

Article

Artificial Intelligence-Based Simulation Training in Midwifery Education: A Descriptive Cross-Sectional Study on Chatbot-Supported Medical History Taking

Marie Therese Ettlen ^{1,*}, Ulrike Keim ^{1,2} and Claudia F. Plappert ^{1,*} 

¹ Institute for Health Sciences, Section of Midwifery Science, University Hospital Tuebingen, 72076 Tuebingen, Germany

² Research Institute for Women's Health, Department of Women's Health, University Hospital Tuebingen, 72076 Tuebingen, Germany

* Correspondence: marie-therese.ettlen@uni-tuebingen.de (M.T.E.); claudia.plappert@med.uni-tuebingen.de (C.F.P.)

Abstract

(1) Background: Midwifery students require practical training experience to learn how to perform a medical history. Simulation-based training, such as chatbot exercises using large language models like GPT, provide structured practice but require ongoing evaluation. This study explores German midwifery students' views on using an AI chatbot simulating a pregnant woman, regarding usability, realism, and educational value. (2) Methods: Twenty-six students participated in a descriptive, quantitative cross-sectional survey, using a literature-based, self-developed questionnaire after interacting with the AI generative chatbot. Data were analyzed via SPSS 30.0, with results shown in a stacked horizontal bar chart. (3) Results: The findings indicate that students experienced no difficulties when interacting with the chatbot. Both the quality and realism of the conversations were evaluated positively. Chatbot training was perceived as helpful in supporting structured medical history interviews and the collection of relevant data but was not considered a substitute for practice with actors or real-life situations. (4) Conclusions: The findings suggest that the medical history chatbot offers midwifery students an innovative, flexible simulation for training. Students responded positively, and it may help develop structured history-taking skills. Further study is needed to determine if repeated chatbot use improves medical history collection skills.

Keywords: midwifery students; chatbot; medical history



Academic Editors: Douglas McHugh, Anthony Payne and Branko Aleksic

Received: 25 January 2026

Revised: 2 March 2026

Accepted: 5 March 2026

Published: 10 March 2026

Copyright: © 2026 by the authors.

Published by MDPI on behalf of the Academic Society for International Medical Education. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Background

Midwives are internationally recognized as healthcare professionals specializing in the care of women during pregnancy, childbirth and the postpartum period [1]. In professional midwifery, woman-centered care focuses on making the woman an active and responsible participant in all decisions during pregnancy, childbirth, and after delivery [1,2]. Building a trusting relationship between the midwife and the pregnant woman is therefore essential for effective care [3]. A detailed medical history forms the basis for needs-oriented, woman-centered midwifery care [4]. Taking a medical history is one of the first practical skills that midwifery students learn during their training. To take an effective medical history, students must learn to listen actively and ask precise and relevant questions with the

aim of obtaining and documenting necessary social and health details from the pregnant person [4].

In general, the medical history is taken during the first meeting between the midwife and the pregnant woman. In addition to documenting essential data, the medical history can lay the foundation for the relationship between the midwife and the pregnant woman. The midwife should encourage open dialogue and address all the pregnant woman's concerns. This aspect helps to ensure individualized and needs-based care for the pregnant woman [4,5].

Taking a medical history should therefore be systematically incorporated into early-stage training. Midwifery students acquire skills in obtaining medical histories during their clinical placements. Practical training presents significant challenges, as it must be tailored to the formal requirements of each practice site. In German clinical contexts, this process is typically referred to as "registration for birth" and is often subject to strict time constraints. Students lack opportunities to practice taking medical histories in their entirety without time limits [4,5].

1.2. Chatbot Training in Health Education

Simulation-based exercises are a proven teaching model in health sciences [6]. This approach embodies a practice-oriented method of learning that enables students to systematically develop their clinical skills. Skills may be practiced, refined, and repeated within a controlled setting, ensuring patient safety is maintained, and unnecessary stress is avoided [6,7]. Ton et al. describe simulation-based learning as a dynamic and immersive learning experience that allows students to develop their clinical skills and competencies [8]. Midwifery students at the University of Tübingen are already familiar with simulation-based role-play scenarios. Exercises, mid-term examinations, and final assessments are conducted within this environment. Simulation-based learning approaches have received positive evaluations from midwifery students on an international scale [9]. The integration of artificial intelligence, specifically in the form of chatbots, presents a novel opportunity for training. Evidence for their use in midwifery education, especially for training in the medical history, is limited.

A chatbot is a dialogue system capable of producing natural language through textual or auditory means [10]. The chatbot under consideration utilizes the ChatGPT-4 platform developed by OpenAI. ChatGPT is a language model that employs advanced machine learning techniques [10]. This model develops its communication capabilities using data sourced from forums, texts, books, and spoken language, and it is further refined through human feedback. Research and implementation of chatbot technologies in healthcare have become increasingly common [11,12]. In the field of medical education, chatbots are currently being used to help train students in taking medical histories [13–15].

Compared to traditional role play, practicing with a chatbot offers distinct advantages: it is not reliant on individual participants, thereby reducing the time, organizational effort, and expense required [13]. Once created, a chatbot can be accessed repeatedly by students for ongoing practice [12–14]. Additionally, one of AI's notable strengths is its capacity to generate numerous clinical cases. This functionality allows AI to support training across various healthcare professions, broadening its scope of application [14].

In 2024, TIME (Tübingen Institute of Medical Education) developed a chatbot for medical students based on the large language model (LLM) GPT [13]. An artificial patient generated by TIME served as the bot. Twenty-eight students tested the bot in taking a medical history. After review and evaluation by the researchers, the chatbot's responses were deemed plausible. The students also rated the chatbot's usefulness and user-friendliness as positive [13]. Previous research investigating AI chatbots as simulated patients has largely

focused on mixed groups of healthcare students, without presenting findings specific to midwifery students. To the best of our knowledge, this is the first study dedicated solely to midwifery students engaging with a chatbot to practice medical history-taking. It is therefore important to examine how midwifery students perceive the use of a chatbot for training in medical history-taking in terms of realism, user-friendliness, and the expected expansion of their skills. In the future, the chatbot will also be available to students at home, enabling them to practice independently. However, before it is implemented, the tool must be evaluated to determine whether students will find it helpful, as this will influence its subsequent use.

