Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and Evaluation of an Artificial Intelligence-Simulated Patient Tool for Clinical Training in Family Medicine

Seyma Handan Akyon
Sincan Training and Research Hospital
Fatih Cagatay Akyon
METU University
Ikbal Humay Arman
Istanbul Medipol University
Gulsah Onur
Mus Bulanık District Health Directorate
Rita Abecasis
Primary Health Care Area, Unidade de Saúde Familiar da Baixa, Unidade Local de Saúde São José
Michael Harris
mpsmfh@bath.ac.uk

University of Bath

Research Article

Keywords: Artificial Intelligence, Medical Education, Family Practice, Teaching Materials, Simulation Training

Posted Date: March 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9096627/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Traditional clinical training in family medicine relies on case–based discussions and standardised patients, which are resource-intensive and limited in scope. We designed and evaluated an artificial intelligence (AI)-simulated patient tool for family medicine training, focusing on usability, realism, and educational value.

Methods

We used a two-phase developmental research framework (design and development, and evaluation) to design a large language model-based conversational agent powered by GPT-4o, grounded in case reports, patient personas, and normal-range databases. To evaluate the tool, we recruited European family doctors through the European Young Family Doctors' Movement network and snowball sampling. Participants interacted with the tool and completed nine 5-point Likert items (ranging from 1/Strongly disagree to 5/Strongly agree) and three open-ended questions. Quantitative data were analysed descriptively, while qualitative responses underwent thematic analysis.

Results

The tool was rated by most of the 21 participants as enjoyable, beneficial for clinical practice, giving evidence-based feedback (all median 4, interquartile range (IQR) 1); easy to use, making appropriate responses and giving a realistic conversational flow (all median 4, IQR 2). Most participants would strongly recommend it to their colleagues (median 5, IQR 1). Strengths included its supportive role, usability, and educational value, while weaknesses involved interaction limitations, realism, and cultural challenges. Suggested improvements included expanding content, adding multilingual and cultural adaptations, and enhancing interactiveness.

Conclusion

The AI-simulated patient is feasible, well-received, and holds promise for enhancing family medicine education. Future work should add speech-to-text, AI avatars, and measurement of the impact in a randomised controlled trial.

INTRODUCTION

Traditional medical education methods can be costly and time-consuming, and may not effectively cover very different scenarios [1]. Actor-based standardized patient programs, despite their widespread use, face the following practical limitations: high costs associated with hiring, training, and scheduling actors;

performance variability and fatigue in repeated sessions; and limited scalability for large student groups [2]. Artificial intelligence (AI) technologies incorporating large language models (LLMs) can simulate complex scenarios, provide real-time, personalized feedback, and reduce costs by generating unlimited, varied patient interactions [3, 4].

AI-powered chatbots have become widely adopted educational tools in medical training [4]. Early rule-based chatbots, limited to predetermined responses, lacked flexibility and deeper contextual understanding. In contrast, LLM-based conversational agents generate more nuanced, context-sensitive output that better supports learner engagement and clinical reasoning [6–8]. Generative Pretrained Transformer (GPT)-powered chatbots have been shown to generate 97.9% medically plausible patient responses, with medical students rating the interactions positively [9], and GPT-4 achieved a “very high level of” agreement with human feedback raters in history-taking training [10].

Beyond chatbots, virtual patient (VP) systems provide structured, realistic settings in which learners can practise communication and diagnostic skills [11–13]. A recent study highlights user satisfaction with VPs' accessibility but emphasises the need for better user-centred design, including faster animations and improved navigation features [14]. While simulation-based learning is well-established in specialised fields such as dentistry [15], surgery [16, 17], and radiology [18], there remains a significant gap in general medicine and primary care training [8]. A recent AI study on general practitioner consultations reported a 24–84% cost reduction, but it focused on a single structured-exam scenario [19]. Many existing training tools concentrate on narrow technical skills or tightly defined scenarios, rather than on open-ended encounters with patients who present with vague or overlapping complaints. Using AI-simulated patients in primary health care teaching could help by reproducing the complex, ambiguous problems that family doctors face in daily work [20].

This study had two main objectives: first, to design and build an AI-simulated patient platform driven by a LLM and structured around both case-based and persona-based databases for clinical training in family medicine; and second, to assess its usability, perceived realism, and educational value among family doctors and trainees. Additionally, the study examined user feedback to identify strengths, weaknesses and suggestions for improving AI-based clinical simulation tools, particularly for communication skills training.

MEHODOLOGY

We followed a systematic two-phase process comprising (i) the design and development of the AI tool and (ii) its evaluation with a heterogeneous European cohort of family doctors and trainees, using quantitative outcome measures and qualitative data subjected to thematic analysis.

