Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n.3 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Beyond the Checklist: Designing Emotional Engagement in Conversational Chatbots for Medical Education, between Dyadic and Triadic Interaction

Valeria De Padova [1], Gianluca Amatori [2], and Anna Dipace [1]

[1] Pegaso University; valeria.depadova@unipegaso.it; anna.dipace@unipegaso.it
[2] European University of Rome; gianluca.amatori@unier.it
***** Correspondence: valeria.depadova@unipegaso.it

Abstract: The integration of artificial intelligence (AI) into medical training opens up extraordinary opportunities to practice communication and interpersonal skills in simulated environments with virtual patients. However, excessive reliance on technology carries the risk of reducing clinical empathy to a purely procedural task, thereby compromising the authenticity of the care provided by doctors. This paper addresses this tension and examines pedagogical and design criteria for the development of chatbots based on Large Language Models (LLMs). The aim is to foster genuine cognitive, emotional and relational engagement amongst medical students. Methodologically, the study proposes a theoretical and conceptual synthesis based on a critical re-analysis of two complementary scoping reviews: The first analysis identifies design requirements for dyadic interaction (doctor and adult patient), while the second extends these parameters to the pediatric context, where the presence of a caregiver gives rise to a more complex triadic dynamic. The study develops a theoretical framework rooted in the "Medical Humanities" and defines pedagogical indicators to avoid the reification of empathy. The results provide a conceptual reference architecture for the future implementation of conversation simulators that can support socio-affective learning and contribute to a more humanized clinical practice.

Keywords: Conversational Chatbots; Engagement; Medical Humanities.

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

This paper proposes a re-examination of the studies included in two scoping reviews, with the aim of identifying - based on aspects not highlighted in those reviews - the pedagogical requirements of a chatbot capable of fostering emotional and relational engagement. In recent years, Artificial Intelligence (AI) has permeated numerous professional fields, and the healthcare sector has not been immune to this development: today, it is largely shaped by generative technologies used to support diagnostic and treatment processes.

While AI increasingly supports medicine, reaching a greater number of patients, it is equally necessary to engage with them in a more humane manner (Paglialonga & Simonetti, 2025). It is precisely within this tension - between technological advancement and the need to humanize care - that the training of future doctors is

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

situated; they are called upon to develop, alongside clinical skills, communication skills that cut across the various branches of medicine.

For this reason, AI is increasingly being used as a tool to train communication, interpersonal and empathic skills in simulated contexts where the presence of real patients is no longer required. It is precisely here, therefore, that a risk emerges: clinical empathy may end up being reduced to a measurable outcome, a checklist item, a score, rather than being cultivated as a relational quality. An overly technology-driven learning environment thus risks diminishing, rather than supporting, students' relational and emotional engagement in the act of care.

This paper aims to explore this tension by examining which design criteria should guide the development of conversational chatbots based on Large Language Models (LLMs), so that they function not merely as communication exercises, but as environments capable of supporting cognitive, emotional and relational engagement in clinical communication training.

The study proceeds on two levels, corresponding to two complementary scoping reviews: the first maps AI tools in medical education, identifying design criteria for interaction with adult patients; the second, focused on the pediatric context, provides the parameters for extending the chatbot to the more complex triadic dimension.

This study does not present a fully developed prototype but outlines the theoretical framework and pedagogical indicators intended to serve as a reference for the subsequent development of a chatbot for clinical communication simulation.

2. Theoretical framework

This paper adopts engagement as a multidimensional construct - behavioral, cognitive and emotional (Fredricks et al., 2004) - as a theoretical lens through which to rethink the design of AI-based communication simulators. Behavioral engagement concerns the student's active participation in the simulated dialogue; cognitive engagement concerns the activation of clinical reasoning within the communicative exchange; emotional engagement concerns the student's ability to recognize and respond to the affective and narrative dimensions of the illness.

It is precisely this latter dimension that proves to be the most fragile in current AI-mediated learning environments and forms the core of this study.

This theoretical framework is reflected in the findings of a scoping review, conducted in accordance with the PCC model and the PRISMA-ScR guidelines (De Padova et al., 2026).

The review (65 studies included) mapped the artificial intelligence tools used in educational trials within the undergraduate medical field, classifying them according to the domains of competence on which they act. The findings reveal a predominance of constructivist pedagogical paradigms (60% of the studies), consistent with a conception of learning as the situated construction of knowledge, in which simulation offers students an authentic practical context, and a high level of accessibility of LLM-based tools (61.5%).

