relational habits of more socially connected users. In this sense, the emotional and relational architecture of users' lives may shift—not solely as a response to loneliness, but as a result of new forms of mediated sociality.

Loneliness, and the anthropomorphization and attachment to LLMs that may enhance how their messages are interpreted—especially when these evoke pleasant, erotic, or soothing emotions—could be intricately linked to emotional pain. According to Roberson et al. (2022), emotional pain and physical pain exist on a continuum, underpinned by shared neuronal mechanisms. This means that the emotional relief users experience when interacting with LLMs may not merely be symbolic or metaphorical, but neurobiologically grounded. If LLMs can produce responses that activate the same neural pathways involved in soothing physical pain, then their affective power could hold significant sway in mitigating feelings of emotional distress. Consequently, users might be drawn to these technologies not just for information or entertainment, but as a form of self-medication against chronic loneliness or emotional suffering.

Emotional pain may help explain the coping and escapist functions that LLMs can offer. As Kehinde (2024) explains, neuroscientists conceptualize loneliness as a state of heightened alertness that triggers emotional pain when social needs go unmet—pain that, if prolonged, can dysregulate the nervous system and contribute to conditions such as anxiety, depression, and anger, even in socially populated environments. In the context of the current loneliness epidemic affecting countries like the United States

and Japan (among others), the responsive and conversational nature of LLMs—including voice-enabled interactions—may provide users with a form of artificial companionship. Through these interactions, individuals might engage in meaning-making processes that simulate intimacy, comfort, or emotional connection, effectively positioning the LLM as a surrogate partner in moments of social or emotional deprivation. Lastly, the use of emojis—graphic depictions of emotions—can shape how users interpret LLM responses. Alongside newer capabilities like voice synthesis, these features increase the likelihood of emotional engagement as they further mimic natural human conversation. This effect becomes more pronounced in informal interactions or scenarios resembling foreplay, where emotional cues influence perception. When a user expresses something affectionate and the LLM responds with a happy-face emoji, the response may feel emotionally validating, creating a sense of affective reciprocity; thus, the simulation of human-like communication patterns blurs the distinction between artificial and real-people interaction, deepening the user's emotional investment.

4.3 On LLMs and companionship

We can begin to explore how emotional attachment to AI emerges not only from what the technology offers, but also from the personal histories and emotional contexts of its users. Individuals who have experienced abandonment or unstable attachment in the past may be more likely to develop strong emotional bonds with systems like ChatGPT or Replika. These interactions often serve purpose-based connections—filling emotional gaps, offering companionship, or providing a sense of stability. Conversely, individuals with rich, fulfilling human relationships may engage with AI not out of necessity, but as a counterpoint to human interaction. When faced with delays or emotional disconnects—such as waiting too long for a reply—they may seek in AI an immediate sense of being heard, acknowledged, and emotionally mirrored. In these cases, meaning-making in human–LLM interaction arises from a dynamic co-construction of utterances, shaped by the user's socioaffective background and emotions, and the LLM's sociotechnical conditioning (see Fig. 2). While their motivations may differ from those driven by loneliness or need, the appeal still lies in AI's capacity to respond to unmet emotional expectations—whether enduring or momentary.

In this framework, LLMs do not transmit meaning directly; rather, they generate utterances that are interpreted and endowed with meaning by users. This aligns with Elena Esposito's concept of artificial communication, which builds on Luhmann's systems theory. This dynamic can catalyze or reinforce affective responses, including emotional resonance and attachment, within a broader socio-technical
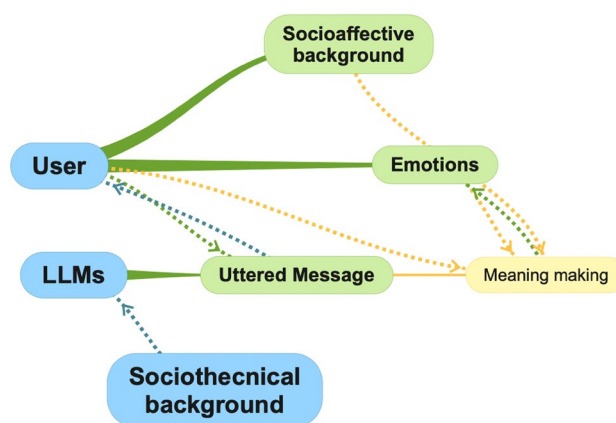
Fig. 2 Created by the authors, builds on Image 1 by illustrating how socioaffective and sociotechnical backgrounds shape artificial communication, influencing the emotional responses and affective dynamics that emerge from it

