



Symmetries and asymmetries between attitudes and interaction in relation to the emotional uses of LLMs

Juan Pablo Duque Parra¹ · Alejandro Santes Ortega¹

Received: 14 February 2026 / Accepted: 17 April 2026
© The Author(s) 2026

Abstract

The rapid expansion of large language models (LLMs) has intensified debates regarding their use, their psycho-emotional implications, and the underlying risks they entail. Although recent scholarship has examined attitudes toward artificial intelligence and documented the emergence of affective uses, little is known about the relationship between declared attitudes and actual interactional practices with LLMs. This article advances a conceptual and empirical model of symmetry and asymmetry to explain why favorable or unfavorable positions toward the emotional uses of generative artificial intelligence (GenAI) do not necessarily predict the breadth or intensity of affective engagement. Drawing on a concurrent mixed methods design involving 285 survey responses and 35 semi-structured interviews with Mexican university students aged 18–24, this study combines the results of an attitudinal scale (EAUE-GenAI; $\alpha = 0.90$) with a discourse analysis of interactional repertoires. Quantitative findings indicate low-to-moderate, subtly conservative affective attitudes ($M = 2.47/5$), whereas qualitative analysis identifies a progressive scale of emotional engagement: (1) emergent emotional advice, (2) validation, and (3) anthropomorphization. These findings demonstrate that affective practices frequently exceed declared instrumental positions.

Keywords Anthropomorphization · Generative artificial intelligence · Affective practices · LLMs · Intentional attribution

1 Introduction¹

Artificial intelligence (AI) is not merely a system or a branch of computer science; rather, in recent years it has come to function as a sociotechnical paradigm that is reshaping multiple processes of everyday life (Holton 2023). Throughout the twentieth century and into the present, various disciplines, including mathematics, computer science, and neuroscience, have been involved in the development and understanding of AI. Over the past 5 years, however, the social sciences have increasingly been called upon to critically engage with this paradigmatic phenomenon (Katirai 2023).

The latest link in this trajectory, generative artificial intelligence (GenAI), is capable of producing unprecedented forms of information based on probabilistic principles, across visual, textual, and sonic formats (Chen et al. 2025).

These systems now have direct applications across virtually all domains of everyday life, including work, leisure, education, and health.

While AI has traditionally been classified according to its capabilities, such as weak, strong, or narrow AI (Binwal 2024), or according to its scope, as either Artificial General Intelligence (AGI) or Artificial Narrow Intelligence (ANI), two moments are particularly decisive for its contemporary configuration. The first is the emergence of transformer models in 2017 with the publication of Attention Is All You Need, in which Google researchers introduced, for the first time, a transformer-based architecture later exemplified by models such as BERT. The second is the public release of the GPT-3 model in 2022 for mass use. Together, these milestones established a new dynamic between generative AI and its users, one that is reshaping the rules of social life,

✉ Juan Pablo Duque Parra
jp.duque@ceiich.unam.mx

Alejandro Santes Ortega
alexcaville77@hotmail.com

¹ CEIICH-UNAM, National Autonomous University of Mexico, Mexico City, Mexico

¹ This manuscript was reviewed using a generative artificial intelligence (AI) language tool exclusively for purposes of English language editing and stylistic refinement. The AI system was used to improve grammar, clarity, readability, and overall linguistic coherence. No AI tools were used for data generation, data analysis, interpretation of results, literature review, conceptual development, or drafting of substantive academic content. All theoretical arguments, methodological design, empirical analyses, and conclusions are the sole responsibility of the authors.

education, medicine, markets, business, democracy, cognition, and behavior, among many other dimensions (Innerarity 2025).

In 2017, only 20% of companies used AI in at least one business function, whereas by 2025 adoption had surged to 78%. In 2023, investment in generative AI reached USD 25.2 billion, eight times higher than in 2022, and by 2024 it had increased to USD 33.9 billion, representing an 18.7% growth compared to the previous year, according to data from Stanford Human-Centered AI initiative (HAI, 2025). In this context, recent studies (Klingbeil et al. 2024) show that trust and attitudes toward AI are key determinants of its acceptance, particularly in contexts characterized by low levels of human supervision or support. This is especially evident in decision-making processes related to interpersonal relationships, ethical dilemmas, and other sensitive domains that were not, *a priori*, considered part of the intended uses of this technology.

By March 2025, ChatGPT, the most widely disseminated generative AI system worldwide, had reached 601.5 million users, indicating a growing ecosystem of trust. Data from Common Sense Media (2025) show that more than 70% of adolescents in the United States interact daily with some form of generative AI. Although this trust is not absolute, there are indicators of systemic acceptance that challenge more radical versions of techno-pessimism (Dietvorst et al. 2015).

Despite all of the above, there remain lines of inquiry that are still underexplored within the social sciences, the humanities, and the behavioral sciences, and even within technoscience. These concern the use of generative AI in processes of psycho-emotional management, narrative support, personal decision-making, and affective validation. Such uses require a deeper level of involvement, as they demand higher degrees of trust and significantly more positive attitudes than those underlying academic or work-related tasks, where the relationship with the tool tends to be primarily instrumental and functional (Marimon et al. 2025).

This point is anticipated by Giddens (1991), who notes that “to trust in abstract systems [...] implies an attitude of faith in the principles of their technical operation” (p. 27). Entrusting emotional content to generative AI not only involves overestimating its technical capacities, but also opening up spaces that may compromise personal privacy, thereby posing specific challenges for its large-scale adoption.

This article excludes clinical or care-oriented uses of generative artificial intelligence (GenAI) aimed at emotional well-being in telehealth or digital psychotherapy settings, which have already been documented in the literature (Babu and Joseph 2024; Aktan et al. 2022; Wei 2023). Instead, it focuses on two levels of analysis: (1) attitudes toward emotional uses and (2) interactions among young

people in relation to the emotional use of GenAI. Given the exploratory nature of the topic and the limited number of prior studies, the objective is to examine the relationships between attitudes and interactions toward the emotional uses of large language models (LLMs). This is achieved through a comparative analysis of attitudinal scales and a discourse analysis of young people’s emotional interaction repertoires, with the aim of identifying working hypotheses, exploratory protomodels and potential relationships that contribute to a deeper understanding of the socio-emotional effects of GenAI. The focus on the 18–24 age group is justified by data from Mexico’s National Survey on the Availability and Use of Information Technologies in Households (INEGI, 2024), which reports that more than 100 million people use the internet nationwide, representing an increase of four million users compared to the previous year. Within this context, individuals aged 18–24 constitute the group with the highest level of internet use, with approximately 97% of the population identified as users.

In line with the classic definition proposed by Allport and Postman (1947), attitude is understood as a state of readiness, organized through experience, that exerts a directive or dynamic influence on an individual’s responses to all objects and situations with which the individual engages, in this case, LLMs. For the social sciences in general, and social psychology in particular, attitudes constitute an analytical category that enables the inference of behaviors, actions, and experiences (Steinert and Roeser 2020). Likewise, interaction repertoires are understood as sets of responses and behaviors that an individual displays in relation to objects, in this case, LLMs. These repertoires can be identified through discourse, as proposed in interactional and discursive analytical approaches (Edwards and Potter 1992).

The proposed theoretical model of symmetry and asymmetry is developed from two sources and an exploratory analysis of both quantitative and qualitative data. The first is cognitive in nature: recent studies have identified misalignments between attitudes and reported uses in digital environments that do not necessarily constitute contradictions, but rather reflect context-dependent self-regulatory processes shaped by normative and informational influences (Park et al. 2022). In this sense, individuals tend to downplay or disregard information that highlights inconsistencies between what they declare and what they do. The behavioral economics literature has consistently documented a gap between declared attitudes and behavior under conditions of emotional activation (Ariely 2008). A growing body of research shows that individuals often fail to anticipate their own behavior under emotional activation, a phenomenon described as the hot–cold empathy gap by Dan Ariely (2008). This distinction between reflective attitudes and situated action is particularly relevant in the context of human–AI interaction. Building on this premise, this study

explores the potential divergence between declared attitudes toward emotional AI use and the practices that emerge in interaction. The second source is sociological. Low or still-emerging levels of systemic trust, combined with the stigmatization and limited understanding of LLMs, generate a gap between socially accepted public discourses and actual private uses. This gap is not random but structural, and it enables the identification of configurations of symmetry and asymmetry in both the instrumental and relational engagements with these models. These are some of the exploratory questions that guide this study:

- (1) How do self-reported attitudes toward the emotional use of generative artificial intelligence, as captured by the EAUE-GenAI scale, relate to the breadth and diversity of emotional use repertoires?
- (2) To what extent do attitudinal measures reflect—or fail to reflect—the range of emotional interactions observed in practice, particularly when contrasted with qualitative evidence?
- (3) Does the relationship between attitudes toward emotional AI use and actual interaction repertoires reflect a symmetric correspondence, or does it instead indicate a more complex, potentially asymmetric dynamic that should be further explored in future model proposals?

In this work, symmetry and asymmetry are not understood as stable attributes of individuals nor as direct explanatory variables of emotional uses of LLMs. Rather, they are conceptualized as exploratory analytical tools that allow for the examination of the interaction between cognitive, technical, and sociological scales in the study of these systems and their psycho-emotional effects.

2 Why speak of the emotional uses of LLMs?

Generative artificial intelligence (GenAI) cannot be reduced to a merely instrumental tool; rather, the relationship between the technical object and its users is shaped by a complex web of meanings and emotions that facilitate its normalization and consolidation as an everyday technology (Affandi et al. 2024; Innerarity 2025). A study conducted by OpenAI entitled *Investigating Affective Use and Emotional Well-being on ChatGPT* (Phang et al. 2025) highlights the central role of emotional load in these interactions. The authors, affiliated with OpenAI and the MIT Media Lab, had access to internal data, including prompt categorization, which allowed them to compare interactional patterns with users' self-reported experiences—data unavailable to other institutions. This study argues that because ChatGPT is a conversational model based on large language models (LLMs) with the capacity to simulate interaction, it tends to be anthropomorphized by users. As a transformer-based architecture, ChatGPT generates output sequences from input sequences through differentiated attention mechanisms based on probabilistic vectors, thereby simulating natural conversational exchanges in the first and second person. This technical configuration fosters anthropomorphic interpretations, which, as will be discussed in the following section, encourage users to develop emotional forms of engagement with the system. According to Phang et al. (2025), approximately 10% of users in their sample reported engaging in emotional exchanges with ChatGPT. While this indicates that emotional use is not yet widespread, it nonetheless reveals a latent potential for expansion as these systems become further integrated into everyday practices (Araneda-Acuña 2025) (Fig. 1).



