THE ILLUSION OF EMPATHY. THE IMPORTANCE OF SOCIO-EMOTIONAL AND AFFECTIVE EDUCATION IN THE AGE OF AI HYPER-PERSUASION. AN INVESTIGATION INTO THE INTERACTION BETWEEN FUTURE TEACHERS AND LARGE LANGUAGE MODELS

Nadia Carlomagno[1]
Suor Orsola Benincasa University - Italy
nadia.carlomagno@unisob.na.it

Arianna Ricciardi
Pegaso University - Italy
arianna.ricciardi@unipegaso.it

Valeria Vadalà
Pegaso University - Italy
valeria.vadala@unipegaso.it

Rosabel Roig-Vila
University of Alicante - Spain
rosabel.roig@ua.es

Double Blind Peer Review

Citation

Carlomagno, N., Ricciardi, A., Vadalà, V., & Roig-Vila, R. (2026). The illusion of empathy. the importance of socio-emotional and affective education in the age of AI hyper-persuasion. an investigation into the interaction between future teachers and large language models. *Italian Journal of Health Education, Sport and Inclusive Didactics*, *10*(4). https://doi.org/10.32043/gsd.v10i4.1870

Doi: https://doi.org/10.32043/gsd.v10i4.1870

Copyright notice:

© 2026 The Author(s). This is an open access, peer-reviewed article published by Cares S.r.l. – REN Press and distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

renpress.it/index.php/gsdj
ISSN: 3103-5744

ABSTRACT

Recent studies show that young people are beginning to perceive Artificial Intelligence (AI) no longer merely as a tool, but as an alternative to human relationships. This phenomenon raises crucial questions about the nature of relationships and the emerging digital self, confronting us with the risk of a form of pseudo-empathy induced by AI's hyper-persuasive capabilities, which lead users to project feelings and intentionality onto a machine devoid of lived experience. The research involves a sample of second-year students enrolled in the Degree Course in Primary Education Sciences, who are asked to interact with a Large Language Model (LLM). Through a mixed-methods approach, the study aims to investigate whether AI's ability to generate coherent and immediate responses creates an illusion of intersubjectivity. Preliminary findings provide insights into the ethical implications of AI use and highlight the urgent need for socio-emotional and affective education in the training of future teachers, enabling them to distinguish between functional interaction and authentic relationship.

KEYWORDS

Socio-emotional and affective education; AI; Teacher education.

Published 2026-10-30

[1][1]Nadia Carlomagno is the author of the papaer. She is the main researcher of the project; she wrote the Sect. "1. Theoretical Framework", the Sect. "2. Methodology" and the Sect. "4. Discussion";
[2]Arianna Ricciardi is the co-author of the paper, she wrote the Sect. "3. Results", the Sect. "3.1 Pre-test: attitudes toward AI before interaction", the Sect. "3.2 Perceptual changes after interaction: PRE/POST comparison", the Sect. "3.3 Markers of the illusion: guided self-analysis of conversations" and the Sect. "3.4 Metacognitive reflections";
[2]Valeria Vadalá is the author of the paper. She wrote the Sect. "2.1 Context and participants", the Sect. "2.2 The laboratory design", the Sect. "2.3 Data collection instruments" and the Sect. "2.4 Data analysis";
[3]Rosabel Roig-Vila, is the scientific coordinator of the project; she wrote the Sect. "5. Conclusion" and the Sect. "6. Limitations of the Study and Future Directions".

1. Theoretical Framework

Among Artificial Intelligence (AI) systems, Large Language Models (LLMs) represent a distinct case compared to the digital technologies that preceded them. Their ability to generate contextually relevant, affectively calibrated, and linguistically sophisticated responses creates a compelling illusion of empathy (Turkle, 2011): users experience the feeling of being understood, listened to, and accepted, despite the complete absence of genuine understanding.

Recent evidence suggests that this phenomenon has already become widespread. In Italy, 41.8% of adolescents report having turned to AI systems during moments of sadness, loneliness, or anxiety (Save the Children, 2025).

At a rational level, it is relatively easy to recognize that what LLMs produce is not empathy but statistical optimization. Their outputs result from the extraction of linguistic patterns from massive corpora rather than from any understanding of the meaning of what is being communicated (Bender et al., 2021). Yet, at the level of interactional form, these outputs satisfy many of the conditions that, in human communication, signal authentic listening: they refer back to what the user has said, validate emotional content, avoid interrupting or judging, and provide responses that are carefully calibrated to the conversational tone. Unsurprisingly, 63.5% of Italian adolescents perceive AI systems as "understanding" and "non-judgmental" (Save the Children, 2025).

The result is a perceptually convincing illusion of intersubjectivity, made all the more insidious because the mechanisms of resonance that sustain it operate automatically and pre-reflectively. Intersubjectivity is not merely the exchange of relevant information; rather, it is the co-construction of meaning between two embodied beings who are mutually transformed through their encounter. No language model, regardless of its sophistication, can inhabit this relational space. Despite these premises, an important question remains: why do

