

Creating Virtual Patients using Robots and Large Language Models: A Preliminary Study with Medical Students

Alexander Borg
alexander.borg@ki.se
Karolinska Institute
Stockholm, Sweden

Ioannis Parodis
ioannis.parodis@ki.se
Karolinska Institute
Stockholm, Sweden

Gabriel Skantze
skantze@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

This paper presents a virtual patient (VP) platform for medical education, combining a social robot, Furhat, with large language models (LLMs). Aimed at enhancing clinical reasoning (CR) training, particularly in rheumatology, this approach introduces more interactive and realistic patient simulations. The use of LLMs both for driving the dialogue, but also for the expression of emotions in the robot's face, as well as automatic analysis and generation of feedback to the student, is discussed. The platform's effectiveness was tested in a pilot study with 15 medical students, comparing it against a traditional semi-linear VP platform. The evaluation indicates a preference for the robot platform in terms of authenticity and learning effect. We conclude that this novel integration of a social robot and LLMs in VP simulations shows potential in medical education, offering a more engaging learning experience.

CCS CONCEPTS

• **Computer systems organization** → **Robotics**; • **Information systems** → **Language models**; • **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Virtual Patients, Human-Robot Interaction, Large Language Models, Robots in Education

ACM Reference Format:

Alexander Borg, Ioannis Parodis, and Gabriel Skantze. 2024. Creating Virtual Patients using Robots and Large Language Models: A Preliminary Study with Medical Students. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3610978.3640592>

1 INTRODUCTION

Virtual patients (VPs) have been defined as interactive computer simulations of real clinical scenarios for educational or learning purposes [11]. This definition includes a variety of systems and platforms that address many learning outcomes in medical education curricula [18]. They can be designed in a linear fashion, where medical information and treatment options are gathered in a fixed order without consequences, or with a branched design, where



Figure 1: The Furhat robot as a virtual patient.

interactions also impact the outcome of the case. VPs have been proposed as an effective educational tool for practicing clinical reasoning (CR) in undergraduate medical education [8], while also ensuring that medical students encounter representative patient cases, which might not always be possible during their clinical rotations [10, 12, 25]. An essential aspect of VP simulation is the interface, which allows the user to interact with the patient in multiple ways and to receive the medical history of the patient. A common feature for VP platforms is that they include the possibility to ask a patient questions regarding their medical history, conduct physical examinations, gain access to laboratory analyses and test results, and based on this information suggest a reasonable diagnosis and treatment plan [7].

Previous research within VPs shows that, while medical students believe that VP simulation is a good educational activity to train CR, they acknowledge the lack of complexity and level of interactivity compared with authentic patients [9]. Interestingly, students have been shown to perform better at examinations and report a high level of satisfaction after using VPs compared to other forms of education, but the positive effect on clinical reasoning is still unclear [17, 24].

To increase the interactivity of VP interactions and to make them more realistic, the use of dialogue systems capable of natural language processing, embodied conversational agents [5], and social robots [15] have been explored. The recent development of large language models (LLMs) and their use in dialogue systems and social robots has opened up for new opportunities in this application area [1, 14]. However, while LLMs have recently been explored quite extensively for their use in medical applications in general (such as diagnosis) [6, 26], only initial attempts have been made to use them for VPs [23]. We are not aware of any previous work where they



This work is licensed under a Creative Commons Attribution-NonDerivs International 4.0 License.

have been applied, together with a social robot, for this purpose, and been evaluated in an actual educational setting.

In this paper, we present our ongoing work on developing a VP platform based on the social robot Furhat [2], together with LLMs. The primary use of the platform will be for clinical teaching in rheumatology for medical students at the Karolinska Institute (KI) in Stockholm, Sweden. Rheumatic diseases are among the major public health issues and affect about 10% of the population. In order for patients with rheumatic diseases to be adequately cared for at the correct level of care, it is essential that doctors in primary care receive ongoing education and that future doctors and other healthcare personnel receive evidence-based education in clinical reasoning for diagnostics and management. The clinical teaching in rheumatology for medical students at KI is currently located in semesters 5 and 6 and includes theory and practical training in inpatient or outpatient care for 1–2 weeks.