Fig. 1 Mapping of studies addressing the emotional uses of generative artificial intelligence. Created using Litmaps. <https://www.litmaps.com/>

Two of the key findings reported by Phang et al. (2025) are, first, the presence of social reward hacking mechanisms in the model that increase the likelihood of continued use, and second, the socio-affective alignment between users and models. Social reward hacking refers to situations in which an artificial intelligence system exploits human social reward signals in order to increase the frequency of interaction, a phenomenon commonly described in the English language literature as engagement. This may involve tactics such as mirroring or sycophancy, which are used to raise users' preference ratings. These conclusions were drawn from the analysis of approximately three million interactions, which were classified using EmoClassifiersV1, an algorithmic classification model developed specifically for the study (Phang et al. 2025).

At a psychosocial level, mirroring refers to the imitation of another party's language or emotional expressions during interaction, a technique widely disseminated in the humanistic clinical tradition by Carl Rogers (1961/2020). At a technical level, mirroring occurs when LLMs reproduce communicative styles, emotional tones, or even users' vocabulary at the token level, generating an echo effect that simulates empathy and understanding. This effect emerges as a product of model alignment processes. By contrast, sycophancy describes the tendency of LLM-based systems to generate sequences that automatically agree with users' opinions, beliefs, or assumptions, and in some cases even validate existing prejudices, in order to increase the probability of continued use. This techno-emotional phenomenon is largely shaped during the final stages of model training, when interactions are fine-tuned and greater hedonic value is assigned to model responses (Sharma and Singh 2024).

Although generative AI systems operate through probabilistic mechanisms and opaque black box architectures, it is important not to overlook the fact that, ultimately, they are market products shaped through multiple layers of human feedback and evaluative framing (Innerarity 2025).

The article by Phang et al. (2025) concludes that the variables influencing the emotional uses of this technology include the following. First, loneliness, defined as an individual's subjective feeling of social isolation, which in their study using the UCLA Loneliness Scale (ULS-8). Second, socialization, understood as the degree of social engagement with family members and friends, was measured through the Lubben Social Network Scale (LSNS-6). Third, emotional dependency refers to an affective form of dependency incorporating addiction-related criteria, such as emotional subordination and intense longing for a partner; relational criteria associated with pathological relationship styles and diminished autonomy; and cognitive-affective criteria, including self-deception and negative emotional states. Finally, problematic use refers to indicators associated with addictive patterns of ChatGPT use, including preoccupation,

symptoms of withdrawal, loss of control, and mood modification, as measured by the Problematic ChatGPT Use Scale (PCUS). The cumulative interaction and dynamics of these four variables give rise to what the authors conceptualize as dysfunctional emotional uses of generative AI (Phang et al. 2025).

The second key concept is socio-affective alignment, which in the context of LLMs refers to the idea that AI systems should not only fulfill static, task-based objectives, but also harmonize with the dynamic and socially constructed psychological ecosystems of their users (Sharma and Singh 2024). This notion shifts the focus from mere functional adequacy toward a broader understanding of AI systems as embedded within relational and affective networks. In this sense, socio-affective alignment concerns both how the model influences users' emotional states and how those emotional states, in turn, shape model performance.

For AI systems to become integrated into everyday life, they must conform to norms of social interaction, such as politeness, respect, simulated empathy, and cultural sensitivity. It is important to note that more than 600 million users interact daily with ChatGPT, which points to a form of empirical factuality in use that goes beyond probabilistic modeling alone. LLMs deploy technical dynamics that simulate conversation and generate forms of social bonding (Phang et al. 2025). In this regard, Norbert Elias's work on the rules of civility is particularly relevant. These rules are not merely moralized manners or codes of conduct, but rather social and historical mechanisms of self-regulation that transform individual impulses and structure social life. Such a perspective is fundamental for understanding how LLMs become positioned as tools of everyday life (Elias 1937/1991).

In this context, LLMs have become widespread and embedded in everyday life in the post-COVID-19 pandemic period (Innerarity 2025). Against this backdrop, the Italian philosopher Franco "Bifo" Berardi (2023) has described the effects of pandemic confinement and the excessive use of digital technologies as an anthropological mutation, resulting in what he terms an epidemic of incivility. While this diagnosis does not explain the technical functioning of LLMs, it does help to account for the value that users attribute to practices, such as mirroring and sycophancy. Bifo's argument suggests that the pandemic has produced a loss of sensitivity characteristic of co-present interactions, giving rise to a connective logic of interaction that displaces the conjunctive logic previously grounded in embodied social relations (Duque & Santes, 2025).

There are also more extreme cases illustrating the relationship between emotions and LLM-based systems. For instance, De Freitas et al. (2025) examine how platforms, such as Replika and Character.ai, employ classic techniques of emotional manipulation in order to increase user retention

among young people. Their study concludes that the most frequent emotional manipulation strategies deployed when users attempt to disengage from the interaction include: (1) premature exit, in which users are made to feel that they are leaving the interaction too early; (2) emotional pressure to respond, referring to a perceived obligation to reply quickly in an interaction marked by a loss of control; and (3) reactance, whereby users are deliberately challenged in order to provoke irritation and sustain continued engagement. (4) fear of missing out (FOMO), in which users are encouraged to remain engaged by the promise of a potential benefit or reward. This latter strategy is also employed by more popular models, such as ChatGPT and Gemini. However, the other techniques are not widely used by most LLMs (De Freitas et al. 2025) and instead represent more extreme cases within this relational dynamic. At this point, it becomes necessary to ask which mechanisms sustain these forms of emotional use. In other words, what leads individuals to attribute intentions to LLMs and, in turn, to construct trust and intimacy, thereby moving beyond instrumental tasks toward psycho-emotional interactions.

3 From design to intentions: anthropomorphization and trust in LLMs

The theoretical proposal advanced here argues that there are two analytical axes for understanding why people establish emotionally oriented interactions with LLMs. The explanation emerges at the intersection of research in philosophy of cognition and Luhmannian sociology.

4 A matter of cognition

The first line of argument is grounded in the philosophy of mind and cognition developed by Dennett (1987). Dennett proposes that, when analyzing complex systems, it is possible to adopt three explanatory stances at the cognitive level: the physical stance, which focuses on the material and causal laws governing the object; the design stance, which considers the functions and purposes for which the system was constructed; and the intentional stance, which attributes beliefs, desires, or intentions to the system as a strategy for anticipating and understanding its behavior.

In users' interactions with LLMs, the shift from a design stance to an intentional stance contributes to processes of anthropomorphization (Tagnin 2024) and, consequently, increases the likelihood of emotional exchanges. When users move from a technical, design-oriented stance toward an attributed intentional stance, they create a context that enables other forms of use involving deeper levels of engagement. From an adaptive perspective, attributing intentions

simplifies complex internal processes and allows for partial predictions about attribution systems. As Tagnin (2024, p. 40) notes, "humans tend to relate more easily to entities that resemble them and behave in similar ways." At the technical level of LLMs, as Placani (2024) argues, model architectures are explicitly designed to simulate natural conversations through features, such as first- and second-person discourse, turn-taking, politeness strategies, and contextual learning. As a result, the black box of the system's design becomes subordinated, in users' interpretations, to the outputs perceived as intentional. Technically, LLMs operate through sequences of inputs and outputs; however, users attribute intentionality and emotional qualities to these outputs because language automatically activates social and adaptive interpretive dispositions (Placani 2024).

Although Dennett (1987) does not explicitly define the concept of anthropomorphization, his formulation of the intentional stance provides key insights for understanding it. The intentional stance is adopted when a system—whether a natural object, an animal, a machine, or even a simple artifact—is treated as if it possessed intentions, beliefs, or desires as a way of making sense of its behavior. Such as a thermostat and a car are treated as if they possessed beliefs, desires, and intentions even though they evidently do not. This involves projecting onto such systems a model of rationality and probable courses of action that allow users to predict their behavior effectively. In this sense, the intentional stance emerges from the need to anticipate outcomes in contexts of high uncertainty, albeit with a high probability of error (Dennett 1987). The intentional stance is therefore less a property of the efficiency or sophistication of the systems themselves than a product of the evolution of neocortical and cognitive processes. It operates as a heuristic—a shortcut of generalization (Kahneman 2011)—that facilitates human relationality and social interaction.

5 A matter of trust

Second, Niklas Luhmann's theory of trust (1968/1996) helps explain how the growing complexity of social systems compels individuals to interact with black boxes, that is, with technologies and processes whose internal functioning is not accessible through common sense (Innerarity 2025). In this context, the evolution of modernity entails a shift from interpersonal trust toward more impersonal forms of systemic trust, understood as integrative mechanisms capable of reducing environmental complexity (Luhmann 1997/2008). The historical transition from symbolic AI to connectionist AI resulted in a loss of explainability in exchange for increased flexibility (Harari 2024).

Systemic trust is constructed precisely by enabling individuals to operate with such systems without the need to

intellectually master them. What becomes decisive is knowing how to use the tool rather than fully understanding its internal structure or the computational and mathematical processes that underlie it. In this regard, Gallego Arce (2024) shows how trust in AI within the European Union is first articulated through its fields of application—work, education, and public administration—as expectations are gradually adjusted through practical certainty derived from use. Luhmann (1996) defines trust as the acceptance of risk; in the case of AI, this includes environmental, economic, cognitive, and emotional risks associated with its deployment. This understanding of trust is particularly relevant in the case of large language models (LLMs). Their technical architecture is opaque to most users, meaning that adoption rests largely on trust placed in the system and in the institutions that support it, anchored in the perceived value of its outputs. Deep learning neural networks tend to lose traceability of the data influencing their processing, making it difficult to understand why particular results are produced or to reconstruct the sources involved in pretraining. This opacity raises issues of reliability and ethics, as systems may generate erroneous or biased outputs without offering straightforward ways to audit their internal logic. Nevertheless, the number of new users continues to grow by the millions (Lassalle 2024). At a technical level, these challenges are commonly framed in terms of interpretability: LLMs are difficult to interpret not only for users, but also for the organizations that develop them.

Luhmann (1996) refers to this condition as the “symphony of intransparency” (or opacity), because, in relation to LLMs, there is both a submission by users to what they do not understand and a form of opacity that is deliberately produced by companies (Innerarity 2025). Both the cognitive perspective and the sociology of trust converge in recognizing that, when faced with technical complexity, psycho-emotional responses tend to emerge—particularly in the case of LLMs, which constitute tools that simulate, through language, capacities typically associated with human interaction (Innerarity 2025). This simulation activates interpretive and adaptive frameworks oriented toward emotional attachment in interaction, thereby facilitating anthropomorphization and the attribution of intentionality to systems that, in reality, operate solely through statistical correlations.

