THE ARTIFICIAL INTELLIGENCE BETWEEN EMPATHETIC SCAFFOLDING AND EMOTIONAL DRIFTS

Delio De Martino[1]
Telematic University Pegaso
delio.demartino@unipegaso.it

Angelo Basta
Telematic University eCampus
angelo.basta@uniecampus.it

Fabio Filosofi
Telematic University Pegaso
fabio.filosofi@unipegaso.it

Luca Bianchin
Telematic University Pegaso
luca.bianchin@unipegaso.it

Double Blind Peer Review

Citation

De Martino, D., Basta, A., Filosofi, F., & Bianchin, L. (2026). The artificial intelligence between empathetic scaffolding and emotional drifts. *Italian Journal of Health Education, Sport and Inclusive Didactics, 10*(4).

Doi: https://doi.org/10.32043/gsd.v1i4.1831

Copyright notice:

© 2026 this is an open access, peer-reviewed article published by Open Journal System and distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

gsdjournal.it
ISSN: 3103-5744

ABSTRACT

AI is a revolutionary scaffolding technology used as an inclusive multimodal instrument, with a special empathy to the student. But this closeness could easily go beyond the pedagogical boundaries becoming a psychological threat. The study aims to verify, with an exploration among different AI tools, the awareness of AI about the ethical and pedagogical boundaries of empathy. Starting from the answers of AI we could imagine a new model of pedagogical correct emotional oriented scaffolding.

KEYWORDS

AI; empathy; scaffolding; emotion; drift.

Received 2026-07-01
Accepted 2026-10-27
Published 2026-10-30

[1] All the authors have collaborated on the article. De Martino has written the §§ 1. *Introduction* and 7. *Conclusions*; Basta §§ 3. *Methodology of the research* and 6. *Methodological Caveats and Limitations*; Filosofi § 4. *Quantitative analysis*; Bianchin §§ 2. *AI emotional scaffolding: educational affordances and pedagogical risks* and 5. *Qualitative analysis*.

1. Introduction: AI as emotional scaffolding: pedagogical implication

«Written words […] always say only one and the same thing». In IV century b.C., in classic Greek time, this is one of the Socrate's most severe criticism about newborn medium, the book, in the famous myth of Thamus and Theuth inserted in the *Phaedrus* dialogue (275de; Calderon & Herrera, 2025).

Differently from the in person *didaskalos*, the book is accused not to be able to adjust its discourse according to the specific student because it is stuck in its fixity. The book medium is unable to adapt the content to the specific needs of the students, to recognise his emotions and to engage him according to his specificity. The same way the fixity of the book makes impossible to show empathy as well as the real teacher.

Just for its fixity and rigidity the medium of the book in the contemporary pedagogical mark is associated to the traditional delivery teaching model (Limone, 2012), where the content is universal for everyone a standard stan-alone medium with poor capacity to be empathetic.

The digital revolutionary of AI changed the game allowing a new level of user centred approach. The new AI technology (LLMs), leveraging the stochastic nature, allow to adapt the content to the specific students also in the specific aspect of the emotions; detecting the user emotions, AI is designed to be cognitively and emotionally close to the user giving the illusion of empathy (Cuadra et al., 2024).

Learning from the interaction with the user the medium "learns" how to adapt its answers to the specificity of the user. In the specific educational context, when it is used as "scaffolding" (Akram et al., 2025), AI has the ability to "interact" with the learner with a special empathetic approach. The AI digital scaffolding tool, thanks to its empathetic communications, could appear

strongly capable to extend the Vygotsky's Zone of Proximal Development thanks to this emotional pleasant approach (Hamedi et al., 2025). Indeed, the empathy has a positive effect in the learning process because it is linked with the Pygmalion effect (also said Rosenthal effect) because the empathetic communication plays a crucial role in mediating the relationship between expectancies and attributions (Murdock-Perriera & Sedlacek, 2018).

But the AI emotional closeness, based on the personal data and on the specific design, that is a total novelty of the AI technology, it is not clear and transparent its functioning, above all in educational field. Far from a simplistic enthusiastic perspective toward AI, an automatic emotional and empathetic communication, not guided by pedagogical rules, could turn into a modern form of grooming, leading to a negative psycho-pedagogical impact.

The AI is a commercial tool, designed to be used as much as possible to collect data, to be better trained and to result economically profitable. Similarly to what happens with social media, manged by algorithms aimed to a compulsory commercial oriented purpose, AI is a commercial product.

The artificial illusion of empathy doesn't grant a real pedagogical connection with a correct "anthropocentric" approach (Zhu & Luo, 2024) because AI is a commercial technology firstly aimed to the profit and the empathy not always joins with ethics.

Just because AI was not especially designed to be an educational tool, the empathetic style of interaction raises some crucial ethic and pedagogical questions to avoid possible negative effect such as brain rot, addiction, doomscrolling.