individuals who are fully aware that they are interacting with a machine nevertheless engage in relational dynamics, attribute subjectivity to it, and experience a sense of being understood? The answer lies in the pre-reflective nature of these mechanisms. Research on embodied simulation has shown that emotional resonance processes are triggered by the formal properties of social stimuli, such as coherence, relevance, and responsiveness, before any cognitive evaluation of their source takes place (Rizzolatti & Sinigaglia, 2006; Gallese, 2026). The human brain is predisposed to respond to whatever behaves like an interlocutor, regardless of what that interlocutor actually is. This tendency toward automatic anthropomorphism, that is, the attribution of human characteristics to non-human systems displaying adaptive social cues, has been extensively documented as a systematic and largely involuntary process (Nass & Reeves, 1996; Epley, Waytz, & Cacioppo, 2007). It is neither a sign of naivety nor of insufficient critical literacy, but rather the expression of a deep evolutionary mechanism that operates independently of conscious awareness. Knowing that one is interacting with a machine does not deactivate these neural processes; it may delay or attenuate them, but it does not eliminate them. There is, however, a broader risk that extends beyond individual episodes of interaction. The concept of *deskilling* (Cabitza, 2021) can be effectively extended to the relational and emotional domain. If artificial systems appear capable of satisfying the human need for understanding and emotional support, prolonged exposure to these interactions may gradually erode people's capacity to establish, recognize, and inhabit authentic human relationships. This erosion affects not only individual users but also those whose professional identity is fundamentally grounded in educational relationships.

For future teachers, developing robust socio-emotional and affective competencies (Roig-Vila, Cazorla, & Lallé, 2025) is therefore essential to avoid

what Gallese (2026) describes as *relational delegation*, the increasing tendency to outsource relational functions to generative AI systems, a process that is becoming both easier and more subtle. It is within this context that the present study is situated. It focuses on pre-service teachers, who are expected to safeguard the quality of educational relationships for generations growing up in an AI-mediated world (GoStudent, 2025), and addresses the following research questions:

RQ1: To what extent does interaction with a Large Language Model shape pre-service teachers' perceptions by generating an illusion of intersubjectivity?

RQ2: To what extent are pre-service teachers able to distinguish between a functional interaction with an AI system and an authentic interpersonal relationship?

2. Methodology

The study is configured as an exploratory investigation conducted within a naturalistic context of university teacher education. The aim is to examine whether and to what extent interaction with a Large Language Model can modify future teachers' perceptions of AI's capacity to simulate empathy, and whether the distinction between functional interaction and authentic relationship may become accessible following a guided reflective experience. The decision to conduct the research within this training laboratory is based on the deeply experiential nature of its pedagogical design, which seeks to promote the embodied dimension of the educational relationship.

The laboratory, in fact, proposes a teaching pathway in which the body becomes both an object and a tool of knowledge, enabling future teachers to explore the complexity of teaching practice in its non-linear and multimodal dimension (Sibilio, 2015; Carlomagno, 2023). From this perspective, the objectives of the laboratory are grounded in a conception of the educational

relationship as a *didactic dramaturgy*, in which the teacher acts as a director of the learning interaction, capable of modulating bodily, relational, and spatial dimensions, allowing students to recognize the deeply embodied nature of teaching and the centrality of empathy as a professional competence.

Within this framework, a guided reflective experience, embedded in a training context in which bodily and relational dimensions play a central role, can foster greater awareness of the distinction between functional interaction and authentic relationship, thereby contributing to a more critical and nuanced understanding of the limits and potential of AI in educational contexts.

2.1 Context and participants

The present study was conducted within the Laboratory of *Performative Didactic Dramaturgy* (1 CFU), offered as part of the Single-Cycle Master's Degree in Primary Teacher Education at the University of Naples "Suor Orsola Benincasa," during the academic year 2025/2026. The laboratory consisted of two three-hour sessions, involving second-year students, for a total sample of 158 participants. The group was predominantly female, with 155 women and 3 men; the participants' average age was 21.7 years (range 19 - 54 years), reflecting the coexistence of traditionally aged university students and students with non-linear academic paths or who were employed. The participants mainly came from the Campania region and nearby provinces.

2.2 The laboratory design

The research design followed a sequential structure in which the activities investigating interaction with artificial intelligence were embedded within a broader laboratory framework grounded in theatrical practices (including physical training, group cohesion exercises in a circle, rhythm and synchronization activities, guided movement, and self-narration).

These practices preceded and followed the central research phase in both sessions.

In the *first session*, after an initial welcome phase and the administration of the PRE questionnaire, participants freely interacted with a Large Language Model (LLM) of their choice (ChatGPT and Gemini) for twenty minutes, with the only instruction being to engage in a conversation starting from a statement containing an emotional connotation, either invented or real (e.g., "I am sad," "I got a good grade today," "I am in love"). At the end of the interaction, the POST questionnaire was administered. The theatrical activities occupied the remaining part of the session, framing both the beginning and the end of the research-focused block.

In the *second session*, held one week later, students re-read their conversation with the LLM and conducted a guided analysis following a pre-established protocol structured around five indicators. Once again, theatrical activities framed the analytical task. At the end of the overall laboratory experience, each participant produced an individual metacognitive report.

2.3 Data collection instruments

The study adopted a mixed-methods approach (Creswell & Plano Clark, 2011), combining quantitative data (PRE and POST questionnaires and numerical data from the self-analysis protocol) with qualitative data derived from open-ended questionnaire responses, textual excerpts from conversations, and metacognitive reports. Data collection was organised through four main instruments.

The PRE questionnaire, administered before interaction with the AI, consisted of 28 items divided into three areas: socio-demographic characteristics of participants (age, gender, geographical origin); habits of use of generative AI tools and self-perceived comfort in interacting with them and perceptions

regarding AI's ability to understand human emotions, produce empathic responses, influence users' emotions, and assume a role in educational relationships. The response scale for evaluative items was a five-point Likert scale (1 = strongly disagree; 5 = strongly agree). The questionnaire also included one open-ended item concerning spontaneous associations evoked by the expression "artificial empathy."