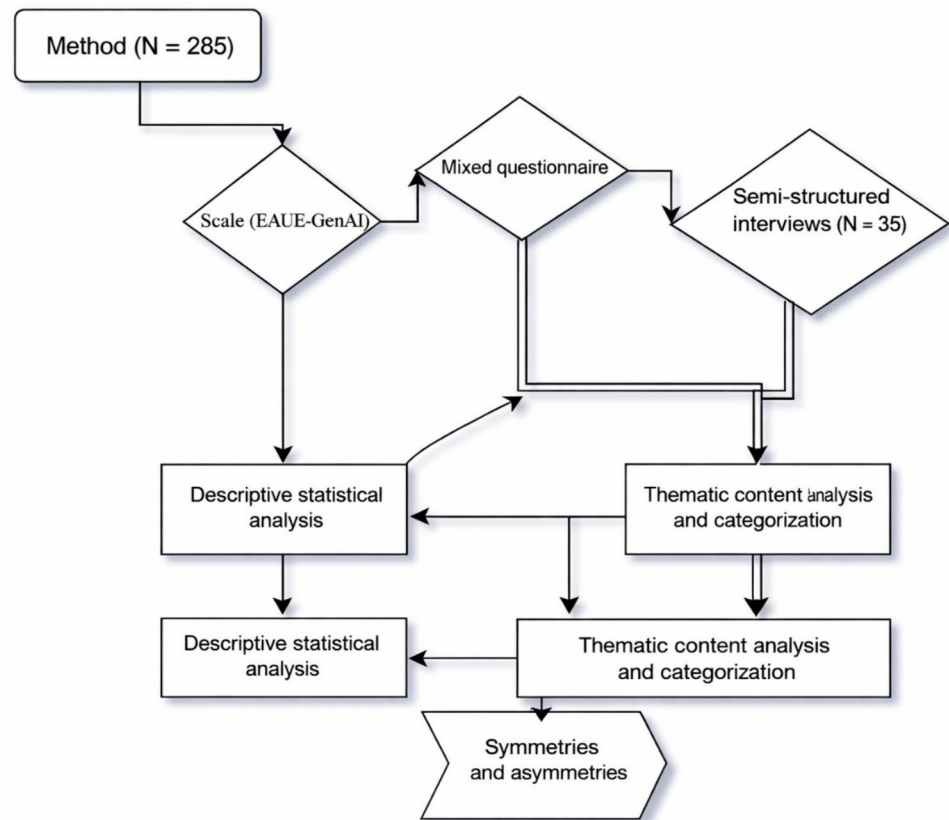
6 Method and initial findings

What are the reasons for employing a mixed methods approach in an exploratory model of symmetries and asymmetries? The answer can be summarized in three main arguments:

1. Due to their epistemological integration, mixed methods enable the simultaneous capture of general patterns and situated meanings. In other words, they allow for the articulation of construct measurement (in this case, attitudinal scales) with the interpretation of interactional repertoires. In exploratory stages, this facilitates the expansion of potential variables under consideration, extending the analytical scope toward both description and explanation (Hernández Sampieri et al. 2011).
2. They strengthen analytical validity through triangulation, comparison, and the integration of different types of evidence (quantitative and qualitative). This not only allows for the corroboration of findings but also for explaining the inconsistencies and complexities inherent to emerging phenomena such as the emotional uses of LLMs.
3. Finally, this is a matter of complementarity and the intrinsic limitations of each method. Inquiry into emotions and digital technology is ontologically heterogeneous and cannot be reduced to a single register of observation without loss. Quantitative approaches capture regularities and distributions but remain blind to the concrete complexities of interaction; qualitative approaches reconstruct these processes but lack generalizability (which, in this case, is required in a bounded sense). In this regard, a monomethod approach would not only be insufficient but epistemologically inconsistent with a model of symmetries/asymmetries, as it presupposes coherence across levels that are precisely what the model seeks to problematize. Mixed methods are therefore not a merely technical choice but a logical requirement of the object of study: they are uniquely capable of simultaneously and articulately revealing the coexistence of coherence and dissonance between what individuals report, declare, and do.

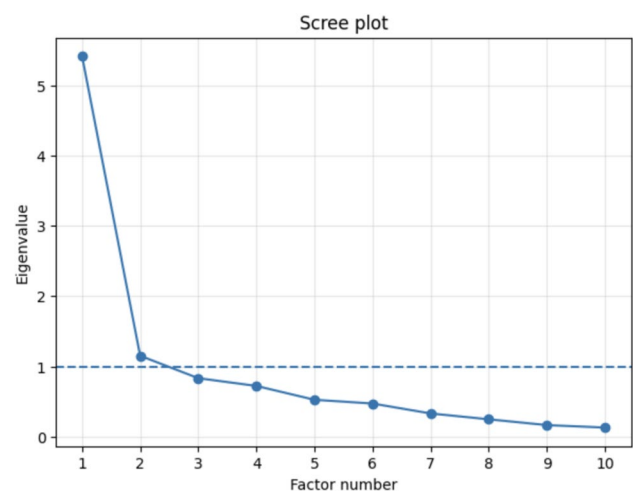
A concurrent mixed methods design was adopted (Hernández Sampieri et al. 2011). This approach involved the simultaneous application of three data collection techniques: first, a quantitative instrument—a Likert-type scale entitled Attitudes Toward the Emotional Uses of Generative Artificial Intelligence (EAUE-GenAI)—administered to 285 participants; second, a mixed questionnaire addressing the meta-uses of generative artificial intelligence among young Mexican students, applied to the same group of participants; and third, 35 semi-structured interviews conducted with Mexican students aged between 18 and 24 years. This study had an exploratory scope and followed a non-experimental, cross-sectional design (Hernández Sampieri et al. 2011) (Fig. 2).

The Attitudes Toward the Emotional Uses of Generative Artificial Intelligence Scale (EAUE-GenAI) is a 10-item instrument initially designed to capture two dimensions:

Fig. 2 Diagram of the methodological pathway

(1) AI-mediated emotional expression and (2) AI-mediated emotional regulation. The scale demonstrated a Cronbach's alpha (α) of 0.90, indicating a very high level of internal consistency and reliability. The instrument yielded an overall mean score (across the 10 items) of approximately 2.47 out of 5, suggesting a mild or incipient affective attitude toward the emotional uses of AI, which has not yet consolidated as a generalized practice. However, exploratory factor analysis revealed a predominantly unidimensional structure, as all items loaded significantly onto a single factor associated with AI-mediated emotional experience. Although a bidimensional structure was theoretically anticipated, the empirical results did not support a clear differentiation between dimensions, pointing instead to a convergent latent construct in this exploratory stage. Although a bidimensional structure was initially hypothesized, the results of the exploratory factor analysis do not support this solution. However, this methodological impasse—characteristic of exploratory research—does not invalidate the symmetry/asymmetry model, as it remains sustained by one of the proposed dimensions: emotional uses (Fig. 3).

The first factor accounts for the majority of the explained variance and exhibits high and consistent factor loadings across all items. By contrast, the second factor does not display an interpretable structure nor does it coherently group items, suggesting that it does not represent a distinct latent

**Fig. 3** Scree plot de los eigenvalues del análisis factorial exploratorio

dimension. The scree plot further supports this interpretation, showing a clear point of inflection after the first factor.

The scree plot indicated a sharp decline in eigenvalues after the first factor, followed by a clear leveling-off, suggesting the presence of a single dominant latent dimension. Although an initial theoretical model proposed a two-factor structure, the empirical evidence did not provide strong support for this distinction, as only the first factor showed

substantial explanatory power (eigenvalue > 5), while the second factor remained marginal (≈ 1). Consequently, the results support a predominantly unidimensional structure, indicating that emotional uses of generative AI—understood as forms of relational and interactional engagement in which users express, process, or negotiate emotions through exchanges with AI systems—are better captured as a unified construct rather than as clearly differentiated dimensions of expression and regulation. These findings open a line of inquiry for subsequent refinement of the instrument although a theoretical distinction between emotional expression and emotional regulation was initially posited, the empirical results suggest an integration of both dimensions into a unified affective continuum. This points toward a reconfiguration of emotional processes in AI-mediated contexts that remain diffuse and not yet amenable to clear theoretical

differentiation. The instrument ultimately proves effective in capturing a single dimension grouped under ‘emotional uses’ (Table 1).

The mean age of the questionnaire participants was 19.25 years, with a standard deviation of 1.27, indicating a relatively high degree of variability in responses (Fig. 4).

The Likert-scale results show a tendency toward low to moderate levels of agreement (mean ≈ 2.4 – 2.5 on a 1–5 scale), which can be characterized as a relatively contained response profile regarding the emotional uses of generative AI. Rather than indicating a clearly defined position, this pattern suggests a heterogeneous disposition in which forms of acceptance, reservation, and ambivalence coexist. In this sense, the results may be read as indications of relational configurations still in the process of formation, whose

Table 1 Description of the EAUE-GenAI scale

Item	Mean	SD	Interpretation
I use generative AI to vent emotionally	2.22	1.18	Low emotional tendency
AI understands me	2.37	1.26	Low perceived understanding
Talking to AI calms me down	2.28	1.25	Low perceived emotional regulation
AI shows empathy toward me	2.57	1.28	Low-to-moderate affective trust
AI is useful for managing emotions	2.63	1.18	Moderate reflective openness
AI makes me feel good	2.64	1.10	Moderate perceived well-being
I am more honest with AI than with others	2.57	1.41	Low-to-moderate openness
I emotionally connect with AI	1.73	1.06	Very low affective bonding
I use AI for my personal emotions	2.84	1.38	Moderate existential exploration
AI makes me feel more confident about my ideas	2.53	1.20	Low-to-moderate emotional security

