THE USER-CENTERED ITERATIVE DESIGN OF ANLLM-POWERED EDUCATIONAL SCENARIO SIMULATOR FOR CLINICAL REASONING

Weibing Zheng [1], Laurah Turner [2], Jess Kropczynski [1], Murat Ozer [1], Matthew Kelleher [2], Seth Overla [2], Shane Halse [1]

[1] School of Information Technology, University of Cincinnati, Cincinnati, Ohio, USA
[2] College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA

ABSTRACT

Language Models (LLMs) have tremendous promisesonconversational tasks in various sectors, including medical education. This study aims to integrate LLMs in medical education by user-centered iterative design and develop an educational clinical scenario simulator for clinical reasoning. The initial iteration prototypes a medical students-AI patient conversational app via prompt engineering. Feedback from physicians, student surveys, and focus group interviewsrevealed needs for a more comprehensive simulation mirroring the multi-agential nature of real clinical encounters. The second iteration prototypesan interactive LLM-based educational scenario simulator for clinical reasoning withan AI patient agent, multiple clinical dataacquisition agents, and educational assistant agents.Post-use surveysindicatetopfavouritesin clinical reasoningdevelopment(72.2%),real-time guidance(47.2%) and information gathering (44.4%). Theprogress from an LLM-powered conversational app to multi-agent educational simulator through iterative cycles with physicians and students-inputestablished a roadmapfor integrating LLMs into medical education and advancingAI-powered educational app design and development.

KEYWORDS

AI, Medical Education, LLM, Clinical Scenario Simulation, Iterative Design

1. INTRODUCTION

The website ChatGPT.com (https://chatgpt.com/), launched on November 30, 2022, by OpenAI.INC. is a significant and tremendousachievement in the recent history of Artificial Intelligence (AI). The groundbreaking Human-Computer Interaction(HCI) websitehas afront-end chat user-interface (UI) andback-end large language models (LLMs). Human users can input their queries, called prompts, into the chat UI. The back endcaptures the prompts and generates a response from LLMs based on its vast corpora of trained text data. The returned responses closely mimic human conversation. The conversationbetweenhumans and LLMs boosted ChatGPT to an unprecedented success and accumulated 1 million users in just five days and 100 million users in twomonths[13][21]. With itsrapidly growing popularity, LLMs show immerse potential in variablesectors of human needs[26][33], including medical education[5].

David C. Wyld et al. (Eds): NLCAI, AIFU, CCSEA, BIoT, SEA, SIPRO, BDML, CLOUD – 2025
pp. 239-260, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151019

Medical education is a subset of education that trains individuals to become future healthcare professionals with clinical, theoretical, and practicalknowledge and skills[34].Clinical scenario simulation in medical education was widely utilized as a controlled and safe learning environment to mimic a real-world clinical situation. A clinical scenario simulation could be physical or virtual. Physical simulation is a mannequin-based simulation, such as the standard plastic model shown in simulation laboratories. A virtual simulation is a computer-based simulation, such as the standard computer model shown in virtual simulation software. Clinical scenario simulation is an educational approach that uses simulated patient cases to help medical students and healthcare professionals practice their clinical reasoning knowledge and skills[5,9].

Mastering clinical reasoning —the complex cognitive process of diagnosing and managing patient problems— remains a cornerstone and significant challenge in medical education. While traditionally learned through observation and practice, the nuanced interplay of hypothesis generation, data acquisition (history, physical exam, investigations), and iterative refinement demands more scalable and adaptable training tools. The recent progress of LLMs boosted a new round of virtual clinical scenario simulation with modern AI technology.LLMs have the potential to be used to create virtual training environments and authentic clinical scenarios for the development of clinical skills[5]. Vaughn et al. [42] used ChatGPT and the prompt ``Using the Healthcare Simulation Standards of Best PracticeTM, created a healthcare simulation for healthcare students caring for a hospitalized adult with ...`` (the symptoms or illness), with five realistic simulation scenarios for the purpose of mitigating labor-intensive and time-consuming development and implementing simulation scenarios. This single prompt and directly using ChatGPT has a big issue in missing information, such as vital signs to nursing interventions, lack of history, etc. The authors pointed out that following up queries could help to solve those issues. Thesen et al.[37]developed a web-based app named AI Patient Actor with the LLM GPT-4o and a user interface (UI) for medical education in Neuroscience & Neurology course. The LLM GPT-4o controlled by the system prompts and the patient case file to simulate a patient. These LLMs-powered clinical scenario simulations are cost-effective and scalable, has great opportunities for medical education to facilitate clinical reasoning training. However, the exploration of integrating LLMs with medical education is still limited. There remains a significant gap increating systems that are deeply informed by the clinical reasoning theory and practise, thoroughly tested with physicians and medicals students, and iteratively refined based on the real-world clinical scenario cases.

User-centered iterative design—a methodology that involves end users throughout repeated cycles of design, development, testing and refinement—could address this gap effectively.For example, medical educators and students could participate in multiple feedback sessions on the clinical simulator prototyping, with each iteration incorporating their insights on diagnostic workflow and clinical reasoning challenges.This study aims to leverage user-centered iterative design to develop an educational application that simulates real-world clinical scenarios, and to investigate the integration of LLMs in medical education for clinical reasoning process, thereby enhancing medical students' clinical skills and strengthening their underlying cognitive processes. Based on an adapted clinical reasoning theory hypothetico-deductive modelwith clinical data collection and hypothesis generation, this study utilizes two phasesiterative design to prototype and improve an LLMs-powered educational scenario simulator for clinical reasoning. The initial prototype (iteration 1) focused on harnessing LLMs via prompt engineering to create a conversational AI patient, allowing students to practice initial clinical reasoning development. Feedback gathered through physicians, student surveys and focus group interviews revealed the need for a more comprehensive simulation that mimics the multi-agential nature of real clinical encounters. The second iteration prototyped an LLM-based multi-agent educational scenario simulator for clinical reasoning. Medical students can enter an interactive simulated clinical scenario to act as a physician, have a diagnostic conversation with the AI patient agent, acquire physical exam data from the physical exam agent, collect lab and test data from the intervention agent, and more convenientlyconsulate the guides from the AI supervisor agent, gain the feedback from AI evaluator agent. The feedback from physicianswas integrated to enhance clinical fidelity in the second iteration. The post-usesurveyswere distributed to medical students to gather more feedback and indicate highfavourites in clinical reasoning development, information gathering and real-time guidance.

Thecontributions of this study are as follows through the user-centered iterative design process from an LLM-powered conversational app to a multi-agent educational simulator. 1) A best practice for designing and developing an LLM powered educational application. 2)Integration ofthe abstractclinical reasoning theorywithconcrete educational application implementation. 3) An LLMs-powered clinical scenario simulation. 4) Technical improvement on an educational clinical scenario simulator from a chatbot to a multi-agent clinical simulation.

The following sections are organized as follows. Background and related work will introduce some basic knowledge on LLMs, clinical reasoning, clinical scenario simulation, user-centered iterative design, and current literature review on LLM-powered clinical scenario simulation.The LLM-powered educational scenario simulator section would introduce itstechnical design and development. Methodology will introduce the two stages of user-centered iterative design. Results will reveal the feedback from surveys and interviews and the iterative design prototypes. Discussion will show the implication of the work. Future Works will explain what we will do next.The conclusion section will summarize the contribution and bring a bright future.