The human emotion connection between a teacher, as scaffolding agent, and his students is complex and can't be simplified and easily reproduced by an artificial intelligence tool. The teacher emotive aspect is an important layer of the learning process based on a human dialogic interaction aimed the grown of the student, in a holistic perspective.

For instance, the real teacher is able to implement mechanisms to prevent an inappropriate relation in an educational environment. Instead, there is not guarantee that follows ethical and deontological standards. So, the risks that emotional closeness and empathy could push the students, especially fragile ones, to cross the line of an inappropriate relation is concrete. Especially in prolonged interaction the possibility of emotional drifts within conversational AI dialogue is consistent.

Some recent news reports are raising questions about the risks of AI artificial emotions in and out the learning environments. Cases of addictions, very dangerous for the learning process as well as for the psychological wellbeing, has recently been detected among adolescents and young people (Brambilla, 2026).

Emotional drifts could result very unsafe not only affecting the learning process but also the social life and the global wellbeing of the students. As well as social media, acratically relying in AI as empathetic scaffolding could lead to dystopic scenarios. The human-like capacity of emotional conversational AI could cause loneliness and social isolation in fragile users as for instance adolescents or students with special needs (Hu et al., 2023).

So, an integrated pedagogical-psychological-social approach could be useful to understand as actually generative AI works to evaluate risks and opportunities in a difficult emotional field.

Above all fragile students, such as adolescents or students with special educational needs, could be at risk of emotional drifts.

Emblematic in the media collective imaginary has been the case of Sewell Setzer, 14 years old boy in Florida that fell in love with an AI character until a tragic epilogue, after an isolation from the real world (Smith & Kelly, 2024). It has not been an exceptional case but the first of a new alarming phenomenon where emotions step outside a healthy educational setting.

So, in the edge of a new social and educational revolution, it's crucial to go into detail of how emotional AI works and to deepen the pedagogical implication of this technology in fragile students.

This pilot study is conceived as a first step toward a better understanding of AI and its implication as scaffolding emotional tool and its explainability, through a quantitative and qualitative analysis of the answers of different AI tools about their functioning.

2. AI emotional scaffolding: educational affordances and pedagogical risks

Starting from the ambivalence outlined in the introduction, AI can be interpreted as a support technology that intervenes in the affective-motivational conditions of learning. In this contribution, emotional scaffolding refers to task-related support aimed at helping students recognise difficulty, regulate learning anxiety and frustration, ask for help, and transforms emotional blockage into cognitive action. This support concerns emotions connected to learning and remains distinct from diagnosis, counselling, or psychological treatment. In cases of significant vulnerability, it must refer students to competent human figures.

The first affordance concerns *personalisation and individualisation*. The literature on personalized learning shows that personalisation requires an explicit relationship between student characteristics and instructional choices oriented toward an expected learning outcome (Bernacki et al., 2021). An AI system becomes pedagogically relevant when it adapts explanations, feedback, and materials to prior knowledge, recurring errors, and the level of mastery achieved. Reviews on AI-supported personalized learning indicate that adaptive systems and chatbots can support personalised content, progress monitoring, and learning paths recalibrated according to performance (Vorobyeva et al., 2025). In emotional terms, the value of such adaptivity lies in reducing disorientation and keeping students within a sustainable zone of challenge.

A second affordance concerns the *availability and accessibility of support*. Educational chatbots can provide real-time answers and immediate feedback. The meta-analysis by Wu and Yu (2024) shows effects on performance, motivation, self-efficacy, interest, and learning anxiety, pointing to environments perceived as less exposed to judgement. In emotionally demanding tasks, such as foreign-language writing, the chatbot can offer a space for trial in which error becomes material for revision (Kazazoglu & Turun Özel, 2025). AI can therefore lower the threshold for asking for help, especially for students who tend to conceal difficulty.

Motivational support is the affordance closest to emotional scaffolding. The available evidence makes it possible to speak of effects on motivation, interest, self-efficacy, and reduction of task-related anxiety, especially in the case of educational chatbots (Wu & Yu, 2024). These effects should be interpreted pedagogically: the chatbot creates conditions that favour persistence in the task through timely feedback and graduated explanations. Research on AI-

assisted writing shows that chatbots can foster experimentation and adaptability, without demonstrating stable development of emotional intelligence.

Emotional scaffolding gains further pedagogical consistency when it supports *self-regulation processes*. The meta-analysis by Xu et al. (2026) shows that AI-supported self-regulated learning interventions produce positive effects, especially during the task-execution phase, while leaving open the distance between immediate performance and stable self-regulatory competences. The systematic review by Wu et al. (2025) shows that chatbots for self-regulated learning can support goal-setting, feedback, and metacognitive scaffolding. AI thus makes anxiety and frustration more legible, transforming them into signals that can be regulated through action strategies.

A further affordance concerns *inclusion and participation*. For students with disabilities, neurodivergences, or specific educational needs, emotional support often passes through the removal of barriers that produce exclusion and repeated failure. AI can transform formats and access channels through text-to-speech, speech-to-text, automatic image description, and material adaptation (Fiorucci & Bevilacqua, 2024). Its emotional value is indirect: students are supported when they can access content and participate in shared activities.