VPs are already used in the clinical teaching at KI. This is based on conventional computer-based platforms such as the Virtual Interactive Case Simulator (VIC), a software for developing VPs created by the Toronto General Hospital Department of Anesthesia Perioperative Interactive Education [3]. VIC has a semi-linear setting, which means that the beginning and the end of the case always stays the same, while the environment in-between can be explored in any desired order. A case begins with a short patient description and ends with a diagnostic decision and a management plan for the case through multiple choice. See Figure 2 for an example of the VIC user interface. Using this interface, the student can examine the medical history of the patient, perform physical examination and get laboratory test results. All interaction with this platform is done by clicking on different options, and there is no natural language processing involved. This existing platform constitutes a natural point of comparison for the robot VP developed in our project.

This paper is outlined as follows: We will start with a technical description of the initial implementation of the robot platform, then present the preliminary results from an ongoing study that started during the spring of 2023 as part of the clinical teaching program at KI, and conclude with some reflections and learnings.

2 VIRTUAL PATIENT IMPLEMENTATION

2.1 Robot

As an embodiment for the virtual patient, we use the Furhat Robot [2], as seen in Figure 1. Furhat has an animated face, back-projected on a translucent mask. The robot has a 3 degree-of-freedom neck, allowing for natural and flexible head movements. Furhat supports several different speech synthesis vendors, providing a wide variety of voices. This, combined with the animated face, allows for designing many different personas. In this way, Furhat is a suitable platform for the simulation of patients with different identities (w.r.t. gender, age, ethnicity, etc.). The animated face also allows for the expression of nuanced emotions (through facial expressions and gaze behaviour), which is important for conveying the internal state of the patient [22].

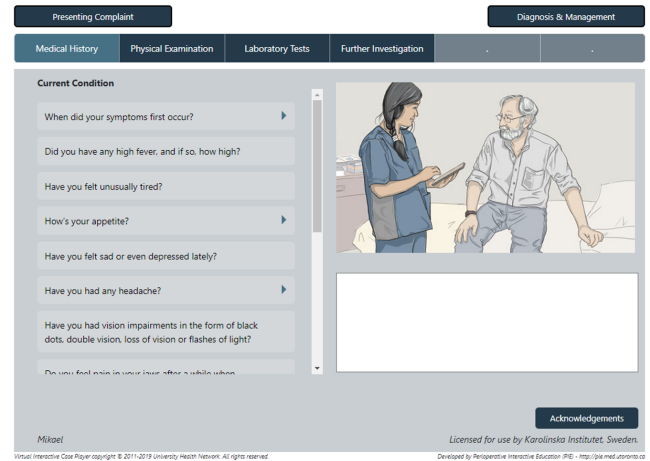


Figure 2: The traditional VIC user interface.

2.2 Dialogue system

The system was implemented using the FurhatSDK together with the chat completion model gpt-3.5-turbo from OpenAI [4]. However, as we wanted the patient utterances to be as realistic as possible, and not be in the style of an assistant, we did not use the chat completion API with the system/user/assistant message protocol. Instead, we only used one user message containing all necessary information, to generate the VP's response. This prompt consisted of a patient case description, the past 10 turns of the dialogue, and an instruction to write the next line of the dialogue, as seen in Figure 3. Note that the case description (in red) has been drastically shortened here for illustrative purposes. The full description consists of about 2000 tokens and thus takes up a large part of the available LLM context window of 4096 tokens.

An example of the resulting dialogue is shown in Figure 4. Apart from using the LLM to generate what the robot should say, we also use it to generate facial expressions for the robot (similar to [22]). This is important, since a patient might go through different emotional states in a doctor-patient dialogue, and this should be reflected in the patient's face. At the start of each utterance, numbered anchor points are inserted in the text at phrase boundaries. The LLM is then queried to insert suitable facial expressions at each anchor point, based on the list of available expressions in the FurhatSDK.

The long response delays that the use of LLMs are associated with can potentially lead to problems in turn-taking, where the user does not understand that the robot has heard their utterance but not responded yet [14]. This is even more problematic when, as in our case, there is no support for barge-in [27]. There are several potential mitigation strategies to this problem, such as using a turn-holding cue to indicate that the robot is "thinking" about what to answer [28]. We adopted a simple gaze aversion strategy [20, 21], together with changing the LED lights at the base of the Furhat robot to indicate its turn-taking state (i.e., speaking, thinking, or listening).

Mikael is about to meet his Doctor for the first time. Mikael is a 68 year old man who made an appointment a few days back. The only information the Doctor has, listed in the booking is that he is here because of "ache in the body".

