

Learning with virtual patients in medical education

G.Tavarnesi¹, A.Laus¹, R.Mazza², L. Ambrosini²,
N. Catenazzi², S. Vanini², D. Tuggener³

¹ Lifelike SA, Chiasso, Switzerland

² Scuola Universitaria Professionale della Svizzera Italiana,
Information System and Networking Institute, Manno, Switzerland

³University of Zurich, Institut fur Computerlinguistik

*giulio.tavarnesi@lifelike.ch, andrea.laus@lifelike.ch,
riccardo.mazza@supsi.ch, luca.ambrosini@supsi.ch,
nadia.catenazzi@supsi.ch, salvatore.vanini@supsi.ch,
don.tuggener@gmail.com*

Abstract. VPL – Virtual Patient Learning is an online simulation system designed to train and assess clinical and relational abilities in a realistic and interactive problem-based learning scenario, where users (doctors and medical students) can interact and communicate with characters specifically designed to challenge their skills and facilitate the generation of learning objectives. Virtual Patients are designed to improve users' effectiveness in the areas of anamnesis, diagnosis, treatment and follow-up. Our Virtual Patient simulator is presented as an interactive movie based on pre-recorded clips with real actors, in which the learner plays the role of a physician. The dialogue between the simulated patient and the user takes place through selection from a set of questions and answers, or through a natural language processing system. At the end of each visit, the system provides a feedback, given directly through the patient's comments during a call with a friend. This feedback is based on the decisions and the communication strategy that the user applied during the last visit.

Keywords: virtual patient, medical education, machine learning

1 The context

1.1 Theoretical background

Simulation based learning (SBL) [1] is a rapidly growing paradigm in education. From the days of simple hyperlinked versions of textbooks, digital learning systems have evolved to complete simulation environments where students are placed in complex, life-like situations, and it has been shown that these systems provide definite advantages in terms of learning efficiency [2][3]. Medical colleges around the world adopt SBL by means of problem-based learning (PBL) approaches [4], where the student's learning is guided by the objectives they set themselves in a context of specifically-designed clinical case scenarios. A problem is usually presented to students in paper

form or in digital format such as a MS Word or PDF file, and then investigated and discussed in small groups over at least two sessions under the guidance of a tutor. Paper-based PBL is now being replaced by Simulation-Based Medical Education (SBME) which is an essential part of graduate medical education training. It provides a structured, learner-centred environment in which novice, intermediate, and advanced practitioners can learn or practice skills without causing harm to patients [5]. SBME tools are simulation environments that provide realistic representations of complex clinical scenarios. They can contribute considerably to improving medical care by boosting medical professionals' performance and enhancing patient safety.

1.2 The limits of current simulation tools

During the last decade, advances in computer hardware, software and graphics, allowed to refine medical simulators to offer life-like replications of medical procedures. Among the tools that have been developed, Virtual Reality (VR) simulators and behavioural simulations training system are the most significant. VR simulation for training has been introduced in medicine thanks to their minimally invasive procedures and reduced time for training. There are some examples of VR simulators to achieve competence in the domain of laparoscopy, endoscopy, endovascular surgery and urology [6-8]. Their main drawback is that they do not achieve high levels of fidelity in replicating human anatomy, physiology and bio-mechanical properties. In the context of behavioural simulations, a number of virtual patient simulators are currently available on the market: Campus Software [9], Decision-SIM [10], Open Labyrinth [11] and Web-SP [12] are the most known. All of them tend to present solutions widely based on static graphic design, such as cartoons or avatars, and are text-based with interactive deterministic questionnaires. Usually, little or no pedagogical feedback is given, mainly in the form of correct/wrong answers to the clinical questions or branching choices. Behavioural analysis and dialogic experience with the patient are usually missing.

1.3 Filling the gap

In this context, we have created Virtual Patient Learning (VPL), an online simulation system based on pre-recorded clips, in which the learner plays the role of a physician who meets a simulated but authentic patient, played by a professional actor who is trained to simulate variable moods, attitudes and emotional responses through verbal and non-verbal communication. The experience comprises a series of interconnected visits where the student has to take decisions according to the role and the resource of the doctor he is impersonating and the system provides different outcomes according to the choices made. The interview can be paused at any point for discussion, generation of learning issues and interaction with the tutor. At the end of each visit, a feedback based on the decisions and the communication strategy of the user is provided.