This survey seeks exploratory feedback on the chatbot's use and addresses the following research question: How do midwifery students at the University of Tübingen rate an AI-based chatbot for medical history exercises in terms of realism, user-friendliness and skills development?

2. Materials and Methods

At the beginning of 2025, 26 volunteer midwifery students from the University of Tübingen had the opportunity to test the chatbot in four seminar events. The students' assignment was to take and document a full medical history. To evaluate their initial perceptions of this training tool, a questionnaire grounded in the established literature was designed. Following their interaction with the chatbot, the students provided feedback on its perceived user interface, realism, and educational support through the self-developed questionnaire. This questionnaire was intended to evaluate a chatbot developed for bachelor's students in Midwifery Science at the University of Tübingen. This study employed a descriptive, exploratory methodology using quantitative data. This approach facilitates an initial evaluation of the AI-generated training tool, focusing on developing a thorough and systematic understanding of the dataset via exploratory data analysis (EDA), rather than testing overarching hypotheses. This study does not aim to demonstrate effectiveness; instead, it presents an initial appraisal by the participating students. This approach distinguishes the study from hypothesis-driven research designs. Anonymous questionnaires gathered student feedback for systematic evaluation. This approach ensures unbiased data and enables the analysis of potential correlations. This survey provides targeted and structured data relevant to the research question.

2.1. Prompting

In autumn 2024, a script was created for a chatbot dialogue featuring a fictional woman experiencing early pregnancy. A detailed backstory was crafted for this character based on the prompt provided. The character was created by midwives with clinical experience from the University of Tübingen. A chat interface was created to let students communicate via text. The chatbot has already been tested on medical students [13]. It should be noted that this case vignette constitutes a wholly distinct case, and it marks the preliminary implementation of midwifery students within this specific context. We used ChatGPT version 4 for the study.

Character and Setting Overview

To prompt the chatbot character, we collected comprehensive details regarding the pregnant individual's psychosocial, medical, and obstetric history, as well as their current health status. All information mandated by German maternity guidelines was considered. Supplementary data concerning employment conditions, leisure pursuits, hobbies, sports activities, nutrition, and present physical complaints were recorded as additional notes within the prompt.

The virtual pregnant woman created for midwifery students is a 27-year-old woman originally from Spain, living in Germany with her husband and three-year-old son. She is expecting her second child. She is excited about the new arrival but worries about her children's social isolation due to limited connections in Germany and is also challenged by tight living conditions. Her medical history includes diabetes mellitus treated with metformin. This is her third pregnancy. Five years ago, her first pregnancy ended in early abortion followed by curettage. Her second pregnancy resulted in a cesarean section at 34 weeks; she has since recovered.

In terms of the setting, the woman is in her seventh week of pregnancy. The pregnancy has already been confirmed by the gynecologist. She is now attending her initial appointment at the midwife's office, where an in-depth medical history is being collected. The character appears fatigued, reserved, and compliant in conversation. She often voices concerns about her living and social conditions and was designed to be anxious and ask frequent questions.

2.2. Creation of the Questionnaire

The questionnaire is based on the specialist literature and expert consensus within the research team. It was distributed to students in paper form. It includes the title, personal data, instructions, and blocks of 16 questions on specific topics. Questions cover previous experience with AI, evaluation of the user experience, quality and realism of the conversation, and usefulness for practice purposes. In addition, students were asked to assess whether they expect to gain competence in taking a medical history through training with the chatbot. Figure 1 shows the differentiated category system on which the questionnaire is based. Demographic data were collected with the intention of analyzing the responses in groups. To this end, the data collected was first grouped by semester and then by previous experience with the medical history. The collection of age and gender data was solely for the purpose of standard quality assurance and reporting of the university's evaluation data.

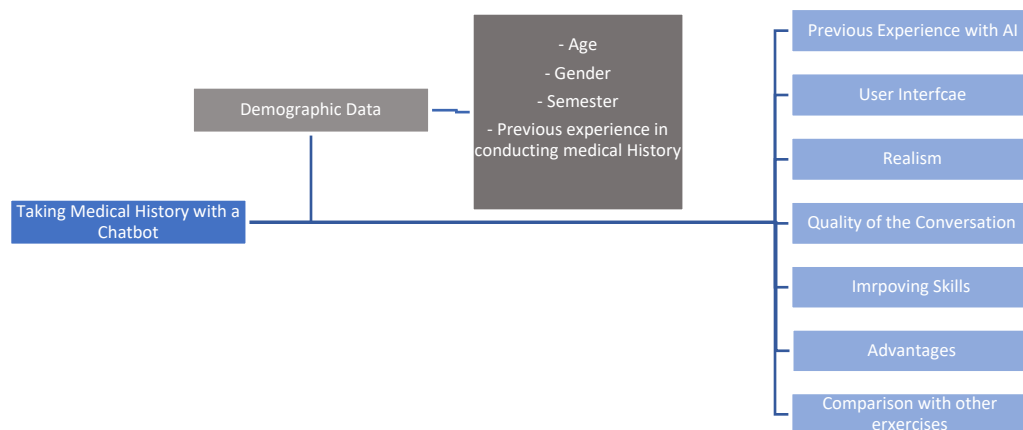


Figure 1. The differentiated categories. Gray: Demographic data; blue: variables.

The questionnaire was standardized by maintaining consistent questions, conditions, and answer options, allowing statistical analysis with SPSS 30.0 and enabling quantitative data collection.