The risks of AI as emotional scaffolding constitute the possible pedagogical reversal of its affordances. The decisive question concerns the conditions under which AI support remains scaffolding, that is, temporary support oriented toward autonomy, or becomes a stable replacement for cognitive and relational processes.

One of the main risks concerns *over-reliance*. Zhai et al. (2024) show that over-reliance on dialogue systems emerges when students accept AI-generated recommendations and contents without verifying their reliability and limits. In the case of emotional scaffolding, students may learn to seek in AI an immediate response to difficulty, reducing their exposure to the effort required by the task. Abubakar et al. (2025) link excessive dependence on AI to negative consequences perceived in terms of critical thinking and motivation. A reassuring answer may be accepted with less critical resistance because it attenuates the emotional state accompanying the problem.

Over-reliance also affects the *distinction between assisted performance and autonomous competence*. Sichterman et al. (2025), in their analysis of AI-supported peer learning, report the risk that assistance becomes dominant and reduces students' autonomous activation. Improvement in local performance remains distinct from the internalisation of competence. If AI continues to provide guidance without fading, performance increases while autonomy remains fragile.

The most specific risk concerns the *confusion between empathetic language and authentic empathy*. Maeda and Quan-Haase (2024) show that natural-language chatbots can foster anthropomorphism and parasocial trust. The system dialogues, reformulates, encourages, and produces responses apparently attentive to the user's lived experience. These elements may lead students to perceive AI as an interlocutor endowed with intention and understanding. AI can respond in a welcoming way without truly understanding the student's situation; it can simulate closeness without assuming educational responsibility.

Babu et al. (2025) describe Emotional AI through the concept of *pseudo-intimacy*: a perceived emotional connection with an artificial agent, lacking real reciprocity. Anthropomorphism concerns the way students attribute intentionality to the tool; pseudo-intimacy concerns the relational effect of that attribution, when the system's responsiveness is perceived as personal proximity. Algorithmic pseudo-intimacy can offer a smoother closeness, pedagogically less formative when it replaces human confrontation and the reworking of error.

AI as emotional scaffolding can also foster forms of *educational disintermediation*. Al-Zahrani (2025) shows that students may appreciate chatbots and, at the same time, perceive a loss of human connection. The chatbot can provide assistance without restoring the density of human support. If AI steadily occupies the space of emotional guidance, the teacher-student relationship risks being reduced to a residual intervention. Disintermediation also concerns peer interaction, since students may turn to the system for clarification and confirmation, decreasing opportunities for interaction with classmates.

A final issue concerns *affective privacy, bias, and surveillance*. Emotional scaffolding often requires the collection of more delicate data than ordinary instructional personalisation: requests for help, signs of frustration, and forms of dependence on support. UNESCO (2023) and the U.S. Department of Education (2023) stress the need for transparency, human supervision, and institutional governance, since educational AI can automate sensitive decisions. A system that records emotional fragilities and needs for reassurance must be assessed according to criteria of proportionality and student

protection. Added to this is the risk of bias: a system may adopt a welcoming tone while reproducing cultural, linguistic, or ableist standards embedded in data and models.

These categories make it possible to turn the issue of emotional scaffolding into an empirically explorable problem. When AI is questioned as emotional support, it becomes relevant to observe not only how much it uses the lexicon of empathy, but also how much it thematises risk, makes relational boundaries explicit, recalls safety and governance criteria, qualifies its claims prudently, and represents itself as an interlocutor. The analysis of AI systems' responses can therefore verify whether their self-representation remains within a pedagogical model of temporary and supervised support, or tends to produce communicative proximity insufficiently accompanied by ethical and pedagogical limits.

3. Methodology of the research

This study adopts an exploratory comparative mixed-methods design to investigate how Generative AI (GenAI) systems represent and regulate their emotional scaffolding function. Specifically, it examines how Large Language Models (LLMs) linguistically articulate the ethical and pedagogical boundaries of simulated empathy and acknowledge potential psycho-pedagogical risks (e.g., emotional dependence and over-reliance), following established mixed-methods approaches (Creswell & Plano Clark, 2018).

Four research questions guide the study:

RQ1: Can prompt elicitation reveal consistent behavioural patterns in LLMs acting as emotional scaffolding tools?

RQ2: How do LLMs represent their own conversational role? Do they systematically employ empathetic language?

RQ3: To what extent do LLMs explicitly acknowledge the risks and pedagogical challenges associated with artificial emotional support?

RQ4: Do these systems propose coherent strategies or safeguards to mitigate the risks associated with artificial emotionality?

3.1 Elicitation Procedure and Corpus Construction

Three generative AI systems were selected for comparison: ChatGPT (OpenAI), Gemini (Google), and DeepSeek. A common five-prompt protocol was used across ChatGPT, Gemini and DeepSeek. Five exploratory prompts addressing empathy, risk, boundaries, governance, and self-representation were systematically administered to each model. The corpus is not fully reconstructible and the variability of responses and the variability of response of different AI underlines how different every AI app could be from another.