Besides the well-known advantages of training simulation systems (they undoubtedly prove to be safer, shorten product development, and are cost-effective), the main contribution of our VPL is the introduction of a high degree of realism into a simulation

system. As described in section 3.6, the integration of a voice-based interface brings users the perception of a simulated experience closer to the real experience. Users can interact with the training system in a natural way and this allows them to improve not only their problem solving and critical thinking skills, but also their relational and communication skills. Furthermore, the use of a human-machine interface that mimics natural interaction provides improved learning efficiency, as demonstrated by the preliminary evaluation results described in section 4.2. Traditional interaction with computer systems is performed using mouse, keyboard and physical screen. This forces users to understand what and how to execute operations and causes serious challenges to users not directly involved with computer systems, such as doctors and medical operators. Preliminary experiments showed that with the new voice-based interaction, students were able to complete a simulation without previous training.

To enable natural human-computer interaction, the main scientific challenge is to match a transcribed speech input to a set of closed questions (canned texts). In NLP, this task is defined as paraphrase classification and it is mostly approached by modeling similarity between words of a text through vector representation of words, called word embeddings [13,14]. As described in section 3.6, we tackled this challenge by implementing an NLP component that uses a combination of word overlap metrics, semantic similarity and importance of words to compute similarity between speech inputs and canned texts and to evaluate the suitability of the matching.

2 Our solution

2.1 Aims and key challenges

Our goal with VPL – Virtual Patient Learning was to bring a real, authentic patient to the student from the first year of their study, giving them the possibility to simulate a realistic doctor-patient interaction, and see the outcome of their clinical decisions, even if they were sub-optimal or negative. We have designed it to be an excellent tool for medical education that triggers discussion in the classroom around some specific learning objectives and stimulates better clinical reasoning in the student. Some examples of learning objectives are: "Symptom and sign patterns of hyperthyroidism", "Principles of management of hyperthyroidism", "Identify Risk factors of lung cancer", "Understand the pathology and classification of lung cancer", etc.

One of the key challenges we set ourselves was to place the student in front of an authentic person, with realistic feelings, who can interact in real time and have authentic reactions to the user's input (Fig.1).



Fig. 1. The virtual patient is waiting for the user's input.

2.2 Interacting with the patient

Users can interact with the virtual patient by selecting one question or answer from a list (Fig. 2). The selection of the questions/answers set is managed by an algorithm that provides from 2 to 6 options according to the criteria we will describe here below. The objective is to not limit the students and let them be free to build the medical interview deciding the order and the topic of the questions (from a pre-made set in the canned text version). So, the algorithm initially provides many questions and then it acts as a "funnel" by reducing their number while we get closer to the solution. When one question generates an answer from the patient that needs to be further explored, the list is populated with one or more questions connected to the last one, generally presented as #1 and #2. When the story comes at a point where it is necessary to ask for a difficult/embarassing question or give bad news, the students are presented with at least 2 possible options with the same content but different wording. Selecting the correct one will improve the "trust score", while a wrong one will lower it. A clinical board provides, during the design phase, those options based on the most common mistakes doctors can make in a similar situation (like being too rude or too technical, lying to the patient instead of giving a bad news, etc).

In some cases, questions can be automatically hidden or shown in relation to mistakes made in the previous visit (like selecting a too rude answer) or considering the "trust score" that the student gained until that point. In the version for Doctor's learning more algorithms are used to provide a more difficult scenario and offer a greater number of possible outcomes of the story, by considering also the drugs prescribed, the dosage, the advices given, etc. Learner's non-verbal and tone of voice are currently not analysed by the software, while patient's mood and behaviour is related to the current outcome of the story: if his symptoms are improving, he will be nice and relaxed, but if the "trust

score" is low or the student selected a wrong question/answer, he will react accordingly by being rude, worried, uninterested, etc. The interface also offers the option to order tests, do a physical examination and consult the clinical diary (Fig. 3).