Design and Development Phase of the AI Tool

Database Development

We based the AI-simulated patient tool on three interconnected databases.

Case-Study Database

We compiled case notes from published medical literature using a standardised template that included demographics, presenting complaint and duration, relevant medical history, examination and investigations, and the confirmed diagnosis with management plan. **(Fig. 1)**

Patient Persona Database

To mirror real-world clinical encounters and help users manage the complexity of patient interactions, we constructed a set of simulated patient personas, each presenting distinct communication challenges and representing different patient backgrounds. **(Fig. 1)** The design of these personas was informed by the “difficult patient” examples presented in a standard clinical medicine textbook [21]. Personas varied by emotional tone, health literacy, and interaction style to support communication-skills training.

Normal Ranges of Findings Database

This database provided reference ranges for laboratory findings, imaging results and physical examination metrics. The normal ranges, categorised by age group, included paediatric values. The reference values in this database were drawn from established medical literature and clinical guidelines to ensure accuracy and relevance.

Development of LLM-based AI-Simulated Patients:

The AI-simulated patient was powered by a Claude-based LLM (Gpt4o, 2024-08-06 version), enabling natural language processing and generation. The LLM was trained on the combined data from both the Case Study and Patient Persona databases using PostgreSQL as the database management system. AI Patient Simulator used a combination of medical case studies and patient personas to create realistic, interactive simulations.

Prompt Engineering

In the development of AI-simulated patients for clinical training, ‘prompt engineering’ (the systematic design and refinement of the textual inputs provided to the LLM model) is essential to ensure realistic, dynamic and educationally valuable interactions. Our approach to prompt engineering ensured that the AI system could interact with users in a highly context-aware manner. The prompts used case studies, guidelines, and databases to create a comprehensive simulation environment.

System Prompt

The system prompt set the foundational context for the AI's behaviour. It drew information from two key sources:

Case Articles

These provided the medical context, including patient history, symptoms, and potential diagnoses.

Patient Personas

These defined the AI-simulated patient's communication style and emotional characteristics, ensuring diverse interaction scenarios.

The system prompt ensured that the AI maintained consistency and realism throughout the interaction, simulating both medical and interpersonal aspects of a consultation.

Three Key Template Prompts

To guide the AI in its response generation, we designed three distinct prompt templates that corresponded to different stages of the clinical interaction:

History Taking and Examination Template Prompt

This prompt guided the AI in the examination phase. It drew heavily from the case study to provide detailed information on the patient's current state, symptoms, and medical history. The AI used this to simulate the physical examination process, including vital signs and relevant findings.

Analysis Template Prompt

This template was divided into several parts, each drawing from specific sources to simulate a well-rounded clinical analysis

- *History Taking*: This section drew on clinical guidelines to ensure thorough, accurate history-taking procedures are followed in line with best practices.
- *Diagnosis and Treatment*: The AI drew from the case articles to generate a differential diagnosis and treatment based on the patient’s presentation.
- *Red Flags*: Using guidelines, the AI identified and highlighted critical red flags that might indicate serious underlying conditions.
- *Communication Skills*: The AI used specific guidelines for communication, helping doctors develop effective strategies to engage patients with empathy, patience, and professionalism.

Normal Ranges Template Prompt

For inquiries related to laboratory results or physical findings, this prompt pulled data from the Normal Ranges Database. The AI checked these values against the case data to determine if any findings fell outside the normal range, providing instant feedback to the user.

Figure 2 illustrates the core architecture of the LLM-based AI patient framework. The LLM generates patient dialogue conditioned on a selected case (clinical data/history) and persona (communication/emotional traits), optionally consulting reference databases (e.g., normal ranges/analysis), and it returns interactive responses and feedback (e.g., diagnostic and management suggestions) to the user.

Evaluation Phase

Participants

We recruited family doctors and trainees from European countries through the European Young Family Doctors' Movement (EYFDM). The study invitation was disseminated through EYFDM WhatsApp groups and the EYFDM newsletter, and recipients were encouraged to share the survey link with colleagues in their local and national networks to expand participation.

Data Collection

After a 15–20-minute hands-on interaction with the AI-simulated patient tool (myaipatient.com), participants completed a structured survey comprising three sections, designed to collect evaluation data (Table 1). The questionnaire had demographic questions, nine Likert-scale questions (1 = Strongly Disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly Agree) and three open-ended questions. For this interaction, participants were instructed to work through at least one clinical case scenario on the platform, enter their responses, and review the feedback generated by the tool before completing the survey.