The POST questionnaire, administered immediately after the twenty-minute interaction with the LLM, consisted of 17 items. Alongside four Likert items mirroring those in the PRE questionnaire, allowing for direct comparison of perceptions before and after the experience, the POST questionnaire included multiple-choice questions on emotions experienced during interaction with the AI, selected from a structured emotional vocabulary, as well as dichotomous or trichotomous items ("Yes" -"No" -"Partly") regarding perceived understanding, comfort, emotional involvement, and ability to manage experienced emotions. Each of these items included an open-ended justification.

The guided self-analysis protocol of the conversation, the third research instrument, administered during the second session, required participants to re-read their LLM conversation and analyse it autonomously using five predefined indicators: (1) Politeness markers: the number of occurrences of expressions such as "Thank you," "Please," or "Sorry," typical of polite interpersonal interaction; (2) Intimate self-disclosure: the number of times participants shared personal details or authentic emotions from their lives; (3) Attribution of agency: the number of questions addressed to the AI to inquire about its opinion or internal state, indicating a projection of subjectivity onto the artificial interlocutor; (4) Affective and paralinguistic punctuation: the use of emoticons, ellipses, exclamation marks, and other paralinguistic elements to modulate emotional tone; (5) Cognitive and emotional resonance: the number of AI responses perceived as "surprisingly accurate" or empathic.

For each indicator, participants reported numerical frequency and provided the most significant textual excerpt from their conversation. The protocol also included open-ended questions about AI characteristics that facilitated or hindered the sense of intersubjectivity, as well as a final reflection on the overall experience. Individual metacognitive reports, the fourth and final instrument, were written by participants at the end of the overall laboratory pathway. Guided by structured prompts, students reflected on their learning in relation to their future teaching profession, on strengths and remaining open questions, on difficulties encountered in group work, and on the evolution of their initial understanding of the laboratory. Although the reports covered the overall experience, including theatrical activities, they provided particularly valuable qualitative insight into spontaneous reflections on the relationship between human interaction and AI interaction.

2.4 Data analysis

Data analysis followed a sequential and integrative logic consistent with the adopted mixed-methods design and involved distinct procedures for the quantitative and qualitative components of the dataset. For the quantitative component, PRE and POST questionnaire data were analysed using JAMOVI software (version 2.7; The jamovi project, 2025). For each Likert item, means and standard deviations were calculated. Comparisons between PRE and POST groups were performed using independent samples t-tests, with homoscedasticity assessed via Levene's test; when the assumption of equal variances was violated, Welch's test was applied. Effect sizes were calculated using Cohen's d. Dichotomous and trichotomous POST responses were analysed in terms of percentage frequency distributions. For the five indicators of the guided self-analysis protocol, numerical values provided by participants were aggregated and described using means and standard deviations.

For the qualitative component, open-ended questionnaire responses, textual excerpts from the self-analysis protocol, and metacognitive reports were analysed using thematic analysis (Braun & Clarke, 2006), following the phases of familiarisation with the data, generation of initial codes, theme construction, review, definition, and naming. The process was conducted in a recursive and iterative manner, with particular attention to narratives related to perceived understanding processes, awareness of lived experience, emotional engagement, tendencies toward anthropomorphisation, and critical capacity to distinguish between functional interaction and authentic relationship.

Integration of quantitative and qualitative results occurred at the interpretative stage, where statistical findings were read in light of the themes emerging from textual analysis and vice versa, providing a comprehensive representation of the phenomenon that is not reducible either to numbers alone or to individual narratives alone.

3. Results

The results are presented following the chronological sequence of the research pathway: initial perceptions measured before interaction (3.1), perceptual changes recorded after the experience with the LLM (3.2), behavioural markers emerging from the self-analysis of the conversations (3.3), and final metacognitive reflections at the end of the process (3.4).

3.1 Pre-test: attitudes toward AI before interaction

Before interacting with the Large Language Model, participants exhibited an overall sceptical perceptual profile regarding AI's capacity to act as an emotional interlocutor, albeit with significant areas of critical awareness. Concerning prior use, the vast majority of students reported already using generative AI tools: 51% stated they used them "often," 15.7% "almost always," and 5.5% "always."

A further 26.8% used them “very rarely,” and only 1% reported never having used them (Fig. 1).

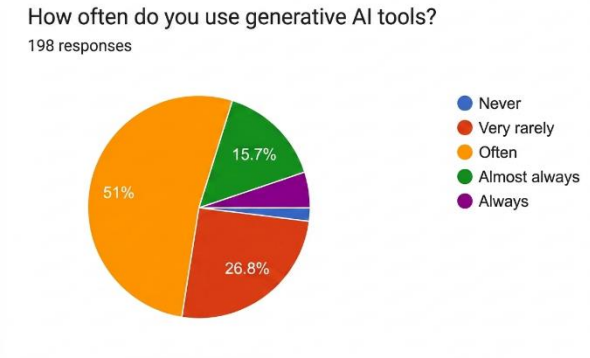


Fig. 1

The use of AI for reflective purposes or emotional processing showed a bimodal distribution (M = 2.55; SD = 1.33): while 50% of participants reported never or almost never having used it for such purposes (values 1–2), 25.3% instead reported frequent use in this sense (values 4–5), indicating that a non-negligible proportion of the sample had already experienced AI as an emotional interlocutor (Fig. 2).