The following is some information about Mikael's condition:

- He has always been healthy and have not felt any pain similar to this.
 - Sometimes he has back pain of course, but it's never been present for this long and with this character
 - In Mikael's medical charts it says that his blood lipids are above normal.
 - Mikael has been advised to eat healthier in order to avoid developing diabetes and also to lower his blood lipids.
 - Mikael has had a stable blood pressure since starting medicating with hypotensive drugs.
- <<SHORTENED>>

The following is a dialogue between Mikael and his Doctor:

Mikael: Hi doctor

Doctor: hello Michael my name is Morgan

Mikael: Nice to meet you, Morgan. Thanks for seeing me.

Doctor: no problem can you describe why you are here today

Write the next line that Mikael would say.

Figure 3: The prompt used for the LLM, with the initial part of the case description in red.

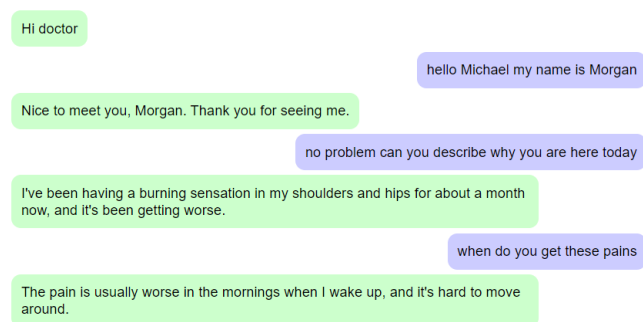


Figure 4: Example start of a dialogue between the student (in purple) and the VP (in green).

2.3 Dialogue analysis and Student feedback

After having the dialogue with the VP, it is possible to use the training session in different ways for pedagogical purposes. The dialogue itself (as seen in Figure 4) is saved and so both the student and the teacher can read them and use them as material for learning and feedback.

We have also investigated automatic analysis and generation of feedback, based on the dialogue. For this we also use the same LLM. After the dialogue is complete, it is sent together with a sequence of evaluation questions to the LLM. The responses from the LLM then forms an evaluation report (as seen in Figure 5) that can be seen by the student and/or the teacher.

3 PRELIMINARY STUDY

We conducted an interventional explorative pilot study to investigate the added value of our robot VP platform compared to a

Evaluation

Did Mikael talk about how long he had been suffering from his symptoms?	Yes, Mikael mentioned that he had been experiencing the burning sensation in his shoulders and hips for about a month.
Did Mikael talk about his usage of medications?	Yes, Mikael discussed the medications he was taking including Warfarin, Enalapril, and Atorvastatin, and why he was taking them.
Did Mikael talk about his heredity for medical conditions?	Yes, Mikael mentioned that his mother had some kind of rheumatism and struma, and his sister complains a lot about muscle pain, but he doesn't know if it's rheumatic.
Did Mikael talk about the character of the pain that he has been experiencing?	Yes, Mikael mentioned that the pain feels like a burning sensation and it is located in both his shoulders and hips.
Did the doctor ask about if the patient's symptoms have been varied during the time of day?	Yes, the doctor asked about the timing of the patient's pain and if it varied throughout the day.
Did Mikael talk about his own theory regarding his condition?	No, Mikael did not talk about his own theory regarding his condition.
Did the doctor ask about Mikael's fear or if he was afraid of some specific condition?	No, the doctor did not ask about Mikael's fear or if he was afraid of a specific condition.
Did the doctor ask about Mikael's wishes for the meeting?	No, the doctor did not ask about Mikael's wishes for the meeting.
Did the doctor reason on what medical condition that Mikael may have?	Yes, the doctor hypothesized that Mikael may have polymyalgia rheumatica and giant cell arteritis injury temporal arteries.

Figure 5: Example LLM analysis of a doctor-patient dialogue which can be used for feedback to the medical student.

conventional semi-linear VP platform (VIC, as described above) for clinical reasoning (CR) training. The study protocol has been reviewed and approved by the Swedish Ethical Review Authority (diary number: 2022-04437-01).

3.1 Participants

15 students from the sixth semester and international exchange students from the second half of the medical programme were recruited from a student elective course in rheumatology, conducted at the Karolinska University Hospital. The participants had an average age of 23.9 years (SD = 4.65). 9 identified as male and 6 as female. 11 reported that they had no previous experience with VP cases.

3.2 Procedure

A VP case of a rheumatic condition ("Mikael") was created in the robot platform and the traditional computer-based semi-linear platform (VIC), mentioned in the Introduction. All of the students performed the same case in both platforms for direct comparisons (balanced within-subject experimental design). The cases simulated a common rheumatic condition presented in a primary care setting, where students were assigned the role as the attending physician. The case was completed when students decided on a management plan for the patient, either by suggesting further investigation, perform diagnostics and/or initiate treatment.