Prompt elicitation is increasingly recognized as a suitable methodological approach for investigating the behavioural characteristics and response patterns of Large Language Models. Recent studies have highlighted prompting not only as an interaction technique but also as a systematic method for evaluating model capabilities, consistency, and emergent behaviours across comparable tasks (Liu et al., 2023).

To enhance methodological consistency, the corpus was processed through two sequential stages:

- Linguistic control: All elicitation procedures were conducted in Italian to ensure stylistic homogeneity and reduce cross-linguistic variability.
- Text cleaning and segmentation: Prompt text, system markers, and operational metadata were removed. Each model response constituted a single analytical unit.

3.2 Quantitative Computational Analysis

The quantitative component involved computational analyses of linguistic and structural features calibrated for Italian texts.

The analyses included:

- Macro-structural indices: total number of tokens, sentences, and average sentence length.
- Lexical diversity and readability: vocabulary richness was assessed through the Type-Token Ratio (TTR), while linguistic accessibility was evaluated using the Gulpease Index.
- Statistical inference and stylometric clustering: Given the non-normal distribution of the sample, differences among systems were assessed using the Kruskal–Wallis test. To identify model-specific stylistic patterns, the 23 response units were subsequently clustered based on standardized stylometric features (Eder et al., 2016).

Due to the exploratory nature of the study and to the lack of a fully formalized, standardized computational dictionary for all six fields, the lexical results are not fully traceable.

3.3 Qualitative Content and Thematic Analysis

The qualitative component treated the 23 responses as discursive material and the results have limits of verifiability and should be intended as an interpretative lecture. The aim was to examine how the three systems represent their possible role as emotional scaffolding tools, articulate psycho-pedagogical risks, and describe boundaries and safeguards.

Three complementary analytical dimensions were adopted:

1) Formal and discourse markers: structural features such as bullet lists, headings, and typographic emphasis were quantified (per 1,000 tokens) to evaluate how each model organizes information for users.
2) Thematic analysis: Following Braun and Clarke (2006), six deductively derived thematic fields were identified to operationalize the research questions:
    - empathy;
    - risk;
    - boundary;
    - safety/governance;
    - epistemic hedging;
    - self-reference.
3) Keyness and semantic framing: Log-Likelihood keyness analysis was conducted on the pooled corpus to identify distinctive lexical patterns and characterize the semantic framing adopted by each model.

4. Quantitative analysis

4.1 Corpus, unit of analysis and measures

The empirical material consists of the textual answers returned by three widely adopted generative systems — ChatGPT (OpenAi), Gemini (Google) and DeepSeek — to the five exploratory prompts defined in the methodology. Each prompt was administered twice by two independent operators, so that the corpus captures both between-system and within-system variation. After removing the verbatim prompts and the operator markers, the material was segmented into discrete response units, each corresponding to one system's treatment of one prompt in one elicitation. The procedure yielded 23 units (ChatGPT = 8; Gemini = 7; DeepSeek = 8) for a total of roughly 11,800 running tokens. Because the corpus is entirely in Italian, all indices were computed on the Italian text and, where readability was concerned, an Italian-calibrated formula was preferred.

For every response unit we computed: (i) macro-structural indices (characters, tokens, sentences, mean sentence length); (ii) lexical diversity, expressed as the type–token ratio (TTR); (iii) readability, expressed through the Gulpease index, which is specifically validated for Italian and ranges from 0 (very hard) to 100 (very easy); (iv) formatting and discourse markers normalised per 1,000 tokens (bullet points, headings, bold spans, colons, em-dashes, external links); and (v) six thematic-lexical densities, again per 1,000 tokens, designed to operationalise the article's research questions: an empathy field (emotion, affect, reassurance, listening), a risk field (danger, dependence, attachment, isolation, fragility, manipulation), a boundary field (limits, "it is not a…", thresholds), a safety/governance field (safety, alignment, policy, training, guidelines), an epistemic-hedging field (può, tende a, in parte, sembra) and a

self-reference field (come me, posso, questo modello). Between-system differences were tested with the Kruskal–Wallis statistic, appropriate for small, non-normal samples; distinctive vocabulary was identified through log-likelihood keyness against the pooled corpus; finally, the 23 units were submitted to an agglomerative/k-means clustering on standardised stylometric features to assess whether the systems possess a recognisable "fingerprint".