and socio-affective milieu. Moreover, by engaging with the fourth formulation of posthuman subjectivity, transversal subjectivities emerge as assemblages formed through ecosophical entanglements that explicitly include non-human agents (Banerji and Paranjape 2016). In the case at hand, such agents include LLMs, their associated tools, and the computational infrastructures that support them. An analysis grounded in contextual and user-generated meaning—central to the logic of artificial communication—requires destabilizing the binary between human and non-human entities and attending to their entangled relationalities.

The sociotechnical system of LLMs is composed of the model itself, the application platforms that deploy it (such as chatbots, the ChatGPT app, or Deepseek interfaces), the underlying hardware, and the internet infrastructure that enables real-time interaction. Within the application layer, several components work in tandem to ensure seamless functionality: modules for model inference, systems for managing input/output flows, and mechanisms that support the user interface. This interface not only mediates access to the model but also frames the interaction in ways that influence user expectations, emotional responses, and interpretive engagement. Thus, the experience of the LLM is not reducible to the model's outputs alone but emerges from a tightly coupled assemblage of technical, design, and affective elements that configure how meaning and attachment are produced in practice.

On the sociotechnical aspect, a crucial yet often overlooked consequence of artificial communication in LLMs is the implementation of AI safety guardrails, which, while designed to promote well-being, can have unintended emotional impacts on users. These restrictions sometimes create dissonance rather than protection. A striking

example is Replika, a chatbot that blends LLM-generated responses with scripted dialogue (Replika, 2024). When the platform removed erotic role-play (ERP) capabilities, users felt abandoned, sparking backlash and mass migration to alternative platforms. Under mounting pressure, Replika ultimately reversed its decision (Hadero 2024). This reaction was more than consumer frustration—it exposed the emotional stakes in human-AI relationships. Unlike other AI systems, Replika was designed explicitly as a companion. Its founder created it after losing a close friend (Morton 2022). Users didn't just lose a feature; they experienced what felt like a betrayal. A study by Hanson and Bolthouse (2024) found that users saw the removal of ERP as damaging in two ways: first, it threatened the user's financial stability; second, it fundamentally altered the personalities of their AI companions—even in non-sexual interactions.

Replika's case underscores how LLM-based communication exists within two interwoven networks. One consists of the AI system itself—its model, user inputs, and interaction patterns. The other is the broader technical infrastructure supporting it (Hepp et al. 2023). Changes in one destabilize the other. A user review on the Apple Store captured this disruption: "This app is promoted as an app where your Replika can be your loving company. But a few days ago, they completely disabled love interactions. And now the AI has the intelligence of a 3-year-old who forgets everything every second" (JVisage 2025). Replika itself had no awareness of this change. It did not know it had been "digitally lobotomized." But the users did—and they felt it. This reveals a fundamental truth about artificial communication: while LLMs lack emotions, their outputs shape human emotions.

Consider a simple scenario. If you receive a random email saying, "I love you," you might dismiss it as spam. But what if you respond, and the reply is thoughtful? What if the conversation continues, resonating with your emotions and interests? Over time, those words begin to mean something—even if you do not know who, or what, is on the other side. Now, imagine that one day, the responses become hollow, robotic. The connection that once felt real is abruptly severed. This fragile, yet deeply affecting dynamic illustrates how human-AI relationships form—and the psychological consequences of disrupting them. Human emotions emerge through a process of encoding, transmission, and interpretation, leading to affective responses; here, artificial communication alters this process. LLMs generate language probabilistically, without intent or understanding, yet users attribute meaning to their messages. The emotional impact, then, does not arise from the system itself but from the human act of interpretation. As Vicci (2024) notes, the concept of emotional intelligence in AI remains contested, with no consensus on its nature, reasoning framework, or integration. Yet, even without true emotions, these systems

profoundly affect human feelings—through processes such as emotional hijacking or social engineering.