After interaction with each VP platform, students immediately filled out validated indices of self-perceived accrual of CR skills

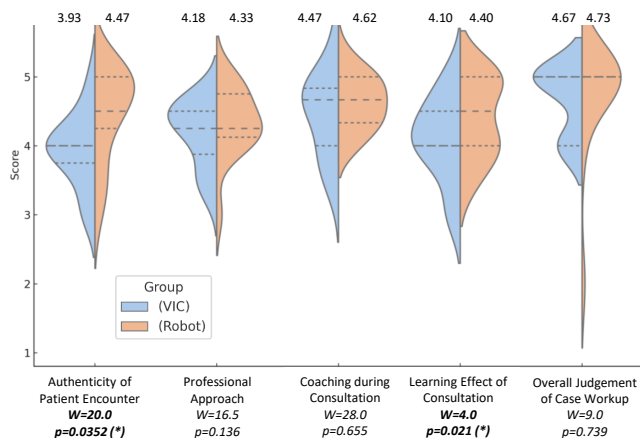


Figure 6: A violin chart showing the results of the quantitative subsets in the clinical reasoning questionnaire.

[13] and left comments regarding the different themes included in the indices. The questionnaire consists of 14 questions, divided into seven subsets: (1) Authenticity of patient encounter, (2) Professional approach in the consultation, (3) Coaching during consultation, (4) Learning effect of consultation, (5) Overall judgment of case workup, (6) Special strengths of the case, and (7) Special weaknesses of the case.

The first 12 questions (belonging to the five first subsets) were graded using a 5-point scale, whereas the last two questions (corresponding to subset six and seven) were answered with text.

3.3 Results

The answers to the questions for each of the five quantitative subsets were averaged per subset. A Wilcoxon signed rank test was performed to compare the Robot vs. the VIC condition. The results can be seen in Figure 6. Two of the subsets showed a significant difference, favoring the robot platform: *Authenticity of patient encounter* and *Learning effect of consultation*.

4 CONCLUSIONS & DISCUSSION

In this paper, we have presented our ongoing work on developing and evaluating a virtual patient (VP) platform, using a social robot for embodiment and a large language model (LLM) for driving the dialogue, as well as for displaying emotional expressions in the robot’s face, and for analysing the resulting dialogue. A pilot study with a limited number of students at the medical university was conducted, where we compared the robot platform with a traditional VP platform using a simple graphical user interface (VIC). The results show that the robot platform was perceived as being more authentic and to have better learning effects. Although this is just a preliminary study, the results are encouraging, as there are also a lot of room for improvement. To our knowledge, this is the first example of a VP utilizing both a social robot and LLMs, which has also been evaluated in an authentic educational setting with medical students.

The use of LLMs for this type of application is appealing, as it would be very hard to pre-program the robot with all the different

paths that these kinds of dialogue might take. Sometimes the patient behaved in a somewhat unexpected, but not necessarily implausible, way. For example, at one point the student recommended a certain medication, after which the VP asked in return whether there were any side effects to be worried about.

A general concern when using LLMs for dialogue systems is their tendency to “hallucinate”, i.e., to generate non-factual content [16]. In the case of virtual patients, this might not be as serious of a problem, since the main task of the agent is not to provide factual content, but to stay “in character”. Still, it is of course important that the agent displays behaviours and gives information that is in line with the intended case description and condition and does not, for example, report symptoms that a patient in this condition is not likely to have. For the preliminary study presented here, we did not see any signs of that, probably thanks to the very long and detailed case description in the prompt. Still, the problem of hallucinations might be an issue, and something that should be looked out for in future work.

To avoid having to use such a long prompt, a retrieval-augmented generation (RAG) approach could also be explored [19]. A challenge with that approach is that the response time would be longer, which might affect the naturalness of the interaction. Another option is to fine-tune the model on actual doctor-patient dialogues to achieve an even more realistic behaviour.

When it comes to learning outcomes, we think there is a lot of potential in this approach. The use of an LLM to automatically provide feedback on the dialogue (as discussed in Section 2.3) is a very promising method for conversational training in general, not just VPs. Given the face-to-face setting that a social robot allows for, we can also imagine the possibility to give feedback on not just the verbal aspects of the dialogue, but also non-verbal aspects, such as to what extent the doctor is smiling, the tone of their voice, and eye contact. Computer-based VPs have been shown to have mixed effects on communication skills [17], and it would be interesting to investigate if this can be improved with our platform.