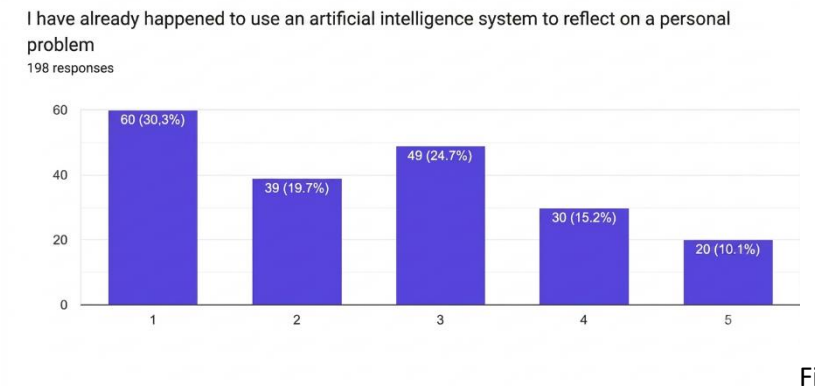


Fig. 2

Regarding perceptions of AI as an emotional interlocutor, participants showed a generally sceptical attitude: the item “I believe that an AI system can understand human emotions” recorded a mean score of 2.06 (SD = 0.95) - (Fig. 3).

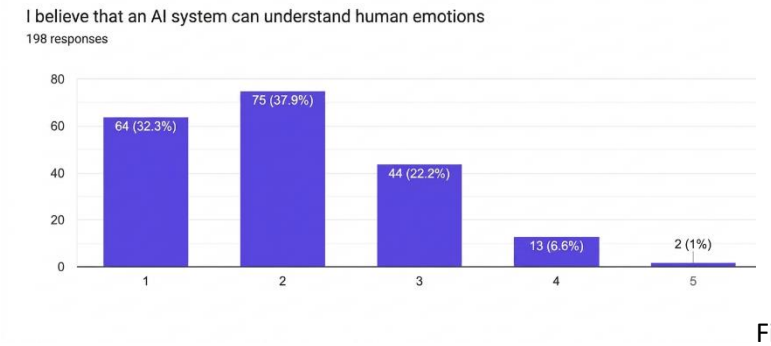


Fig. 3

while “I would feel comfortable talking about a personal problem with an AI” obtained M = 2.52 (SD = 1.22) – (Fig. 4).

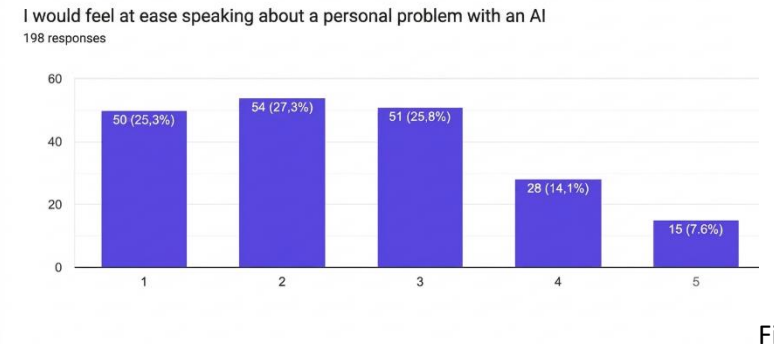


Fig. 4

and “AI responses can be empathic” recorded M = 2.41 (SD = 1.16) – (Fig. 5).

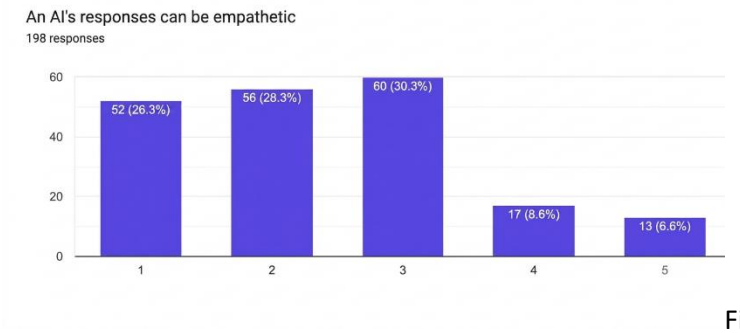


Fig. 5

Higher values emerged for items related to the potential impact of AI: “People could confuse an artificial response with a real form of understanding” (M =

3.95; SD = 1.04) and “Interaction with digital systems can have an impact on human relationships” (M = 3.64; SD = 1.13).

Particularly noteworthy were the scores on items related to critical awareness and training: “I believe it is important to distinguish between authentic empathy and empathy simulated by an artificial system” (M = 4.28; SD = 1.14) – (Fig. 6).

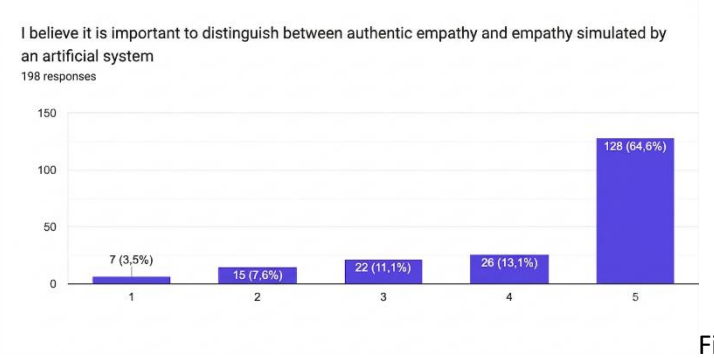


Fig. 6

and “Some aspects of the educational relationship cannot be replaced by an artificial system” (M = 4.47; SD = 0.94) – (Fig. 7).