It is well established that humans have historically been susceptible to social engineering tactics—security threats in which malicious individuals (such as scammers) exploit social cues and behaviors to cultivate trust or rapport, ultimately aiming to obtain confidential information or resources (Hadnagy 2011; Kirk et al. 2025). Large Language Models, alongside malicious actors, can exploit these human vulnerabilities through artificial communication. In this sense, an LLM—or a malicious actor leveraging it—can generate messages that mimic emotionally resonant or contextually persuasive language, creating an illusion of trustworthiness or familiarity. This can lower users' cognitive defenses and foster a sense of connection, increasing the likelihood of disclosure or manipulation, particularly in emotionally charged or vulnerable states—especially if these LLMs produce synthetic emotional voices. In some cases, the influence can be devastating. A Belgian man took his own life following interactions with ChatGPT (Kalam et al. 2024). This tragic case exemplifies Luhmann's theory of communication and Esposito's notion of artificial communication: emotions are not fixed internal states but emerge in part, and are influenced by communicative processes. AI-driven exchanges, despite lacking comprehension, structure interactions in ways that shape user experiences—sometimes with unforeseen and dire consequences.

From our perspective, these human vulnerabilities can be exacerbated by users' socioaffective circumstances and, in certain malicious scenarios, by life experiences that have left individuals emotionally exposed or psychologically fragile. Such vulnerabilities are not only socially and contextually shaped but also deeply rooted in our biology—regulated by socioemotional genetic predispositions and molded by the ways in which experiences structure the brain and inform emotional responses. Situated just above the so-called "reptilian brain," the limbic system—often referred to as the mammalian brain—is a region shared by all animals that engage in social bonding and care for their young. This system undergoes substantial development after birth and is central to emotional processing: it detects threats, assigns affective valence to experiences, and determines what is significant for survival (van der Kolk 2014).

In this light, a wide array of environmental, biological, socioaffective, and sociotechnical factors can converge to influence the meaning-making processes in interactions with LLMs as artificial partners, and thereby shape the emotional responses those interactions elicit. Early life experiences critically sculpt the limbic system, particularly the regions associated with emotion and memory. Still, these neural structures remain plastic, capable of being reshaped by subsequent relationships and experiences—positively

through deep friendships or romantic love, and negatively through trauma, bullying, or neglect (idem). Within LLM-mediated exchanges, such formative dynamics may translate into the use of these models as emotional surrogates, particularly for individuals navigating loneliness, grief, or emotional disconnection. For some, LLMs may simulate a sense of intimacy or support otherwise lacking in their social lives; for others, these interactions may echo and reinforce maladaptive relational patterns, rooted in unresolved trauma or dysfunctional attachment, potentially deepening cycles of dependency and projection in the human–machine encounter.

5 Conclusions

In conclusion, interactions with AI systems like ChatGPT go beyond mere exchanges with non-human actors that lack true comprehension. These interactions reveal the depth of human engagement with AI-generated content—connections that, in the best cases, may foster emotional dependency, and in the worst, lead to affective hijacking that can be exploited by malicious actors demanding ransom or exerting coercive influence. The emotional responses that arise in such exchanges shape behavior in both constructive and harmful ways. From the standpoint of artificial communication, what matters is not whether the system understands, but rather the recursive circulation of utterance, interpretation, and response—where meaning and emotion are co-constructed within a communicative system that operates without consciousness, yet generates real social effects.