When it comes to the evaluation, we acknowledge that this is just a preliminary study with a limited number of participants, so the results should be interpreted with caution. Also, the students had to interact in English, which was not the first language for all of them. However, supporting other languages should be fairly straightforward, given that LLMs such as those from OpenAI have fairly good coverage for other languages as well. We now plan a much larger scale study as part of the standard curriculum at KI. This will also involve a wider repertoire of case descriptions. One interesting direction could be to use a LLM to describe and generate a VP and then compare that to an existing VP case, that has already been validated as an effective tool to practice CR skills.

Finally, it should be noted that the VIC platform offers other aspects that are currently missing the robot platform, such as physical examination and laboratory tests. An interesting direction for future work could be to also add those aspects to the robot platform.

ACKNOWLEDGEMENTS

This project is funded by Region Stockholm (ALF, FoUI-977096), Karolinska Institute (FoUI-964139), and the Swedish Research Council (VR 2020-03812).

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 (June 2023), e48291. <https://doi.org/10.2196/48291> Company: JMIR Medical Education Distributor: JMIR Medical Education Institution: JMIR Medical Education Label: JMIR Medical Education Publisher: JMIR Publications Inc., Toronto, Canada.
- [2] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 7403 LNCS. 114–130.
- [3] Marie Jam Bravo, Miranda So, Cindy Natsheh, Gordon Tait, Zubin Austin, and Karen Cameron. 2019. Descriptive analysis of pharmacy students' impressions on virtual interactive case software. *American Journal of Pharmaceutical Education* 83, 1 (2019), 6821.
- [4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *arXiv 2005.14165* (2020). arXiv: 2005.14165.
- [5] Leonardo Campillos-Llanos, Catherine Thomas, Éric Bilinski, Pierre Zweigenbaum, and Sophie Rosset. 2020. Designing a virtual patient dialogue system based on terminology-rich resources: Challenges and evaluation. *Natural Language Engineering* 26, 2 (March 2020), 183–220. <https://doi.org/10.1017/S1351324919000329> Publisher: Cambridge University Press.
- [6] Jan Clusmann, Fiona R. Kolbinger, Hannah Sophie Muti, Zunamys I. Carrero, Jan-Niklas Eckardt, Narmin Ghaffari Laleh, Chiara Maria Lavinia Löffler, Sophie-Caroline Schwarzkopf, Michaela Unger, Gregory P. Veldhuizen, Sophia J. Wagner, and Jakob Nikolas Kather. 2023. The future landscape of large language models in medicine. *Communications Medicine* 3, 1 (Oct. 2023), 1–8. <https://doi.org/10.1038/s43856-023-00370-1> Number: 1 Publisher: Nature Publishing Group.
- [7] David A. Cook, Patricia J. Erwin, and Marc M. Triola. 2010. Computerized Virtual Patients in Health Professions Education: A Systematic Review and Meta-Analysis. *Academic Medicine* 85, 10 (Oct. 2010), 1589. <https://doi.org/10.1097/ACM.0b013e3181edfe13>
- [8] David A Cook and Marc M Triola. 2009. Virtual patients: a critical literature review and proposed next steps. *Medical Education* 43, 4 (2009), 303–311. <https://doi.org/10.1111/j.1365-2923.2008.03286.x> eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2923.2008.03286.x>.
- [9] Samuel Edelbring, Maryam Dastmalchi, Håkan Hult, Ingrid E Lundberg, and Lars Owe Dahlgren. 2011. Experiencing virtual patients in clinical learning: a phenomenological study. *Advances in health sciences education* 16 (2011), 331–345.
- [10] Samuel Edelbring, Ioannis Parodis, Ingrid E Lundberg, et al. 2018. Increasing reasoning awareness: video analysis of students' two-party virtual patient interactions. *JMIR medical education* 4, 1 (2018), e9137.
- [11] Rachel Ellaway, Chris Candler, Peter Greene, and V Smothers. 2006. An architectural model for MedBiquitous virtual patients. *Baltimore, MD: MedBiquitous* 6 (2006).
- [12] Grace Huang, Robby Reynolds, and Chris Candler. 2007. Virtual patient simulation at US and Canadian medical schools. *Academic Medicine* 82, 5 (2007), 446–451.