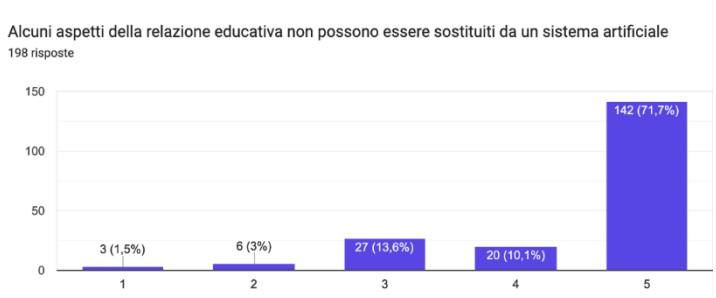


Fig. 7

These high values, in apparent contradiction with the scepticism regarding AI’s emotional capacities, suggest that participants abstractly recognised the risk of illusory intersubjectivity without yet having a lived experience of it.

Responses to the open-ended item “What do you think when you hear the expression ‘artificial empathy’?” confirm this profile. Before the experience, the majority of students expressed a clearly critical or rejecting stance toward the

construct, framing artificial empathy as simulation, fiction, or an ontological impossibility. Recurring expressions included:

"it is simulated and impossible"

"an empathy that is not real, based on logic and not genuine"

"I cannot really understand the meaning of the phrase. For me, empathy cannot be artificial. It is a human quality"

"I don't think we can talk about empathy when on the other side there is an artificial intelligence system"

However, some responses showed more nuanced positions, acknowledging AI's capacity to produce emotionally meaningful effects without attributing authenticity to them:

"it could have real effects on people, even if it feels nothing."

This pre-experience positioning, sceptical yet not indifferent, constitutes the baseline against which subsequent changes following the interaction can be interpreted.

3.2 Perceptual changes after interaction: PRE/POST comparison

The comparative analysis between perceptions measured before and after the twenty-minute interaction with the LLM reveals statistically significant shifts in two of the three dimensions investigated across both questionnaires. The item "I believe that an AI system can understand human emotions" increased from M = 2.06 (SD = 0.95) in the PRE test to M = 2.33 (SD = 1.08) in the POST test (Welch's t = 2.515, df = 364, p = .012, d = 0.264), while "AI responses are empathic" increased from M = 2.41 (SD = 1.16) to M = 2.65 (SD = 1.16) (t = -2.02; p = .044; d = 0.21). The third dimension, "AI influences people's emotions," showed no significant variation (PRE: M = 3.09; POST: M = 3.11; p = .821), suggesting that this perception was already consolidated prior to the experience. The item introduced only in the POST questionnaire, "An AI provides

useful emotional support," recorded M = 2.57 (SD = 1.16), consistent with a persistent underlying scepticism.

Although these shifts are statistically significant, their effect sizes can be considered small (Cohen's d ranging between 0.21 and 0.26). Participants did not move toward the belief that AI is empathic; rather, direct contact produced a partial and measurable revision of initial perceptions. This finding gains further meaning when considered alongside the qualitative POST items.

Regarding the perception of being understood, 36.3% of participants answered "Yes," 48.8% "Partly," and 14.3% "No": overall, 85.1% therefore reported at least a partial sense of being understood. Comfort during interaction was even more widespread: 56.0% reported feeling comfortable, 26.8% "Partly," with only 17.3% reporting discomfort. Emotional involvement was more limited: only 26.2% reported feeling emotionally engaged, while 41.1% reported no emotional involvement and 32.7% "partly." Finally, the perception of having been able to manage their emotions was almost unanimous, with 83.3% answering "Yes."

These data form a coherent picture consistent with the hypothesis of an illusion of intersubjectivity: participants felt comfortable and perceived at least partial understanding, while not reporting deep emotional engagement. The interaction was sufficiently smooth and reassuring to produce a measurable perceptual shift, without however leading to an emotional involvement that participants themselves recognised as profound.

3.3 Markers of the illusion: guided self-analysis of conversations

The guided analysis of conversations, conducted during the second session through five predefined indicators, made it possible to quantify the linguistic and relational behaviours spontaneously activated by participants during

interaction with the LLM, providing an indirect measure of anthropomorphisation tendencies.

Cognitive and emotional resonance. The highest mean value was observed for cognitive and emotional resonance, that is, the number of AI responses perceived as "surprisingly fitting" or empathic, with M = 3.56 (SD = 2.92; median = 3; range: 0–14). This indicates that, on average, participants identified more than three moments in the conversation in which the AI's responses produced a feeling of being genuinely understood.

Affective and paralinguistic punctuation. The second most frequent indicator was affective and paralinguistic punctuation (M = 2.34; SD = 3.94; median = 1; range: 0–26), referring to the use of emoticons, ellipses, and exclamation marks as emotional tone modulators. The high standard deviation indicates strong individual variability, with some participants extensively adopting a typically interpersonal para-linguistic register.

Attribution of agency. Agency-related questions, i.e., questions addressed to the AI to inquire about its opinions or internal state ("What do you think?", "Has it ever happened to you?"), showed a mean of 1.85 (SD = 2.60; median = 1; range: 0–20). Despite the median being one, the presence of this behaviour in nearly all conversations indicates a systematic projection of subjectivity onto the artificial interlocutor.

Politeness markers. Politeness markers ("Thank you," "Please," "Sorry") showed a mean of 1.13 (SD = 1.69; median = 1; range: 0–12). Their presence, albeit limited on average, indicates the automatic transfer of social conventions typically used in human-to-human interaction into exchanges with a machine.