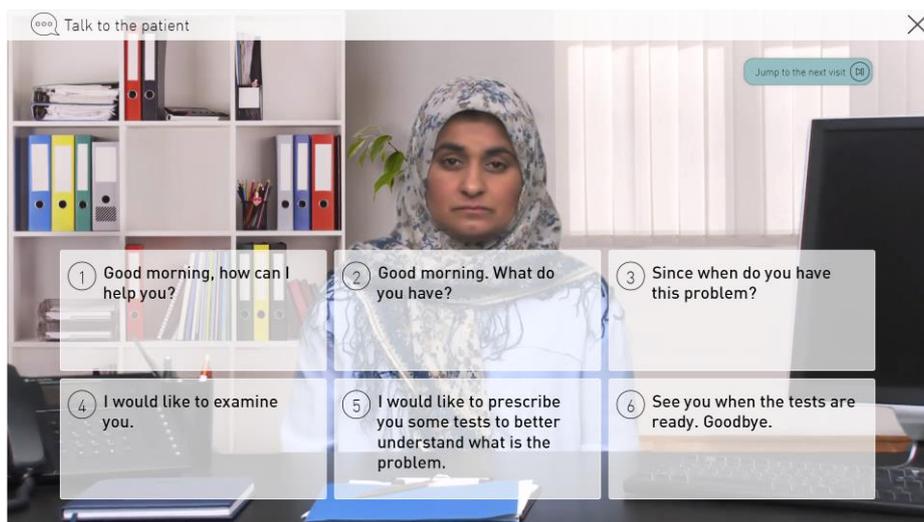


Fig. 2 A list of pertinent questions or statements is presented by the A.I.

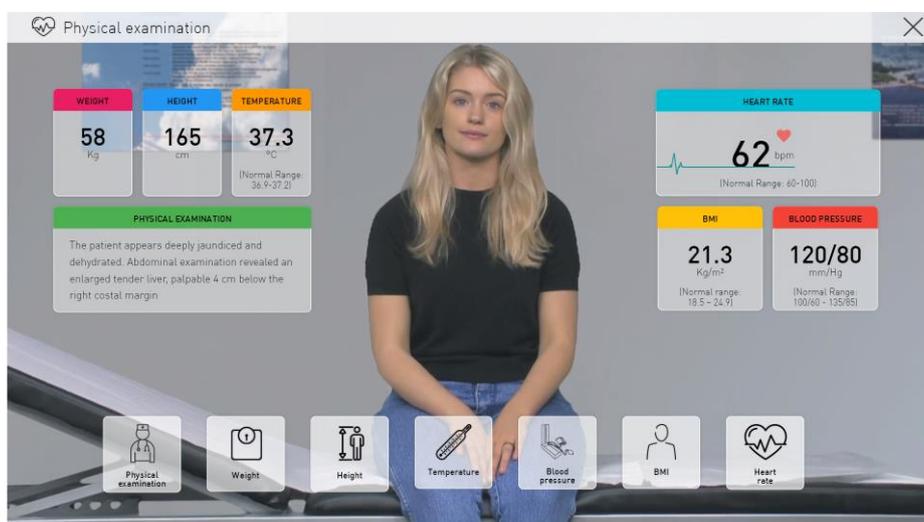


Fig. 3 Carrying out a physical examination on a virtual patient.

2.3 A life-like scenario to train clinical and relational skills

Virtual Patients in the VPL environment are designed to improve the user's effectiveness in the areas of anamnesis, diagnosis, treatment and follow-up. There is also a strong focus on establishing a trusting relationship with the patient, as a part of the

treatment and healing of the patient. Each simulation is completed by a series of interconnected visits, where the user can impersonate a General Practitioner or a Specialist, according to the learning objectives of the case. In each visit, the student is required to take decisions according to the role and the resources of the doctor he is impersonating. Differently from the paper-based version, where the students just read the story and have to imagine the situation, the discussion with the patient and his reactions, here everything happens in front of them, in a first-person view and in a (simulated) real-time scenario, where also the time limitations have to be considered (for example they can't have a laboratory test result like a blood test immediately: they will have to send the patient home, wait for him to do the test and analyse it then, he comes back for a second appointment. This means that they will not have this information in the first visit and will have to think about what decision they can take without it, just how this would be in real life).

Being a tool based on Problem Based Learning, the objective is not solving the problem, but learning from the problem, under the guidance of the tutor. So, every scenario is designed to generate discussion around specific topics, and the contents does not change according to the year of the curriculum, mainly because the curriculum usually differs from country to country. The tutor selects the scenario to use according to the learning objectives it is tagged with, and/or the disease of the patient. This makes possible to use this tool not only under a PBL curriculum, but also in similar models like Case Based Learning or Team Based Learning. Since every case basically tells the story of a patient, it can be also used to support classical lectures.

Currently there is not an authoring tool, but we are developing a solution that will let each college of medicine adapt the cases to their needs by modifying test results, questions etc.