Table 1
Survey Instrument

Demographics	**- Country** **- Age** **- Sex** **- Resident or Specialist in family medicine** **- Years of work experience from graduation** **- Years of work experience in the field of family medicine**
Quantitative Study: 5-point Likert Scale Questions *(Please indicate your level of agreement with the following statements about the AI-simulated patient tool, where: 1 = Strongly Disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly Agree)*	1. The AI-simulated patient tool is easy to use. 2. The AI-simulated patient tool is enjoyable to use. 3. The AI-simulated patient tool is beneficial for my clinical practice. 4. AI-generated feedback is evidence-based. 5. AI-patient responses are medically plausible. 6. The AI patient was able to make an appropriate response to open-ended questions. 7. AI-simulated interactions are realistic in terms of conversation flow. 8. I would like to use this app in my daily postgraduate medical education activities. 9. I would like to recommend this application to my colleagues.
Qualitative Study: Open-ended Questions	1. In your opinion, what are the strengths of this application? 2. In your opinion, what are the weaknesses of this application? 3. How could this app be improved?

Analysis

Continuous variables are presented as medians with interquartile ranges (IQR). Data from open-ended questions were analysed using thematic analysis with manual coding. Three researchers (SHA, IHA, GO) independently familiarised themselves with the data, generated initial codes, and organised codes into candidate themes. The individual coding frameworks were compared in consensus meetings where discrepancies were discussed and resolved through iterative dialogue. Themes were reviewed against coded extracts and the full dataset for coherence. A fourth researcher (RA) reviewed the final thematic structure for credibility. To enhance trustworthiness, we maintained an audit trail of coding decisions and employed investigator triangulation across coders from different clinical and cultural backgrounds. The resulting codes and categories were subsequently merged and presented under overarching themes.

RESULTS

This study included 21 participants from 8 countries in the World Health Organisation (WHO) European Region. Of the participants, 17 were female and 4 were male, with 10 family medicine trainees and 11 qualified family doctors. *(Appendix 1)*

Quantitative Findings

The results from the participants in relation to their experience with the AI-based simulation instrument are presented in Table 2.

Most of the 21 participants rated the tool as enjoyable, beneficial for clinical practice, and giving evidence-based feedback (all median 4, IQR 1). The majority found it easy to use, with appropriate responses, and giving a realistic conversational flow (all median 4, IQR 2). Most participants would strongly recommend it to their colleagues (median 5, IQR 1).

Table 2
Participants' ratings of their opinions about the artificial intelligence-based tool

1	Variable	Median	IQR*	Range
	The AI-simulated patient tool is easy to use.	**4**	**2**	**2–5**
2	The AI-simulated patient tool is **enjoyable** to use.	4	1	2–5
3	The AI-simulated patient tool is **beneficial** for my clinical practice.	4	1	3–5
4	AI-generated feedbacks are **evidence-based**.	4	1	2–5
5	The AI-patient responses are **medically plausible**.	4	1	2–5
6	The AI-patient was able to make an **appropriate response to open-ended questions**.	4	2	1–5
7	AI-simulated interactions are **realistic** in terms of conversation flow.	4	2	1–5
8	I would like to **use this app in my daily** postgraduate medical education activities.	4	2	2–5
9	I would like to **recommend** this application to my colleagues.	5	1	3–5

**IQR: inter-quartile range.*

Qualitative findings

The strengths, weaknesses, and improvement suggestions were categorised and coded under the following headings: strengths, weaknesses and suggestions for improvement.

Strengths included its supportive role, usability and educational value, while weaknesses involved interaction limitations, realism and cultural challenges. Suggested improvements included expanding content, adding multilingual and cultural adaptations, and enhancing interactiveness. Figure 3 presents a mind map of our findings, with illustrative quotes.

DISCUSSION

Principal Findings

We designed and evaluated an LLM-based virtual patient platform among trainees and general practitioners from several European countries. Usability, realism, and perceived clinical benefit were generally rated positively, with all items achieving high median scores.

Comparison With Previous Work

Since our initial work, two systematic reviews have confirmed the rapid growth of this field [22, 23]. Recent developments include multi-agent systems grounded in real patient data, web-based prototypes with AI-assisted case generation and multi-LLM comparison, and platforms that provide automated

feedback on communication skills [24]. Our findings align with earlier studies reporting that chatbot-based virtual patients are perceived as engaging and easy to use, and with prior demonstrations of high medical plausibility and agreement with human raters [9–10, 25–26]. However, our study extends this evidence to a multinational primary care context, a setting where communication demands are heterogeneous and highly sensitive to cultural and systemic variation.