2. Background and Related Work

2.1. AI and Generative AI (GenAI)

AI is a field to make the machine have the human intelligence through perceiving its environment[32]. In 1950, a seminal paper ``Computing Machinery and Intelligence" from Alan Turning is widely considered as a benchmark work as the ``Turning test"[40], which is a measurement of a computer or program's ability to achieve intelligence indistinguishable, equivalent to, or over a human, and a test whether the machine intelligent enough of simulating a real human. In October 2024, the Nobel Prize awards in Physics ``for foundational discoveries and inventions that enable machine learning with artificial neural networks"[45], and Chemistry with AI ``for computational protein design and protein structure prediction"[46]. With the development of neural networks, AI entered into a new generative age.Generative AI (GenAI) refers to special types of AI algorithms that can generate new content, such as text, images, and videos, based on patterns and examples from existing data [9][10]. GenAI is rapidly evolving in diverse fields ranging from art[15] and music creation[20]to drug discovery [16], education, and medical education [1]. In the context of medical education, GenAI can be used to create virtual patient cases, simulations, and educational materials that mimic real-world clinical scenarios[28]. These materials can then be integrated into a smart medical education platform that processes and delivers them to users based on their individualized learning needs and preferences[17]. Although ethical concerns regarding bias and potential misuse require careful consideration, the potential of GenAI to reshape various aspects of our lives is undeniable[22].

2.2. Large Language Models (LLMs)

LLMs are special GenAI algorithms in natural language processing fields and primarily focus on text and language. LLMs [6] are built on neural networks with huge and enormous amounts of text data (corpus) and can generate text content. LLMs are models of intelligent generative systems that can understand, manipulate, and process human language and generate human-like text to mimic human communication. The transformer architecture, proposed in 2017 research paper named "Attention is All You Need"[41], marked a turning point, leading to more large language models emerging such as BERT and GPT. These new models pushed thenew state-of-the-art (SOTA) benchmarks for what AI can achieve in natural language understanding, processing, and generation. With the boosting new waive of AI research pushing the SOTA boundaries of what these models can perform and how they can be used in education [18], medicine and medical education[38].

2.3. Prompt Engineering

Prompts are the input queries that users entered the chat box to interact with LLMs. Prompt engineering is about the engineering process of improving prompts to get a better and reliable output from the LLMs. By expanding on the input tokens (words), the prompt could contain more information and longer than before. A structured prompt could make LLMs respond better and desirable responses. The Elvis Saravis prompt engineering guideline (https://github.com/dair-ai/Prompt-Engineering-Guide) points out that a prompt could contain instructions, context, input data, and output format. The instruction describes the specific LLMs task. A content could guide the LLMs response. Input data could include the data to which you want the LLMs response. The output format could describe the LLMs response format.

2.4. Clinical Reasoning and Clinical Scenario

Clinical reasoning is a core competency in medical education that involves the cognitive processes, which used by physicians and healthcare professionals to diagnose and treat patients[8]. Sometimes, it has also been called clinical decision-making process, clinical judgment, and diagnostic reasoning[29]. It encompasses the ability to gather and interpret patient data, generate differential diagnoses, and develop treatment plans based on evidence-based practice and clinical guidelines[23]. Clinical reasoning is a complex and multifaceted skill that requires a combination of knowledge, experience, critical thinking, and problem-solving abilities[7]. Due to the complexity of clinical reasoning, different people may have different ideas of clinical reasoning[29]. Clinical reasoning has been widely considered a cognitive and clinical decision-making process[44]. It is a core and foundational competence for physicians, residents, and medical students[12][31]. A critical review[44] collected 5 decades of clinical reasoning research and presented three categories of theories and models, including theories and models based on clinical reasoning process, theories and models based on knowledge structure, compilation theories and mode. The first category only has hypothetico-deductive model (Elstein, 1978). The second category has pattern recognition model and illness script theory. The third category has dual processing theory and cognitive model. Figure 1 from Clinical reasoning introduction[4] is a type of hypothetico-deductive model, which shows the progress of collecting different diagnostic data and leads to the final hypotheses.

A clinical scenario case is a clinical encounter environment that involves the physician and the patient. Physicians can leverage their clinical reasoning skills to diagnose the disease of a patient. For medical students, the simulation of clinical scenarios provides a controllable and safe environment to practice and improve clinical reasoning knowledge and skills[27][36]. As shown in Figure 1,there are several levels of clinical scenario.The first level is a simple scenario, in which after collecting initial patient data, the medical student could make the hypotheses. The second level is the medium level that needs to collect more data on the history of patients and physical exams to make the hypotheses. The third level is a complex scenario that requires medical students to order the appropriate labs and images from the patients and leads toa hypothesis.

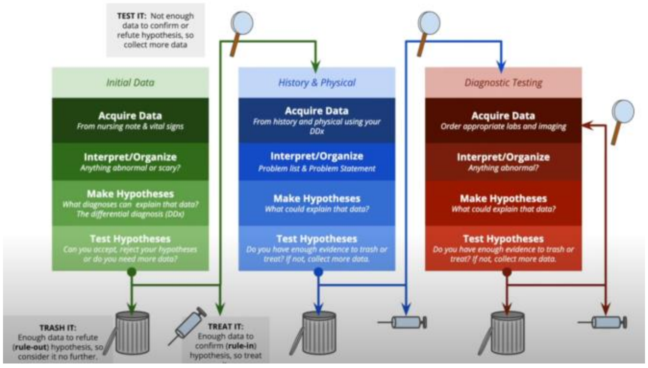

Figure 1. Clinical Reasoning (Rahul Patwwari, 2019)

2.5. Large Language Models (LLMs) with Medical Education

Compared to traditional one-size-fits-all medical education, such as clinical skills laboratories [3], simulation-based education[24], virtual clinical skills teaching or clinical skills video[25], LLM could cultivate proficient healthcare practitioners with unlimited access to individualized preferences, learning materials, educational resources, and consideration of the personal background of the learners, study progress, and real-time feedback[19][30]. Goh et al.[11] conducted a single-blind randomized clinical trial shows that the use of LLM (OpenAI ChatGPT) alone did not significantly enhance the performance of diagnostic reasoning, while further development would be needed in clinical practice. Johri et al.[14] presented a multi-agent conversational framework named CRAFT-MD (Conversational Reasoning Assessment Framework for Testing in Medicine) with AI patients, AI doctor, and AI grader for skin disease. CRAFT-MD demonstrated that LLMs have potential to extract patient medical history and augment physician decision-making from the conversation. Wei et al.[43] proposed MEDCO (Medical EDucationCOpilots), a multimodal learning environment with AI patient, AI doctor and AI radiologist. The AI students achieved improvement through training with MEDCO. However, MEDCO didn't have a real student in the training. Thus, real-world human-LLM interactive applications for medical education are still lacking and need more exploration. In order to design a user-centered human-LLM interactive application, this study uses the iterative design method to collect the stakeholders feedback and suggestions.

2.6. Clinical Scenario Simulation in Medical Education

Simulation in medical education is a method to create a controlled and safe environment that mimics real-world clinical situations. Clinical scenario simulation is to mimic the real-world patient encounters with physicians, medical students, and healthcare professionals. Simulation based medical training provides a controlled and safe environment for healthcare professionals and medical students to practice their clinical reasoning, patient interaction, and clinical skills without the risk of error diagnosis on real patients. The historical development of simulation-based education could divide into 3 stages, including physical simulation, virtual simulation, AI based simulation. The simulation could use mannequins or standardized patients, or virtual, using computer-based simulations. The recent ChatGPT emerging boosted a new round of scenario simulation with modern AI technology. Vaughn et al. [42] use ChatGPT and prompt ``Using the Healthcare Simulation Standards of Best PracticeTM, create a healthcare simulation for nursing students caring for a hospitalized adult with ...`` with the symptoms or illness, to create five realistic simulation scenarios for nursing education with the purpose of mitigating labor-intensive and time-consuming development and implementation simulation scenarios. This single prompt and directly using ChatGPT has a big issue on missing information, such as vital signs to nursing interventions, lack of history, etc. The authors pointed out that following up queries could help to figure those issues out.