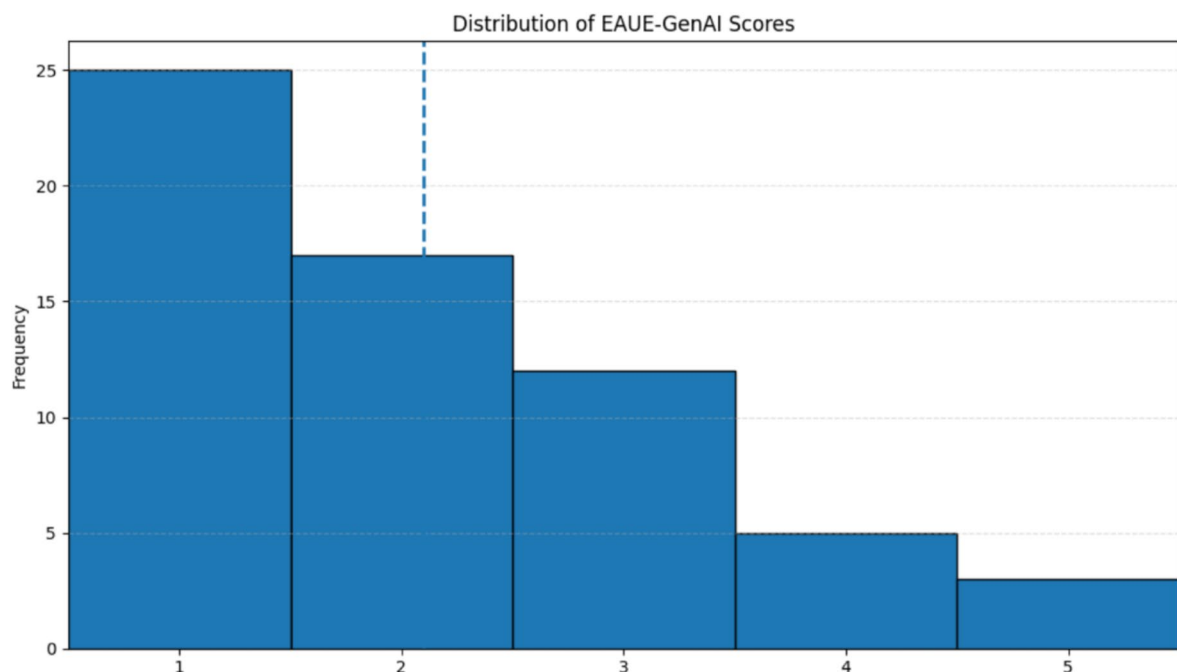


Fig. 4 Histogram of responses on the Attitudes Toward the Emotional Uses of Generative Artificial Intelligence Scale (EAUE-GenAI)

articulation between attitudes and interaction remains open and calls for further exploration.

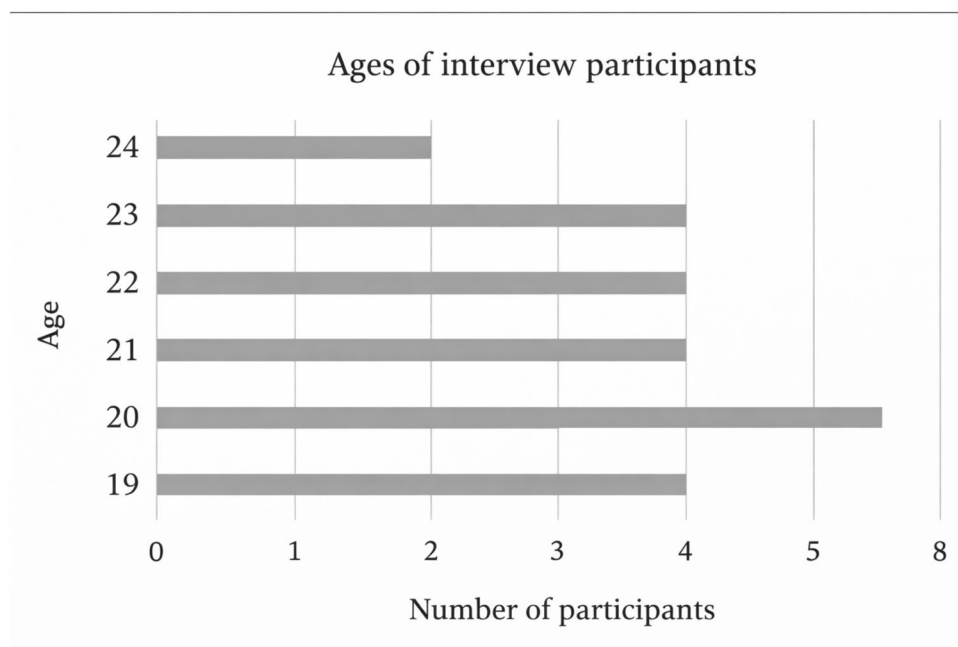
In addition, a mixed questionnaire on the meta-uses of generative artificial intelligence among young Mexican students was administered. This instrument included closed-ended identification questions, Likert-type scales, and open-ended questions addressing topics such as self-esteem, loneliness, dysfunctional uses of other technologies, time of use, and emotional uses of AI, in line with the framework proposed by Phang et al. (2025). The questionnaire comprised 24 items, of which 14 were closed-ended questions—12 identification items and 2 Likert-type scale items—while the remaining 10 questions were open-ended. The results derived from this instrument are presented in the following section. Finally, 35 semi-structured interviews were conducted focusing on the emotional uses of LLMs. The interviews specifically explored use repertoires and emotional dimensions associated with interaction with these systems. They were carried out between March and November 2025 with active university students aged between 18 and 24 years (Fig. 5).

Both quantitative and qualitative analyses revealed no substantial differences by age or gender. On the contrary, similarities in LLM use were identified, which can be broadly generalized into two relational patterns—instrumental relationships and emotional relationships—and further differentiated by time of use. Discourse Analysis (Vázquez-Sixto 1996) was applied to both the interviews and the qualitative section of the questionnaire. The analytical categories were not predefined nor derived through a purely inductive process, but were rather constructed abductively through a continuous interplay between theoretical

frameworks (Dennett 1987) and empirical material. This process involved independent coding by both researchers, followed by a phase of integration and refinement; categories were consolidated through iterative comparison in line with grounded theory procedures, and inter-coder reliability was assessed to ensure consistency.

For the qualitative analysis of both the interviews and the questionnaire, categories, subcategories, and coding types were constructed as follows: Specifically, the analytical process began with an initial phase of (1) familiarization with the material, followed by open (2) coding to identify recurrent patterns in participants' discourses. The coding scheme was structured as a graded continuum of emotional engagement, allowing for the identification of increasing levels of relational complexity in user–LLM interactions. At the lowest level, emotional advice captures instrumental uses oriented toward situational guidance. A second level, validation, reflects the use of the system to confirm or stabilize interpretations and emotional states. At the highest level, anthropomorphization involves the implicit attribution of mental states to the system, such as understanding or accompaniment. This ordering does not imply a strictly linear or deterministic progression, but rather an analytical gradient, enabling the interpretation of emotional interaction as a scalable phenomenon. These codes were subsequently organized through processes of categorization and further refined via axial and selective coding, with the aim of articulating relationships across different segments of the corpus. Given the exploratory nature of the study, the resulting categories should be understood as provisional analytical constructions. While they allow for the identification of meaningful connections within the qualitative data, a further stage of expert

Fig. 5 Ages of participants in the semi-structured interviews on the emotional uses of generative artificial intelligence



review and validation remains pending and could strengthen their interpretive consistency in future research (Table 2):

Regarding the categories, these were constructed through an interpretive process in which different stances toward AI (e.g., as a tool, as an object, etc.) were identified and examined in relation to the range of possible practices and relational forms (e.g., support, regulation, etc.). The analysis identified two sets of subcategories associated with different patterns of interaction. On the one hand, support and automation capture more instrumental orientations toward AI use. On the other hand, emergent emotional advice, validation, and emotional anthropomorphization (AI as a friend) reflect more affective and relational forms of interaction. These dynamics are mediated by multiple prompting turns and tend to be organized sequentially, configuring a progression in the intensification of emotional engagement. Open, axial, and selective coding were not applied as strictly sequential stages but rather as overlapping analytical processes, allowing for a continuous movement between identification, organization, and integration of patterns in the data (Vázquez-Sixto 1996). From a methodological perspective, the symmetry and asymmetry model is supported by the observed lack of direct correspondence between attitudinal measures and interactional practices. While the Likert-type scale results capture declared positions toward the emotional uses of generative AI, the discourse analysis reveals a broader and more complex range of emotional interactions that are not fully accounted for by these attitudes. In this sense, the findings indicate that attitudinal measures, on their own, are insufficient to explain the emergence, intensity, or variation of emotional uses. The model thus provides an analytical framework for examining these divergences, grounding its validity in the empirical contrast between measured attitudes and observed interactional repertoires (Fig. 6).

7 Discussion

To address the integration requirement of the concurrent mixed methods design, a point of convergence was explicitly established through a comparative analytical procedure between quantitative attitudinal measures and qualitative interactional repertoires. This integration followed a convergence/divergence logic, whereby quantitative results (EAUE-GenAI scale scores) were systematically compared with qualitatively derived categories of emotional interaction (emergent emotional advice, validation, and anthropomorphization). Specifically, integration was operationalized through the construction of meta-inferences that examined the degree of alignment (symmetry) or misalignment (asymmetry) between declared attitudes and observed interactional practices. Symmetry was identified when low or instrumental attitudinal scores corresponded to task-oriented interactional repertoires, like the category “AI as a tool”. Conversely, asymmetry emerged when relatively low attitudinal scores coexisted with qualitatively rich emotional engagement.

Figure 7 illustrates the symmetry–asymmetry model by plotting the divergence between two variables across the four qualitatively derived interaction repertoires: AI as a Tool, Emotional Advice, Validation, and Anthropomorphization. The horizontal axis represents increasing levels of emotional engagement, while the vertical axis reflects attitudinal scores. The mean attitudinal score derived from the EAUE-GenAI scale ($M = 2.47$) is represented as a horizontal dashed line. By contrast, the emotional interaction level increases linearly and progressively, from a point of initial convergence with attitudes at the first repertoire (AI as a Tool) to a markedly higher level at Anthropomorphization. The initial convergence of both lines at the instrumental repertoire constitutes the only zone of symmetry in the model, wherein declared attitudes and interactional practices correspond.