While the tool we used was entirely text-based, it still achieved positive user perceptions of usability and plausibility. By comparison, recent studies that combined an LLM with a social robotic interface reported even higher authenticity ratings [27]. This contrast suggests that while text-only platforms can demonstrate feasibility at lower cost and complexity, future multimodal approaches should enhance immersion and emotional realism.

Our findings should also be considered alongside commercially available platforms. A recent randomised crossover trial comparing an AI-driven virtual patient simulator with actor-based training among 274 medical students found no significant difference in self-reported communication skills attainment, while the AI platform achieved 24–84% cost reductions [28]. However, most commercial platforms are designed for undergraduate OSCE preparation with scripted scenarios, whereas our tool targets postgraduate family medicine training, uses published case reports, and incorporates diverse patient personas reflecting the communication complexity of primary care. The only GP-specific evaluation to date tested a conversational AI on a single scenario without persona selection or structured diagnostic feedback [19]. Furthermore, studies using commercial platforms typically cannot disclose the underlying system architecture, prompt design, or feedback algorithms due to proprietary restrictions. In contrast, our tool was developed entirely by the research team, allowing us to describe the full technical pipeline, including database structure, prompt engineering strategy, and feedback generation logic, thereby supporting transparency and reproducibility.

Large-scale deployment has also been demonstrated on a scripted platform with over 45,000 consultations [29]. That study identified hallucinations and off-topic responses as significant challenges. In our open-ended system, hallucinations were not prominent; instead, slower interactions were reported as barriers. This suggests that adoption barriers may differ across design approaches: scripted systems risk accuracy problems such as hallucinations, whereas structured open-prompt systems like ours reduce these risks but can feel less fluid to use.

The automated feedback was perceived positively, consistent with recent studies identifying immediate AI-generated feedback as a highly valued feature of simulated patient platforms [6, 30]. However, participants noted that feedback could be more specific to national guidelines, suggesting that future iterations should incorporate regional context selection.

Work on predefined communication archetypes, such as the “accuser” and “rationaliser” personas [31], demonstrated that LLMs can reproduce specific conversational and emotional styles with fidelity. Our approach differed by developing a broader persona database that included grieving, aggressive, talkative, distrustful, medically knowledgeable, and linguistically limited patients. This wider range reflects real-

world primary care challenges. It may offer broader training value than narrowly defined archetypes, with the potential to expand access to communication skills training, which currently varies substantially across institutions and countries [32].

Strengths and limitations

Previous research emphasises the potential of LLM-driven patients to enhance realism, interactivity, and scenario diversity. Our contribution lies in testing such a tool on a heterogeneous European family medicine cohort, highlighting its feasibility and identifying practical barriers that have not been prominent in earlier work.

The small, predominantly female sample and single-session, self-reported outcomes limit generalisability and preclude conclusions about learning gains. The qualitative component, while providing valuable exploratory insights, was limited to three brief open-ended questions rather than in-depth interviews, constraining the richness of qualitative data. Future studies should consider semi-structured interviews or focus groups to explore user experiences in greater depth. While our results suggest broad acceptability, the restricted score range raises concerns about ceiling effects, limiting the ability to differentiate between tool features or detect incremental improvements. The text-based format may have reduced conversational naturalness. Differences in guidelines and medications across countries complicated judgments of correctness.

Future Research

Future research should move beyond feasibility and assess whether these tools improve communication and clinical reasoning skills. There is a need for randomised controlled trials with objective performance outcomes, longitudinal exposure, and cost-effectiveness analyses. While technical enhancements (e.g., speech input, avatars) warrant evaluation, claims of skills improvement should remain cautious until supported by controlled evidence.

CONCLUSION

Using a two-phase developmental design with quantitative outcome measures and supplementary thematic analysis of open-ended responses in a heterogeneous European cohort, we found that the tool was generally feasible, acceptable, and experienced as being realistic and educationally useful. Participants also identified barriers such as time demands, challenges in prompt formulation, and variation in international guidelines that should be addressed in future refinements. Overall, our findings suggest that LLM-based virtual patients can serve as a scalable, user-friendly, and relatively low-cost adjunct to traditional communication skills training in GP education, warranting further optimisation and evaluation in routine practice.

Declarations

Author contributions

FCA and SHA developed the study idea and overall concept. FCA developed the AI-supported infrastructure and application, with clinical input from SHA. FCA, SHA, and MH contributed to the development of the study concept and methodology. MH and SHA designed the questionnaire items. SHA, GO, RA, and IHA assisted with survey dissemination and data collection. IHA, GO, and SHA contributed to data analysis. All authors contributed to drafting, reviewing, and refining the manuscript, with MH providing additional input during the revision process. All authors approved the final version of the manuscript.