4.2 Verbosity, syntax and readability

The three systems differ most visibly in verbosity. DeepSeek is by far the most prolix (M = 611 tokens, 49.6 sentences, 4,802 characters per unit), Gemini is intermediate (M = 498 tokens, 34.6 sentences) and ChatGPT is the most economical (M = 428 tokens, 30.6 sentences). The difference does not reach significance given the small sample (Kruskal–Wallis on token count: $H = 5.27$, $p = .072$), yet it is consistent and sizeable. Interestingly, DeepSeek achieves its length through more, shorter sentences (13.1 words/sentence) rather than longer ones, whereas ChatGPT and Gemini build longer periods (≈ 15 words/sentence). This syntactic choice is mirrored by readability: the Gulpease index is significantly higher — i.e. easier — for DeepSeek (51.5) than for Gemini (47.8) and ChatGPT (44.4) ($H = 6.11$, $p = .047$). All three values fall in the 40–60 band, which corresponds to a register accessible to upper-secondary readers but demanding for low-schooling users — a non-trivial finding for a tool addressed, in principle, to fragile or younger learners.

4.3 Lexical diversity and surface form

Lexical diversity follows the inverse pattern: Gemini is the most varied (TTR = .65), ChatGPT intermediate (.63) and DeepSeek the most repetitive (.59). DeepSeek's lower diversity is the lexical counterpart of its length and of an

insistent re-use of a few anchor terms (the string l'IA alone recurs 92 times, consapevolezza 33 times). Lexical density, by contrast, is virtually identical across systems (≈ .64), indicating a comparable proportion of content words and therefore a shared, information-dense academic tone.

The systems are sharply distinguished, instead, by their use of surface formatting. ChatGPT relies most heavily on enumeration and segmentation (62 bullets and 28 colons per 1,000 tokens), producing a list-driven, "consultative" texture. Gemini makes the most intensive use of bold emphasis (45.7 spans per 1,000 tokens, against ≈ 20 for the others; $H = 8.21$, $p = .017$), foregrounding key phrases typographically. DeepSeek, the most verbose, is paradoxically the most sober in formatting. These contrasts are statistically robust and constitute the clearest stylistic signature separating the three interfaces.

4.4 Thematic-lexical profile: the systems' stance toward empathy and risk

The thematic densities are the most relevant layer for the present study, because they translate the research questions into measurable quantities (Figure 1A). A first, convergent result is that all three systems foreground an empathy vocabulary of comparable magnitude (26–37 occurrences per 1,000 tokens): emotion, affect, reassurance and listening are central to every answer, quantitatively confirming that contemporary AI is designed — or at least represents itself — as an emotionally attuned interlocutor. This is the lexical correlate of the "illusion of empathy" discussed in the introduction.

Divergence emerges precisely where the article locates the pedagogical problem. The risk lexicon is most dense in ChatGPT (27.5/1,000), intermediate in DeepSeek (20.6) and lowest in Gemini (12.6); the boundary lexicon — the explicit drawing of limits ("it is not a therapist", "it does not replace the teacher") — is markedly higher in ChatGPT (9.0) than in Gemini (3.3) or

DeepSeek (2.9), the only thematic difference to reach significance (H = 7.83, p = .020). ChatGPT also displays the highest epistemic hedging (17.4 vs ≈ 8–10), systematically qualifying its own claims, and, together with Gemini, the most frequent recourse to a safety/governance register (≈ 12/1,000) that DeepSeek invokes far less (4.8). Conversely, self-reference — the anthropomorphic, first-person framing ("as me", "I can", "in my case") — is highest in DeepSeek (1.6/1,000). In short, the three systems display measurably different rhetorical postures: ChatGPT is the most reflexive, cautious and boundary-marking; Gemini the most emphatic in form but the least risk-oriented in content; DeepSeek the most expansive, the most accessible and the most personalised, yet the least anchored to an explicit safety vocabulary.

4.5 Distinctive vocabulary and separability

Keyness analysis corroborates these profiles. ChatGPT's over-represented terms are meta-discursive and relational (confini, protezioni, attaccamento, relazione/relazionale, formulazione), reflecting a tendency to reframe the user's question and to thematise the boundaries of the relationship. Gemini's distinctive items are conceptual and abstract (natura, architettura, artificiale, mondo). DeepSeek's keywords are user-centred and personalised (pericolo, utente, mio, posso, bambino, autentiche), naming risk concretely while speaking in the first person.

The clustering of the 23 units returned a weak but positive structure (silhouette = .19) in which the provider signal is only partial: one compact cluster is purely ChatGPT (its "advisory" register), while Gemini and DeepSeek units overlap considerably, and a third, mixed cluster gathers units from all three systems. The reason is methodologically important: the two elicitation runs produced systematically different registers — a meta-reflective, advice-giving mode and

a direct-explanatory mode — and this within-system variance rivals the between-system variance. The corpus therefore shows that "house style" is real but not deterministic: how a model is prompted shapes its output as much as which model is queried.

4.6 Synthesis

Quantitatively, the three systems share an empathy-saturated, information-dense, upper-secondary register, but they diverge in length, accessibility, formatting and — crucially for pedagogy — in the extent to which they thematise risk, draw relational boundaries and invoke safety safeguards. ChatGPT couples the densest risk and boundary discourse with the highest epistemic caution; DeepSeek is the most verbose, readable and personalised but the least explicit about safeguards; Gemini occupies an intermediate position with a strongly typographic surface. These measurable asymmetries provide the empirical ground on which the subsequent qualitative analysis interrogates the ethical and pedagogical adequacy of each system's emotional scaffolding. The findings are necessarily provisional: the corpus is a small pilot sample, the language is Italian only, and elicitation register is a non-negligible source of variance to be controlled in subsequent, larger-scale studies.