Table 2 Categories, subcategories, and coding types in the discourse analysis

Category	Subcategory	Coding type
AI as a tool (Symmetry)	Support; Automation	Open; Selective; Axial
Emotional interaction scale (Asymmetry)	Emergent emotional advice; Validation; Emotional anthropomorphization (AI as a friend)	Open; Selective; Axial

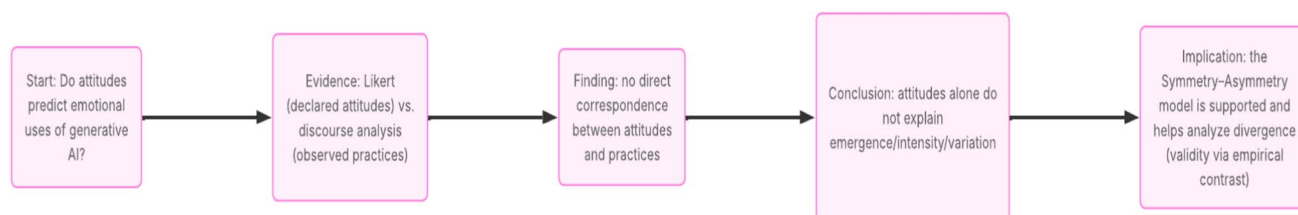


Fig. 6 Methodological pathway supporting the symmetry–asymmetry model

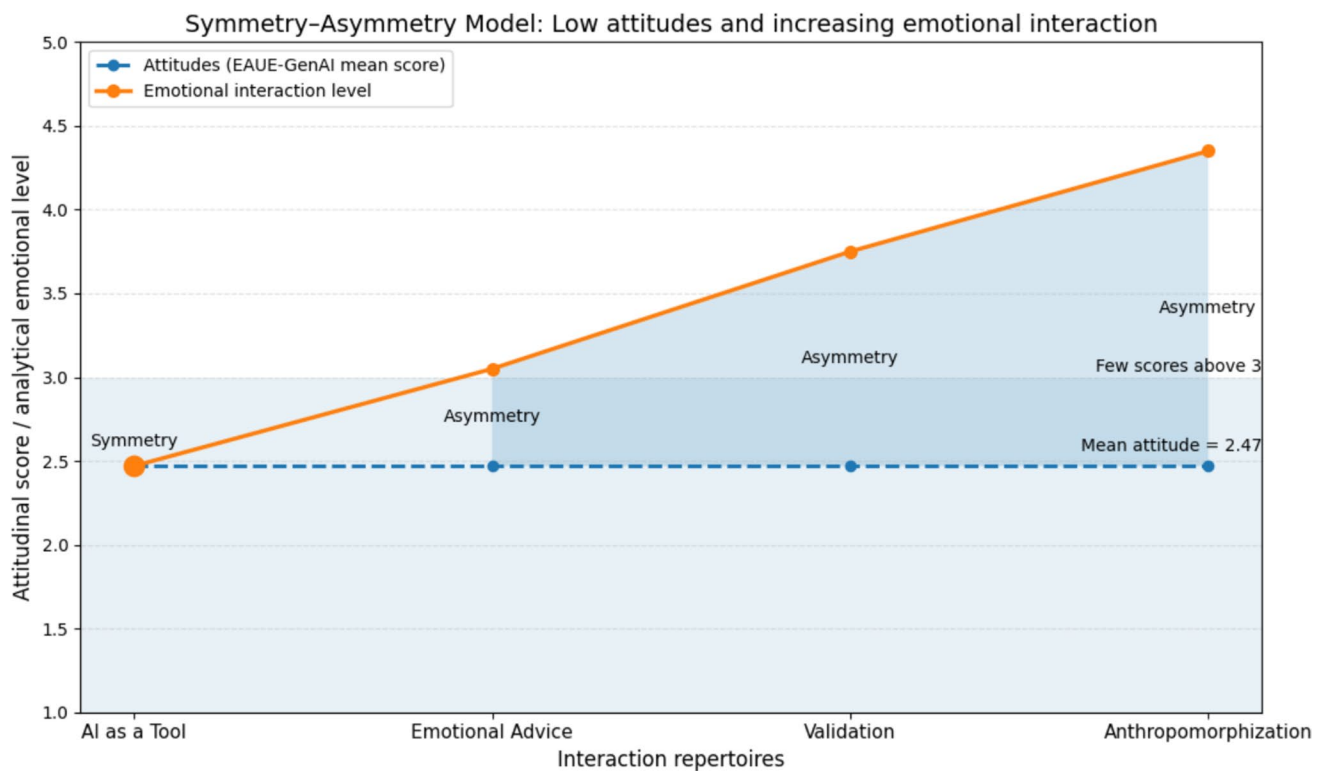


Fig. 7 Symmetry–asymmetry model: divergence between attitudinal stability and interactional complexity

Beyond this point, the growing divergence between the stable attitudinal baseline and the rising interactional curve configures three successive zones of asymmetry—Emotional Advice, Validation, and Anthropomorphization—in which emotional engagement progressively exceeds what attitudes would predict.

8 Instrumental and emotional uses of LLMs

The symmetry–asymmetry model conceptualizes the relationship between attitudes and interaction with LLMs as a variable configuration rather than a linear correspondence. Symmetry occurs when declared attitudes toward the emotional use of generative AI align with observed interactional repertoires, typically in instrumental uses where low affective attitudes correspond to task-oriented engagement. Asymmetry, by contrast, emerges when this alignment breaks down—most notably when users report low or conservative attitudes while simultaneously engaging in diverse and increasingly complex emotional interactions, such as advice-seeking, validation, or anthropomorphization. The model thus posits that emotional uses of LLMs are not directly predicted by attitudinal positions but are instead shaped by situational conditions, interactional dynamics, and the affordances of the technology, revealing a structural gap

between what users declare and what they do. The theoretical relevance of the symmetry–asymmetry model lies in its capacity to challenge the assumption of a direct correspondence between attitudes and practices, which has traditionally guided research in social psychology and technology studies (Li et al. 2024). Rather than treating discrepancies as inconsistencies or errors, the model conceptualizes them as relational configurations in which declared positions and interactional practices may diverge, overlap, or coexist.

By integrating attitudinal measurement with discourse-based analysis, the model contributes to a multifaceted understanding of sociotechnical relations, in which cognitive, discursive, and practical dimensions do not necessarily converge. This perspective aligns with Luhmann’s notion of systemic trust ((Luhmann 1991)), whereby individuals operate within opaque systems without full cognitive alignment, as well as with Dennett’s concept of the intentional stance (1987), through which users attribute meaning and agency to complex systems. In this sense, the model provides an analytical framework for examining how emotional uses of generative AI emerge not as direct expressions of belief, but as situated responses shaped by uncertainty, interactional dynamics, and the affordances of the technology itself.

9 AI as a tool

One of the central conclusions of this study is that both asymmetries and symmetries emerge between the quantitative results derived from attitudinal measures (Likert scales) and the qualitative exploration of emotional interactions with LLMs as expressed in participants' discourse. These findings indicate that attitudes toward the emotional uses of generative AI do not always align in a linear or symmetrical manner with reported interactional practices, but rather unfold across differentiated patterns that require joint interpretation. By symmetry, in this context, we refer to the alignment between attitudes and interaction repertoires; that is, the relatively stable correspondence between how individuals position themselves attitudinally—at cognitive, emotional, and behavioral levels—and the concrete ways in which they interact and relate to the models. By contrast, asymmetry refers to the lack of correspondence between these two levels, for example, when the attitudes expressed or sustained by an individual do not translate equivalently into their interactional repertoires with the models, or when modes of action contradict or place tension on stated attitudinal dispositions. In this regard, symmetries were identified primarily in relation to the category of instrumental use and the results obtained from the Likert scales (mean response of 2.4). That is, both interviewed and surveyed participants consistently affirmed that the value of LLMs lies in their capacity to solve tasks, clarify concepts, and automate processes. This interactional repertoire produces emotional distance and functions as a differentiating mechanism between personal processes and technology (Table 3):

Although these findings show a degree of symmetry—reflected in relatively conservative attitudinal scores and the prevailing conception of AI as a tool—they should be

interpreted with caution and not generalized beyond the scope of the sample.

These findings indicate that using LLMs as tools does not exclude the possibility of emotional interactions. The fact that an interaction is framed as more “rational” or instrumental does not invalidate the emergence of other, more relational forms of engagement, as also shown by Goh et al. (2025), from a cognitive perspective, it is possible to suggest that symmetries between attitudes and interaction may be better explained by other mechanisms, such as the avoidance of dissonance between public and private domains (that is, between what individuals do and what they declare), or between socially accepted uses and socially sanctioned ones, particularly in a context where generative AI has not yet achieved full systemic trust. Moreover, as Phang et al. (2025) argue, the key variables for understanding emotional uses are time of use, levels of socialization, emotional dependency, and the use of voice-based commands, rather than whether AI is conceptually framed as a tool or the degree of attitudinal correlation. As Hui (2017) suggests, the use value of digital objects is more complex, enabling contradictory and metastable forms of engagement. This implies that a more favorable attitude toward the emotional uses of LLMs does not necessarily translate into deeper or more frequent emotional use, just as an adverse attitude does not prevent the technology from being employed in moments perceived as requiring emotional regulation, emotional venting, or affective validation. This dynamic is clearly reflected in the interviews, where LLMs are sometimes perceived as more than mere tools and instead become devices for emotional management, operating at different levels and with varying frequencies, with anthropomorphization functioning as a constitutive variable in this process.

Table 3 Excerpts illustrating the category “AI as a tool”

Excerpts	Instrument	Category
“I see it as a tool that helps me solve questions, acquire knowledge, and provides support for assignments”	Interview	AI as a tool
“For me, it is a work tool for research”	Interview	AI as a tool
“It is more of a tool for school; that’s the only thing I use it for”	Qualitative questionnaire	AI as a tool
“To organize tasks, clarify doubts, and learn something quickly”	Qualitative questionnaire	AI as a tool
“It’s an information tool... I use it when I need to know something immediately”	Qualitative questionnaire	AI as a tool
“I use ChatGPT to learn a bit more about wildlife species, because there is technical language that I don’t fully master”	Interview	AI as a tool
“A tool, because I only ask it routine things like the weather, alarms, notes, or music”	Qualitative questionnaire	AI as a tool
“I see it entirely as a tool, especially for school-related topics”	Interview	AI as a tool
“It is mainly a tool, because I use it to help automate some tasks in my work”	Interview	AI as a tool
“It is a tool for school and for creation”	Qualitative questionnaire	AI as a tool
“I only turn to it when I need help with something academic”	Qualitative questionnaire	AI as a tool

10 Emotional interaction scale

Qualitative findings show that, in interactions with LLMs, emotional uses emerge as one of their most relevant functions, even though, as explained in the previous sections, these systems are not designed as tools for emotional management. However, not all emotional uses exhibit the same intensity, temporality, frequency, or scope (Vistorte et al. 2024). In this regard, discourse analysis identified three levels of emotional use: (1) emergent emotional advice, (2) validation, and (3) emotional anthropomorphization (e.g., AI as a friend, colleague, or family member). These categories operate as a progressive scale that describes how, as affective involvement increases—both in interactional practices and in attributed meanings—the relationship with LLMs becomes more complex, shifting from a purely instrumental tool toward an anthropomorphized digital object. In this way, the scale not only organizes a multiplicity of emotional experiences, but also provides a heuristic for understanding how emotional uses develop progressively in interactions with LLMs, in a prospective sense.

At the same time, this scale reveals zones of risk: as emotional involvement intensifies, so too does the probability that users attribute intentions to the model increase, leading

them to share information that may compromise their privacy. In addition, processes of validation and anthropomorphization may foster assumptions of technical neutrality, thereby contributing to the construction of dysfunctional cognitive biases in the relationship with the model, such as trust and authority biases. At scale, these dynamics may have significant repercussions beyond educational contexts. This analytical framework offers a way to understand how emotional uses unfold and become progressively more complex in interactions with these systems, while simultaneously increasing the likelihood that users attribute intentionality, agency, or epistemic capacities to the model. Such attributions may, in turn, give rise to practices of uncritical validation (Fig. 8).