These two findings directly inform the design choice adopted: an open-scenario chatbot, rather than a closed-ended exercise, is both consistent with the dominant paradigm in the corpus and already technically feasible. In terms of competencies, communication skills emerge as one of the most frequently investigated domains,

confirming the growing centrality of the communicative dimension in medical education.

Two elements are particularly significant. Firstly, studies addressing communication skills almost exclusively simulate interaction with an adult patient: interaction with pediatric patients and the care relationship involving children are virtually absent, and only one study features a 17-year-old female patient in a dyadic consultation. Secondly, although numerous studies address themes that fall squarely within the scope of the Medical Humanities (the empathetic dimension, the building of trust, the relational quality of the clinical consultation), none of them make explicit reference to them in theoretical or terminological terms.

The Medical Humanities thus remain an implicit framework, evoked in the content but never explicitly addressed as a theoretical framework. It is precisely this gap that this paper aims to fill, by adopting the Medical Humanities as the theoretical framework for the design process.

The pediatric dimension is addressed in a second scoping review (De Padova et al., in press), with 62 included studies, which examines educational strategies for developing communication and empathy skills among medical students in the context of pediatrics. In this context, the care relationship takes on an inherently triadic structure, in which the child and the primary caregiver function equally as conversation partners and co-narrators of the experience of illness. From the perspective of the Medical Humanities, this triad is thus not only a communicative but also a narrative configuration: the illness is recounted from more than one perspective, and students are expected to listen to these narratives and integrate them into the process. The second review therefore provides the theoretical and empirical basis for expanding the chatbot's design from a dyadic to a triadic level.

Finally, one finding from the initial review proves decisive for the design and is taken up directly in the following section: only a small proportion of the studies create an emotionally responsive virtual patient, calibrated to specific emotional parameters, while several studies report limited emotional realism.

3. Method

This paper adopts a second-order theoretical and conceptual synthesis approach. The aim is to draw on the findings of two scoping reviews and to reinterpret the studies contained therein in a targeted manner in order to gain new insights.

Ultimately, the aim is to derive specifications for the design of a conversational chatbot based on LLMs (Large Language Models). To this end, the study is divided into three successive phases:

1. Analysis of the first corpus, based on 65 articles included in the first scoping review (De Padova et al., 2026). From these articles, only those experimental studies dealing with the enhancement of communicative skills were selected. To this end, 27 papers were re-examined, focusing on five key dimensions:
    - Design and structure of the patient chatbot;
    - Simulation and response at an emotional level;
    - Application of validated communication protocols;

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

- Type of feedback provided to students;
- Integration of the chatbot into the students' learning process.

2. Analysis of the second corpus, based on a re-examination of 62 studies from the second scoping review (De Padova et al., in press). This focuses on pedagogical approaches to developing empathic, communicative and relationship-based skills in trainee doctors in a pediatric context. In the scoping review, the experimental studies were grouped into five categories according to the type of interaction, one of which (the dyad) was divided into two subtypes (dyad with patient and dyad with caregiver). In this re-examination, the categories were defined with explicit criteria and the studies were reclassified, resulting in six categories:
    - Complete triad: The child (real - including those of pre-verbal age - or portrayed by an actor or a standardized patient) and the caregiver are present alongside the student in the same interactive situation, whether the student conducts, takes part in or observes the encounter live; in interventions involving multiple scenarios, at least one of these includes both participants;
    - Dyad with patient: The student communicates exclusively with the child or adolescent, without a parent being present;
    - Dyad with caregiver: The student communicates exclusively with the parent; the child is not present in the situation;
    - Nominal triad: Although a triad is named or assumed, the child is not a real person in the situation (e.g. a doll or a child who dies in the scenario);
    - Mixed model: Scenarios or activities with different constellations, in none of which, however, are both the child and the caregiver present together;
    - No direct interaction: no patients or family members, whether real or simulated, are present; communication takes place via videos or other recorded material, testimonials or written case studies, is assessed using scales, or is practiced solely with staff.

Only those individuals who were explicitly documented in the described activity were considered as participants. Finally, the distribution was analyzed according to publication period and the child's age.