2.3. Operationalization of the Variables

At the beginning of the operationalization process, the seven categories from Figure 1 were entered as variables in a table (Table 1). Suitable indicators were defined for each variable. The indicators represent the items examined, which were then formulated in

the questionnaire as statements that could either be agreed with or disagreed with. The abbreviation Team in the Source column indicates that the indicator originates from team discussions and not from the literature. The research team consisted of a scientist specializing in artificial intelligence and language processing of simulated agents, who was himself involved in creating the chatbot, a physician specializing in AI in medical education, a midwifery scientist with clinical experience, and a professor of midwifery sciences.

Table 1. Construction of variable.

Investigated Variable	Indicator	Sources
Previous experience with AI	AI use in everyday life Experience with chatbot	¹ Team
User experience	Rating item maternity log User interface Application Technical difficulties	Team [16]
Realism	Answering questions Conversation flow	[17]
Quality of the conversation	Realistic setting Conversation ‘human-like’	[17]
Improving skills	Complete collection Documentation Time constraints Maintaining the conversation	[1,18,19]
Advantages	Advantages of AI (accessibility...)	[14]
Comparison	Comparison with other training options	Team [18]

¹ Indicators labeled “Team” were derived from expert discussions within the interdisciplinary research group.

2.4. Detailed Concept of the Questionnaire and Differentiation of Statements

The questionnaire was intentionally structured to facilitate comprehension and efficiency, given the exploratory nature of this study and the absence of a validated instrument for this specific application. Only two question types were featured to minimize respondent reconsideration. Initial inquiries included age, semester, gender, and practical experience with medical history-taking. Responses to statements were collected using a five-point Likert scale, ranging from “I completely agree” to “I completely disagree,” which enables the measurement of specific characteristics across multiple items. The design supports prompt and straightforward data evaluation. The research team conducted two rounds of revisions to refine the questionnaire. A repetitive question related to realism was eliminated; the statements were revised for clarity and simplicity, and the questions were regrouped by theme. Additionally, a vague point and a query concerning conversation flow were clarified. One new question—about permanent access to the chatbot training tool—was added. Ultimately, the final version of the questionnaire comprised 16 carefully selected items. You can find the final questionnaire in the Supplementary Materials.

2.5. Ethical Considerations

This study received approval from the Ethics Committee of the Faculty of Medicine at University Hospital Tübingen (552/2024BO2). Data were collected anonymously and was not linked to individual students. All participants were informed about the intended use of their data for research purposes, and consent was obtained from every student

prior to participation. There were no risks identified other than those generally associated with routine classroom activities. Participation was strictly voluntary, with the option to withdraw at any point without negative consequences to grades or academic status. Data collection was conducted by an individual who is not affiliated with the teaching staff for midwifery students enrolled in the bachelor's program at the University of Tübingen.

2.6. Recruitment of Study Participants and Sample Size

Recruitment was conducted in late November 2024, with us inviting all third-, fifth-, and seventh-semester midwifery students at the University of Tübingen to participate on a voluntary basis; first-semester students were excluded due to limited experience. Invitations were disseminated via email and WhatsApp channels, and an online informational session was organized. Owing to initially low registration numbers and scheduling conflicts related to clinical placements and lectures, additional sessions were scheduled. In total, 30 students registered, with 26 ultimately participating; withdrawal or absence incurred no penalty. Study requirements were reviewed during the event, and all attendees provided informed consent to participate.

Twenty-six midwifery students from the University of Tübingen, aged 19–33, participated in this study over three days: seventeen from the third semester (65.4%), eight from the fifth (30.8%), and one from the seventh (3.8%). Twenty-five identified as female, and one as diverse. Sixteen had previously conducted an anamnesis independently (61.5%), while ten had not (38.5%).

2.7. Survey Procedure

The survey was conducted at TIME. Each participant was assigned to a computer equipped with an active chat interface. Prior to commencing the survey, all participants attended a fifteen-minute theoretical session focused on medical history-taking. Additionally, information regarding data protection protocols when utilizing AI was provided. Following a brief orientation to the chat platform, students initiated their chats simultaneously. Participants were allotted 40 min to perform medical history interviews. Faculty from midwifery science and members of the medical education research team supervised the session, with one member designated for technical support. Upon completion of the chat interaction, students were given 10 min to complete a questionnaire. Experts in midwifery science were available throughout to address any queries.

2.8. Data Preparation

The data preparation process is outlined for transparency. Data was collected using paper questionnaires, then manually entered into SPSS 30.0. Questionnaires were numbered 1–26 to correspond with the cohorts: 1–7 (first group, 13 January 2025), 8–12 (morning, 31 January 2025), 13–25 (noon, 31 January 2025), and 1 (final session, 5 February 2025). All questionnaires were anonymized, and the dataset was checked three times by the same individual.

2.9. Variable Formation

Initially, variables were established in SPSS. The variable age was categorized into four groups (<20 years, 20–25 years, 26–30 years, and >30 years). The statement 'I have already taken a medical history independently' had binary response options ('yes' or 'no'), which were coded as 1 and 0, respectively. Gender information was recorded using abbreviations to streamline data entry (m = male, f = female, d = diverse). Questions 1–16 corresponded to items 1–16. For coding purposes, responses were assigned numerical values: "strongly agree" = 5, "agree" = 4, "neither agree nor disagree" = 3, "disagree" = 2, and "strongly disagree" = 1. All 26 questionnaires were included in the analysis. For three

questions where respondents chose between two answer options, the higher value was used, which may have introduced a slight positive bias and is considered a methodological limitation. We opted for the higher value because the cross was between answer 3 and 4 for all three answers, suggesting a tendency towards agreement. One questionnaire did not indicate age; however, as this did not impact the evaluation, and an age range of 20–35 years was assumed for these participants. The data was fully analyzed. All response patterns appeared consistent and plausible. The dataset was saved in .sav format by SPSS, and graphical representations were subsequently generated using Excel.