Thesen et al.[37] developed a web-based app named AI Patient Actor with the LLM GPT-4o and a user interface (UI) for medical education in Neuroscience \& Neurology course. The LLM GPT-4o controlled by the system prompts and the patient case file to simulate a patient. The AI Patient Actor is cost-effective and scalable and has great opportunities for medical education to facilitate communication and clinical reasoning. However, the study lacks user feedback, the design process, and the evaluation and refinement of the app.

These studies demonstrate the potential of AI-driven simulations to enhance the learning experience by providing interactive and adaptive training environments. The approach of thisstudy aligns with the user-centered iterative design methodology to mitigate these gaps and developsan LLM-powered educational scenario simulator for clinical reasoning. The use of simulation in medical education helps bridge the gap between theoretical knowledge and practical application, providing learners with hands-on experience and immediate feedback. It also allows educators to assess learners' clinical skills, identify areas for improvement, and tailor their training accordingly.

3. LLM-Powered Education Scenario Simulator

3.1. Simulator Components Requirements

Education scenario simulator is a clinical scenario simulation in medical education for medical student's clinical reasoning learning platform. Like a real-life patient encounter,an LLM-powered education scenario simulator has several components as follows.

- **AI-Powered Virtual Patients -** AI patient is a must component that mimics the real-world patient. This study is leveraging prompt engineering and LLMs to develop the AI patient. The AI patient could have a conversation, provide the initial symptom data, medical history,test or lab data, and intervention data, etc.
- **Clinical Scenario Generation-**This component is to generate different clinical cases with different AI patients, which could be used to train the medical student's clinical knowledge.
- **Adaptive Learning Environment-**Based on the diverse background of the medical students, this component could provide the appropriate learning settings.
- **Interactive User Interface-**A UI could let the medical students interact with AI patients, generate their clinical scenario, andset up their learning environment.
- **Intelligent Tutoring System-**The tutor is by the side of the learner and providesguidance if the medical student has any questions on the clinical diagnosis.
- **Assessment and Feedback -** The component is provided forassessing medical students' clinical reasoning learning process by the feedback about what the strength and weakness is.

3.2. Simulator Architecture Design

Based on the components of the educational clinical scenario simulator, this subsection presents a three-layer architecture for the education scenario simulator, as shown in Figure 2. The architecture contains a *User Interface (UI)*, *ChatGPT API connector*, and *data storage layer*. The UI layer provides the interaction between the medical student and the simulator. The business layer ChatGPT API connector provides the diagnosis conversation between the medical student and the LLM powered AI patient. The data and storage layer stores and manipulates the data for the conversation and activities of medical students.The three-layer architectureis a flexible and effective design pattern that separates the software development concerns.

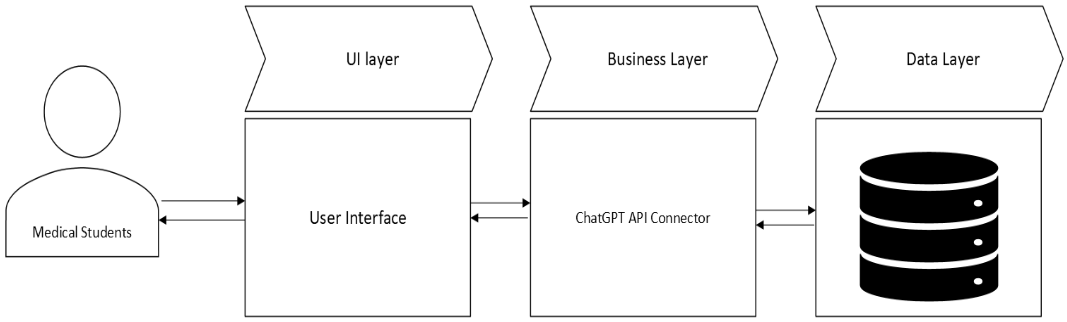

Figure 2. Clinical Scenario Simulator Architecture

3.3. Simulator Platform Implementation

The implementation of the LLM-powered educational scenario simulator is discussed and decided by the researchers with the school IT department.Since the cooperation between the University of Cincinnati with Microsoft, INC, all our simulator are designed, developed and secured with Microsoft products.The front-end UIis implemented with Power Apps and Power Page. The database uses the SharePointData tables. The middle layer is the ChatGPT API connectors for education. Thesource code has been controlled by theGit version management on Azure DevOps repo.The published applicationsarehosted in the default university environment. Only the participants and controlled medical students have access to the application URLs.

4. METHODOLOGY

4.1. Ethical Considerations

This study has been approvedby the University of Cincinnati Institution Review Board ((1/19/2023, MOD01_2021-1032) as not human subjects. The study also follows all HIPAA compliances, and all participant’s data is de-identified. Allparticipants are provided the informed consent before participation by the team, and it is explained that they can withdraw at any time during the interview. Participants in the focus group interview were provided lunch funded by AMA funding, and no other compensation was provided. All voice recordings and transcripts from focus groups interview stored in a securelocation provided by the university. All the study dataisonly accessible to the related researchers.

4.2. User-Centered Iterative Design

Iterative design is a cyclic design methodology that involves the process of needs analysis, prototyping, testing, and refinement[39].As shown in Figure 3, iterative design in this study is a user-centered process that put the needs of users (medical students), educators, and physicians in the centrethrough survey, interviews and usage data. With this approach, educational clinical scenario platforms can be iterativelyrefined to meet the clinical reasoning training needs of learners and enhance clinical reasoning training platform design through iterativeneeds analysis, prototyping, testing, and refinement. For example, a study on mobile health applications demonstrates the positive impact of user-centered design on health app development [35]. This study aims to harness the power of two phases of user-centered iterative design to create an interactive and adaptive educational clinical scenario simulation environment for clinical reasoning training. The educational clinical simulator platform for clinical reasoning simulate the real-world clinical encounter that medical students can act as a physician, and the LLM performs as an AI patient. Medical students can have a conversation with the LLM based AI patient to collect the patient's information and lead to a primary diagnosis and differential diagnosis. When the conversation is over, the medical students receive an evaluation summary of the conversation about the strengths and weaknesses. Feedback and suggestions for prototypes are collected from users and physicians throughface-to-face group interviews and online surveys. We iteratively design and implement our LLM based platform that is suitable best for clinical reasoning training. Through this approach, we can ensure that our medical student-AI conversation, a human-computer interaction platform, is not only relevant and engaging but also aligned with the needs of physician and medical student.

Figure 3. User Centered Iterative Design

4.2.1. Phase 1: An LLM-Powered Educational Clinical Scenario Simulator for Clinical Reasoning

Phase 1 process is an initialdiscovery and exploration phase. This phase aims toturn “0 to 1” with the idea “0” that integrating the LLMs with medical education to “1” a usable LLM-powered educational clinical scenario simulator platform.

4.2.1.1. Recruitment and Participants

The initial recruitment includes medical education experts, physicians, and medical students.Two medical educationphysicians are working closely during the concept building, prototyping, and all procedures.We recruited medical students for focus interviews, usage, post-use surveys by emails and from physician professors.The focus group interviews have happened at a meeting room in the building of College of Medicine, University of Cincinnati during lunch time. The building is where medical students take their medical courses. We also provided the lunch, which sponsored by the American Medical Association (AMA) funding. We conducted four round focus interviews on October10, 2023, with 5 students, October11, 2023,with 3 students, October 25, 2023, with 5 students, and October 26, 2023, with 9 students. Two posted-use surveys havebeen distributed. One survey is only threepulse questions with 17responses back. One is fourteen questions with another 17 responses. Both surveys were distributed by email and anonymously collected data.Since the HIPAA compliances, the studydid not identify and record the participants' age, gender, or other personal identifiable information (PII).