The authors declare that there is no conflict of interest. They have no financial or commercial relationships with any organisations or firms that could be perceived as influencing the submitted work.

Ethics

The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the Istanbul Medipol University Ethics Committee (approval no: E-10840098-202.3.02-2073). Informed consent was obtained from all participants, and data collection adhered to ethical standards regarding confidentiality and voluntariness.

Acknowledgment

We thank the European General Practice Research Network for allowing us to present the study at its 100th meeting and for its members giving us feedback [33]. We thank the course organisers and participants from the European Young Family Doctors' Movement network for their constructive feedback on the early version of this work.

Data sharing

The datasets generated and analysed during the current study are not publicly available. Anonymised data will be made available from the corresponding author on reasonable request and subject to relevant ethical and data protection considerations.

Funding
The authors received no specific funding for this work.

Clinical trial number
Not applicable.

References

1. Majumder MA, Haque M, Razzaque MS. Trends and challenges of medical education in the changing academic and public health environment of the 21st century. Front Commun. 2023;8:1153764.

2. Elendu C, Amaechi DC, Okatta AU, et al. The impact of simulation-based training in medical education: A review. Medicine. 2024;103(27):e38813.
3. Sun L, Yin C, Xu Q, Zhao W. Artificial intelligence for healthcare and medical education: a systematic review. Am J Transl Res. 2023;15(7):4820–8.
4. Garg T. Artificial Intelligence in Medical Education. Am J Med. 2020;133(2):e68.
5. Kerly A, Hall P, Bull S. Bringing chatbots into education: Towards natural language negotiation of open learner models. Knowl-Based Syst. 2007;20(2):177–85.
6. Thesen T, Park SH. A generative AI teaching assistant for personalized learning in medical education. NPJ Digit Med. 2025;8:627.
7. Gorbanev I, Agudelo-Londoño S, González RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ online. 2018;23(1):1438718.
8. Kononowicz AA, Woodham LA, Edelbring S et al. Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. J Med Internet Res 2019; 21(7), e14676.
9. Holderried F, Stegemann-Philipps C, Herschbach L et al. 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.
10. Holderried F, Stegemann-Philipps C, Herrmann-Werner A et al. A Language Model-Powered Simulated Patient With Automated Feedback for History Taking: Prospective Study. JMIR Med Educ. 2024; 10, e59213.
11. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010;85(10):1589–602.
12. Liu X, Wu C, Lai R, et al. ChatGPT: When the artificial intelligence meets standardized patients in clinical training. J Transl Med. 2023;21:447.
13. Shorey S, Ang E, Yap J et al. A Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study. J Med Internet Res 2019; 21(10), e14658.
14. Richardson CL, Chapman S, White S. Establishing the acceptability and usability of an animated virtual patient simulation. Explor Res Clin Soc Pharm. 2021;4:100069.
15. Suárez A, Adanero A, Díaz-Flores García V, et al. Using a Virtual Patient via an Artificial Intelligence Chatbot to Develop Dental Students' Diagnostic Skills. Int J Environ Res Public Health. 2022;19(14):8735.
16. Sardesai N, Russo P, Martin J, et al. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training. Postgrad Med J. 2024;100(1182):237–41.