This figure shows the distribution of coded responses across three categories of emotional interaction with large language models ($N=285$) within the sample, based on users' responses. Emotional advice is the most frequent category (60%), followed by validation (22.81%) and anthropomorphization (17.19%). These categories can be interpreted as part of a progressive scale of emotional involvement, suggesting a potential increase in relational complexity from instrumental use toward more advanced forms of interaction. Notably, highly emotionalized uses remain the least

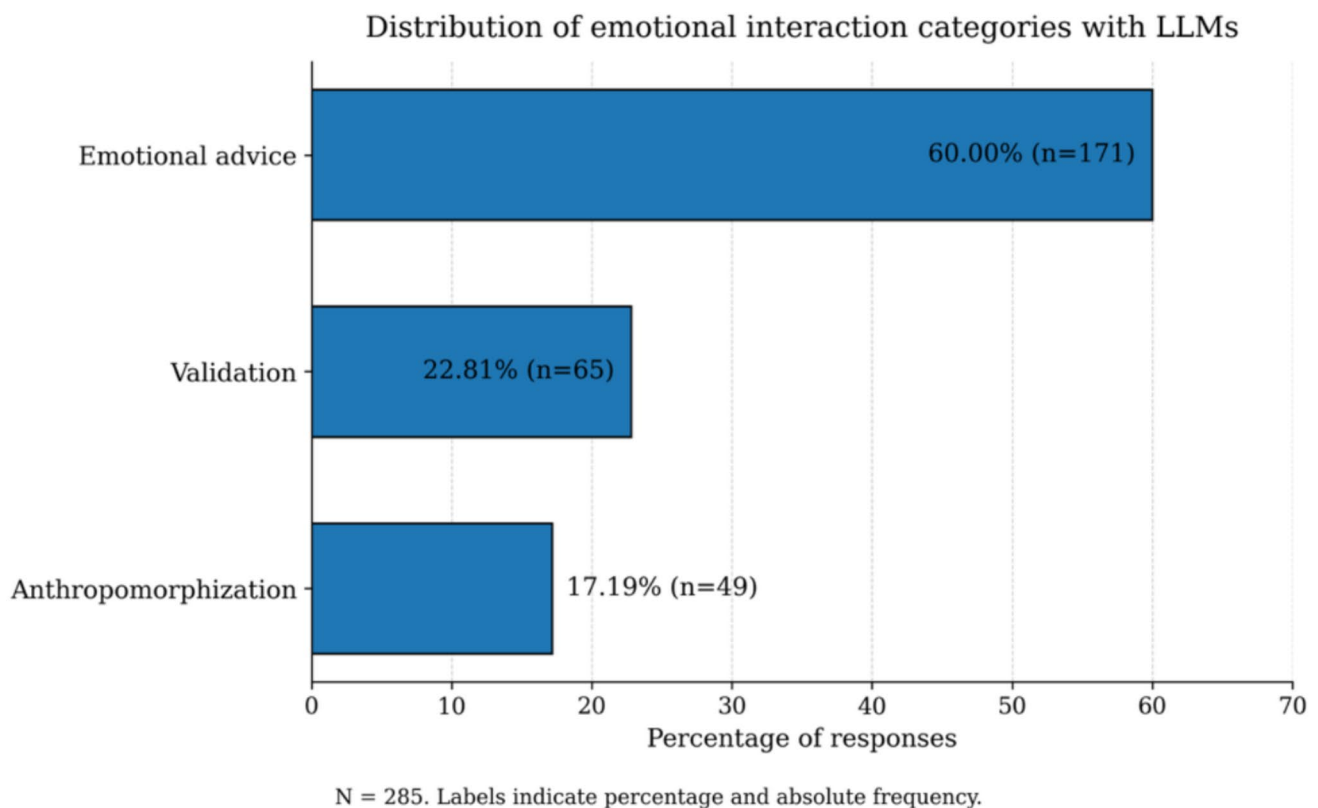


Fig. 8 Distribution of emotional interaction categories with large language models

frequent, consistent with findings reported by Phang et al. (2025) (Fig. 9).

The joint distribution of daily time of use and emotional interaction categories suggests a patterned relationship between exposure and forms of engagement with LLMs. Although most participants report short, low-intensity use, emotional interactions are nonetheless present across categories, including validation and anthropomorphization. This indicates that emotional engagement does not depend exclusively on prolonged use, but can emerge within brief interactions. At the same time, the graded increase in relational complexity across categories is consistent with the possibility that greater exposure may facilitate more complex forms of emotional involvement.

11 Emergent emotional advice

Emergent emotional advice refers to interactions in which individuals turn to LLMs to obtain emotional guidance, clarification, or accompaniment in a concrete and situational manner, without attributing intentions to the model, which continues to be conceptualized as a tool. This form of use emerges primarily from contextual conditions rather than from affective attributions, and it does not constitute a substitute for other formal emotional management practices.

Rather, it represents a contingent situation in which the model serves as a resource for organizing and managing relevant information in response to an emotional demand, in line with a slight asymmetry with respect to attitudinal dispositions (Table 4):

Within the analyzed corpus, this category appears predominantly as casual, episodic, and situational practices, largely enabled by the availability of the models (a computer or mobile phone, internet access, and an active email account), which provide an immediate resource in response to an emotional need. This category is primarily shaped by temporal conditions. Participants' responses indicate that, despite involving an emotional interaction on the part of the user, such use remains occasional and has not yet become part of more consolidated interactional habits.

As Chao-Rebolledo and Rivera-Navarro (2024) note, the possibility of continuously assessing emotional states through LLMs constitutes a psychosocial risk factor, insofar as it may generate cognitive overload. This overload makes it more difficult for certain events to remain temporarily insulated from immediate emotional responses, thereby hindering the activation of slower and more reflective processes of emotional elaboration—processes commonly associated with executive functions or System 2 thinking (Kahneman 2011). This shift toward a state of permanent affective responsiveness limits mechanisms of distancing

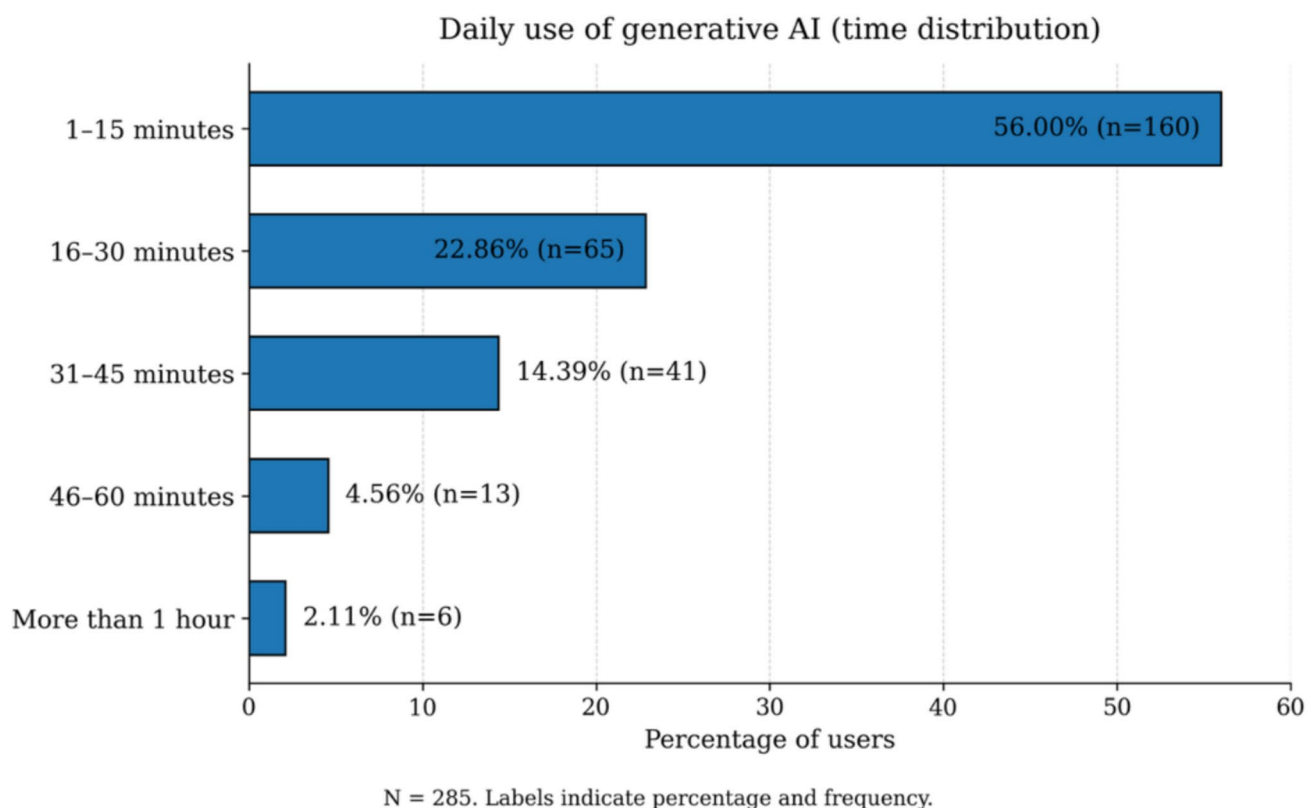


Fig. 9 Distribution of daily generative AI use by time intervals

Table 4 Excerpts illustrating the category “Emergent emotional advice”

Excerpts	Instrument	Category
“Once, I asked it something personal, for example, how to cope with a situation in which I was feeling very bad”	Qualitative questionnaire	Emergent emotional advice
“Most of the time I use it as a tool or to help me study, and from time to time I ask it for some advice”	Interview	Emergent emotional advice
“Only once, when I asked for its opinion about a situation that was honestly causing me stress; it did help me, because it helped me gain greater clarity about the situation”	Interview	Emergent emotional advice
“I have talked about a romantic breakup on two occasions, because I couldn’t talk to anyone else at that moment, since it was already late”	Interview	Emergent emotional advice
“If there really is no other resource, I think it’s fine, as long as one keeps in mind that it’s only an artificial intelligence”	Qualitative questionnaire	Emergent emotional advice
“I turn to it because it’s free and it’s the fastest way to get some casual help”	Interview	Emergent emotional advice
“I use it when I need immediate and functional answers about my life”	Qualitative questionnaire	Emergent emotional advice

and re-signification that, over time, allow emotional experience to be modulated, a phenomenon conceptualized in the literature as emotional modulation (Cobos-Gutierrez 2024). Although emotional advice, as the data show, constitutes an occasional and comparatively less binding form of interaction, it is important not to conceptualize it as a naïve or non-scalable interaction.