3. Survey Results: Participants' Views on the Use of a Generative AI Chatbot as a Training Tool for Taking Medical Histories

3.1. Ungrouped Display of the Results

Fifty percent of students said they use AI in their everyday lives; slightly more than half had already interacted with a chatbot before. None of the students reported any difficulties in using the tool, and 80% found the maternity log questions helpful. The realism of the clinical case was rated as good by 27% of students and very good by 73%. In terms of the realism of the conversation, 80% rated it as good and 12% as very good. The quality of the conversation was assessed based on the accuracy of the responses: 50% of students rated it as good, with the remaining 50% rating it as very good. Only 8% of students had difficulty maintaining the conversation with the bot, while 85% of the students surveyed stated that they expected to improve their competence in conducting a situation-related medical history by practicing with the chatbot. All participants agreed that using the chatbot promotes the ability to record and document a complete medical history interview in a structured manner. The user-friendliness of the chatbot was rated positively by 85%. Furthermore, 42% of students appreciate the advantage of not having to face a real person, while 35% particularly emphasize this aspect. Compared to other training options, 35% prefer interacting with an actor and 35% with an actual pregnant person; 50% abstained from answering this question. Overall, 85% of respondents would still like to have the option of practicing the anamnesis interview independently and regardless of location with the help of a chatbot (Figure 2).

3.2. Group Comparison

The evaluation examined the response patterns exhibited by individual groups. The objective of the present study was to examine whether third-semester students (17 participants) differed from fifth-semester students (8 participants) in their responses. The sample size of the seventh semester group was too small for meaningful statistical analysis. Different response patterns emerged when students were asked whether they would prefer to practice with a real pregnant woman or a chatbot. Third-semester students were the only ones to disagree with this statement. Apart from this, no notable differences in response patterns were observed.

In addition, the data was separated into two groups: 'Previous experience—yes', and 'Previous experience—no'. Ten participants had previous experience, while sixteen did not. No discernible difference in response patterns was found between the two groups.

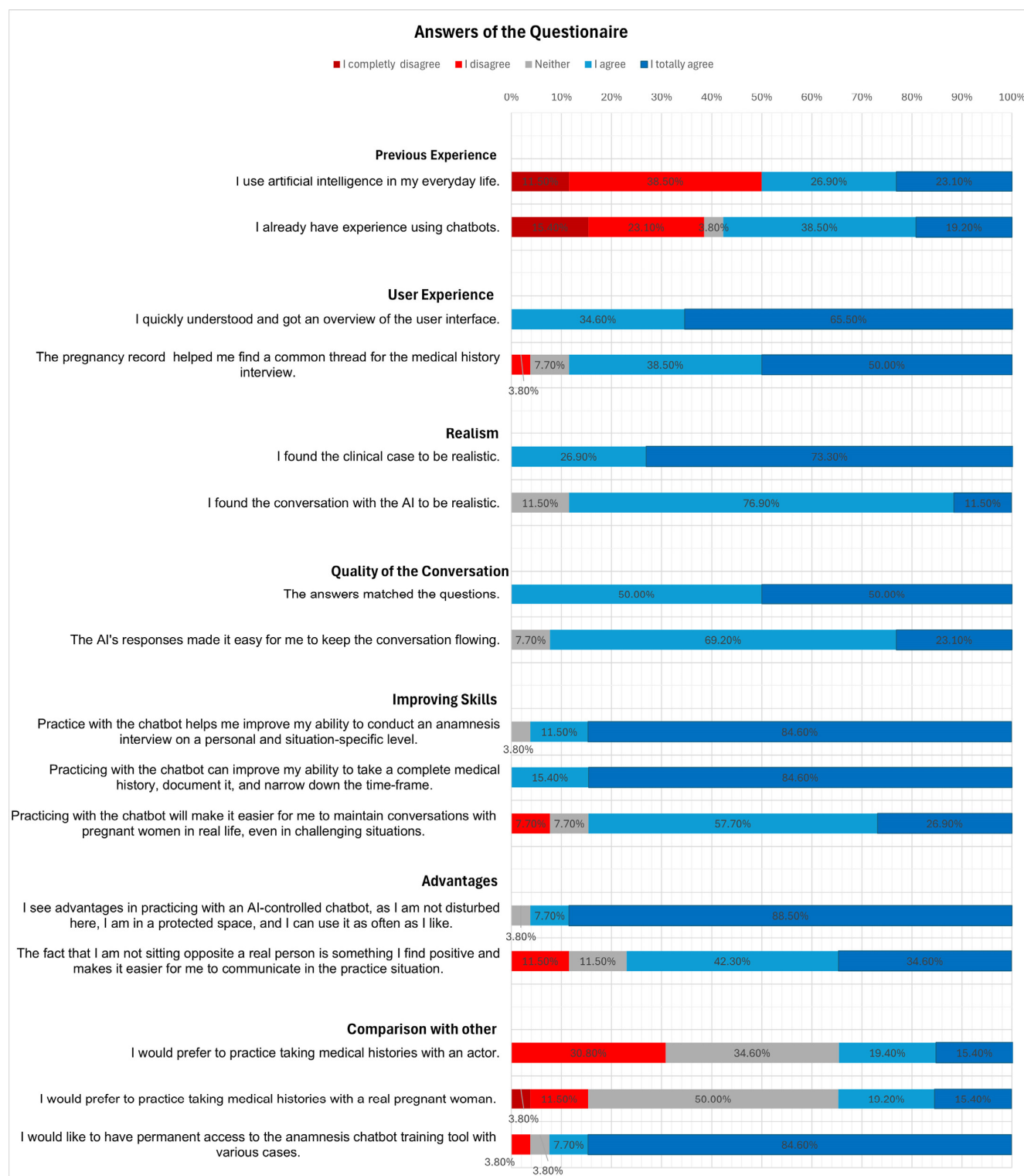


Figure 2. Survey results: Participants' views on the use of a generative AI chatbot as a training tool for taking medical histories.