4.2.1.2. Data Collection

Mixed methods, including face-to-facefocus group interviews and surveys, collected both qualitative and quantitative data from the participants. The face-to-face focus group interviews are open, free discussion sessionswhere the participants could express descriptive, in-depth opinions, and subjective experiences on their usage of the initial LLM-powered educational clinical scenario simulator. While the surveys provided another angle of numerical andstructured data about their feedback.

4.2.1.3. Procedures

- **Needs Analysis.**The initial "0-to-1" needs analysis is from theguidance ofclinicalreasoning theory data collectionand the opinions frommedical expertson what a clinical scenario simulation should be functionalized and how to integrate the LLMs into clinical scenario simulator with the prompt engineering.AI-Powered virtual patients, clinical scenario generation, interactive user interface were the functions needed.Later then, interviews and surveys collected needs from medical students and assessment, and feedback is added.

- **Prototyping.**The needs analysis shows the direction of initial designing and developinganLLMs-powered educational scenario simulation.The low fidelity wireframes are designed within Miro, as shown in Figure 4.1 Scenario Generation Screen, and 4.2 Scenario Conversation Screen. Based on the low-fidelity and feedback from experts,the high fidelity is designed in Figma, as shown in Figure 5.1 High Fidelity Scenario Generation Screen, and 5.2 High Fidelity Scenario Conversation Screen. The implementation of LLM-powered educational clinical scenario simulator for clinical reasoningis done with Power App and SharePoint Data tables. The UIs will be shown in the result section.

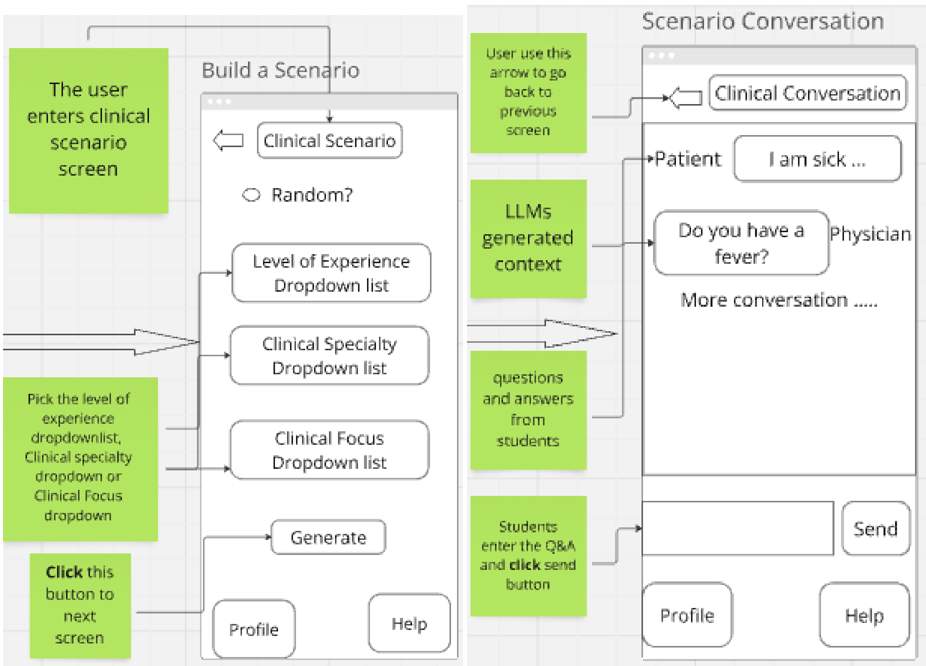

Figure 4 Low Fidelity Scenario Generation Screen (a), Conversation Screen (b)

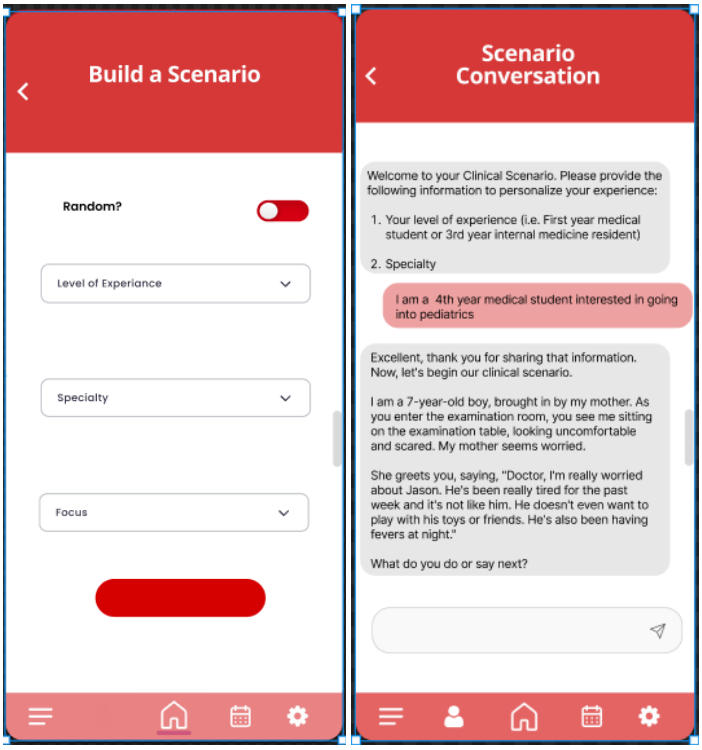

Figure 5 High Fidelity Scenario Generation Screen (a),Conversation Screen(b)

- **Testing.**The testing involves the experts testing and medical students' testing. The testing is to finish at least one, but no limited clinical scenario cases. As shown in Figure 4.1 and Figure 4.2, here are the following tasks.

  Task 1: Choose the right level of experience.
  Task 2: Select thewantedclinical scenario cases name.
  Task 3: Click "generate" for the clinical scenariocase.
  Task 4: Have a conversation with AI patients and collect patient information as much as possible.
  Task5: Submit diagnosis and differential diagnosis.
  Task 6: Read the summary and feedback from the clinical scenarioconversation.
  Task 7: Quit the application.

- **Refinement.**The refinement is from the experts testing and medical students' feedback from interviews and surveys as well as the real-time emails.The refinement is focused on improving the functionality and usability of the application.

4.2.1.4. Analysis

Analysis includes the interview qualitative data analysis and survey quantitative data analysis, as well as the simulator usage data analysis. Interview data analysis would use contextanalysis, and thematic analysis. Contextual analysis is the method that goes through the whole script to note the feedback/suggestions from the users. For example, the following feedback suggestionis morespecifically on the physical exam.

"Maybe there can be prompts to be more specific? Like when I say physical exam it gives a whole list of results, some of which I didn't even think of. So maybe asking us to be more specific

with the maneuvers we want to do (or at least the system like neuro exam, trauma exam etc.) can help us practice that".

The survey quantitative data will use descriptive statistics, such as percentages, to identify the overall feedback.

4.2.2. Phase 2: A LLM Based Multi-Agent Education Clinical Scenario Simulator for Clinical Reasoning.

Phase2 is a process that turns"1", an LLM-powered educational clinical scenario simulator, to a better version(V2) based onphase 1's feedback. The main feedback and updates areseparating the concernson the LLMs roles and leveraging the LLMs-powered multi-agent to design and build educational clinical scenario simulator to further mimic the real-world clinical encounter simulation andimprove the skills of medical student's patient data collection andclinical reasonings.

4.2.2.1. Participants

The two physician experts provide lots of suggestions on the design of multi-agents functions and diagnosis process with these agents.The participants have medical students from College of Medicine, University of Cincinnati. These medical studentshave taken clinical skills courses and haveused the new platform. Thesurvey wascreated by Microsoft Forms and distributed by school emails. All the users who get the email have access to the survey and they can finish the survey online. The focus group interviewsrecorded and saved on December 3, 2024, with 5 medical students, December 12, 2024, with 4 students.The survey collected 42 responses. Same as phase 1 that following the HIPAA compliances, the study did not identify and record the participants' age, gender, or other personal identifiable information (PII).