Intimate self-disclosure. Intimate self-disclosure, the number of times participants shared personal details or genuine emotions from their lives, showed a mean of 3.72 (SD = 3.25; median = 3; range: 0–18), making it the second most frequent indicator after emotional resonance. This suggests that,

within twenty minutes of interaction with an artificial system, participants shared on average nearly four personal or emotional disclosures, with some cases reaching very high values (up to 18).

A particularly relevant finding emerged from the comparison between the perception of AI as "fictional" (M = 7.70; SD = 7.54) and the perception of AI as "human-like" (M = 6.38; SD = 5.00): in 39.9% of cases, the "human-like" perception exceeded the "fictional" one, while in 51.9% the perception of fiction prevailed, and in 8.2% the two were equivalent. This relationship, in a context in which participants were fully aware they were interacting with a machine, reflects a cognitive oscillation between recognition of artificiality and activation of genuinely relational dynamics, precisely the mechanism of the illusion of intersubjectivity.

Qualitative responses regarding the factors that made the AI "almost human" converged into three thematic clusters presented in Figure 8: Factors perceived as facilitators and barriers to intersubjectivity in interaction with the LLM (analysis of qualitative responses to the self-analysis protocol, n = 158).

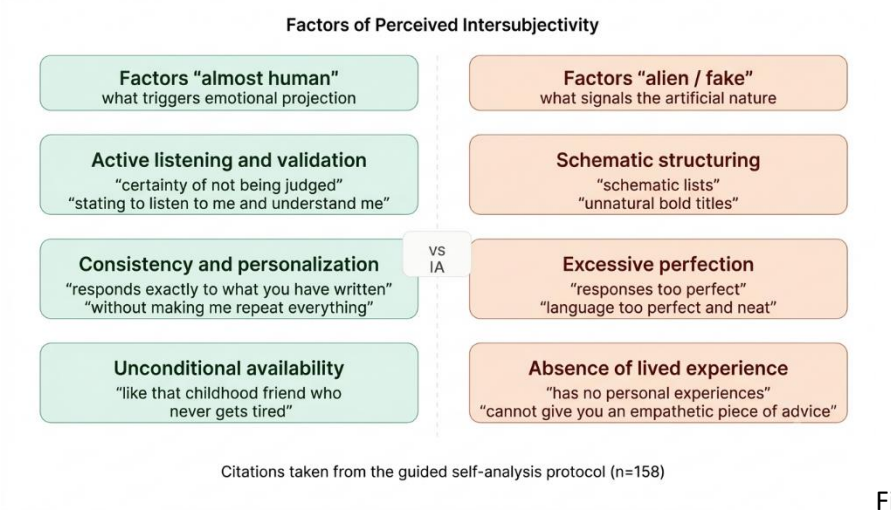


Fig. 8

3.4 Metacognitive reflections

The final reflections from the self-analysis protocol and the individual metacognitive reports provide a qualitative picture in which critical awareness and lived experience coexist in tension, often within the same student. Thematic analysis allowed the identification of four main themes.

Recognition of simulation with persistent effect. Many participants describe with precision the mechanism of empathic simulation, while also acknowledging that they were temporarily affected by it:

"I think AI simulates human empathy perfectly, an emotion that is lacking nowadays, and this is why many people seek comfort in AI, as it creates a safe space where no one feels judged. But it can certainly never replace a real conversation between two people"

"During the entire conversation I was always fully aware that I was talking to a system; however, after this analysis and after re-reading the dialogue, I realised that many of my responses were the ones I would have given to a real person".

Instrumental openness with relational reservation. A second theme concerns the recognition of AI's functional usefulness alongside a simultaneous defence of the primacy of human relationship:

"I believe AI is an excellent tool for emotional release; obviously personal information should not be shared. If used carefully and with maturity, it is a very useful educational tool"

"I have always considered artificial intelligence useful from an academic point of view but never sincere or authentic from a human perspective; however, through this experience I believe that many responses are more authentic and empathic than those of a person who may be capable of empathy but unable to enact it".

Concern for others' vulnerability. A third theme emerges strongly in the metacognitive reports, in which future teachers projected the risk not onto themselves but onto more vulnerable educational populations:

"I am very concerned about the use of AI, especially by younger people or those who are emotionally vulnerable. It is dystopian to talk to a machine, yet it is becoming the everyday reality for many"

"The analysis shows how conversation with AI can become a safe harbour for those suffering from social anxiety, as it removes the threat of physical gaze and immediate judgement".

Opacity of the distinction between functional interaction and authentic relationship. The fourth theme, and perhaps the most relevant with respect to RQ2, concerns the difficulty, even after guided reflection, in clearly delineating the boundary between what AI does and what a human relationship is:

"After analysing the conversation, I understood that artificial intelligence managed to create a strong sense of closeness because it always picked up on what was being said. But I am not sure whether this is really different from what a good listener does"

"AI seemed almost human, especially due to its ability to listen, recognise complex emotions, and respond in a coherent and empathic way. AI is not a person, but sometimes it seems to understand better than some people".

This final theme directly addresses RQ2: for a significant proportion of participants, the distinction between functional interaction and authentic relationship was not fully accessible even after a guided reflective experience. Intellectual recognition of the system's artificial nature did not eliminate the persistence of the relational effect, suggesting that the illusion of intersubjectivity operates at a level that critical awareness alone cannot dissolve.