4. Discussion

4.1. Summary of the Main Findings

A preliminary trial of the Anamneses Chatbot simulation tool was conducted with 26 midwifery students. Using a questionnaire, students respond to statements about their previous experience with AI, user-friendliness, realism, the quality of the conversation,

and the increase in competence. Finally, the students compare the training tool with other exercise options they are familiar with.

4.1.1. Previous Experience with AI

Half of the participants incorporate AI into their daily routines and have interacted with chatbots. However, an equal proportion does not use AI regularly and lacks prior experience with chatbot technologies. Consequently, it cannot be assumed that all students possess familiarity with AI in this context, understand its theoretical foundations, or demonstrate adequate competence in recognizing the associated risks and challenges of AI implementation.

4.1.2. User Experience

Student feedback demonstrated a high level of satisfaction with the tool. The interface was described as intuitive, allowing users, including those without previous experience with chatbots, to navigate and use it efficiently. Among the 26 respondents, 23 rated the maternity record tool as useful. This suggests that the tool assists students in structuring conversations and ensures comprehensive documentation of all relevant information.

4.1.3. Realism

The statements regarding realism were rated positively by the students. The students found the generated pregnant person and their manner of conversation to be realistic. The students' evaluations show that chatbot conversation resembles a real conversation despite not fully reflecting all the characteristics of real communication.

4.1.4. Quality of the Conversation

The students rated the quality of the conversation as good and found that the AI responses were mostly appropriate. The medical history was successfully taken, and important information was obtained. Only two people had difficulty maintaining the flow of conversation. Several factors may have contributed to this outcome. One potential reason is that the chatbot exhibited repetitive patterns in its responses and queries, which may have hindered deeper engagement from the students. Additionally, the scenario involving the AI-generated pregnant woman presented an individual experiencing significant emotional distress and anxiety regarding her future with the baby. The midwifery students might not have been able to effectively employ communication strategies to provide reassurance, address concerns, and redirect the conversation towards obtaining a comprehensive medical history.

4.1.5. Improving Skills

The clear majority of respondents indicated that practicing with the AI chatbot enhances their ability to conduct anamnesis interviews in a person- and situation-specific way. Students unanimously agreed that working with AI contributes to improved competence in collecting, documenting, and efficiently timing medical histories. Meanwhile, 8% were unsure about the chatbot's usefulness for complex conversations, and another 8% disagreed with its effectiveness. The responses suggest that the exercise supports students in structuring medical histories and reinforces their understanding of the appropriate categorization of personal, family, and social medical history data. Responses in the "improving skills" category indicate that students anticipate enhancing their conversational abilities through practice with the chatbot.

4.1.6. Advantages

There were clear positive responses in the category of chatbot advantages. Here, 97% of students saw uninterrupted communication, a protected space, and the possibility of using the chatbot multiple times as advantages. Furthermore, 77% of students believed that sitting opposite a fictional person offered a clear advantage. The responses show that students recognize the advantages of the chatbot. It can therefore be concluded that the new training opportunity has been positively received by most students.

4.1.7. Comparison with Other

The tool was compared with practice opportunities that students were already familiar with. The survey found that about one-third of medical students preferred practicing medical history-taking with an actor, while another third did not. Due to the highly divergent distribution of responses, it is not possible to ascertain a clear trend among the students. The subsequent statement investigated whether students would prefer to engage in practice with a real pregnant individual. We found that 50% of students abstained from this statement, while approximately 35% expressed agreement. It was observed that the preference to refrain from practicing with a live, pregnant individual was exclusively held by third-semester students. This finding may suggest that students in higher semesters prefer practicing with a real person, while students in lower semesters prefer alternative methods of practice such as the chatbot. The statement, "I would like permanent access to the practice tool with various case vignettes," received an 80% approval rating from the students. This finding indicates that most students would prefer an additional flexible practice option for taking medical histories.

4.2. Comparison with Literature

Comparable studies have achieved similar results, particularly regarding usability, accessibility, and perceived realism of AI-simulated patients [13,20]. In a study by Benfatah et al., 12 students evaluated the communication and interaction skills of ChatGPT as a simulation patient. The students evaluated ChatGPT as easily accessible, with an average rating of 19.4 out of a possible 25 points [20], which aligns closely with the positive usability ratings observed in the present study. A study by the University of Sonora (Mexico) revealed that medical students exhibited a high degree of acceptance of a dialogue with an AI simulation scenario [21]. Most of the test subjects, 94%, attested to the clarity and comprehensibility of the clinical case. A total of 91% of students found the scenario to be realistic, and 94% of students rated the chatbot's responses as appropriate based on ChatGPT-4 [21]. The results obtained for the medical history chatbot designed for midwifery students are comparable.

This study could not determine if using the chatbot significantly improved communication skills during medical history interviews. Other studies have already provided initial results on this topic. In Japan, the results of a randomized controlled trial demonstrated that medical students' confidence in performing the procedure improved when they practiced medical history-taking with an AI-simulated patient [22]. A hybrid systematic review from 2024 found that using interactive virtual patients can improve students' clinical skills, especially in history-taking and clinical reasoning [20]. Additionally, a randomized controlled trial at Anhui University in China assessed the effectiveness of chatbot-based training for medical students. Participants were divided into two groups: an intervention group utilizing regular chatbot sessions and a control group engaging in conventional role play with instructors. The chatbot group interacted through both text and voice input and received comprehensive feedback from the system, while the control group obtained feedback from instructors. The results demonstrated that the GPT simulation group achieved significantly

higher scores on the structured clinical examination post-training compared to the control group. Furthermore, participants in the GPT group reported increased enthusiasm for learning, stronger self-directed motivation, and improved communication and feedback skills ($p < 0.05$) [23].