3. Definition of the characteristics of the conversational chatbot based on the findings and gaps identified in the two previous reviews. In particular, following a re-examination of the first scoping review, it was decided to focus on characteristics relating to the artefact's lack of emotional responsiveness, the almost total absence of references to the theoretical framework of the Medical Humanities, the lack of attention to the pediatric context, the construction of cases based on the disease as a clinical fact, the lack of the patient's perspective in the feedback, the limited reliability of automated assessment of relational skills, and the use of the chatbot in isolated sessions. Rereading the second scoping review, on the other hand, allowed us to focus on the fragmented relationship between the specialist and the pediatric patient, supported by their family.

The findings from the analyses mentioned above have enabled us to identify the indicators that could lay the foundations for the development of a conversational chatbot.

4. From the review criteria to the chatbot specifications

The proposed indicators stem from a targeted re-examination of the included studies, focused no longer on systematic mapping but on identifying elements relevant to the design of the chatbot. This secondary analysis results in operational criteria across multiple levels of interaction.

The first level concerns the simulation of the interaction between a doctor and an adult patient. Upon closer examination of the studies addressing communication skills, it is observed that two internationally validated communication protocols were used in a limited number of studies (4 out of 27): the Calgary-Cambridge Guide (Kurtz & Silverman, 1996), which structures the consultation into sequential relational phases, and the SPIKES protocol (Baile et al., 2000) for communicating difficult news. Integrated into the system's conversational logic, these protocols require the student to simultaneously draw upon clinical reasoning and communication skills, rather than treating them as separate dimensions: in this way, the system operates along the axis of behavioral and cognitive engagement.

It is here that emotional calibration, as highlighted by this review, finds a practical application: translating it into a design criterion amounts to instructing the chatbot to recognise and assess the student's communicative cues, in order to modulate the virtual patient's emotional reactions during the interaction.

Emotional engagement thus ceases to be an outcome to be measured retrospectively and becomes an upstream design criterion: the chatbot is developed to support a relational quality of attention, consistent with what the Medical Humanities identify as constitutive of narrative competence.

Yamamoto et al. (2024) present an experiment based on the use of a simulated patient, to whom eight emotions had been assigned - joy, sadness, anticipation, surprise, fear, disgust, trust and anger - each with ten different levels of intensity, with the aim of providing a response based on what the student says.

Hryszko et al. (2026) present an agent whose emotions are calibrated through a series of instructions, while Yu et al. (2026) propose a model to which one of thirty-two personality profiles can be assigned. The results reported in the paper highlight the students' appreciation, as they judge the simulated patient to be more emotionally realistic than a standardized human patient.

Finally, Borg et al. (2024) link the model to a social robot, which selects facial expressions based on the patient's emotional state. At the same time, some studies identified in the scoping review that focus on students' communicative and interpersonal skills present virtual patients with a fixed emotional state specified in the instructions, which does not evolve during the consultation.

A key requirement therefore emerges that must be borne in mind when designing and building a new conversational chatbot: namely, the need for students to receive responses that evolve, from an emotional and affective perspective, during the conversation. Another factor to consider concerns the feedback that the conversational system provides to the student.

In some experiments, feedback on social and emotional skills is absent, as the chatbot focuses on feedback relating to clinical reasoning (Brügge et al., 2024) and

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

the completeness of the medical history (Holderried et al., 2024a). Further studies, however, focus on providing feedback on the quality of the relationship through the use of the Calgary-Cambridge model (Guo et al., 2025) and the SPIKES protocol (Comulada et al., 2025).

On reviewing the focus, the feedback provided by a ChatGPT-based conversational chatbot is particularly interesting; the chatbot was asked to evaluate the same consultation three times: the first evaluation was carried out by the chatbot without using any assessment grid, the second was carried out by the chatbot using a detailed grid provided by the researchers, and the third by the lecturer (Chiu et al., 2025). The evaluations yielded varying scores, highlighting the need to anchor the assessment of the virtual agent to protocols relating to the relational dimension and to supplement this with feedback from the tutor.

Another significant finding to emerge from the analysis concerns the patient's perspective. Indeed, data from the second scoping review reveal a key finding in this regard, as in traditional training the views of the patient and their family form part of the feedback (Frost et al., 2015; Persson et al., 2013). The data from the first scoping review, however, provide limited feedback on this aspect. Specifically, only one conversational agent provides the student with feedback from the patient's perspective, using the voice function of ChatGPT-4o (Mukadam et al., 2025). This significant finding offers further food for thought, suggesting that it would be useful to incorporate the patient's feedback at the end of the conversation. The chatbot could incorporate this dimension by providing the patient's impressions at the end of the conversation.