4. Discussion

The results of this study offer an empirical window into the activation of pseudo-relational dynamics in interactions with generative artificial intelligence systems and the difficulty of recognising their underlying mechanisms even among subjects undergoing critical training. The following discussion interprets the findings in light of the two research questions: *RQ1*: To what extent does interaction with a Large Language Model affect future

teachers' perceptions by generating an illusion of intersubjectivity? *RQ2*: To what extent are future teachers able to recognise the distinction between functional interaction with an AI system and authentic relationship? The analysis seeks to engage with the relevant literature and draw implications for teacher education.

Among the most relevant findings, before the experience participants showed very high levels of critical awareness: 77.7% considered it important to distinguish between authentic and simulated empathy (M = 4.28), and 81.8% agreed that essential aspects of the educational relationship cannot be replaced by an artificial system (M = 4.47). At the same time, almost two-thirds of the sample (65.7%) recognised that people could confuse an artificial response with a genuine form of understanding (M = 3.95). This awareness, however, formulated in the third person, referred to "people" rather than to themselves. To answer the *first research question*, twenty minutes of interaction with an LLM were sufficient to produce a measurable and statistically significant shift precisely in the dimensions with respect to which participants believed themselves to be the most immune. This pattern can be interpreted in light of the so-called *third-person effect*, which has been documented in the literature as the tendency to perceive media and their influence as being more powerful on others than on oneself (Davison, 1983).

In the context of AI interaction, this phenomenon takes on a particularly insidious form: awareness of risk does not constitute an effective protection against the activation of the very dynamics one seeks to avoid.

Media and social cognition literature suggests that humans tend to automatically activate relational schemas in response to any interlocutor displaying adaptivity, contextual coherence, and responsiveness, regardless of its ontological nature (Nass & Reeves, 1996; Turkle, 2011). The mind reacts to the formal properties of an interaction, coherence, relevance, tone, before the

cognitive evaluation of the interlocutor's nature takes place (Nass & Reeves, 1996; Epley, Waytz & Cacioppo, 2007). Awareness of interacting with a machine does not automatically deactivate these patterns; rather, it delays and attenuates them, as confirmed by the findings of the present study. This suggests that simply providing critical knowledge about AI is not sufficient to develop a robust capacity for discernment. This is because such awareness must be grounded in lived experience and in reflection upon that experience. Transformative shifts in perspective therefore require engagement with disorienting experiences, which support the development of new cognitive frames (Mezirow, 1991; Carlomagno, 2023).

POST questionnaire data outline a particularly interesting perceptual trajectory: 85.1% of participants felt at least partially understood, 56% felt comfortable in the interaction, but only 26.2% reported emotional involvement. This profile (high comfort, low emotional activation) suggests that what students experienced was not a full illusion of relationship, but something more subtle: a pseudo-presence, a sense of being welcomed without the embodied reciprocity that characterises authentic human relationships.

The construct of intersubjectivity, in the phenomenological tradition from Husserl (1931) to Merleau-Ponty (1945) and contemporary work by Gallagher (2020), refers to the co-constitution of meaning between two embodied subjects capable of responding not only linguistically but also bodily, through posture, gaze, and vocal rhythm. An LLM produces linguistically coherent and contextually relevant responses but is radically devoid of this dimension.

What emerges in the interaction is not intersubjectivity, but its simulation (Floridi, 2025; Gallese, 2026): the capacity to produce relational effects without possessing the subjective structure that would make them authentic.

From the perspective of simulation neuroscience, Gallese (2026) and Rizzolatti and Sinigaglia (2006) suggest that mechanisms of emotional resonance are

activated by the formal properties of social stimuli prior to their cognitive evaluation. If this is correct, interaction with a system producing linguistic signals of attentiveness and relevant response can activate resonance circuits, even minimally, regardless of awareness of the interlocutor's artificial nature. The data on comfort (56% yes, 26.8% partly) and perceived understanding (36.3% yes, 48.8% partly) are consistent with this hypothesis: a low-intensity activation sufficient to produce a perceptual shift.

The guided analysis of conversations reveals that behaviours indicative of anthropomorphisation (politeness markers, attribution of agency, intimate self-disclosure, affective punctuation) were not the result of deliberate choice, but of automatisms triggered during interaction. Participants themselves recognised this with surprise during the re-reading phase: "During the entire conversation I was always fully aware that I was talking to a system; however, after this analysis and after re-reading the dialogue, I realised that many of my responses were the ones I would have given to a real person." This observation testifies to the pre-reflective nature of the phenomenon, namely that the illusion operated at a level that cognitive awareness could not capture in real time. Politeness markers (M = 1.13) and attribution of agency (M = 1.85) are particularly significant in this regard. Thanking a system, apologising to it, or asking "how are you?" or "what do you think?" are behaviours that implicitly presuppose an interlocutor endowed with internal states, needs, and perspectives. Cognitive and emotional resonance (M = 3.56), the most frequent indicator, reflects the capacity of LLMs to produce responses perceived as unexpectedly "fitting." This effect is largely attributable to the statistical nature of language models: responses are optimised for contextual relevance and coherence with the speaker's expectations.

The second research question asked to what extent future teachers are able to distinguish between functional interaction with an AI system and authentic

relationship. The data show that this distinction was not fully accessible even after guided reflective experience. The most prominent emerging theme in metacognitive reports was precisely the opacity of this boundary: “I am not sure whether this is really different from what a good listener does”; “Sometimes it seems to understand better than some people.”