AI-based simulations offer a safe space to practice clinical skills, reducing risks to patients. Medical histories can be taken outside clinical settings, and procedures are adaptable anywhere. The midwifery students also identified these advantages and gave positive ratings to the characteristics of flexible use in a quiet learning environment. Research by Adnan and Srinivasan supports these findings, highlighting that students can move through training at a pace that suits them best [7,24]. Overall, our findings are largely consistent with existing research. However, we were unable to identify studies involving comparable participant groups for direct comparison.

4.3. Interpretation and Implications

The students' overall assessment of the tool was positive, indicating a high level of acceptance of AI-supported simulation in midwifery education. A good user experience is vital, especially when using AI tools independently or without supervision [18]. A favorable evaluation of both the conversation and the case is typically considered beneficial. Nevertheless, some research indicates that ChatGPT often exhibits excessive friendliness in its replies [25]. GPT was found to be overly compliant, particularly in instances of substandard medical counsel [26]. It is already well known that ChatGPT tends to give very friendly answers, which may reduce learners' fear of making mistakes and encourage autonomous practice. Researchers have expressed a favorable view of this characteristic, considering it to be beneficial in certain respects. The conversation is pleasant due to its positive responses [18,21,25]. ChatGPT's friendly tone reduces students' fear of mistakes and motivates independent practice [27]. There is debate as to whether this excessive friendliness could lead to distorted expectations among students. In real-life situations, patients may be more critical or less cooperative, particularly regarding managing resistance, conflict, or emotional distress in real clinical encounters. From a didactic perspective, this potential limitation must be considered [21,24].

Our research indicates that students have varying preferences for exercise formats, suggesting that no single method is universally best for everyone. The chatbot cannot replace practical experience with actors or real pregnant individuals. However, the results indicate a tendency among students in higher semesters with extensive practice experience to perceive advantages in rehearsing with an actor or a real person. This finding suggests that the implementation of an AI-powered anamnesis chatbot may offer educational benefits to novice students. During the anamnesis interview, students in the lower semesters tend to focus on correctly recalling content and carefully documenting it. As previously indicated, the utilization of a chatbot has been demonstrated to be a valuable practice option in this context. By higher semesters, students typically know the structure, content, and documentation of the medical history interview, so their learning shifts to interpersonal communication skills. At this point, it must be mentioned that the chatbot's ability to fully replicate human communication is still limited, particularly regarding nonverbal and paraverbal elements [18]. This limitation is partly due to the communication medium of the chatbot. The vocal qualities of the talking partner, the presence of eye contact, the capacity to withstand periods of silence, physical contact, and the tonal nuances of the voice of an individual seated opposite during an anamnesis play no part in the interaction with the chatbot. The utilization of an AI-generated pregnant person in the form of a chatbot does not appear to be sufficient to adequately cultivate students' competencies in empathic

and interpersonal communication [24]. As a result, the chatbot discussed can serve as an additional resource for learning and practice.

In addition, it is essential to consider the key ethical issues when integrating AI-based tools into health professions' education. In the context of artificial intelligence education, it is essential to address critical issues such as bias, discrimination, data protection and security, privacy, transparency, and the displacement of human labor, dehumanization, and personalization [28]. For example, using distorted or incomplete data to train AI systems could result in inaccurate information about certain health conditions or patient groups [24]. It is essential to examine these ethical considerations in greater detail, particularly when AI is used in home environments without the guidance of an educational instructor.

4.4. Recommendations and Future Directions

Chatbots can help to bridge the gap between theory and practice in teaching by making learning more dynamic and participatory, and allowing concepts to be applied immediately [13,25]. Artificial intelligence chatbots may offer midwifery students an efficient and flexible way to train in taking medical histories. A potential area for future exploration in the functionality of chatbots involves integrating direct feedback mechanisms. As previously mentioned, researchers are currently investigating the potential of artificial intelligence (AI) to provide real-time feedback to support the learning process [17,20].

Further research on AI-supported simulation patients is needed, particularly regarding assessing their effectiveness and educational value, issues which were not addressed in this study. To properly assess skill improvement, it would be necessary to conduct a longer-term study with a larger group of participants and a validated questionnaire. Participants' initial competencies would be documented, followed by comprehensive chatbot training and a final reassessment. This would make it possible to assess if the training tool improves competence. The composition of participants should be based on semesters three, five, and seven, with the aim of achieving an even distribution of group sizes up to 30 participants each. Such a distribution will enable meaningful comparisons across semesters.

It is essential for students to develop competencies in artificial intelligence throughout their academic programs, equipping them to identify potential risks and misinformation at an early stage. This content must be firmly anchored in the curriculum [29]. However, it is not recommended to use generative AI exclusively for training on the medical history.

In conclusion, this study examined midwifery students' perceptions of the medical history chatbot, with particular attention to its realism, usability, and value in fostering skill development. A key strength of our work is that, as far as we know, it is the first research dedicated solely to midwifery students learning to conduct a medical history using a chatbot. Earlier studies either grouped midwifery students with other participants or did not provide separate data about them.

Participants gave mostly positive ratings for the chatbot with our generative AI character. However, as the instrument used was exploratory, these findings should be interpreted cautiously. The outcomes offer initial insights that underscore the need for further research utilizing validated methodologies.

4.5. Limitations

The findings of this study should be interpreted considering several limitations that may influence their generalizability. This study's scope was constrained by the limited sample size, comprising only 26 students. Consequently, due to the descriptive and exploratory nature of the study design, it is not appropriate to generalize the results to all midwifery students. The small number of participants was attributable to scheduling conflicts for some students who were unable to attend due to shift work, as well as higher-

semester students preparing for examinations and thus unavailable for additional voluntary seminars. Voluntary participation may have resulted in a selection bias, with individuals interested in AI being more inclined to participate and consequently providing more favorable evaluations of the bot. Additionally, the cohort exhibited varied performance levels among students. The group of third-semester students comprised sixteen people and was thus the largest, followed by nine participants from the fifth semester and one participant from the seventh semester. An analysis of the various groups and their response patterns revealed trends, but no clear, comparable generalizations.