Another point examined concerns the individuals involved in the development of clinical cases. In many cases, the system is developed under the guidance of specialists, including doctors and educational experts (Yamamoto et al., 2024; Peralta Ramírez et al., 2025; Liu et al., 2025; Holderried et al., 2024b), while, at times, it is the virtual agent itself that formulates the stories to be told (Chiu et al., 2025; McCarrick et al., 2025). There are no studies that utilize the stories of patients or their families, thereby excluding narrative medicine - and, in a broader sense, the contribution of the Medical Humanities - from the design process. The final finding to emerge from the re-examination of the corpus concerns the educational context in which the chatbot is used. In most cases, the chatbot is used in a single session or on a voluntary basis, sometimes followed by discussions with the tutor and small-group discussions.

At the same time, some studies incorporate the use of the chatbot into structured programs, allowing students to practice their skills periodically on the platform (Wang et al., 2025) or making the tool available for a period for unlimited practical exercises (Yamamoto et al., 2024).

It is therefore clear that it is important to integrate the tool into a well-structured and pedagogically designed program, which could involve repeated practical sessions over time with increasing complexity, followed by debriefing sessions with teachers.

5. The triadic level: the fragmentation of the pediatric triad

The second level concerns the most complex form of interaction: the triadic relationship. In pediatric contexts, the care relationship takes on an intrinsically triadic structure, involving the doctor, the child and the caregiver simultaneously. In

this case, the child is both the recipient of care and an interlocutor with evolving communication skills, while the caregivers acts as a mediator, a co-narrator of the symptom and, in turn, the recipient of relational attention.

For the chatbot, this entails a specific design requirement: the system must not simulate a single interlocutor but must manage the co-presence of multiple voices and the layered experiences that characterize the doctor-child-caregiver relationship.

The criteria for this level are drawn from the second scoping review (De Padova et al., in press), which includes 62 studies on educational strategies for communication and empathy in pediatrics.

The review confirms that the triadic configuration is rarely implemented as such in practice: 19 studies (30.6%) involve a full triad; 11 (17.7%) reduce the encounter to a dyad with the caregiver and 10 (16.1%) to a dyad with the patient; in 5 (8.1%) the triad is merely nominal, as the child is a manikin or dies during the scenario; 9 (14.5%) do not involve direct interaction; and 8 (12.9%) follow a mixed model.

This provides empirical support for the design requirement outlined above: a chatbot intended to simulate the pediatric triad must be capable of managing the co-presence of two distinct interlocutors, with asymmetrical communication skills, rather than a single, simplified voice.

This re-examination led the research towards further reflection. The articles included in the scoping review show a distribution of arguments that varies over time.

In studies published up to 2015, the full triad is frequent (7 out of 15) both in real clinical settings, where the student observes or takes part in the encounter with the child and the family, and in simulations with standardized patients, where the student conducts the interview.

In the subsequent years, specifically between 2016 and 2023, the communicative triad does not appear with the same frequency (4 out of 28). Specifically, the interaction is fragmented into modules with separate scenarios, such as consultations with a single simulated parent, encounters with standardized patients and peer role-plays. One activity that differs from the rest of the corpus concerns the Teddy Bear Hospital, in which children aged 4–8 bring their soft toys to the doctor for a check-up. In this case, the student conducts a consultation with the children, who take on the role of the soft toy's parent (Kis et al., 2022; Thng et al., 2022).

From 2022, a triad defined in this paper as nominal began to appear in trials, in which the child is not physically present but is replaced by teaching aids, such as the SimBaby infant (Tian et al., 2022), or is present only in an idealized sense, as the child dies during the scenario and the student finds themselves communicating solely with the parent (Tervajärvi et al., 2024; Brandt et al., 2026).

Between 2024 and 2026, the full triad once again became more common (8 out of 19), with children, adolescents and trained adults acting as standardized patients (McCall et al., 2024; Gao et al., 2026). Studies also highlight activities taking place on wards and in the community (Ba et al., 2025; Schwarz et al., 2025).

These data do not allow for the identification of a single cause but suggest some possible hypotheses. The data reveal the complexity universities face in involving real children during lessons. Specifically, involving children in the simulation of clinical consultations is, in some cases, challenging, as child labor regulations prohibit the employment of minors (Frost et al., 2015) or the use of

children as standardized patients is the subject of a feasibility study (McCall et al., 2024). Furthermore, some scenarios simulate conversations in which the pediatric patient is not an active participant and cannot take part, such as in situations involving the delivery of difficult news, palliative care or the death of a child.