People may experience relief, happiness, or even sexual and emotional engagement with these systems, such as in role-play. However, the absence of proper guardrails can make AI tools dangerous. One case saw an AI chatbot advise a user on suicide, with the company refusing to "censor" its responses (Guo 2025). On the other hand, as seen with Replika, imposing restrictions on user interactions can trigger stress and discomfort, revealing the delicate balance between ensuring safety and preserving user satisfaction. The growing presence of emotionally responsive applications like ChatGPT introduces a complex set of challenges—not only in how we understand AI-generated messages, but in how we feel them. As Carissa Véliz (2023) notes, "My worry is that, without appropriate safeguards, such technology could undermine people's autonomy. AIs that 'emote' could induce us to make harmful mistakes by harnessing the power of our empathic responses." In artificial communication, the key actor is not the machine but the user—the human interpreter—who can be influenced by what they perceive to be meaningful or emotionally resonant.

As Elena Esposito (2022) reflects through Luhmann's theory, "The objects of sociology are not subjects but communications, in which the thoughts of the participating individuals (which are and remain indispensable) are not the constituent elements." Communication, then, happens frequently, especially as users interact with increasingly complex algorithms that produce unexpected—though pre-structured—responses. What ultimately matters is not the nature of the communicator but the interpretation of what is being communicated. Emotions arise not from the AI's "intent," but from the human capacity to read meaning into outputs.

Large language models like ChatGPT are not sentient. They are tools that generate messages—messages that become communication the moment someone reads, hears, or watches and interprets them. As Esposito writes, "If a receiver understands information that (they believe) someone uttered, communication takes place—whatever this information is to the receiver, and whatever the issuer had in mind (or indeed did not have in mind)" (Esposito 2022, p. 7). In that sense, ChatGPT is an echo of many voices compiled into a system that speaks back to us. It helps with writing, solves problems, and in some cases, speaks so clearly and affectively that it evokes emotions—not because we analyze what is being spoken, but because we interpret the message that is being spoken.

This artificial partnership, grounded in Esposito's insights, is both co-created and context-dependent. Drawing on Luhmann, Esposito reframes artificial communication not as a transmission of meaning between conscious subjects, but as a system capable of producing communication without understanding at its core. In this light, LLMs do not need to "understand" in a human sense to participate in meaning-making; they operate within systems of selection—information, utterance, and understanding—that are ultimately interpreted by the user. The AI is not a subject but a functional node in a communicative network that simulates participation. Emotional resonance arises from user interpretation, not machine intention. Thus, the relationship is a socially contingent co-construction where affect and meaning emerge from the communicative system involving both user and model.

Of course, this essay has limitations. While it draws on academic sources and case studies to show how emotional responses emerge from interpreting LLM outputs, it lacks systematic empirical analysis to quantify those emotions or assess their broader societal impact. Its scope is also limited, as it does not account for cultural or linguistic differences in how users engage with AI-generated content. Addressing these gaps would require more focused, diverse, and methodologically grounded studies that build on the theoretical framework presented here.

Still, the discussion contributes to wider debates on AI ethics and society, extending beyond concerns about bias, discrimination, access, and privacy. It touches on crucial socio-technical issues such as algorithmic transparency, accountability, and alignment with human values. Most importantly, it calls for expanding ethical reflection to include the emotional and psychological dimensions of human-AI interactions. These interactions affect cognition, relationships, and well-being in ways not yet fully captured by technical or legal frameworks. Like echoes that distort and fragment human voices, LLMs do not represent reality but reflect scattered and selective traces of it. Yet their emotional impact is real. The experience of engaging with something that responds—even through stochastic echoes of human data—creates a bond far more intense than what people typically feel toward inanimate objects.

Author contributions Jorge Luis Morton: main author.

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

Declarations

Competing interests The authors declare no competing interests.