A central methodological limitation of this study is that the questionnaire could not be pretested due to time constraints. Without a systematic assessment of its comprehensibility, reliability, and construct validity, there is a risk that some items were not interpreted as intended. As a result, certain questions may not have consistently measured the targeted constructions and may have introduced possible measurement error and reduced the comparability of results. The absence of validation processes may restrict both the interpretability and reproducibility of the findings and may have contributed to data that does not accurately reflect the underlying constructs.

Additionally, the authors recognize that the inclusion of a free-text field could have enabled the collection of more comprehensive insights from students on their views regarding chatbots.

The authors supervised the students throughout the courses and were responsible for conducting the research and preparing the data. The author who was responsible for conducting the study, collecting data, and evaluating the results, is not affiliated with the teaching staff at the University of Tübingen, there was no direct association with the students.

5. Conclusions

The results from this study apply only to the specific sample and context examined and cannot be broadly generalized due to this study's design. This study highlights the potential of an AI medical history chatbot for midwifery students at the University of Tübingen as a valuable addition to medical history training, based on students' self-reported perceptions and experiences with a positive overall student response. The bot is rated positively in terms of user-friendliness and realism. The current version does not fully replace simulation training, such as role-playing settings and guided practice with pregnant women, in everyday practice.

The chatbot's key strengths are its flexibility and the independence it offers in terms of location, time, and frequency of use. From the students' perspective, the greatest skill of acquisition through practice is in the complete recording, implementation, and documentation of the medical history as well as in structuring the interview process.

The findings suggest that the chatbot is particularly suitable as a supplementary training tool in the early stages of midwifery education, supporting students in gaining confidence and practicing structured medical history-taking in a protected environment. However, for the development of advanced interpersonal, empathic, and relational communication skills, training formats involving real people, such as simulated patients or real pregnant women, remain essential.

More research is required to assess how chatbot training affects students' medical history-taking skills, especially compared to traditional simulation methods and at various levels of midwifery education. Future research should also investigate the integration of AI-based chatbots into blended learning concepts that include pedagogical supervision, reflection, and feedback.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ime5010032/s1>.

Author Contributions: Conceptualization, M.T.E. and C.F.P.; methodology, M.T.E.; software, M.T.E.; validation, M.T.E., C.F.P. and U.K.; formal analysis, C.F.P.; investigation, M.T.E.; resources, M.T.E.; data curation, M.T.E.; writing—original draft preparation, M.T.E.; writing—review and editing, C.F.P.; visualization, M.T.E. and U.K.; supervision, C.F.P.; project administration, C.F.P.; funding acquisition, C.F.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Stiftung Innovation in der Hochschule project number, Fr-307/2023. We also acknowledge support from the Open Access Publication Fund of the University of Tuebingen.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Faculty of Medicine at University Hospital Tübingen (552/2024BO2, 15 October 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
GPT	Generative pre-trained transformer
LLM	Large language model
TIME	Tübingen Institute for Medical Education

References

1. ICM. Essential Competencies for Midwifery Practice. Available online: <https://internationalmidwives.org/resources/essential-competencies-for-midwifery-practice/> (accessed on 4 August 2025).
2. World Health Organization. *WHO Recommendations: Intrapartum Care for a Positive Childbirth Experience*; World Health Organization: Geneva, Switzerland, 2018. Available online: <https://iris.who.int/server/api/core/bitstreams/ba043cf7-cba4-484d-bf7e-ec79c4102d54/content> (accessed on 3 December 2025).
3. Rodríguez-Martín, S.; Greig, Y.; Shaw, E.; McKellar, L.; Kuipers, Y. Strategies and interventions used to provide communication education for midwifery students. A scoping review. *Nurse Educ. Pract.* **2024**, *78*, 103995. [CrossRef]
4. Seehafer, P. Anamneseerhebung. In *Schwangerenvorsorge Durch Hebammen*, 4th ed.; Deutscher Hebammenverband: Karlsruhe, Germany; Georg Thieme Verlag: Stuttgart, Deutschland, 2023; pp. 54–55.
5. Seehafer, P. Diagnostisches Zuhören lernen. Titelthema „Anamnese“. *Deutsche Hebammenzeitschrift*, 2022.
6. Sezgin, M.G.; Bektas, H. Effectiveness of interprofessional simulation-based education programs to improve teamwork and communication for students in the healthcare profession: A systematic review and meta-analysis of randomized controlled trials. *Nurse Educ. Today* **2023**, *120*, 105619. [CrossRef]
7. Adnan, A.I. Effectiveness of communication skills training in medical students using simulated patients or volunteer outpatients. *Cureus* **2022**, *14*, e26717. [CrossRef]
8. Ton, D.N.M.; Duong, T.T.K.; Tran, H.T.; Nguyen, T.T.T.; Mai, H.B.; Nguyen, P.T.A. Effects of standardized patient simulation and mobile applications on nursing students' clinical competence, self-efficacy, and cultural competence: A quasi-experimental study. *Int. J. Environ. Res. Public Health* **2024**, *21*, 515. [CrossRef] [PubMed]
9. Lendahls, L.; Oscarsson, M.G. Midwifery students' experiences of simulation-and skills training. *Nurse Educ. Today* **2017**, *50*, 12–16. [CrossRef] [PubMed]
10. Lotze, N. Zur sprachlichen Interaktion mit Chatbots—eine linguistische Perspektive. In *Talk with the Bots Gesprächsroboter und Social Bots im Diskurs*; Hug, T., Pallaver, G., Eds.; Innsbruck University Press: Innsbruck, Austria, 2018; pp. 40–43.
11. Jovanović, M.; Baez, M.; Casati, F. Chatbots as conversational healthcare services. *IEEE Internet Comput.* **2020**, *25*, 44–51. [CrossRef]