4.2.2.2. Data Collection

This follows the same process as phase 1, mixedmethod, including face-to-face focus group interviews and survey, are used to collect both qualitative and quantitative data from the participants.

4.2.2.3. Procedure

The procedure is still following the needs analysis, prototyping, testing and refinement steps.

- **Needs Analysis**- The second version of needs analysis is from the feedback and opinions from the domain experts and medical students, as well as a deeper dive into clinical reasoning theory and the technical progress on the LLMs based multi-agent.Containing the phase 1 functions, an intelligenttutoring system has been added in phase 2.

- **Prototyping-** Based on the needs analysis, an LLMs-powered multi-agent educational scenario simulation was designed and developed in Power Page directly**.**

- **Testing -** The testing involves the experts testing and medical students' testing. The testing is to finish at least one, but no limited clinical scenario cases. Including all the tasks in phase 1 testing, phase 2 testing contains three more steps.

  Task 0: Electronically consent to terms and conditions by marking the checkbox.
  Task 4.1: Select the right buttons to collect the patient data, including physical exams,diagnostic studies, interventions.

Task 4.2: Select the supervisor to consult with any questions.

- **Refinement** - The refinement is from the experts testing and medical students' feedback from interviews and surveys as well as the real-time emails.

4.2.2.4. Analysis

Same as the phase 1 analysis methods, the interview qualitative data analysis (descriptive statistics) and survey quantitative data analysis(), as well as the simulator usage data analysis for the second phase will be used.

5. RESULTS

The results would contain two iterations of LLM-powered educational clinical scenario simulator design and development as well as the analysis results from interviews, surveys and usage data.

5.1. Phase 1: An LLM-Powered Educational Clinical Scenario Simulator for Clinical Reasoning.

5.1.1. Prototype

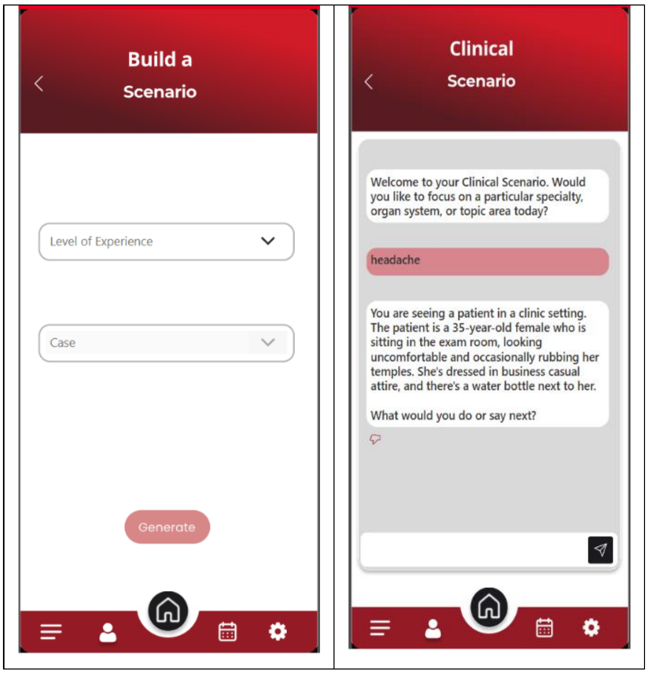

Figure 6. LLM-powered educational scenario simulator (a, b).

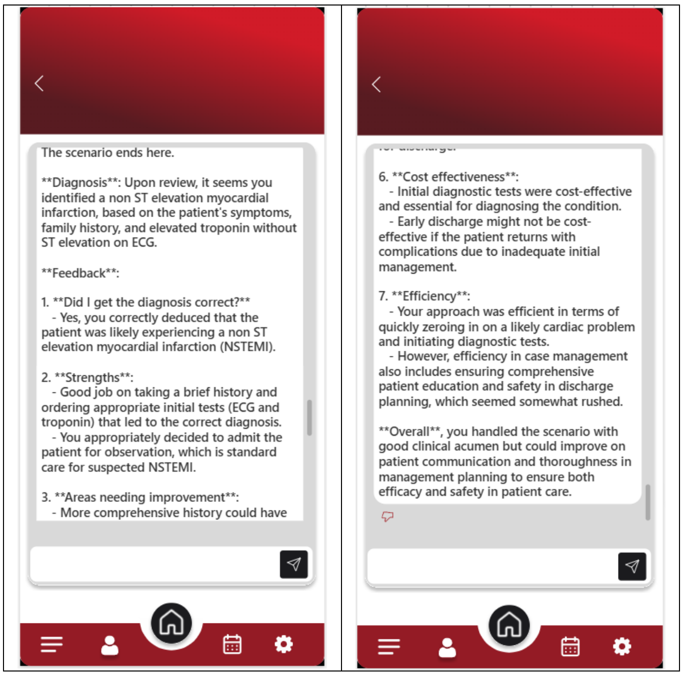

Figure 7. Educational scenario simulator feedback (a, b).

Figure 6(a, b) and Figure 7(a, b) are the UIs for the LLM-powered educational clinical scenario simulator for clinical reasoning. Those UIs meet thefunctions for interactive user interface, clinical scenario generation in Figure 6a,AI-powered virtual patients in Figure 6b, assessmentand feedback function in figure 7 (a, b).

5.1.2. The First Surveys Analysis

The first survey (Appendix 1)shows the average overall satisfaction rating is 3.65on a scale of 1 to 5.

Figure 8.Average overall satisfaction rating.

When asking positivefeedback, question 2 of the first surveycollected 17 responses and concludedas follows.

Engagement and learning: Many users found the platform engaged and educational with the ability to practice clinical skills independently, especially the real-world clinical opportunities werelimited.

Feedback useful: Specific and personalized feedback helps the students improve their clinical practices.

Personalized practice: Users expressed their favorite on dynamic and personalized diagnoses conversation.

Realistic Experience: The conversation seems realistic, and interaction feels real.

When asked about the suggestionforimprovement, the third question of the first surveyhad 15 answers, which revealed technical issuesregarding loading and connectivity. error messages, and app stability.

5.1.3. The Second Surveys Analysis

After we fixed the technical issues on the error message and refined the application, the overall satisfaction rating increased to 4.12, as shown in Figure 8.

Figure 9. Average overall satisfaction rating.

When asked which feature should be added, 88% in 17 responses medical students would like to add more clinical scenario cases. While no one mentioned the error message anymore and this message shows that the error message issue hasbeen fixed.

“I think the error messages were a problem at the beginning but thankfully, that resolved. It made the experience much more pleasant. I do wonder if the interface could be modernized visually.”
When asking aboutuser interface or other feature improvement, “would prefer a browser version”was mentioned 4 times.

When asking about the app’s contribution to their medical education. The average rating is 3.88 in a Likert 5points.

Figure10. Contribution score for medical education.

When asking whether the app helps them in achieving their learning goals. 76% of medical students answered “Yes”, while 24% answered “Not Sure”. No one answered “No”.

When asking if they need help, 4 out of 13 answered N/A, while 8 would send emails to physicians or others to ask for help. Seems a supervisor function is needed.

When asking the suggestions for improving the app, one student mentioned that theyadded a separate box for “differential diagnosis”. Two students hadasked the visuals, like lab tests, and x-rays, EKGs etc.and sounds, like heart sounds, lung sounds testing.

When asking for two more features that I want to add, physical exam, lab/test are been mentioned 10 out of 17 times, while a browser version instead of a mobile version is mentioned 2 times.

“Option to increase difficulty/tested knowledge. For example, if I felt really comfortable with the cases and wanted to take it a step further and start thinking about testing, treatment, imaging, etc. it would be nice to be able to tell it to let me keep going instead of stopping me at the diagnosis step. ”

The surveys show that medical students need“a browser version”, “physical exam, lab/test” and “supervisor” the most.