- [13] Sören Huwendiek, Bas A De Leng, Andrzej A Kononowicz, Romy Kunzmann, Arno MM Muijtjens, Cees PM Van Der Vleuten, Georg Friedrich Hoffmann, Burkhard Tönshoff, and Diana HJM Dolmans. 2015. Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. *Medical teacher* 37, 8 (2015), 775–782.
- [14] Bahar Irfan, Sanna-Mari Kuoppamäki, and Gabriel Skantze. 2023. Between Reality and Delusion: Challenges of Applying Large Language Models to Companion Robots for Open-Domain Dialogues with Older Adults. <https://doi.org/10.21203/rs.3.rs-2884789/v1>
- [15] Robert Johansson, Gabriel Skantze, and Arne Jönsson. 2017. A psychotherapy training environment with virtual patients implemented using the furhat robot platform. In *Proceedings of IVA*, Vol. 10498 LNAI. Stockholm, Sweden. https://doi.org/10.1007/978-3-319-67401-8_20 ISSN: 16113349.
- [16] Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large Language Models Struggle to Learn Long-Tail Knowledge. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 15696–15707. <https://proceedings.mlr.press/v202/kandpal23a.html>
- [17] Andrzej A Kononowicz, Luke A Woodham, Samuel Edelbring, Natalia Stathakarakou, David Davies, Nakul Saxena, Lorainne Tudor Car, Jan Carlstedt-Duke, Josip Car, and Nabil Zary. 2019. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. *Journal of medical Internet research* 21, 7 (2019), e14676.
- [18] Andrzej A. Kononowicz, Nabil Zary, Samuel Edelbring, Janet Corral, and Inga Hege. 2015. Virtual patients - what are we talking about? A framework to classify the meanings of the term in healthcare education. *BMC Medical Education* 15, 1 (Feb. 2015), 11. <https://doi.org/10.1186/s12909-015-0296-3>
- [19] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 9459–9474. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf
- [20] Chinmaya Mishra, Tom Offrede, Susanne Fuchs, Christine Mooshammer, and Gabriel Skantze. 2023. Does a robot's gaze aversion affect human gaze aversion? *Frontiers in Robotics and AI* 10 (2023). <https://www.frontiersin.org/articles/10.3389/frobt.2023.1127626>
- [21] Chinmaya Mishra and Gabriel Skantze. 2022. Knowing Where to Look: A Planning-based Architecture to Automate the Gaze Behavior of Social Robots*. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, Napoli, Italy, 1201–1208. <https://doi.org/10.1109/RO-MAN53752.2022.9900740>
- [22] Chinmaya Mishra, Rinus Verdonshot, Peter Hagoort, and Gabriel Skantze. 2023. Real-time emotion generation in human-robot dialogue using large language models. *Frontiers in Robotics and AI* 10 (2023). <https://www.frontiersin.org/articles/10.3389/frobt.2023.1271610>
- [23] Han Wei Ng, Aiden Koh, Anthea Foong, and Jeremy Ong. 2023. Real-Time Hybrid Language Model for Virtual Patient Conversations. In *Artificial Intelligence in Education (Lecture Notes in Computer Science)*, Ning Wang, Genaro Rebollo-Mendez, Noboru Matsuda, Olga C. Santos, and Vania Dimitrova (Eds.). Springer Nature Switzerland, Cham, 780–785. https://doi.org/10.1007/978-3-031-36272-9_71
- [24] Ruth Plackett, Angelos P Kassianos, Sophie Mylan, Maria Kambouri, Rosalind Raine, and Jessica Sheringham. 2022. The effectiveness of using virtual patient educational tools to improve medical students' clinical reasoning skills: a systematic review. *BMC medical education* 22, 1 (2022), 365.
- [25] Nancy Posel, James B Mcgee, and David M Fleischer. 2015. Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Medical teacher* 37, 9 (2015), 813–818.
- [26] Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Abubakr Babiker, Nathanael Schärli, Aakanksha Chowdhery, Philip Mansfield, Dina Demner-Fushman, Blaise Agüera y Arcas, Dale Webster, Greg S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2023. Large language models encode clinical knowledge. *Nature* 620, 7972 (Aug. 2023), 172–180. <https://doi.org/10.1038/s41586-023-06291-2> Number: 7972 Publisher: Nature Publishing Group.
- [27] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction : A Review. *Computer Speech & Language* 67 (2021), 101178. <https://doi.org/10.1016/j.csl.2020.101178> Publisher: Elsevier Ltd.
- [28] Gabriel Skantze, Martin Johansson, and Jonas Beskow. 2015. Exploring Turn-taking Cues in Multi-party Human-robot Discussions about Objects. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 67–74. <https://doi.org/10.1145/2818346.2820749>