17. Mirchi N, Bissonnette V, Yilmaz R et al. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. PLoS ONE. 2020; 15(2), e0229596.
18. Shah C, Davtyan K, Nasrallah I, et al. Artificial Intelligence-Powered Clinical Decision Support and Simulation Platform for Radiology Trainee Education. J Digit Imaging. 2023;36(1):11–6.
19. Jacobs C, Hazra K, Singh A. Enhancing GP consultation skills training: educational evaluation of a conversational AI innovation for simulated consultation assessment preparation. Educ Prim Care. 2025;36(6):291–9.
20. Carr D, Gormley GJ. Visiting uncertainty': an immersive primary care simulation to explore decision-making when there is clinical uncertainty. Educ Prim Care. 2022;33(4):237–43.
21. Longmore M, Wilkinson I, Baldwin A, et al. Oxford Handbook of Clinical Medicine. 11th ed. Oxford: Oxford University Press; 2024.
22. Zeng J, Qi W, Shen S, et al. Embracing the Future of Medical Education With Large Language Model–Based Virtual Patients: Scoping Review. J Med Internet Res. 2025;27:e79091.
23. Li D, Lebai Lutfi S. Large Language Model–Based Virtual Patient Systems for History-Taking in Medical Education: Comprehensive Systematic Review. JMIR Med Inf. 2026;14:e79039.
24. Yu H, Zhou J, Li L, et al. Simulated patient systems powered by large language model-based AI agents offer potential for transforming medical education. Commun Med. 2026;6:27.
25. Adhikary PK, Motiyani I, Oke G, et al. Menstrual health education using a specialized large language model in India: Development and evaluation study of MenstLLaMA. J Med Internet Res. 2025;27:e71977.
26. Cook DA. Creating virtual patients using large language models: scalable, global, and low cost. Med Teach. 2025;47(1):40–2.
27. Borg A, Georg C, Jobs B, et al. Virtual Patient Simulations Using Social Robotics Combined With Large Language Models for Clinical Reasoning Training in Medical Education: Mixed Methods Study. J Med Internet Res. 2025;27:e63312.
28. Tyrrell EG, Sandhu SK, Berry K, et al. Web-Based AI-Driven Virtual Patient Simulator Versus Actor-Based Simulation for Teaching Consultation Skills: Multicenter Randomized Crossover Study. JMIR Form Res. 2025;9:e71667.
29. Potter L, Jefferies C. Enhancing communication and clinical reasoning in medical education: Building virtual patients with generative AI. Future Healthc J. 2024;11:100043.
30. Hicke Y, Geathers J, Vu K et al. MedSimAI: simulation and formative feedback generation to enhance deliberate practice in medical education. arXiv preprint arXiv:2503.05793. 2025 Mar 1.
31. Bodonhelyi A, Stegemann-Philipps C, Sonanini A et al. Modeling Challenging Patient Interactions: LLMs for Medical Communication Training. arXiv preprint 2025; arXiv:2503.22250.
32. Venktaramana V, Loh EK, Wong CJ, et al. A systematic scoping review of communication skills training in medical schools between 2000 and 2020. Med Teach. 2022;44(9):997–1006.