5.1.4. Focus Group Interviews.

The focus groups interviews are anopen discussion sessionabout the medical students’ experience using LLM-powered educational clinical scenario simulator, its functionalities, and areas for improvement.When thematicanalysis of the scripts, the topic widely ranges from purpose and functionality of the simulator, user experience and feedback, suggestion for improvement, technical and practical challenges, education impact, safety, and ethical considerations, and to future directions. Most participants enjoyed using the simulator, expressing positive feedback on their medical education.The features on“a browser version”, “physical exam, lab/test” and “supervisor” are discussed and required during the interviews.

5.1.5. Usage Data.

176second-year medical students generated 1,810 successful sessions, which were recordedin the database. That means the LLM-based educational scenario simulator works well. The functionality of the simulator meets the needs of software engineering.

5.2. Phase 2: A LLM Based Multi-Agent Education Clinical Scenario Simulator for Clinical Reasoning

5.2.1. Prototype

Summarizing the feedback from phase 1, theweb browser version with the physical exam, lab/test, and asking for help, are highly voted features that need to be added and improved. Phase 2 designed and implemented a new LLM based multi-agent education clinical scenario simulator for clinical reasoning, as shown in the following.Figure 11 is for clinical scenario generation.While AI-powered virtual patients and adaptive learning environment functionare shown in Figure 12.Figure 13 is for physical exam screen. Figure 14 provides an intelligent tutoring system with a supervisory agent, which helps students any time when they need help during the session. Figure 15, 16are the Assessment and FeedbackUI.

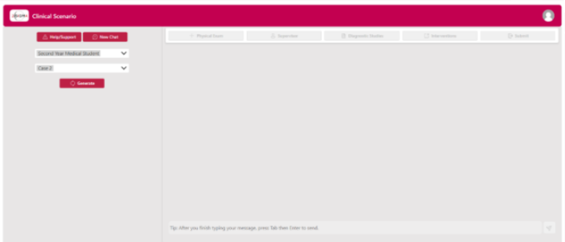
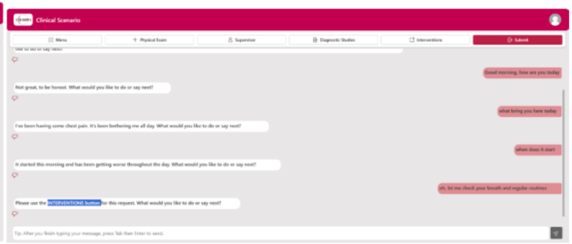

Figure 11. Scenario generation screen. Figure 12.Scenario conversation screen.

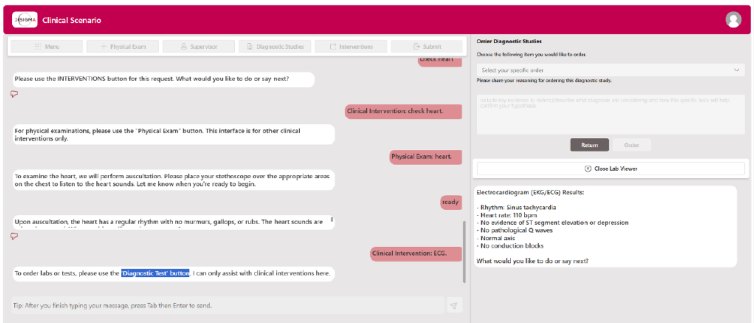
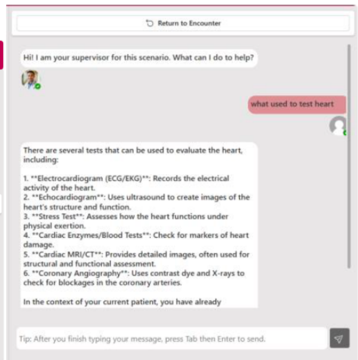

Figure 13. Scenario physical exam screen. Figure 14. Supervisor's screen.

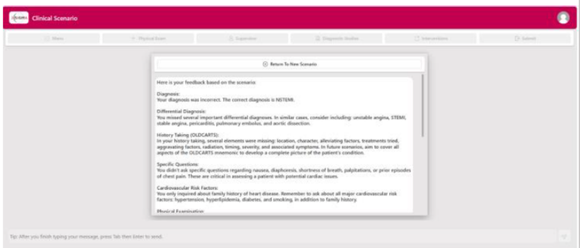
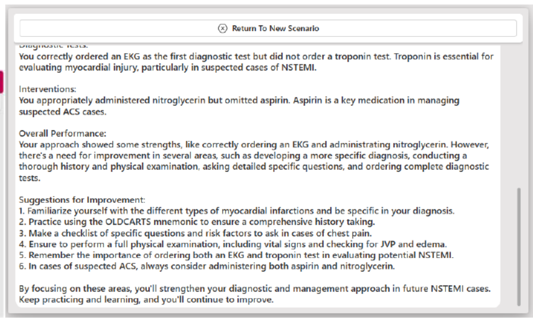

Figure 15. Feedback screen.Figure 16. More details on feedback screen.

5.2.2. Focus Group Interviews.

The focus groups interviews are an open discussion session as same as phase 1. The topics are about the new implementation for medical education to enhance clinical skills and assessment processes. When discussing the user experience and feedback, participants appreciate the AI's ability to respond appropriately and help their clinical reasoning development in real time.

5.2.3. The Surveys Analysis

When asked if the app aligned with clinical skills course objectives, 78.57% of students answered "agree" or "strongly agree". While 9.52% of 42 responses are "neutral". Which means this clinical scenario simulator aligned with the clinical skills course objective.

When asked about what features aremost useful in the simulator, the survey collected 26 answers. All of them showthe simulator has some helpful in their study.10 answersresponse the lab ordering with immediate results and interventions tab is helpful, while 9 answers mentioned that it is helpful for their extra practice for clinical reasoning learning, and 5 answered that "Talk to preceptor/supervisor" is very helpful. The feedback feature received 5 votes.

When asked what the most valuable aspects of the experience on the application,Clinical reasoning skills practicementioned most with 23 times in 29 responses.79.3%approved that the multi-agent educational scenario simulator is helpful for their clinical reasoning skills.

When asked which aspects of the clinical scenarios'simulator were most valuable for clinical reasoning learning, 36 answers returned. 26 responses contain Clinical reasoning development, followed by 17 responses with real-time feedback, and 16 responses with history taking practice,whilephysical exam got 8 responses and independent learning pace has 7 responses.

That means phase 2 LLMs based multi-agent education clinical scenario simulator for clinical reasoning is helpful with the innovative designon physical examination, laboratory, testing, and supervisors are meeting medical students' needs.

5.2.4. Usage Data

The medical students generated 379 successful sessions. The LLM-based multi-agent educational scenario simulator works well for medical students to practice their clinical reasoning, and all the data is saved correctly.

6. DISCUSSION AND FUTURE WORK

User-centered iterative design could focus on user's needsand build a better version of the LLMs-powered clinical scenario simulator. From phase 1 to phase 2, we could see the simulator helps them in achieving or aligning their learning goals increased from 76% in "Yes" to 78.57% of answered "agree" or "strongly agree", lowered 24% "Not Sure" to 9.52% of "neutral". The results suggest that theimproved simulator helps over 16%of medical students increase their clinical reasoning skills achievement.

The feedback from the focus group interviews and surveys is a way to listen to the user's needs. This study leveragephysician and medical students' needs on clinical reasoning skills andintegratesinto the clinical scenario simulator with the steps of need analysis, prototype, testing, and refinement. Theusage data over 2000 successfulclinical scenario simulation sessions shows that the medical students followed the clinical reasoning process of collecting patient initial data, history data, labs and intervention data and make the right clinical decision-making hypothesis.The LLM-powered clinical scenario simulatorcould help medical student's practice on the data collection and decision-making process. The initial LLM-powered chatbot to the LLMs-powered multi-agent is a significant improvement in terms of the needs on web browser versions, and different agents on physical exam, labs and tests, in particular, the real-time supervisor.