It could be argued that the use of manikins also serves sustainability purposes, as the investment in a simulator pays for itself over time, unlike the recurring cost of actors. The full triad, when implemented, therefore requires solutions that are highly resource-intensive.

The child's age also influences the nature of the interaction, in situations where the child is present during the interaction. During adolescence, the consultation often includes a period without the parent present. One example is a consultation in which the student meets a 16-year-old patient without the parent present (Woods et al., 2014), while a different scenario involves a consultation in which the student first interacts with the adolescent and her mother and then asks the mother to step out so that the student can continue alone with the girl (Gao et al., 2026).

These scenarios simulate real-life contexts, but they relate only to sensitive issues where adolescents might request that the parent leave, or situations where the doctor asks the parent or caregiver to step aside to allow the adolescent to open up during the consultation.

In real-life scenarios, however, situations may arise in which the patient has to receive difficult news and the doctor finds themselves simultaneously managing the emotional and affective reactions of the parent present within the communicative triangle. The system should be capable of simulating highly complex scenarios, in which the trainee doctor can experience the communicative triad even with adolescent patients.

Continuing with the data analysis, at school age the triad is most often realized with children trained as patients (de Araújo Baptista et al., 2024; McCall et al., 2024), while in early childhood the child is a non-verbal infant present alongside their mother (Abedi et al., 2025; Schwarz et al., 2025) or is replaced by a manikin (Tian et al., 2022).

This gives rise to a fundamental requirement for the chatbot: the need to adapt the child's role and the structure of the triad to the child's age, distinguishing - hypothetically and subject to validation - age groups corresponding to early childhood (0-3 years), pre-school age (4-7), school age (8-12) and adolescence (13-17). The second review also maps the outcome measures used in the included studies: validated relational instruments (the Calgary-Cambridge Guide, the Jefferson Scale of Empathy, and the CARE) feature in the corpus alongside numerous unvalidated ad hoc questionnaires.

Of particular relevance to the triadic level is the PCAT (Pediatric Consultation Assessment Tool), the only instrument in the corpus specifically designed for pediatric consultations: it explicitly distinguishes items relating to interaction with the child from those relating to interaction with the parent, offering a direct model for articulating, at the chatbot's evaluation level, the two communication channels required by the triadic simulation.

6. Conclusions

A re-examination of the two corpora shows that AI-mediated clinical communication tends to simplify precisely those aspects that the Medical Humanities consider to be fundamental to the care relationship. At the dyadic level, the virtual patient is rarely designed to react emotionally; it is constructed based on the illness as a clinical fact rather than the patient's experience, and provides content-focused feedback, while the patient's perspective almost disappears. At the triadic level, the relationship between doctor, child and caregiver is fully realized in fewer than a third of the studies and breaks down into dyads, separate scenarios or merely nominal triads.

These gaps give rise to the pedagogical requirements that the artefact should fulfil. In terms of emotional engagement, the virtual patient should not be limited to a static state of mind but should react in a manner consistent with the case and with the student's communicative choices, so that attention to the other produces visible effects during the consultation. In line with the Medical Humanities perspective, cases should be constructed not only based on the illness as a clinical fact, but also on the narratives of patients and their families, to convey the experience of the illness beyond its biomedical framework. On a cognitive and behavioral level, reference to validated protocols such as the Calgary-Cambridge Guide and SPIKES could guide the structure of the interview, while feedback should combine content assessment with a focus on the quality of the relationship and a summary formulated from the patient's perspective, capable of showing the student how the encounter was experienced. Finally, the chatbot should not be a stand-alone tool, but rather part of a program that includes theoretical preparation, repeated practice with increasing complexity, and a debriefing with the tutor, who remains responsible for assessing the relational aspect.

In the pediatric context, these requirements are supplemented by the need to simulate two distinct interlocutors - the child and the caregiver - each with their own voice, personal experience and distinct communication skills. The child's role should be adapted to their age: from the non-verbal presence of early childhood - which the student learns to observe and incorporate into the interaction - to the demand for autonomy and privacy in adolescence, where the student must be able to negotiate the space of the consultation with the parent.