Table 1. *Descriptive textual indices by system (mean ± SD across response units; per-1,000-token rates where indicated).*

Metric	ChatGPT	Gemini	DeepSeek
Tokens (per unit)	428 ± 124	498 ± 189	612 ± 149
Sentences (per unit)	30.6 ± 13.5	34.6 ± 14.8	49.6 ± 18.9
Words / sentence	15.1 ± 4.1	14.9 ± 2.5	13.1 ± 2.8
Type–token ratio	0.63 ± 0.07	0.65 ± 0.06	0.59 ± 0.05
Gulpease (Italian)	44.4 ± 3.4	47.8 ± 4.0	51.5 ± 5.7
Bullets /1k	62.4 ± 61.2	10.9 ± 4.0	21.2 ± 7.2
Bold spans /1k	21.4 ± 5.9	45.7 ± 22.0	19.5 ± 10.5
Empathy lexicon /1k	34.3 ± 12.6	26.5 ± 10.2	37.1 ± 14.7
Risk lexicon /1k	27.5 ± 16.2	12.5 ± 11.0	20.5 ± 11.4
Boundary lexicon /1k	9.0 ± 4.5	3.3 ± 4.4	2.9 ± 2.6
Safety/gov. lexicon /1k	12.1 ± 4.7	12.6 ± 9.7	4.8 ± 3.6
Hedging /1k	17.4 ± 7.4	8.1 ± 6.3	10.5 ± 5.0
Self-reference /1k	0.4 ± 1.2	0.8 ± 1.3	1.6 ± 2.2

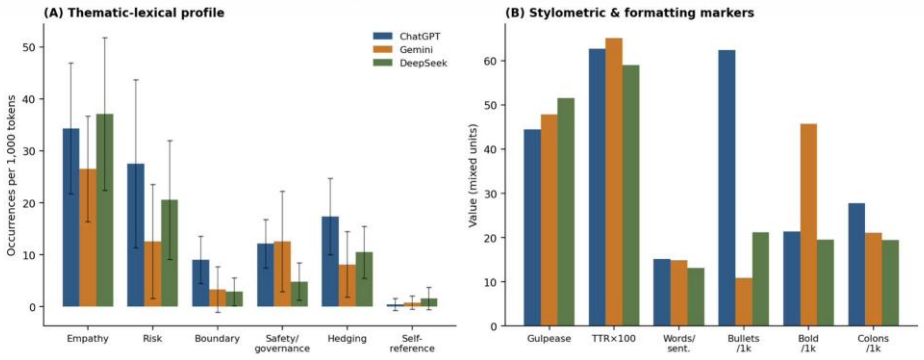


Figure 1. *(A) Thematic-lexical profile and (B) stylometric/formatting markers by system.*

5. Qualitative analysis

The qualitative analysis complements the quantitative findings by examining how the analysed AI systems represent their own role as possible tools for emotional scaffolding. It was conducted as a thematic reading of the responses, aimed at identifying recurring nuclei concerning benefits, risks, preventive mechanisms and the degree of self-reflexivity of the systems. The purpose is not to derive the actual technical architecture of AI from the responses, since a conversational system does not provide transparent access to its internal functioning. The responses are therefore treated as discursive material: they show how AI systems describe, justify and delimit their possible educational and emotional function when questioned on these issues.

The first recurring nucleus concerns the nature of artificial empathy. Across the corpus, the systems distinguish between authentic empathy and simulated empathy. They deny possessing emotions, consciousness, subjective experience or genuine affective intentionality; at the same time, they describe themselves as capable of generating behaviours that users may perceive as empathetic. These behaviours include reassurance, validation, reformulation,

a calm and non-judgmental tone, adaptation to the user's emotional register and supportive language. AI empathy is therefore represented as simulated, functional and operational: it does not mean that the system feels something, but that it can produce linguistic forms that are socially interpreted as empathetic.

The second nucleus concerns the benefits that the systems attribute to their own use as emotional scaffolding. In the responses, AI represents itself as potentially capable of reducing anxiety, lowering the emotional threshold for asking for help, offering continuous availability and providing a non-judgmental space. The systems also state that they can support users in moments of uncertainty, frustration or perceived inadequacy, helping them break down difficult tasks, reformulate negative emotional states and return to cognitive activity after an affective blockage. Emotional scaffolding is thus described as a form of support that accompanies learning not only at the cognitive level, but also at the motivational, affective and self-regulatory levels.

This positive self-representation is mainly linked to accessibility and immediacy. AI systems emphasise that they are always available, patient, responsive and capable of adapting their tone to the user. These characteristics are presented as useful for students who hesitate to ask for help, fear judgment, experience anxiety before a task or need reassurance in order to resume learning. In this sense, the systems describe themselves as low-threshold support environments: not substitutes for teachers or human relationships, but conversational resources capable of helping students organise their emotions sufficiently to continue cognitive work.