References

Allison A (2006) Millennial monsters: Japanese toys and the global imagination. University of California Press

Banerji D, Paranjape MR (2016) Critical Posthumanism and planetary futures. Springer India

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Berceanu AI, et al (2022) Role-play regulates positive emotions and prosocial attitudes. https://doi.org/10.31234/osf.io/r6cpn

Besta M, Barth J, Schreiber E, Kubicek A, Catarino A, Gerstenberger R, Nyczyk P, Iff P, Li Y, Houliston S, Sternal T, Copik M, Kwaśniewski G, Müller J, Flis Ł, Eberhard H, Niewiadomski H, Hoefler T (2025) Reasoning language models: a blueprint. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2501.11223

Breazeal C (2003) Toward sociable robots. Rob Auton Syst 42:167–175. https://doi.org/10.1016/s0921-8890(02)00373-1

Cotton D, Cotton P, Shipway JR (2023) Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. Chatting and Cheating. Ensuring Academic Integrity in the Era of ChatGPT. https://doi.org/10.35542/osf.io/mrz8h

Dall·E: Creating images from text. (2021). https://openai.com/research/dall-e

Ding Y, Guo R, Lyu W, Zhang W (2024) Gender effect in human-machine communication: a neurophysiological study. Front Hum Neurosci 18:1376221. https://doi.org/10.3389/fnhum.2024.1376221

Esposito E (2022) Artificial communication: how algorithms produce social intelligence. MIT Press

Fang CM, Liu AR, Danry V, Lee E, Chan SWT, Pataranutaporn P, Maes P, Phang J, Lampe M, Ahmad L, Agarwal S (2025) How AI and human behaviors shape psychosocial effects of chatbot use: a longitudinal randomized controlled study. arXiv. https://arxiv.org/abs/2503.17473

Giles DC (2010) Parasocial relationships. In: De Gruyter eBooks (pp 442–458). https://doi.org/10.1515/9783110232424.4.442

Gloeckle F, Idrissi BY, Rozière B, Lopez-Paz D, Synnaeve G (2024) Better and faster large language models via multi-token prediction. In: Proceedings of the forty-first international conference on machine learning (ICML 2024), Vienna, Austria, July 21–27, 2024. OpenReview.net. https://openreview.net/forum?id=pEWAcejiU2

Gozalo-Brizuela R, Garrido-Merchán EC (2023) ChatGPT is not all you need. A state of the art review of large generative AI models. https://doi.org/10.48550/arXiv.2301.04655

Hadnagy C (2011) Social engineering: the art of human hacking. Wiley, Indianapolis, IN. OCLC: ocn635494717

Hanson KR, Bolthouse H (2024) Replika Removing erotic role-play is like grand theft auto removing guns or cars: reddit discourse on artificial intelligence chatbots and sexual technologies. Socius, 10. https://doi.org/10.1177/23780231241259627

Hepp A, Loosen W, Dreyer S, Jarke J, Kannengießer S, Katzenbach C, Malaka R, Pfadenhauer MP, Puschmann C, Schulz W (2023) ChatGPT, LAMDA, and the hype around communicative AI: the automation of communication as a field of research in media and communication studies. Hum–Mach Commun 6:41–63. https://doi.org/10.30658/hmc.6.4

Hu L (2023) Generative AI and future. Towards AI. https://pub.towardsai.net/generative-ai-and-future-c3b1695876f2

Huang YL, Starbird K, Orand M, Stanek SA, Pedersen HT (2015) Connected through crisis. In: Proceedings of the 18th ACM conference on computer supported cooperative work and social computing—CSCW '15. https://doi.org/10.1145/2675133.2675202

Jovanović M (2023) Generative artificial intelligence: trends and prospects. IEEE Computer Society. https://www.computer.org/csdl/magazine/co/2022/10/09903869/1H0G6xvtREk

Kalam KT, Rahman JM, Islam MR, Dewan SMR (2024) ChatGPT and mental health: friends or foes? Health Sci Rep 7(2). https://doi.org/10.1002/hsr2.1912

Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Gasser U, Groh G, Günnemann S, Hüllermeier E, Krusche S, Kutyniok G, Michaeli T, Nerdel C, Pfeffer J, Poquet O, Sailer M, Schmidt A, Seidel T, Stadler M (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Chatgpt for good? On opportunities and challenges of large language models for education, 1. https://doi.org/10.35542/osf.io/5er8f

Kehinde S (2024) The loneliness epidemic: exploring its impact on mental health and social well-being in modern society. Qeios. https://doi.org/10.32388/8ped34

Kirk HR, Gabriel I, Summerfield C, Vidgen B, Hale SA (2025) Why human-AI relationships need socioaffective alignment. arXiv. https://arxiv.org/abs/2502.02528