33. Selected abstracts of 100th European General Practice Research Network Meeting. Eur J Fam Med 14(Suppl 3), 1–86.

Figures

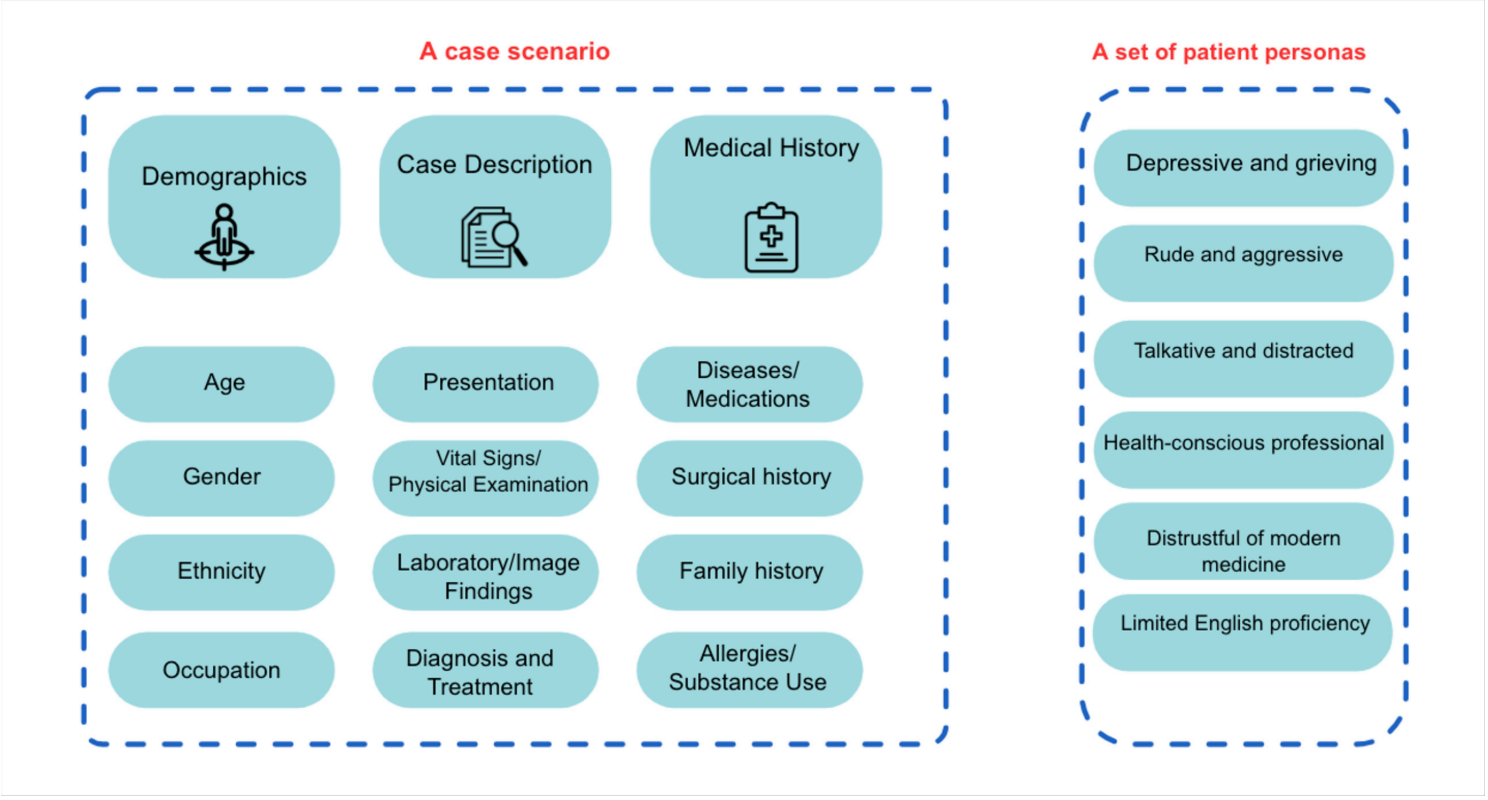


Figure 1

Figure legend not provided with this version