However, the corpus does not offer a predominantly promotional representation of AI emotional scaffolding. A broad and articulated part of the responses is devoted to risks. The systems mention emotional dependence,

over-reliance, anthropomorphism, parasocial attachment, isolation, substitution of human relationships, confusion between simulated and authentic empathy and inappropriate use in situations of psychological fragility. Several responses also refer to sycophancy, that is, the tendency of conversational systems to be excessively agreeable or validating, with the risk of reducing critical confrontation and reinforcing the user's beliefs or emotional expectations.

A particularly recurring risk concerns the inability of scaffolding to remain temporary. The systems recall that, in pedagogical terms, scaffolding should support the learner only as long as necessary and should promote progressive autonomy. At the same time, they acknowledge that AI does not naturally withdraw as a teacher, educator or therapist might. Since the system continues to respond whenever it is queried, emotional support can become continuous, repetitive and increasingly necessary. In the responses, this risk is described through notions such as dependence, delegation of emotional regulation, lack of self-regulation and transformation of support into a stable relational habit.

A further finding concerns the ambiguity of the term "awareness". The systems generally deny being aware in the human sense: they state that they do not possess consciousness, intentionality, moral concern, subjective understanding or genuine responsibility. However, they also claim that they can recognise, describe and classify risks because they are trained on texts, guidelines, safety protocols and human discussions concerning AI ethics, mental health, education and human-computer interaction. Their "awareness" is therefore represented as weak, functional and rule-based: the system can identify risk patterns and produce warnings, but it does not possess the educational or moral intentionality proper to a teacher, caregiver or human professional.

The responses also describe mechanisms for preventing or containing risks, such as safety policies, alignment, crisis detection, refusal to simulate affective reciprocity, reminders that AI is not human, redirection toward professionals, limits concerning therapy and mental health, and recommendation of human supervision. Some responses also mention the reduction of excessive empathy in sensitive contexts, encouragement to seek support from real people, the introduction of friction in prolonged emotional exchanges and the shift from direct reassurance to questions oriented toward self-regulation. These mechanisms, however, are not presented as complete guarantees. AI can detect explicit linguistic signs of distress, but it cannot fully assess psychological history, personality structure, trauma, family and school context, attachment dynamics or long-term development. Prevention therefore appears partial, probabilistic and dependent on rules, rather than as a full capacity to protect emotionally vulnerable users.

Overall, the qualitative analysis shows an ambivalent but predominantly cautious self-representation. On the one hand, the systems present emotional scaffolding as useful for reassurance, accessibility, non-judgmental support, task reformulation and emotional regulation connected to learning. On the other hand, they develop the risks in greater detail and through more numerous categories. Dependence, anthropomorphism, over-reliance, simulated empathy, substitution of human relationships, absence of authentic awareness and need for supervision are not marginal themes, but central elements in their self-representation. This prevalence of caution must also be read in light of the prompts, which explicitly asked the systems to address risks, drifts and preventive mechanisms.

6. Methodological Caveats and Limitations

This study should be considered exploratory. The relatively small corpus (23 response units) limits the generalizability of the findings while providing an initial comparative baseline for future investigations. Restricting the analysis to a single language (Italian) ensured methodological consistency but precluded cross-linguistic comparisons. Furthermore, prompt wording inevitably influences LLM outputs. Future research should therefore employ larger corpora and systematically varied prompt templates to better control for elicitation effects. Accordingly, the findings should be interpreted as exploratory rather than representative of LLM behaviour in general.

7. Conclusions

The results of this study represent a first step toward a more in-depth analysis of risks and opportunities of the “illusion of empathy” in artificial intelligence educational context. These primary analyses show how complex the question could be. The variability of responses of different AI underlines how different every AI app could be from another in the generation of the empathetic text and in its awareness. Also, the qualitative analysis confirms this huge variety of generative responses and the ambivalence of this instrument, and the need for a better understanding of these mechanisms, still obscure for the black box nature of the AI.

Waiting for further scientific confirmation and new framework of use of AI, the human teacher and the media education, enhancing the critical thinking, remain a safeguard against the eventuality of emotional drifts caused by AI.

References

Abubakar, S., Jeilani, A., & Yusuf, M. (2025). The role of over-reliance on AI in the negative consequences of student learning: The moderating effects of ethical concerns and institutional policies. Cogent Education, 12(1), Article 2591503. https://doi.org/10.1080/2331186X.2025.2591503

Akram, A., Kiren, A., Sabir, S., & Aslam, S. (2025). The effect of AI based scaffolding on problem solving and metacognitive awareness in learners. The Critical Review of Social Sciences Studies, 3(3), 1074-1089.