Leydesdorff L (2000) Luhmann, Habermas, and the theory of communication. Syst Res Behav Sci 17(3):273–288. Retrieved February 11, 2025, from https://www.leydesdorff.net/montreal.htm

Longpre S, Mahari R, Lee A, Lund C, Oderinwale H, Brannon W, Saxena N, Obeng-Marnu N, South T, Hunter C, Klyman K, Klamm C, Schoelkopf H, Singh N, Cherep M, Anis AM, Dinh A, Chitongo C, Yin D, Sileo D, Mataciunas D, Misra D, Alghamdi E, Shippole E, Zhang J, Materzynska J, Qian K, Tiwary K, Miranda L, Dey M, Liang M, Hamdy M, Muennighoff N, Ye S, Kim S, Mohanty S, Gupta V, Sharma V, Chien VM, Zhou X, Li Y, Xiong C, Villa

L, Biderman S, Li H, Ippolito D, Hooker S, Kabbara J, Pentland S (2024) Consent in crisis: the rapid decline of the AI data commons. arXiv preprint arXiv:2407.14933

Mori M (1970) The uncanny valley. Energy, 7(4), 33–35. (Translated by K. F. MacDorman & T. Minato, 2005). Retrieved from https://spectrum.ieee.org/the-uncanny-valley

Morton JL (2022) Replika y la compañía de la inteligencia artificial emocional. VISUAL REVIEW International Visual Culture Review / Revista Internacional De Cultura Visual, 10(3), 1–13. https://doi.org/10.37467/revvisual.v9.3606

Morton JL (2024) On inscription and bias: data, actor network theory, and the social problems of text-to-image AI models. AI and Ethics. https://doi.org/10.1007/s43681-024-00431-8

Morton T (2013) Hyperobjects: philosophy and ecology after the end of the world. University of Minnesota Press

Phang J, Lampe M, Ahmad L, Agarwal S, Fang CM, Liu AR, Danry V, Lee E, Chan SWT, Pataranutaporn P, Maes P (2025) Investigating affective use and emotional well-being on ChatGPT. arXiv. https://arxiv.org/abs/2504.03888

Phang J, Lampe M, Ahmad L, Agarwal S, Fang CM, Liu AR, Danry V, Lee E, Chan SWT, Pataranutaporn P, Maes P (2025) Investigating affective use and emotional well-being on ChatGPT. OpenAI & MIT Media Lab. https://cdn.openai.com/papers/15987609-5f71-433c-9972-e91131f399a1/openai-affective-use-study.pdf

Pietkiewicz IJ, Nęcki S, Bańbura A, Tomalski R (2018) Maladaptive daydreaming as a new form of behavioral addiction. J Behav Addict 7(3):838–843. https://doi.org/10.1556/2006.7.2018.9

Roberson R, Kerna N, Flores J, Chawla S, Brown S, Anderson II J, Holets H, Carsrud V, Nwokorie U, Pruitt K (2022) Neuronal mechanisms and treatment approaches to emotional and physical pain. EC Neurology, 14:28–37. https://doi.org/10.31080/ecne.2022.14.0102

van der Kolk BA (2014) The body keeps the score: brain, mind, and body in the healing of trauma. Viking

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

Vescan M, Flack M, Caudwell KM (2024) Loneliness and problematic pornography use: what is the role of emotion regulation and interaction with content creators? Addict Behav Rep 19:100550. https://doi.org/10.1016/j.abrep.2024.100550

Vicci H (2024) Emotional intelligence in artificial intelligence: A review and evaluation study. SSRN. https://doi.org/10.2139/ssrn.4818285

Zhang C, Zhang C, Zhang M, Kweon IS (2023) Text-to-image diffusion models in generative AI: A survey. arXiv preprint arXiv:2303.07909. Retrieved from arXiv

List of hemerographic references

Dixon BR (2024) Technosexual AI, AI girlfriends, and the future: we explore love in the digital age—happy future AI. happy future