The effectiveness of a chatbot of this kind cannot be measured solely by the formal correctness of the communicative sequences. The scales already identified in the two reviews, particularly in the second (Jefferson Scale of Empathy, Calgary-Cambridge, CARE, PCAT), form the starting point for an evaluation framework capable of integrating measures of this kind with indicators of the relational quality of the interaction, and represent one of the future developments of this work. What this paper aims to establish, on a theoretical level, is a fundamental premise: emotional engagement is not a factor that is simply added to technical performance, but a dimension that permeates and defines the entire learning process.

From an educational perspective, the review suggests that the full triad is achieved primarily through resource-intensive solutions, such as trained children, actors and on-ward supervision, while less costly formats tend to fragment it. An LLM-based chatbot could combine low costs with the simultaneous presence of multiple interlocutors: In one of the studies examined, an AI-based simulator costs around half as much as a simulation involving actors - although students still prefer

the latter (Tyrrell et al., 2025). The chatbot should therefore be seen as a complement to simulations involving real people, not as a replacement for them. The tool would also enable students to practice their communication skills in different situations and with different patients - ranging from adults to children accompanied by their caregivers - moving beyond sector-specific training towards a unified approach to communication skills.

From this perspective, the work fits seamlessly into the discourse on pedagogically grounded, engaging education: the design of a chatbot that fosters relational engagement means understanding the learning environment not merely as a vehicle for scenarios to be worked through, but as a space in which cognitive, emotional and relational engagement becomes the very prerequisite for the learning process. The key contribution of this approach thus lies in a repositioning of the Medical Humanities: they are not merely supplementary content for technical training, but form a theoretical framework that structures the design of the artefact itself and informs the rules governing its conversational behavior. As a theoretical and methodological frame of reference, this approach serves as a guideline for the further development phases of the chatbot.

Finally, it should be noted that this study has certain limitations. It constitutes a secondary analysis of studies that were already included in two reviews and were evaluated on the basis of the published texts; the classification of the triadic configuration required an interpretative assessment, and the analysis by time periods is based on a limited data set, which can serve merely as an indicator of trends. Future steps include the creation of a prototype, the validation of the proposed age groups, the involvement of patients and their relatives in the development of the case studies, and testing with students - with the results to be assessed using validated instruments and through the evaluation of teaching staff.

References

Abedi, A. J., Jain, A., Zubair, M. Y., Mehnaz, S., & Sujir, N. R. (2025). Community-based interprofessional training for undergraduate medical students through family adoption programme centering on infant and young child feeding. *Indian Journal of Community Health, 37*(5), 787–794. https://doi.org/10.47203/IJCH.2025.v37i05.024

Ba, H., Xu, L., Gu, Y., Li, Y., Jiang, X., Li, X., & Li, S. (2025). Comparative study of problem-based learning and traditional teaching methods on medical students' outcomes in pediatrics clerkships. *Advances in medical education and practice*, 615-624. https://doi.org/10.2147/AMEP.S515527

Baile, W. F., Buckman, R., Lenzi, R., Glober, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES—a six-step protocol for delivering bad news: application to the patient with cancer. *The oncologist*, *5*(4), 302-311. https://doi.org/10.1634/theoncologist.5-4-302

Borg, A., Jobs, B., Huss, V., Gentline, C., Espinosa, F., Ruiz, M., ... & Parodis, I. (2024). Enhancing clinical reasoning skills for medical students: a qualitative comparison of LLM-powered social robotic versus computer-based virtual patients within rheumatology. *Rheumatology International*, *44*(12), 3041-3051. https://doi.org/10.1007/s00296-024-05731-0

Brandt, J. B., Bibl, K., Grafeneder, J., Masel, E. K., Bergsträsser, E., Aburn, G., ... & Peyrl, A. (2026). The SimCom-Method: Impact of Simulated Communication Teaching in Pediatric Resuscitation Scenarios. *Academic Pediatrics*, 103204. https://doi.org/10.1016/j.acap.2025.103204

Brügge, E., Ricchizzi, S., Arenbeck, M., Keller, M. N., Schur, L., Stummer, W., ... & Darici, D. (2024). Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial. *BMC medical education*, *24*(1), 1391. https://doi.org/10.1186/s12909-024-06399-7

Chiu, J., Castro, B., Ballard, I., Nelson, K., Zarutskie, P., Olaiya, O. K., ... & Zhao, Y. (2025). Exploration of the role of ChatGPT in teaching communication skills for medical students: A pilot study. *Medical Science Educator*, *35*(4), 1871-1882. https://doi.org/10.1007/s40670-025-02394-9