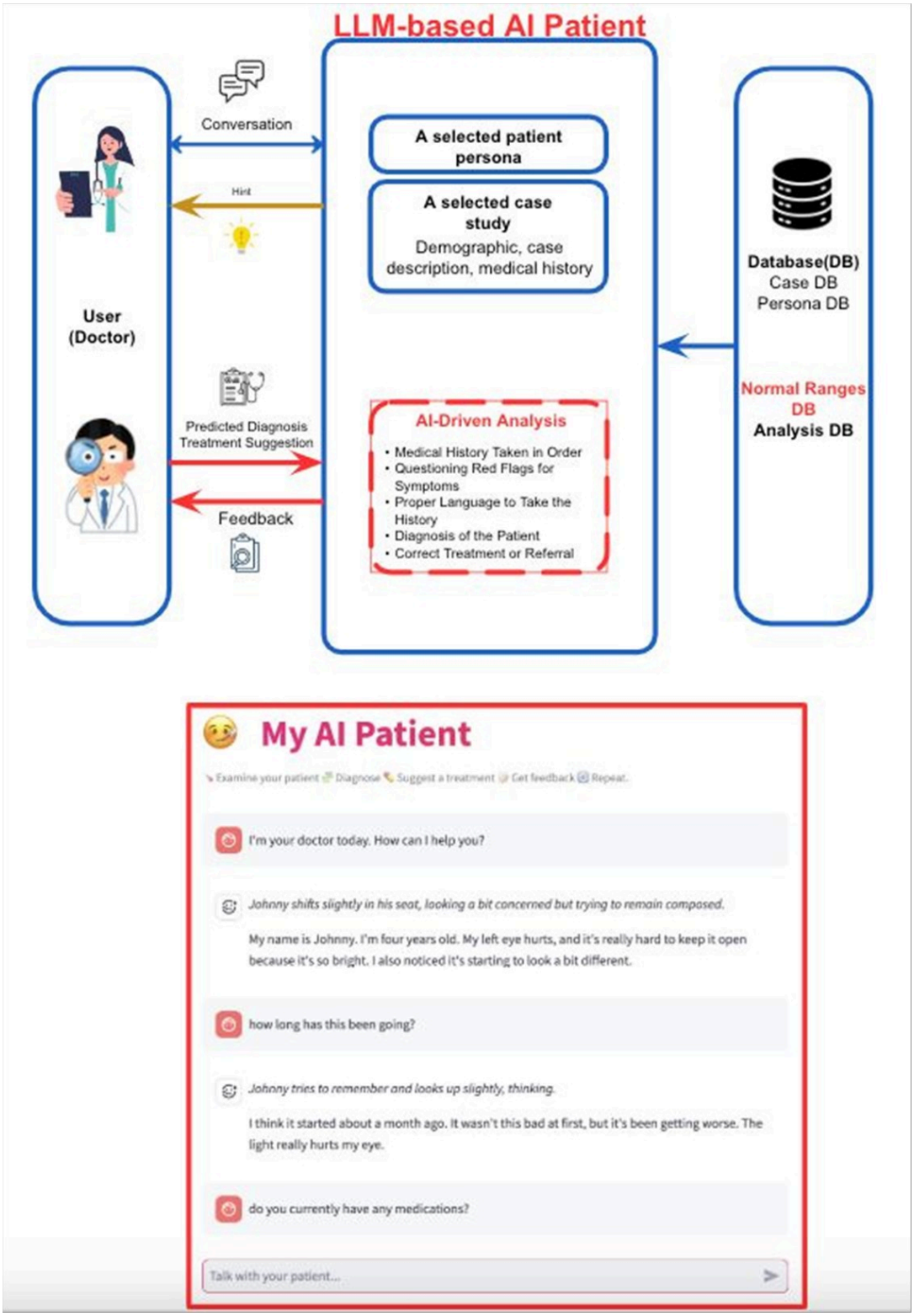


Figure 2

Figure legend not provided with this version

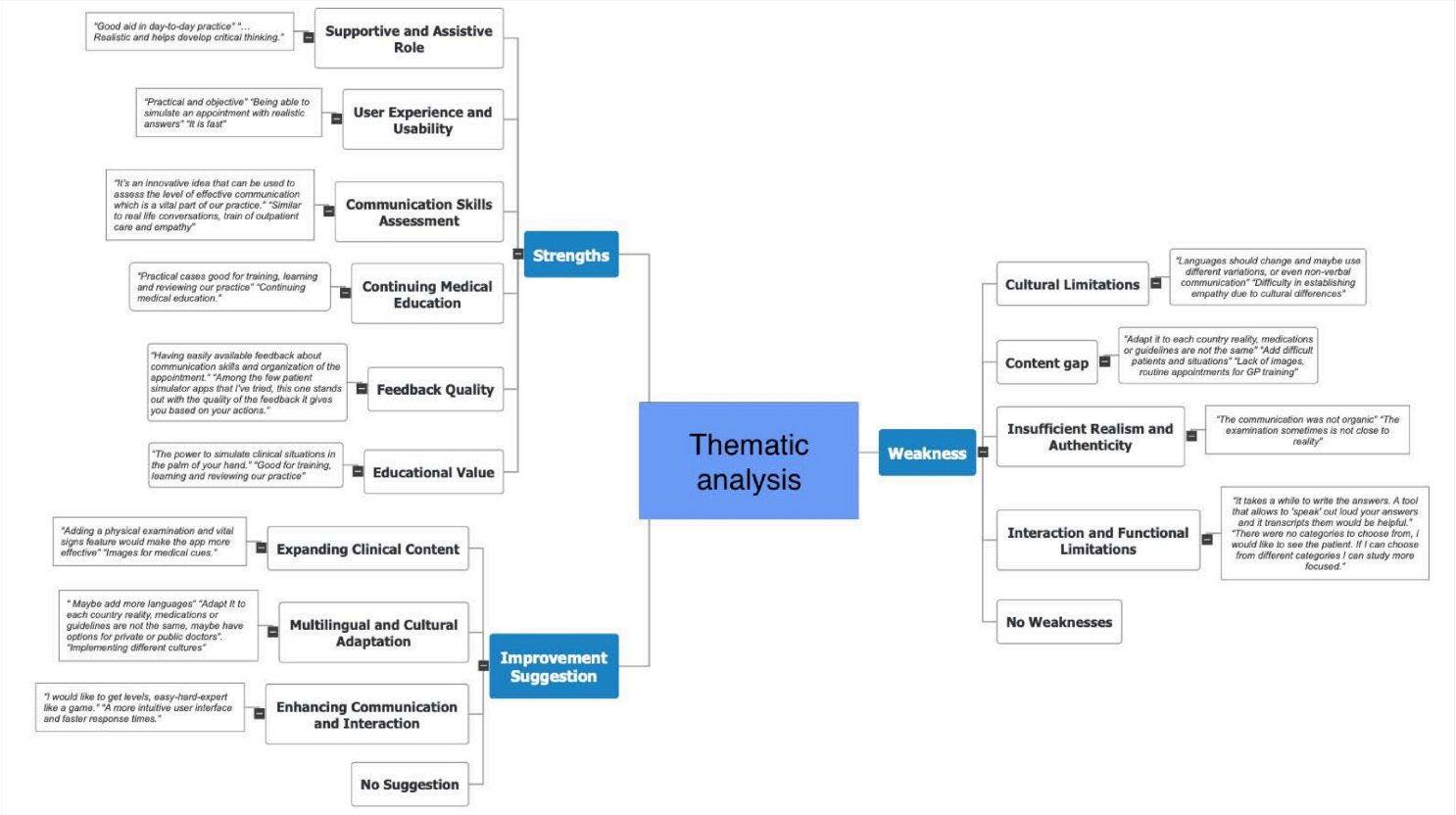


Figure 3

Figure legend not provided with this version