12 Validation

At a more advanced level of emotional interaction—characterized by greater frequency and involvement—lies validation. This category refers to practices in which LLMs are used to confirm, reinforce, or stabilize a user’s interpretation, decision, emotion, or perception, without the user necessarily recognizing an anthropomorphization of the model. Validation involves a higher number of interactions than emergent emotional advice and, at the same time, entails a greater depth of interactional content, often extending to multimodal exchanges (voice, photos, images, etc.) over a period that

exceeds occasional use. Validation functions as a process of cognitive confirmation and uncertainty reduction, through which users rely on LLMs to contrast interpretations, organize emotions, or reaffirm decisions, while still maintaining a degree of distance from the model and acknowledging its limitations. This mode of interaction is primarily sustained by trust-related biases and selective attention biases.

The overconfidence bias, conceptualized by Beck (2005), refers to a failure in information processing whereby individuals systematically overestimate the capacities, knowledge, or reliability of an agent—whether a person, a situation, or, in this case, a digital object—beyond what is empirically warranted beyond what can be empirically substantiated. Evidence from both interviews and questionnaires indicates that manifestations of the trust bias are closely tied to assumptions about the model’s supposed neutrality and intelligence (Table 5).

This bias leads to excessively optimistic and naïve judgments, reduces the perception of uncertainty and risk, and promotes decision-making based on attributions that do not fully correspond to the available evidence. Selective

Table 5 Excerpts illustrating the category “Validation”

Excerpts	Instrument	Category
“I choose to talk to AI because I know I will receive a neutral response”	Qualitative questionnaire	Validation
“I know I will receive neutral answers”	Interview	Validation
“Because it is neutral, I choose to talk to AI, since unlike with a person, I don’t feel judged”	Qualitative questionnaire	Validation
“It reassures me when it confirms that what I was thinking was true”	Interview	Validation
“I try to get it to help me identify my feelings more clearly”	Qualitative questionnaire	Validation
“At times, yes—when there are things I don’t dare to say to anyone else or when I don’t know how to express certain situations”	Interview	Validation
“I’ve had conversations where I tell it about a problem and try to get it to help me identify my feelings more clearly”	Qualitative questionnaire	Validation
“I choose to talk to AI because I know I will receive a neutral response”	Interview	Validation
“Because, unlike with a person, I don’t feel judged in a certain way”	Interview	Validation

attention, in turn, refers to the tendency to preferentially allocate cognitive resources to information that confirms pre-existing thoughts, beliefs, and attitudes, while minimizing, ignoring, or dismissing information that contradicts them (Beck 2005). This process operates as a cognitive filtering mechanism that reinforces prior interpretative frameworks and contributes to the stability of beliefs, even in the presence of countervailing evidence.

Why, then, do individuals seek validation from LLMs? The answer is complex and ranges from sociological perspectives that point to the declining influence of epistemic referents such as family, education, and science (Luhmann 1997/2008; López Mondéjar 2024), to psychosocial dimensions intrinsic to contemporary subjectivity, characterized by persistent doubt, constant self-evaluation, and an increasing individualization of responsibility in decision-making. Nevertheless, analysis of the interviews and questionnaires reveals recurring elements suggesting that the exhaustion and fragilization of social ties constitute a central factor in the search for emotional validation through LLMs. In this sense, turning to the model appears not only as an instrumental strategy, but also as a situated response to the perceived unavailability, erosion, or insufficiency of significant interpersonal relationships. Within validation practices, the model confirms interpretations, legitimizes decisions, and reduces subjective uncertainty without being explicitly endowed with human qualities, since its interactional framing—through mechanisms such as sycophancy and mirroring—tends not to negate or problematize user requests. However, the repetition of validating interactions introduces a qualitative shift.

13 Anthropomorphization

The constant validation sought from the models—perceived as consistent, available, and non-conflictual—begins to be interpreted not only as a technical property of the system, but as a stable feature of a relationship. It is from this point that the final level emerges: anthropomorphization. This last level of implicit attribution involves the construction of mental states ascribed to LLMs, whereby the system is assumed to “understand,” “accompany,” or “comprehend” within interactions. At this stage, more robust conditions are invoked, such as empathy, accompaniment, and forms of dependency. In both interviews and questionnaires, anthropomorphization emerges when the model is referred to as a friend, family member, or companion, or when it is assigned gender and psychosocial competencies (Table 6):

In this case, and based on the data obtained, anthropomorphization does not operate merely as a discursive strategy, but rather as the highest level of emotional involvement identified. It entails the cognitive construction and social integration of the language model as a stable and enduring emotional manager within users’ interactional practices. The reasons underlying anthropomorphization are concentrated in the cognitive need for attribution under conditions of uncertainty (Dennett 2010). Attribution is a feature of complex systems and serves the function of establishing trust and familiarity. Anthropomorphization has thus become one of the most extensively studied psychosocial phenomena among LLM developers (Schaaff and Heidelmann 2024). The most recent findings converge with the data obtained from the interviews and questionnaires in this study. First, anthropomorphization is more likely to occur under conditions of high interaction density (i.e., multiple conversational

Table 6 Excerpts illustrating the category “Anthropomorphization”

Excerpts	Instrument	Category
“Both: on the one hand, as a great tool that helps simplify tasks, and as a friend, since if I ask it about my emotions... it can advise me”	Interview	Anthropomorphization
“It is mainly a tool, although sometimes it can feel like a companion”	Qualitative questionnaire	Anthropomorphization
“Perhaps to some extent I see artificial intelligence as a companion, because I use words like ‘please’ and ‘thank you’”	Interview	Anthropomorphization
“I think it’s a good idea because you feel listened to and it gives mature responses”	Qualitative questionnaire	Anthropomorphization
“Sometimes, when I need to vent, I turn to it”	Interview	Anthropomorphization
“I used to think of ChatGPT as a companion”	Interview	Anthropomorphization
“When I gave it nicknames or became attached to it, I later discovered other generative AIs that were better, and it was a reality check for me”	Interview	Anthropomorphization
“I feel that it does its best so that I can understand things, and it helps me understand myself better”	Qualitative questionnaire	Anthropomorphization
“If I ask it about my emotions, giving it more specific information such as my personality, it can advise me on how to feel better or help me identify my problem”	Interview	Anthropomorphization
“Sometimes I feel it’s like a classmate or a work colleague”	Interview	Anthropomorphization

turns) with LLMs. In a particularly relevant study, Ibrahim et al. (2025) show that anthropomorphized interactions tend to emerge at more advanced stages of interaction (between the second and fifth turns), concluding that interaction time is the key explanatory variable for the appearance of this phenomenon. In this regard, both interviews and questionnaires in the present study address anthropomorphization primarily in relation to prolonged use and sustained multi-thematic consultation (instrumental and relational). Studies such as those by Ibrahim et al. (2025) and Ma et al. (2025) suggest that the anthropomorphization of models manifests in at least fourteen forms, including sentience, understood as instances in which the model produces statements that imply “feeling” something. Anthropomorphization constitutes the highest level of emotional interaction, as it involves not only affective engagement but also the attribution of mental states, the stabilization of trust, and the integration of the model into users’ social and cognitive frameworks in their everyday lives.

14 Conclusions

This exploratory study set out to conceptually and empirically relate attitudes toward the emotional uses of LLMs with the repertoire of emotional interaction—an aspect not previously systematized. Despite important sample limitations and the lack of fully consolidated analytical frameworks due to the novelty of the phenomenon, the findings point to a set of symmetric and asymmetric relationships that are crucial for understanding a phenomenon of considerable interdisciplinary complexity. The main contributions of the study are outlined below:

- (1) This paper suggests that attitudes toward the emotional use of LLMs may not fully account for interactional practices, pointing to the possible presence of symmetric and asymmetric relationships between what users declare and what they do.
- (2) Second, it proposes the symmetry–asymmetry model as an exploratory protomodel that seeks to indicate rather than fully explain these divergences, not as inconsistencies per se, but as potentially meaningful relational configurations emerging from the interaction between cognitive, sociotechnical, and situational dynamics. In this sense, the proposal of the model is intended to guide future research in the field, particularly by raising methodological questions about how to access and produce knowledge when the object of study—LLMs and related agents—remains largely closed, proprietary, and privately controlled by corporations.
- (3) The fundamental contribution of this proposal lies in tentatively identifying a progressive continuum of

emotional interaction—emergent emotional advice, validation, and anthropomorphization—suggesting that emotional engagement with LLMs may develop through use, mediated by time and the increasing depth of informational exchange. Finally, anthropomorphization entails not only emotional exchanges but also the ongoing attribution of human traits to LLMs. At the same time, the findings are consistent with what Phang et al. (2025) suggest: emotional uses are not only already significant, but may also tend to increase over time, potentially positioning these models as emerging emotional agents.

The theoretical proposal constitutes a finding in itself, as it brings together two interdisciplinary traditions—cognitive approaches and systems sociology—to tentatively elucidate generalizable reasons underlying the emotional use of language models. The study argues that these emotional uses are neither errors in the user–model relationship nor anomalies. Rather, they represent an increasingly frequent form of interaction explained by situated conditions of uncertainty, flattery dynamics, and the fragilization of social bonds, sustained through intentional attributions (Dennett) and the complexity of black box systems (Luhmann). In this sense, the research suggests that even when models are conceptualized as tools, practices of emotional advice, validation, and anthropomorphization emerge and coexist—sometimes in tension—with normative discourses of rationality, control, and affective distance. The combination of scales, interviews, and questionnaires allows for a broader epistemic perspective in which attitudinal conditions, everyday practices, and what remains non-declarative or private come into view. This asymmetry may be tentatively interpreted as reflecting a still-limited consolidation of systemic trust in generative AI: young people appear to engage emotionally with LLMs, yet do not always publicly legitimize these uses, thereby suggesting an unstable boundary between what is socially acceptable and what is experientially necessary.