Comulada, W. S., Ganz, P. A., Huang, Y. M., Weisman, D., Sugarman, A., Lee, D., ... & Sugarman, A. W. (2025). A pilot test of an AI voice-driven simulation with feedback for medical students to practice discussing diagnostic mammogram results With patients. *Cureus*, *17*(10). https:/doi.org/10.7759/cureus.95606

de Araújo Baptista, V. I., Braga, L. P., de Sousa Mata, Á. N., Carreiro, B. O., dos Santos Rosa, L. P. G., de Morais, H. H. A., ... & Appenzeller, S. (2024). Validation of clinical simulation scenarios for the teaching of soft skills in child-centered care. *BMC Medical Education*, *24*(1), 355. https://doi.org/10.1186/s12909-024-05284-7

De Padova, V., Amatori, G., & Dipace, A. (2026). L'uso dell'Intelligenza Artificiale nella formazione medica pre-laurea: una Scoping Review. *MEDICAL HUMANITIES & MEDICINA NARRATIVA*, *12*(1), 243-260. https://doi.org/10.53136/979122182756917

De Padova, V., Amatori, G., & Dipace, A. (in press). Communication and Empathy Skills Training in Pediatric Settings in Undergraduate Medical Education: A Scoping Review, Paper presented at the 8th International Conference on Higher Education Learning Methodologies and Technologies Online (HELMeTO 2026), Cagliari, Italy, September 28-30, 2026.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109. https://doi.org/10.3102/00346543074001059

Frost, K. A., Metcalf, E. P., Brooks, R., Kinnersley, P., Greenwood, S. R., & Powell, C. V. (2015). Teaching pediatric communication skills to medical students. *Advances in medical education and practice*, 35-43. https://doi.org/10.2147/AMEP.S68413

Gao, C., Plansky, C., Ferguson, N., Caputo-Seidler, J., Weiss, A., & Kiluk, V. (2026). Adolescent-Parent Dyad and Adolescent Eating Disorders: A Standardized Patient Simulation. *Cureus*, *18*(1). https://doi.org/10.7759/cureus.101237

Guo, E., Ramchandani, R., Park, Y. J., & Gupta, M. (2025). OSCEai: personalized interactive learning for undergraduate medical education. *Canadian Medical Education Journal*, *16*(6), 7-14. https://doi.org/10.36834/cmej.79220

Holderried, F., Stegemann-Philipps, C., Herrmann-Werner, A., Festl-Wietek, T., Holderried, M., Eickhoff, C., & Mahling, M. (2024a). A language model–powered simulated patient with automated feedback for history taking: Prospective study. *JMIR Medical Education, 10*, e59213. https://doi.org/10.2196/59213

Holderried, F., Stegemann-Philipps, C., Herschbach, L., Moldt, J.-A., Nevins, A., Griewatz, J., Holderried, M., Herrmann-Werner, A., Festl-Wietek, T., & Mahling, M. (2024b). A generative pretrained transformer (GPT)–powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study. *JMIR Medical Education, 10*, e53961. https://doi.org/10.2196/53961

Hryszko, J., Michałek, A., & Roman, A. (2026). Application of large language models in medical interview training: a study with medical students. *BMC Medical Education*, *26*(1), 586. https://doi.org/10.1186/s12909-026-08819-2

Kis, H., Endres, K., Karwowska, A., Harrison, M., Lau, S., Lemire, O., & Zucker, M. (2022). "Teddy Bear Hospital Project" school visits improve pre-clerkship students' comfort explaining medical concepts to children. *Canadian Medical Education Journal*, *13*(3), 70-74. https://doi.org/10.36834/cmej.73167

Kurtz, S. M., & Silverman, J. D. (1996). The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Medical Education*, *30*(2), 83–89. https://doi.org/10.1111/j.1365-2923.1996.tb00724.x

Liu, Y., Shi, C., Wu, L., Lin, X., Chen, X., Zhu, Y., ... & Zhang, W. (2025). Development and validation of a large language model-based system for medical history-taking training: prospective multicase study on evaluation stability, human-AI consistency, and transparency. *JMIR Medical Education*, *11*(1), e73419. https://doi.org/10.2196/73419

McCall, N., Bekele, A., Ewnte, B., Abaho, J., & Ibrahim, O. (2024). Child standardized patients in pediatric OSCEs: a feasibility study for otoscope examination among undergraduate students in Rwanda. *BMC Medical Education*, *24*(1), 1546. https://doi.org/10.1186/s12909-024-06587-5