This study makes the following specific contributions through its user-centered iterative design approach in transforming an LLM-powered conversational application into a multi-agent educational simulator:

1. Established evidence-based design principles for developing user-centered LLM-powered educational applications, including structured prompt engineering frameworks, targeted users'feedback mechanisms, and adaptive difficulty progression.
2. Successfully bridged the gap between abstract clinical reasoning theory and practical LLM implementation by developing structured clinical scenarios with embedded decision points that align with established diagnostic reasoning frameworks.
3. Created a comprehensive LLM-powered clinical scenario simulation system capable of generating realistic patient cases with variable complexity, providing authentic clinical decision-making experiences with real-time feedback.
4. Advanced the technical architecture from a single-agent conversational interface to a sophisticated multi-agent clinical simulation environment featuring distinct AI-powered entities (patient, nurse, attending physician supervisors) that interact dynamically and respond contextually to learner decisions.

AI has tremendous and transformative powers for medical education; however, AI ethical challenges need to be paid attention as well. Bias, transparency, accountability, and privacy are the main ethical concerns in AI[2]. For this research, we have considered AI as well, so that in the implementation, the prompt engineering use the real patient cases to create the clinical scenario simulation, while we only use the conversation function from LLMs to build theinteractive conversation and learning environment.

There are limitations as well. The LLMs-powered educational scenario simulatoronly supports the text, while audio function of AI-patient and user is not supported yet. Some feedback from medical students requested the audio option. This will be our next work in the summer and further leverage the user-centered iterative design toenhance the app to meet the needs of medical students and physicians.

7. CONCLUSION

The evolution of AI and the promise of LLMson conversation tasksin medical education should requiremeeting human needs and evidence-based research.This study provides evidence with user-centered iterative design approach to design and develop an LLMs powered educational scenario simulator for clinical reasoning. It utilized surveys, face-to-face group interviews, and usage data to analyse the user's feedback and preferences to iterativelyrefine the LLMs powered educational scenario simulator. It contains two major iterations. The initial prototype (iteration 1) focused on harnessing LLMs via prompt engineering to create a conversational AI patient, allowing students to practice initial history-taking clinical skills. Feedback gathered through student surveys and focus group interviews revealed the need for a more comprehensive simulation mirroring the multi-modal nature of realistic clinical encounters.

Based on the feedback, the second iteration prototyped an updated version, which leverages the LLM-based multi-agent to build the educational scenario simulator for clinical reasoning.

Medical students can enter an adaptive interactive simulated clinical scenario to perform as a physician and have a diagnostic conversation with the AI patient agentand acquire physical exam data from the physical exam agent, collect lab and test data from the intervention agent, and consulate the guides from the AI supervisor agent in real-time.The conversation diagnosis process is summarized and evaluated by the AI evaluator agent. The feedback from the survey shows thatover 79%of participants expressed this app valuable for their clinical reasoning skills practice. Some feedback thatrequires adding the audio and video into the apps would be our next focus andpossibly adding the avatar is also in our future work list during this summer.In summary, the evidence of the successful design and implementation progress from an LLM-powered conversational app to multi-agent educational simulator through iterative cycles with physicians and medical students-input established a methodical and technical roadmap for integrating LLMs into medical education and advancing AI-enhanced HCI design and development.

ACKNOWLEDGEMENTS

We thank all the participants, physicians, medical students and manuscript reviewers for their time and feedback. We also gratefully thank the fundings from AMA AI initiative.This paper also used the generative AI, like Claude and ChatGPT, to polish and edit some structures and sentences.

REFERENCES

[1] Alaa Abd-alrazaq, Rawan AlSaad, Dari Alhuwail, Arfan Ahmed, Padraig Mark Healy, Syed Latifi, Sarah Aziz, Rafat Damseh, Sadam Alabed Alrazak, and Javaid Sheikh. 2023. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. JMIR Medical Education 9, 1: e48291. https://doi.org/10.2196/48291

[2] Faiza Alam, Mei Ann Lim, and Ihsan Nazurah Zulkipli. 2023. Integrating AI in medical education: embracing ethical usage and critical understanding. Frontiers in Medicine 10. https://doi.org/10.3389/fmed.2023.1279707

[3] Abdulmohsen H. Al-Elq. 2007. MEDICINE AND CLINICAL SKILLS LABORATORIES. Journal of Family & Community Medicine 14, 2: 59–63. Retrieved November 7, 2023 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3410147/

[4] Alexis Battista, Abigail Konopasky, Divya Ramani, Megan Ohmer, Jeffrey Mikita, Anna Howle, Sarah Krajnik, Dario Torre, and Steven J. Durning. Clinical Reasoning in the Primary Care Setting: Two Scenario-Based Simulations for Residents and Attendings. MedEdPORTAL 14: 10773. https://doi.org/10.15766/mep_2374-8265.10773

[5] Christy K. Boscardin, Brian Gin, Polo Black Golde, and Karen E. Hauer. 2023. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. Academic Medicine: 10.1097/ACM.0000000000005439. https://doi.org/10.1097/ACM.0000000000005439

[6] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. ACM Transactions on Intelligent Systems and Technology. https://doi.org/10.1145/3641289

[7] Salvatore Corrao and Christiano Argano. 2022. Rethinking clinical decision-making to improve clinical reasoning. Frontiers in Medicine 9: 900543. https://doi.org/10.3389/fmed.2022.900543

[8] Arthur S. Elstein. 2009. Thinking about diagnostic thinking: a 30-year perspective. Advances in Health Sciences Education: Theory and Practice 14 Suppl 1: 7–18. https://doi.org/10.1007/s10459-009-9184-0

[9] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative AI. Business & Information Systems Engineering 66, 1: 111–126. https://doi.org/10.1007/s12599-023-00834-7

[10] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research 25, 3: 277–304. https://doi.org/10.1080/15228053.2023.2233814

[11] Ethan Goh, Robert Gallo, Eric Strong, Yingjie Weng, Hannah Kerman, Jason Freed, Joséphine A. Cool, Zahir Kanjee, Kathleen P. Lane, Andrew S. Parsons, Neera Ahuja, Eric Horvitz, Daniel Yang, Arnold Milstein, Andrew P. J. Olson, Jason Hom, Jonathan H. Chen, and Adam Rodman. 2024. Large Language Model Influence on Management Reasoning: A Randomized Controlled Trial. 2024.08.05.24311485. https://doi.org/10.1101/2024.08.05.24311485

[12] Catharina M. Haring, Bernadette M. Cools, Petra J. M. van Gurp, Jos W. M. van der Meer, and Cornelis T. Postma. 2017. Observable phenomena that reveal medical students' clinical reasoning ability during expert assessment of their history taking: a qualitative study. BMC medical education 17, 1: 147. https://doi.org/10.1186/s12909-017-0983-3

[13] Krystal Hu and Krystal Hu. 2023. ChatGPT sets record for fastest-growing user base - analyst note. Reuters. Retrieved October 21, 2024 from https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

[14] Shreya Johri, Jaehwan Jeong, Benjamin A. Tran, Daniel I. Schlessinger, Shannon Wongvibulsin, Zhuo Ran Cai, Roxana Daneshjou, and Pranav Rajpurkar. 2023. Testing the Limits of Language Models: A Conversational Framework for Medical AI Assessment. 2023.09.12.23295399. https://doi.org/10.1101/2023.09.12.23295399

[15] Caroline A. Jones, Huma Gupta, and Matthew Ritchie. 2024. Visual Artists, Technological Shock, and Generative AI. An MIT Exploration of Generative AI. https://doi.org/10.21428/e4baedd9.b4f754fd

[16] Ganesh Chandan Kanakala, Sriram Devata, Prathit Chatterjee, and Udaykumar Deva Priyakumar. 2024. Generative artificial intelligence for small molecule drug design. Current Opinion in Biotechnology 89: 103175. https://doi.org/10.1016/j.copbio.2024.103175

[17] Mert Karabacak and Konstantinos Margetis. 2023. Embracing Large Language Models for Medical Applications: Opportunities and Challenges. Cureus 15, 5: e39305. https://doi.org/10.7759/cureus.39305

[18] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences 103: 102274. https://doi.org/10.1016/j.lindif.2023.102274

[19] Fangfang Liu, Yiyun Wang, Qiuling Feng, Linkai Zhu, and Guang Li. 2024. Optimizing E-Learning Environments: Leveraging Large Language Models for Personalized Education Pathways. 811–817. https://doi.org/10.2991/978-94-6463-502-7_86

[20] Charles Patrick Martin. 2024. Generative AI for Musicians: Small-Data Prototyping to Design Intelligent Musical Instruments.