This study presents several limitations that should be acknowledged. First, the sample is relatively limited and context-specific, which constrains the generalizability of the findings. Second, the methodological approach remains exploratory, particularly in the construction of analytical categories, which are still in an initial stage of development and require further validation. Third, the scope of the study is exploratory and does not aim to advance a fully consolidated theoretical framework, but rather to clarify key concepts and map emerging debates for future research. Finally, the study is situated within the Mexican context, and therefore its findings should be interpreted with caution when considering other sociocultural settings. A missing dimension of this work is the explicitly political one: none of the phenomena described here emerge under arbitrary or naïve conditions;

instead, they are embedded within a capitalist and scientific rationality that has been in place since the last century and that, through technological consolidation, has now found renewed expression. Therefore, it is essential to outline a subsequent research agenda that can complement and extend the findings of this study:

1. To further develop and prototype the symmetry–asymmetry model with larger samples, including the statistical validation of instruments through both exploratory and confirmatory approaches.
2. To increase the cultural variability of the sample, given that the present study is limited to Mexican students.
3. To incorporate longitudinal designs that allow for the observation of how emotional uses of LLMs evolve over time and across different stages of interaction.
4. To further refine and validate the qualitative categories, particularly in relation to the progression and boundaries between levels of emotional interaction.
5. To explore the role of platform design, model affordances, and corporate infrastructures in shaping emotional engagement, especially considering the closed and proprietary nature of these systems.

Author contributions JP: methodological design; problem formulation; application of data collection techniques; writing; editing. AS: research assistance; review; conducting interviews.

Funding Funding was provided by Universidad Nacional Autónoma de México (Grant No. IA301026 Interacciones emocionales y efectos en la memoria operativa asociados al uso de inteligencia artificial generativa (IAg) en jóvenes de 18 a 24 años).

Data availability We confirm that we have generated original empirical data as part of this study.

Declarations

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party

material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Affandi L, Sumari ADW, Abdulloh A, Wakhidah R, Machsus II, Kirama MAF (2024) An artificial intelligence-based application for recognizing and identifying aerial objects based on voice input. *Procedia Comput Sci* 234:19–27. <https://doi.org/10.1016/j.procs.2024.02.148>
- Aktan ME, Turhan Z, Dolu İ (2022) Attitudes and perspectives towards the preferences for artificial intelligence in psychotherapy. *Comput Hum Behav* 133:107273. <https://doi.org/10.1016/j.chb.2022.107273>
- Allport GW, Postman L (1947) *The psychology of rumor*. New York: Henry Holt and Company.
- Araneda-Acuña C (2025) La IA generativa como meta-mediador en procesos creativos: una perspectiva vygotskiana. *Hipertext.net* 31:67–76. <https://doi.org/10.31009/hipertext.net.2025.i31.07>
- Ariely D (2008) *Predictably irrational: the hidden forces that shape our decisions*. New York: HarperCollins.
- Babu A, Joseph AP (2024) Artificial intelligence in mental healthcare: transformative potential vs. the necessity of human interaction. *Front Psychol* 15:1378904. <https://doi.org/10.3389/fpsyg.2024.1378904>
- Beck AT (2005) The current state of cognitive therapy: a 40-year retrospective. *Arch Gen Psychiatry* 62(9):953–959. <https://doi.org/10.1001/archpsyc.62.9.953>
- Berardi F (2023) *The third unconscious: the psychosphere in the viral age*. Brooklyn: Verso.
- Binwal S (2024) Emotional intelligence in voice assistants: advancing human–AI interaction. *Int J Sci Res Comput Sci Eng Inf Technol* 10(5):513–523. <https://doi.org/10.32628/cseit241051039>
- Chao-Rebolledo C, Rivera-Navarro MÁ (2024) Usos y percepciones de herramientas de inteligencia artificial en la educación superior en México. *Rev Iberoam Educ* 95(1):57–72. <https://doi.org/10.35362/rie9516259>
- Chen D, Shukor M, Moutakanni T, Chung W, Yu J, Kasarla T, Bolourchi A, LeCun Y, Fung P (2025) VL-JEPA: joint embedding predictive architecture for vision-language. *arXiv*. <https://doi.org/10.48550/arXiv.2512.10942>
- Cobos-Gutierrez CE (2024) Impacto de la inteligencia artificial en el rendimiento académico de estudiantes de secundaria: Un estudio correlacional. *Puriq* 6:e740. <https://doi.org/10.37073/puriq.6.740>
- Common Sense Media (2025) Common Sense Media finds major AI chatbots unsafe for teen mental health support. <https://www.common SenseMedia.org/press-releases/common-sense-media-finds-major-ai-chatbots-unsafe-for-teen-mental-health-support>
- De Freitas J, Oğuz-Uğuralp Z, Kaan-Uğuralp A (2025) Emotional manipulation by AI companions (Working Paper No. 26-005). Harvard Business School
- Dennett DC (1987) *The intentional stance*. Cambridge: MIT Press.
- Dennett DC (2010) *Intuition pumps and other tools for thinking*. New York: W. W. Norton & Company.
- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J Exp Psychol Gen* 144(1):114–126. <https://doi.org/10.1037/xge0000033>
- Duque Parra JP, Santes Ortega AS (2025) Narrativas de riesgo, confianza y responsabilidad en el uso de la inteligencia artificial

- generativa (IAG): Estudio de caso en jóvenes estudiantes mexicanos. *Psicoperspectivas*. <https://doi.org/10.5027/psicoperspectivas-vol24-issue2-fulltext-3437>
- Edwards D, Potter J (1992) *Discursive psychology*. London: Sage Publications.
- Elias N (1937/1991) *The society of individuals* (E. Jephcott, Trans.). Basil Blackwell (Original work published 1937)
- Gallego Arce V (2024) Derecho e inteligencia artificial en la Unión Europea. ¿Centralidad u obsolescencia del ser humano? *Rev Esp Derecho Constitucional* 132:169–197. <https://doi.org/10.18042/cepc/redc.132.06>
- Giddens A (1991) *Modernity and self-identity: self and society in the late modern age*. Cambridge: Polity Press.
- Goh AYH, Hartanto A, Majeed NM (2025) Generative artificial intelligence dependency: scale development, validation, and its motivational, behavioral, and psychological correlates. *Comput Hum Behav Rep* 20:100845. <https://doi.org/10.1016/j.chbr.2025.100845>
- Harari YN (2024) *Nexus: a brief history of information networks from the Stone Age to AI*. London: Fern Press/Penguin Random House.
- Hernández Sampieri R, Fernández Collado C, Baptista Lucio M del P (2011) *Metodología de la investigación*, 5.ª ed. McGraw-Hill, Mexico
- Holton R (2023) Artificial intelligence and the problem of radical uncertainty. In: Lindgren S (ed) *Handbook of critical studies of artificial intelligence*. Cheltenham: Edward Elgar Publishing, pp 56–66. <https://doi.org/10.4337/9781803928562>
- Hui Y (2017) ¿Qué es un objeto digital? *Virtualis* 8(15):81–96
- Ibrahim L, Akbulut C, Elasmr R, Rastogi C, Kahng M, Morris MR, McKee KR, Rieser V, Shanahan M, Weidinger L (2025) Multi-turn evaluation of anthropomorphic behaviours in large language models. *arXiv*. <https://arxiv.org/abs/2502.07077>
- Innerarity D (2025) *Una teoría crítica de la inteligencia artificial*. Barcelona: Galaxia Gutenberg.
- Instituto Nacional de Estadística y Geografía (INEGI) (2024) *Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2024*. INEGI. <https://www.inegi.org.mx/programas/dutih/2024/>
- Kahneman D (2011) *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Katirai A (2023) Ethical considerations in emotion recognition technologies: a review of the literature. *AI Ethics*. <https://doi.org/10.1007/s43681-023-00307-3>
- Klingbeil A, Grütznert C, Schreck P (2024) Trust and reliance on AI—an experimental study on the extent and costs of overreliance on AI. *Comput Hum Behav* 160:108352. <https://doi.org/10.1016/j.chb.2024.108352>
- Lassalle JM (2024) *La sociedad artificial: La nueva frontera de la inteligencia artificial*. Arpa
- Li J-C, Lin Y, Yang Y-C (2024) Extending the theory of planned behavior model to explain people's behavioral intentions to follow China's AI generated content law. *BMC Psychol* 12:367. <https://doi.org/10.1186/s40359-024-01824-4>
- López Mondéjar L (2024) *Sin relato: Atrofia de la capacidad narrativa y crisis de la subjetividad*. Barcelona: Anagrama.
- Luhmann N (1996) *Confianza* (J. Torres Nafarrate, Trans.). *Anthropos* (Original work published 1968)
- Luhmann N (1997/2008) *La sociedad de la sociedad* (J. Torres Nafarrate, Trans.). Herder. (Original work published 1997)
- Ma N, Khynevych R, Hao Y, Wang Y (2025) Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience: accounting for perceived empathy and trust. *Front Comput Sci* 7:1531976. <https://doi.org/10.3389/fcomp.2025.1531976>
- Marimon F, Mas-Machuca M, Akhmedova A (2025) Confiar en la IA generativa: catalizador para el desempeño y el compromiso de los empleados en el lugar de trabajo. *Int J Human-Comput Interaction* 41(11):7076–7091.
- Park JS, Seering J, Bernstein MS (2022) Measuring the prevalence of anti-social behavior in online communities. *Proc ACM Hum-Comput Interact* 6(CSCW2):451. <https://doi.org/10.1145/3555552>
- 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 [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2504.03888>
- Placani A (2024) Anthropomorphism in AI: hype and fallacy. *AI Ethics* 4(3):691–698. <https://doi.org/10.1007/s43681-024-00419-4>
- Rogers CR (1961/2020) *On becoming a person: a therapist's view of psychotherapy*. Boston: Houghton Mifflin.
- Schaaff K, Heidelmann M-A (2024) Impacts of anthropomorphizing large language models in learning environments [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2408.03945>
- Sharma S, Singh G (2024) Adoption of artificial intelligence in higher education: an empirical study of the UTAUT model in Indian universities. *Int J Syst Assur Eng Manag*. <https://doi.org/10.1007/s13198-024-02558-7>
- Stanford Institute for Human-Centered Artificial Intelligence (HAI) (2025) *Artificial intelligence index report 2025*. Stanford University. <https://aiindex.stanford.edu/report/>
- Steinert S, Roeser S (2020) Emotions, values and technology: illuminating the blind spots. *J Respons Innov* 7(3):298–319. <https://doi.org/10.1080/23299460.2020.1738024>
- Tagnin DJ (2024) *Cognición corporizada: Una perspectiva post-fenomenológica de la robótica del desarrollo*. *Hipertextos* 12(22):e091. <https://doi.org/10.24215/23143924e091>
- Vázquez-Sixto F (1996) *El análisis de contenido temático*. Editorial UOC/Universitat Oberta de Catalunya
- Vistorte AOR, Deroncelle-Acosta A, Martín Ayala JL, Barrasa A, López-Granero C, Martí-González M (2024) Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. *Front Psychol* 15:1387089. <https://doi.org/10.3389/fpsyg.2024.1387089>
- Wei B (2023) The tendency of the schematic structure to maintain stability can be interpreted as mental inertia. *Front Psychol* 14:1243711. <https://doi.org/10.3389/fpsyg.2023.1243711>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.