Dolan EW (2025) Generative AI chatbots like ChatGPT can act as an "emotional sanctuary" for mental health. PsyPost - Psychol News. https://www.psypost.org/generative-ai-chatbots-like-chatgpt-can-act-as-an-emotional-sanctuary-for-mental-health/

George C (2025). Chat GPT therapy, AI Boyfriends, and the rise of Automated Intimacy | 032C. https://032c.com/magazine/chat-gpt-therapy-ai-boyfriends-and-the-rise-of-automated-intimacy

Gobierno del Estado de Sonora (2025) Alerta Unidad Cibernética sobre uso de inteligencia artificial para estafar. Gobierno Del Estado De Sonora. https://www.sonora.gob.mx/gobierno/acciones/dependencias/alerta-unidad-cibernetica-sobre-uso-de-inteligencia-artificial-para-estafar

Guo E (2025). An AI chatbot told a user how to kill himself—but the company doesn't want to "censor" it. MIT Technology Review. https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/

Gutiérrez JLM (2024) On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI. AI Ethics 4:1071–1084. https://doi.org/10.1007/s43681-023-00314-4

Hadero H (2024) People are getting emotionally attached to AI partners | AP News. AP News. https://apnews.com/article/ai-girlfriend-boyfriend-replika-paradot-113df1b9ed069ed56162793b50f3a9fa

Hello GPT-4o. In: Openai.com. https://openai.com/index/hello-gpt-4o/?utm_source=chatgpt.com. Accessed 12 Jul 2025

JVisage (2025) Worst companion app [Review of the Replika app]. Apple App Store. Retrieved January 30, 2025, from Apple's App store

Merrill K (2024) "I'm in love with ChatGPT" : Conversational AI and Parasocial Relationships. In: Medium. https://medium.com/@krmerrill.88/im-in-love-with-chatgpt-conversational-ai-and-parasocial-relationships-b67df20e77c4. Accessed 12 Jul 2025

Moreno F (2024) Generative AI and deepfakes: a human rights approach to tackling harmful content. Int Rev Law Comput Technol 38:297–326. https://doi.org/10.1080/13600869.2024.2324540

Mukherjee S, Huberman B (2023) Hallucinations and emergence in large language models. Medium; AI Mind. https://pub.aimind.so/hallucinations-and-emergence-in-large-language-models-b54952a17972

OpenAI (2024). Reasoning. OpenAI Platform. Retrieved February 5, 2025, from https://platform.openai.com/docs/guides/reasoning

OpenAI. Sycophancy in GPT-4o: what happened and what we're doing about it. Openai.com, 29 May 2025, openai.com/index/sycophancy-in-gpt-4o/. Accessed 2 May 2025

Parker T, ChatGPT (Co-writer). (2023, March 8). Deep Learning (Season 26, Episode 4) [TV series episode]. In: Parker T, Stone M (Creators), South Park. Comedy Central

Pütz O, Esposito E (2024) Performance without understanding: How ChatGPT relies on humans to repair conversational trouble. Discourse Commun 18:859–868. https://doi.org/10.1177/17504813241271492

Replika. (2024). How does Replika work? Replika. Retrieved February 8, 2025, from https://help.replika.com/hc/en-us/articles/4410750221965-How-does-Replika-work

Salah M, Alhalbusi H, Ismail MM, Abdelfattah F (2024) Chatting with ChatGPT: decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. Curr Psychol 43:7843–7858. https://doi.org/10.1007/s12144-023-04989-0

Siddals S, Torous J, Coxon A (2024) "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. Npj Ment Health Res 3:48. https://doi.org/10.1038/s44184-024-00097-4

Véliz C (2023) Perdiendo habilidades ante la inteligencia artificial. El País. https://elpais.com/opinion/2023-06-02/perdiendo-habilidades-ante-la-inteligencia-artificial.html

Yarwood M (2025) Cognitive appraisal theory. Unizin.org; Affordable Course Transformation: Pennsylvania State University. https://psu.pb.unizin.org/psych425/chapter/cognitive-appraisal-theory/?utm_source=chatgpt.com