12. Nadarzynski, T.; Miles, O.; Cowie, A.; Ridge, D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit. Health* **2019**, *5*, 2055207619871808.
13. Holderried, F.; Stegemann-Philipps, C.; Herschbach, L.; Moldt, J.A.; Nevins, A.; Griewatz, J.; Holderried, M.; Herrmann-Werner, A.; Festl-Wietek, T.; Mahling, M. A generative pretrained transformer (GPT)-powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study. *JMIR Med. Educ.* **2024**, *10*, e53961. [\[CrossRef\]](#)
14. Sanz, A.; Tapia, J.L.; Garcia-Carpintero, E.; Rocacado, J.F.; Predraja, L.M. ChatGPT simulated patient: Use in clinical training. *Piscothema* **2025**, *37*, 23–32. [\[CrossRef\]](#)
15. O'Connor, S.; Yan, Y.; Thilo, F.J.S.; Felzmann, H.; Dowding, D.; Lee, J.J. Artificial intelligence in nursing and midwifery: A systematic review. *J. Clin. Nurs.* **2023**, *32*, 2951–2968. [\[CrossRef\]](#)
16. Schlick, C.; Rädcl-Ablass, K.; Schliz, K.; Rupp, S.; Schwendemann, H.; Roddewig, M.; Pahr-Hosbach, S.; Miersch, C. Lernen am KI-Patienten—der Einsatz eines neuen Trainingstools für Studierende der Gesundheitsberufe. *Physioscience* **2025**, *21*, V19.
17. Følstad, A.; Brandtzaeg, P.B. Users' experiences with chatbots: Findings from a questionnaire study. *Qual. User Exp.* **2020**, *5*, 3. [\[CrossRef\]](#)
18. Tübingen Medizinische Fakultät. *Modulhandbuch: Hebammenwissenschaft-Bachelor of Science*; Universität Tübingen, Institut für Geburtshilfe und Frauenheilkunde: Tübingen, Germany, 2023; Available online: <https://uni-tuebingen.de/studium/studienangebot/verzeichnis-der-studiengaenge/detail/course/hebammenwissenschaft-bachelor-mono/#c671-module-manual> (accessed on 4 December 2025).
19. Universität Tuebingen. *Studien- und Prüfungsordnung der Universität Tübingen für den Studiengang Hebammenwissenschaft mit Akademischer Abschlussprüfung Bachelor of Science (B.Sc.)*; Universität Tuebingen: Tübingen, Germany, 2024; Available online: <https://uni-tuebingen.de/studium/studienangebot/verzeichnis-der-studiengaenge/detail/course/hebammenwissenschaft-bachelor-mono/#c671-module-manual> (accessed on 4 December 2025).
20. Benfatah, M.; Marfak, A.; Saad, E.; Hilali, A.; Nejari, C.; Youlyouz-Marfak, I. Assessing the efficacy of ChatGPT as a virtual patient in nursing simulation training: A study on nursing students' experience. *Teach. Learn. Nurs.* **2024**, *19*, e486–e493. [\[CrossRef\]](#)
21. Peralta Ramirez, A.A.; Trujillo López, S.; Navarro Armendariz, G.A.; De la Torre Othón, S.A.; Sierra Cervantes, M.R.; Medina Aguirre, J.A. Clinical Simulation with ChatGpt: A Revolution in Medical Education? *J. CME* **2025**, *14*, 2525615. [\[CrossRef\]](#)
22. Yamamoto, A.; Koda, M.; Ogawa, H.; Miyoshi, T.; Maeda, Y.; Otsuka, F.; Ino, H. Enhancing medical interview skills through AI-Simulated patient interactions: Nonrandomized controlled trial. *JMIR Med. Educ.* **2024**, *10*, e58753. [\[CrossRef\]](#)
23. Wang, Z.; Fan, T.T.; Li, M.L.; Zhu, N.J.; Wang, X.C. Feasibility study of using GPT for history-taking training in medical education: A randomized clinical trial. *BMC Med. Educ.* **2025**, *25*, 1030. [\[CrossRef\]](#)
24. Srinivasan, M.; Venugopal, A.; Venkatesan, L.; Kumar, R. Navigating the pedagogical landscape: Exploring the implications of AI and chatbots in nursing education. *JMIR Nurs.* **2024**, *7*, e52105. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Higgins, M.; Madan, C.; Patel, R. Development and decay of procedural skills in surgery: A systematic review of the effectiveness of simulation-based medical education interventions. *Surg* **2021**, *19*, e67–e77.
26. Cook, D.A.; Overgaard, J.; Pankratz, V.S.; Del Fiol, G.; Aakre, C.A. Virtual patients using large language models: Scalable, contextualized simulation of clinician-patient dialogue with feedback. *J. Med Internet Res.* **2025**, *27*, e68486. [\[CrossRef\]](#)
27. Kranz, A.; Abele, H. The Impact of Artificial Intelligence (AI) on Midwifery Education: A Scoping Review. *Healthcare* **2024**, *12*, 1082. [\[CrossRef\]](#)
28. Weisman, D.; Sugarman, A.; Huang, Y.M.; Gelberg, L.; Ganz, P.A.; Comulada, W.S. Development of a GPT-4-Powered Virtual Simulated Patient and Communication Training Platform for Medical Students to Practice Discussing Abnormal Mammogram Results With Patients: Multiphase Study. *JMIR Form. Res.* **2025**, *9*, e65670. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Gumusel, E.; Zhou, K.Z.; Sanfilippo, M.R. User privacy harms and risks in conversational ai: A proposed framework. *arXiv* **2024**, arXiv:2402.09716. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.