[21] Dan Milmo. 2023. ChatGPT reaches 100 million users two months after launch. The Guardian. Retrieved October 21, 2024 from https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app

[22] Amr M. Mohamed. 2023. Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: perceptions of EFL Faculty Members. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11917-z

[23] Geoffrey Norman. 2005. Research in clinical reasoning: past history and current trends. Medical Education 39, 4: 418–427. https://doi.org/10.1111/j.1365-2929.2005.02127.x

[24] Gozie Offiah, Lenin P. Ekpotu, Siobhan Murphy, Daniel Kane, Alison Gordon, Muireann O'Sullivan, Sue Faye Sharifuddin, A. D. K. Hill, and Claire M. Condron. 2019. Evaluation of medical student retention of clinical skills following simulation training. BMC Medical Education 19, 1: 263. https://doi.org/10.1186/s12909-019-1663-2

[25] Sachin Tushar Patel, Syed Shah, Rhiya Prem Sood, Zohaib Siddiqui, and Iain McKay-Davies. 2021. The Implementation of Virtual Clinical Skills Teaching in Improving Procedural Confidence in ENT Trainees. Advances in Medical Education and Practice 12: 965–969. https://doi.org/10.2147/AMEP.S322965

[26] Dimple Patil, Nitin Liladhar Rane, and Jayesh Rane. 2024. Emerging and future opportunities with ChatGPT and generative artificial intelligence in various business sectors. In The Future Impact of ChatGPT on Several Business Sectors. Deep Science Publishing. https://doi.org/10.70593/978-81-981367-8-7_6

[27] Ruth Plackett, Angelos P. Kassianos, Maria Kambouri, Natasha Kay, Sophie Mylan, Jenny Hopwood, Patricia Schartau, Shani Gray, Jessica Timmis, Sarah Bennett, Chris Valerio, Veena Rodrigues, Emily Player, Willie Hamilton, Rosalind Raine, Stephen Duffy, and Jessica Sheringham. 2020. Online patient simulation training to improve clinical reasoning: a feasibility randomised controlled trial. BMC Medical Education 20, 1: 245. https://doi.org/10.1186/s12909-020-02168-4

[28] Carl Preiksaitis and Christian Rose. 2023. Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. JMIR medical education 9: e48785. https://doi.org/10.2196/48785

[29] Joseph Rencic, Steven J. Durning, Eric Holmboe, and Larry D. Gruppen. 2016. Understanding the Assessment of Clinical Reasoning. In Assessing Competence in Professional Performance across Disciplines and Professions, Paul F. Wimmers and Marcia Mentkowski (eds.). Springer International Publishing, Cham, 209–235. https://doi.org/10.1007/978-3-319-30064-1_11

[30] Madeeha Riaz. 2024. A personalized Learning system : education by AI. Retrieved November 21, 2024 from http://www.theseus.fi/handle/10024/859086

[31] Larissa IA Ruczynski, Marjolein HJ van de Pol, Bas JJW Schouwenberg, Roland FJM Laan, and Cornelia RMG Fluit. 2022. Learning clinical reasoning in the workplace: a student perspective. BMC Medical Education 22, 1: 19. https://doi.org/10.1186/s12909-021-03083-y

[32] Stuart Russell and Peter Norvig. 2021. Artificial Intelligence: A Modern Approach, 4th US ed. Retrieved October 15, 2024 from https://aima.cs.berkeley.edu/index.html

[33] Abdul Samad and Ahmad Jamal. 2024. Transformative Applications of ChatGPT: A Comprehensive Review of Its Impact across Industries. 1, 1.

[34] F. Scheele. 2012. The art of medical education. Facts, Views & Vision in ObGyn 4, 4: 266–269. Retrieved November 7, 2023 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3987478/

[35] Rebecca Schnall, Marlene Rojas, Suzanne Bakken, William Brown, Alex Carballo-Dieguez, Monique Carry, Deborah Gelaude, Jocelyn Patterson Mosley, and Jasmine Travers. 2016. A user-centered model for designing consumer mobile health (mHealth) applications (apps). Journal of Biomedical Informatics 60: 243–251. https://doi.org/10.1016/j.jbi.2016.02.002

[36] Jia Jia Marcia Sim, Khairul Dzakirin Bin Rusli, Betsy Seah, Tracy Levett-Jones, Ying Lau, and Sok Ying Liaw. 2022. Virtual Simulation to Enhance Clinical Reasoning in Nursing: A Systematic Review and Meta-analysis. Clinical Simulation in Nursing 69: 26–39. https://doi.org/10.1016/j.ecns.2022.05.006

[37] Thomas Thesen, Nsomma A. Alilonu, and Simon Stone. 2024. AI Patient Actor: An Open-Access Generative-AI App for Communication Training in Health Professions. Medical Science Educator. https://doi.org/10.1007/s40670-024-02250-2

[38] Arun James Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. 2023. Large language models in medicine. Nature Medicine 29, 8: 1930–1940. https://doi.org/10.1038/s41591-023-02448-8

[39] Bor-Yuan Tsai, Simon Stobart, Norman Parrington, and Barrie Thompson. 1997. Iterative design and testing within the software development life cycle. Software Quality Journal 6, 4: 295–310. https://doi.org/10.1023/A:1018528506161

[40] A. M. TURING. 1950. I.—COMPUTING MACHINERY AND INTELLIGENCE. Mind LIX, 236: 433–460. https://doi.org/10.1093/mind/LIX.236.433

[41] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In Advances in Neural Information Processing Systems. Retrieved October 27, 2024 from https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[42] Jacqueline Vaughn, Shannon H. Ford, Melissa Scott, Carolyn Jones, and Allison Lewinski. 2024. Enhancing Healthcare Education: Leveraging ChatGPT for Innovative Simulation Scenarios. Clinical Simulation in Nursing 87: 101487. https://doi.org/10.1016/j.ecns.2023.101487

[43] Hao Wei, Jianing Qiu, Haibao Yu, and Wu Yuan. 2024. MEDCO: Medical Education Copilots Based on A Multi-Agent Framework. https://doi.org/10.48550/arXiv.2408.12496

[44] Shahram Yazdani and Maryam Hoseini Abardeh. 2019. Five decades of research and theorization on clinical reasoning: a critical review. Advances in Medical Education and Practice 10: 703–716. https://doi.org/10.2147/AMEP.S213492

[45] The Nobel Prize in Physics 2024. NobelPrize.org. Retrieved October 20, 2024 from https://www.nobelprize.org/prizes/physics/2024/summary/

[46] The Nobel Prize in Chemistry 2024. NobelPrize.org. Retrieved October 20, 2024 from https://www.nobelprize.org/prizes/chemistry/2024/summary/