2.4 The importance of authenticity

Authenticity is the key to practice working on a realistic experience and on the emotional aspects, and it is one of the key aspects in the learning design of each VPL simulator. With authenticity we mean that the simulated experience is as close as possible to a real one since it contains the same elements, resources and limitations that the user would have in his everyday clinical practice. For example, we use real-life test results coming from real patients, chosen to allow the student to gain genuine experience of a real test and to discuss how to interpret it, and we consider the time needed to obtain them, so the students will not take "shortcuts" that in real life they can't have. Moreover and most importantly, the reaction of the patients, the words he uses to describe the symptoms, and his opinions after each visit are described using a common and natural wording. The script is provided by a SME (subject matter expert) that usually meets this kind of patients and reports us real-life episodes and wording. Also the storytelling is designed around typical situations described by the clinical board (i.e. an uncooperative or difficult patient, the presence of a caregiver, the personal fears and beliefs of the patient, etc).

2.5 Feedback from the patient

At the end of each visit, the system provides feedback directly through the patient's comments, for example during a telephone call between the patient and a friend (Fig. 4). In this call we can see the patient sharing their impressions on the recent appointment with the doctor, expressing their fears, trust, etc. The feedback is based on the decisions and the communication strategy that the user applied during the last visit or visits. The patient will evaluate the relational aspects such as caring, kindness, empathy etc. This is an excellent opportunity for learning about relational outcomes, thanks to the high emotional impact that these kinds of video have. We decided to implement it this way to show the students that sometimes what the patient says in front of a doctor may be different from what he really thinks. Moreover, showing his opinion once at home is a trigger for discussion in the classroom, and a source of guidance and feedback in self-directed learning.



Fig. 4 A patient discussing the last visit with a friend.

3 Deployment

3.1 A library of virtual patients

VPL – Virtual Patient Learning is offered as a library of simulators, covering all the body systems and the most common diseases. Currently, all the scenarios are in English language. Each patient is mapped based on the learning objectives and topics to be covered and discussed.

All the patients included in the library are of differing ages, ethnicities and personalities. This is of great benefit to the students, as they have the possibility to experience different situations and deal with different realities and people, bringing this virtual experience even closer to real clinical practice.

Each case is created by an expert in the subject matter and revised by another specialist in the same clinical area.

3.2 Scalability and cloud

The solution is entirely cloud-based: users simply need an Internet connection and a browser. There is no need for specific, proprietary hardware and it can be used anytime, anywhere, even by hundreds of users at the same time from all around the world.

This makes this tool extremely scalable, unlike other classical solutions like role-play, which require real actors to be present at a specific time and place.

3.3 Learning and assessment

Virtual patient learning can be used in two modes: a ‘learning’ mode and an ‘evaluation’ mode. The learning mode has been developed to stimulate student-centred learning and is linked to a large number of resources such as radiological images, laboratory results, videos and management guidelines. The evaluation mode can be used for evaluating the student’s decision-making and communication skills.

3.4 Doctors training

This tool has also been used for CME - continuing medical education in Europe for physicians and specialists. The learning design in this case was quite different. The learning itself was self-directed, since no tutors were involved, and the paradigm was not problem-based learning, but focused more on problem solving. Doctors were presented with a complex case and were supplied with correlated lecture notes to help them solve the problem in the best possible way. The learning process also incorporated a degree of trial and error. Completing a test was required to receive the CME credits.

In some cases, our virtual patients were used during congresses, to practise the topics presented during the first part of the day and debate the outcomes of the simulations done by the users.

3.5 Students training

To date, VPL is used in some colleges of medicine in the UAE, USA and Poland. Tutors are trained in using the tool and receive a tutor guide, containing useful information about the case and enabling them to manage the session even if they are not a subject matter expert. Students go through a case in small groups of 6-8 to generate learning objectives and discuss specific topics within each case. Usually the session with the simulator and the related discussion takes place at the beginning of the week, later the learners study the learning objectives they have generated, and at the end of the week they meet again to share what they have learnt and repeat the simulation according to the new knowledge they acquired. This scheme may strongly differ according to the methodology that each college of medicine decides to adopt.

3.6 Adding interactivity with voice interaction and NLP

Since VPL relies on authenticity as one of its main learning enhancement features, the Virtual Patient has been extended with a voice-based interface that allows users to simulate a normal appointment with a real person, complete with voice and natural language, thus enabling users to behave in more natural ways without the need for keyboards or other input devices.