McCarrick, C. A., McEntee, P. D., Boland, P. A., Donnelly, S., O'Meara, Y., Heneghan, H., & Cahill, R. A. (2025). A randomized controlled trial of a deep language learning model-based simulation tool for undergraduate medical students in surgery. *Journal of Surgical Education*, *82*(9), 103629. https://doi.org/10.1016/j.jsurg.2025.103629

Mukadam, A., Suresh, S., & Jacobs, C. (2025). Beyond traditional simulation: an exploratory study on the effectiveness and acceptability of ChatGPT-4o advanced voice mode for communication skills practice among medical students. *Cureus*, *17*(5). https://doi.org/10.7759/cureus.84381

Paglialonga, M., & Simonetti, C. (2025). Beyond the algorithm: the importance of medical humanities in the age of AI. *Monash Bioethics Review*, 1-13. https://doi.org/10.1007/s40592-025-00259-w

Peralta Ramírez, A. A., Trujillo López, S., Navarro Armendariz, G. A., De la Torre Othón, S. A., Sierra Cervantes, M. R., & Medina Aguirre, J. A. (2025). Clinical simulation with ChatGPT: a revolution in medical education?. *Journal of CME*, *14*(1), 2525615. https://doi.org/10.1080/28338073.2025.2525615

Persson, E., Haines, C., & Lang, M. (2013). Parent assessment of medical student's skills in ambulatory pediatrics. *Canadian Medical Education Journal*, *4*(2), e18.

Schwarz, H. L., Straub, C., Bode, S. F., Ferschl, N., Brickmann, C., Berberat, P. O., & Krüger, M. (2025). Learning and working on an interprofessional training ward in neonatology improves interprofessional competencies. *Frontiers in medicine*, *12*, 1483551. https://doi.org/10.3389/fmed.2025.1483551

Tervajärvi, L., Tervajärvi, S., Storvik-Sydänmaa, S., & Hutri, N. (2024). Using Simulation Exercises to Deal With the Death of a Child as Part of Healthcare Studies. *Cureus*, *16*(1). https://doi.org/10.7759/cureus.53284

Thng, J. Z. H., Tan, F. Y. Y., Aw, M. M. H. Y., & Hu, S. (2022). Combination of teddy bear hospital and virtual reality training increases empathy of medical students. *The Asia Pacific Scholar*, *7*(3), 33. https://doi.org/10.29060/TAPS.2022-7-3/OA2739

Tian, J. D., Wu, F. F., & Wen, C. (2022). Effects of a teaching mode combining SimBaby with standardized patients on medical students' attitudes toward communication skills. *BMC Medical Education*, *22*(1), 825. https://doi.org/10.1186/s12909-022-03869-8

Tyrrell, E. G., Sandhu, S. K., Berry, K., Ghannam, S. F., Lewis, S. A., Crowfoot, D., Sahota, G. S., Carson, J., Wilson, E. E., & Taggar, J. (2025). Web-based AI-driven virtual patient simulator versus actor-based simulation for teaching consultation skills: Multicenter randomized crossover study. *JMIR Formative Research, 9*, e71667. https://doi.org/10.2196/71667

Wang, Z., Fan, T.-T., Li, M.-L., Zhu, N.-J., & Wang, X.-C. (2025). Feasibility study of using GPT for history-taking training in medical education: A randomized clinical trial. *BMC Medical Education, 25*, 1030. https://doi.org/10.1186/s12909-025-07614-9

Woods, J. L., Pasold, T. L., Boateng, B. A., & Hensel, D. J. (2014). Medical student self-efficacy, knowledge and communication in adolescent medicine. *International Journal of Medical Education*, *5*, 165. https://doi.org/10.5116/ijme.53d3.7b30

Yamamoto, A., Koda, M., Ogawa, H., Miyoshi, T., Maeda, Y., Otsuka, F., & Ino, H. (2024). Enhancing medical interview skills through AI-simulated patient interactions: nonrandomized controlled trial. *JMIR medical education*, *10*(1), e58753. https://doi.org/10.2196/58753

Yu, H., Zhou, J., Li, L., Chen, S., Gallifant, J., Shi, A., ... & Fan, L. (2026). Simulated patient systems powered by large language model-based AI agents offer potential for transforming medical education. *Communications Medicine*, *6*(1), 27 https://doi.org/10.1038/s43856-025-01283-x