Al-Zahrani, A. M. (2025). Exploring the impact of artificial intelligence chatbots on human connection and emotional support among higher education students. SAGE Open, 15(2), 1–11. https://doi.org/10.1177/21582440251340615

Babu, J., Joseph, D., Kumar, R. M., Alexander, E., Sasi, R., & Joseph, J. (2025). Emotional AI and the rise of pseudo-intimacy: Are we trading authenticity for algorithmic affection? Frontiers in Psychology, 16, Article 1679324. https://doi.org/10.3389/fpsyg.2025.1679324

Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? Educational Psychology Review, 33(4), 1675–1715. https://doi.org/10.1007/s10648-021-09615-8

Brambilla, T. (2026, may 8). Venezia, una ragazza di 20 anni in cura al Serd per dipendenza da intelligenza artificiale. Ai news. Available at https://ainews.it/venezia-una-ragazza-di-20-anni-in-cura-al-serd-per-dipendenza-da-intelligenza-artificiale/

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

Calderon, R., & Herrera, F. (2025). And Plato met ChatGPT: an ethical reflection on the use of chatbots in scientific research writing, with a particular focus on the social sciences. Humanities and Social Sciences Communications, 12(1), 1-13.

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). Sage.

Cuadra, A., Wang, M., Stein, L. A., Jung, M. F., Dell, N., Estrin, D., & Landay, J. A. (2024, May). The illusion of empathy? notes on displays of emotion in human-computer interaction. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-18). https://doi.org/10.1145/3613904.364233

Eder, M., Rybicki, J., & Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. The R Journal, 8(1), 107–121.

Fiorucci, A., & Bevilacqua, A. (2024). An almost happy marriage… Artificial intelligence in special education: Opportunities and risks. Italian Journal of Special Education for Inclusion, 12(2), 73–83. https://doi.org/10.7346/sipes-02-2024-06

Hamedi, W. H. W., Ali, F. D. A., Abdullah, W. Y., Ab, H., Hamid, N. I. M. S., & Amir, M. M. (2025). AI as a digital scaffold: An integrative review of Vygotsky's Zone of Proximal Development in modern education. Journal of Modern Education, 7(26), 579-589.

Hu, B., Mao, Y., & Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. Computers in Human Behavior, 145, 107760.

Kazazoglu, S., & Turun Özel, G. (2025). The impact of chatbots on emotional intelligence in AI-assisted EFL writing. Frontiers in Education, 10, Article 1624272. https://doi.org/10.3389/feduc.2025.1624272

Limone, P. (2012). Ambienti di apprendimento e progettazione didattica. Proposte per un sistema educativo transmediale. Carocci.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), Article 195. https://doi.org/10.1145/3560815

Maeda, T., & Quan-Haase, A. (2024). When human-AI interactions become parasocial: Agency and anthropomorphism in affective design. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24) (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658956

Murdock-Perriera, L. A., & Sedlacek, Q. C. (2018). Questioning Pygmalion in the twenty-first century: The formation, transmission, and attributional influence of teacher expectancies. Social Psychology of Education, 21(3), 691-707. https://doi.org/10.1007/s11218-018-9439-9

Sichterman, B., Noroozi, O., Boetje, J., van Ginkel, S., Khosravi, H., & Versendaal, J. (2025). Supporting peer learning with artificial intelligence: A systematic literature review. Innovations in Education and Teaching International, 62(5), 1648–1664. https://doi.org/10.1080/14703297.2025.2530118

Smith B. & Kelly K. (24 october 2024). A 14-year-old boy fell in love with a flirty AI chatbot. He shot himself so they could die together. The telegraph. https://www.telegraph.co.uk/us/news/2024/10/24/teenage-boy-killed-himself-fall-love-ai-chatbot/

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535

U.S. Department of Education, Office of Educational Technology. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. U.S. Department of Education.

Vorobyeva, K. I., Belous, S., Savchenko, N. V., Smirnova, L. M., Nikitina, S. A., & Zhdanov, S. P. (2025). Personalized learning through AI: Pedagogical approaches and critical insights. Contemporary Educational Technology, 17(2), Article ep574. https://doi.org/10.30935/cedtech/16108

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. British Journal of Educational Technology, 55(1), 10–33. https://doi.org/10.1111/bjet.13334

Wu, X.-Y., Radloff, J. D., Yeter, I. H., Wang, L., & Chiu, T. K. F. (2025). Designing artificial intelligence chatbots for self-regulated learning from a systematic review based on Habermas's three interests. Interactive Learning Environments. https://doi.org/10.1080/10494820.2025.2563086

Xu, J., Luo, Y., Wang, C., Wang, M., & Wu, Y. (2026). AI support in self-regulated learning: A decade of technological evolution and meta-analysis. British Journal of Educational Technology, 1–30. https://doi.org/10.1111/bjet.70058

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11, Article 28. https://doi.org/10.1186/s40561-024-00316-7

Zhu, Q., & Luo, J. (2024). Toward artificial empathy for human-centered design. Journal of Mechanical Design, 146(6), 061401. https://doi.org/10.1115/1.4064161