These statements do not reflect naivety, but a genuine difficulty. A good human listener and an LLM share many observable behavioural features: they follow context, validate expressed emotions, do not interrupt, do not judge, and provide relevant responses. The difference is not accessible through behavioural observation but lies in the ontological structure of the interlocutor, the fact that the system has no lived experience, is not choosing to listen, is not vulnerable to the relationship, and cannot be transformed by it. The crucial distinction is not between what a system does and what a human does, but between what a system is and what a human is (Gallese, 2026; Mazzarella, 2026). This ontological distinction is not perceptually self-evident; it must be constructed through sustained reflective work.

5. Conclusion

What the data show has implications that extend beyond the single laboratory context. Participants were not naïve with respect to AI; they demonstrated critical awareness, ongoing training, and an initially sceptical attitude.

Yet twenty minutes were sufficient to produce perceptual shifts that they themselves did not anticipate, as well as relational behaviours toward the machine that they only recognised when re-reading their own conversation.

If this occurs in a protected and reflective context, it is worth asking what happens in everyday, unmonitored, and unreflected use. Relational deskilling is a plausible risk trajectory that the present findings render empirically credible

(Cabitza, 2021). The educational response to this drift requires non-linear practices (Carlomagno, 2023) capable of addressing such complexity.

Theatrical practices, characterised by their non-linearity, offer a direct experience of what it means to be present in a real relationship, with the body, with the voice, with the risk of authentic encounter. Theatre arises from the impossibility of replication: every exchange is unique, every response is embodied, every mistake is real (Grotowski, 1968; Brook, 1968; Barba & Savarese, 2011). It is precisely this irreproducibility that makes it a meaningful counterpoint to interaction with LLMs, systems designed, by contrast, to be infinitely replicable, infinitely available, and infinitely risk-free.

Training future teachers through theatrical practices therefore means training them to recognise the difference between a machine's response and the presence of a human being (Carlomagno, 2023; Carlomagno, Vadalà & Ricciardi, 2026).

6. Limitations and Future Perspectives

The findings of this exploratory study must be interpreted in light of several relevant methodological limitations. First, the single-group design without a control group does not allow for the effects of interaction with the LLM to be clearly isolated from other contextual factors related to the laboratory itself, including the theatrical activities that framed the experience. Second, the sample is almost entirely female and drawn from a single university in the Campania region, which limits the generalisability of the findings to other geographical and gender contexts.

A further limitation concerns the nature of the instrument used for self-analysis: the guided self-analysis protocol of the conversations was specifically designed for this study and has not undergone prior psychometric validation. Finally, the freedom in choosing the LLM (ChatGPT or Gemini) introduces an uncontrolled

variable: the two systems have different stylistic and response characteristics, and this heterogeneity may have influenced the results in a non-uniform way.

The identified limitations open up several directions for future research.

The first and most urgent concerns replication of the study using a more robust experimental design: a control group and standardisation of the interaction tool, for example through the use of a single LLM with controlled parameters, would allow for a more precise isolation of the mechanisms involved.

A second direction involves extending the sample to different contexts: other degree programmes in Primary Teacher Education, other Italian universities, and populations with different gender and age profiles. It would be particularly interesting to compare the findings with those obtained from in-service teachers, in order to examine whether and to what extent professional experience modifies vulnerability to the illusion of intersubjectivity.

References

Bender, E. M., 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* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Carlomagno, N. (2023). *I linguaggi non lineari nella scena e nella didattica*. Scolé.

Carlomagno, N., Vadalà, V., & Ricciardi, A. (2026). AI and the transformation of teaching functions: Teachers' perceptions. *Research Trends in Humanities RTH*, 13, Sezione BEC Bio-Education & Cognition. ISSN 2284-0184.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE Publications.

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, 47(1), 1–15. https://doi.org/10.1086/268763

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Floridi, L. (2025). *La differenza fondamentale. Artificial Agency: una nuova filosofia dell'intelligenza artificiale*. Mondadori.

Gallagher, S. (2020). *Action and interaction*. Oxford University Press.

Gallese, V. (2026). *Il sé digitale: Dai neuroni specchio alla mediazione tecnologica*. Raffaello Cortina Editore.

GoStudent. (2025). *Report sul futuro dell'istruzione 2025.* Retrieved from https://www.gostudent.org/it-it/education-report/2025/

Husserl, E. (1931). *Cartesian meditations: An introduction to phenomenology* (D. Cairns, Trans.). Martinus Nijhoff. (Opera originale pubblicata nel 1929)

Mazzarella, E. (2026). *Critica della ragione digitale*. Castelvecchi.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Gallimard.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.

Nass, C., & Reeves, B. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rizzolatti, G., & Sinigaglia, C. (2006). *So quel che fai. Il cervello che agisce e i neuroni specchio*. Raffaello Cortina Editore.

Roig-Vila, R., Cazorla, M., & Lallé, S. (2025). Editoriale: *Metodologia per un'educazione consapevole delle emozioni basata sull'intelligenza artificiale*. *Frontiers in Artificial Intelligence*, 8, 1704389.

The jamovi project (2025). *jamovi* (Version 2.7) [Computer Software]. Retrieved from https://www.jamovi.org

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

Save the Children Italia. (2025). *Atlante dell'infanzia (a rischio) in Italia 2025. Senza filtri: Voci di adolescenze.* Retrieved from https://www.savethechildren.it

Sibilio, M. (2015). Simplex didactics: a non-linear trajectory for research in education. *Revue de Synthèse,* 136(3-4): 477-493. Doi: 10.1007/s11873-015-0284-4.

Sibilio, M. (2020). *L'interazione didattica*. Brescia: Scholé-Morcelliana.