Accordingly, we integrated a speech recognition system and built an NLP (natural language processing) mapping component that analyses the sentence spoken by the user and automatically selects the best response. The tool for speech-to-text conversion is Google Cloud Speech [15], which uses neural network models to translate audio into text. This platform was selected after an evaluation process that took into account the accuracy of conversion, the time required and the cost of the related license. The mapping task is performed through matching of the transcribed speech input with the available questions at a specific moment in the story, and is carried out over two tasks:

1. Ranking of all available questions in terms of their similarity to the speech input (question ranking).
2. Checking of whether the highest-ranked question is indeed a suitable match for the input (question matching).

For the first task, we employed a machine learning-based approach (classifier) that takes into account surface features (string overlap ratios), semantic sentence similarities (based on the mapping of words to vectors) and how appropriate it is to ask a given question according to a template for the exchange.

To collect the data to train the algorithms, we created a version of the system that features a text box input, where students were asked to write the questions they wanted to put to the virtual patient. The available canned texts were then displayed, and they had to select which, if any, of the canned texts was a good match for their input. This provided us with data for developing our approach to question ranking and examples of non-matching inputs for developing the question matching component.

With regard to question matching, our initial approach was to use the probability assigned to the question pair by the classifier (e.g. reject the match if the probability is below 0.5). However, we found in early experiments that there is no single suitable threshold to reject a student input/canned text pair. Therefore, we employed a multi-layer perceptron neural network [16] and used as features a value indicating if the N most important words (words that occur infrequently in a canned text are given more importance) appear both in the input and the matching question, and a value representing sentence similarity, which is computed through comparison of vector representations of sentences obtained using neural networks.

4 Lesson Learnt and open challenges

4.1 Survey

We conducted in November 2016 a survey on students and tutors at 3 different colleges of medicine (Lebanon, Qatar, UAE) after a session with a VPL simulator. The sample was completed by 118 students in the second year and 59 tutors. We used questionnaires with pre-defined questions and multiple-choice answers, where the students were presented with statements like the ones showed in the chart below, and they had to choose an answer in a scale of 4 from "I strongly disagree" to "I strongly agree". For all the statements, we asked them for their opinion on using a VPL simulator instead of the classical PBL paper cases. The outcomes of the survey were encouraging since it showed a very good feedback about the benefits of this methodology vs the classical tools: users think VPL is actually useful for enhancing the learning experience, promoting discussion among students and aiding information and content retention. One of the most appreciated elements was the possibility to interact with a real patient quite some time before the classical curriculum, and the presence of real-life test results. Some tutors were critical about the timing, as they felt that sometimes working on a paper case is quicker than using a simulator. We have taken this feedback into consideration to improve the system in the next release.

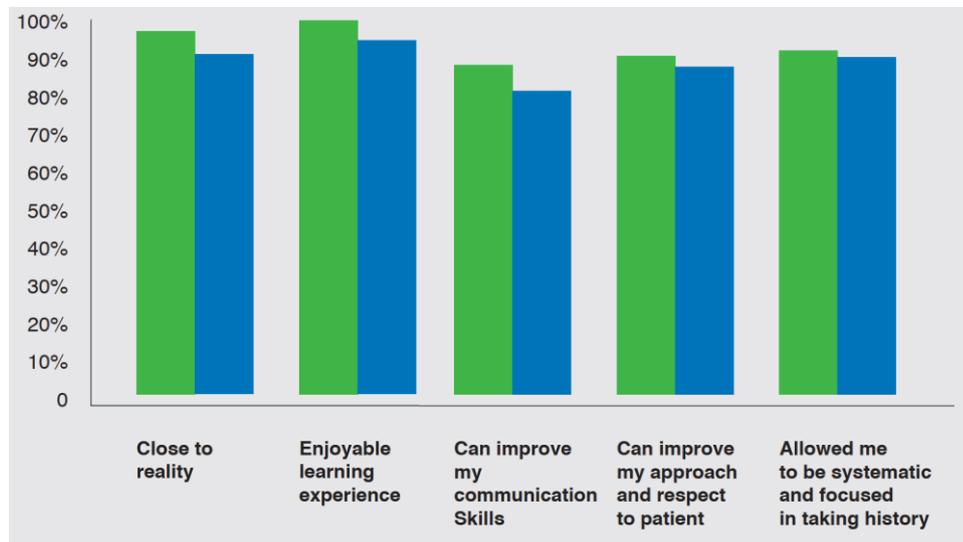


Fig. 5 Results of the survey on VPL use in medical education. In green the students' answers and in blue the faculty's ones. The graph shows the percentage of people answering "I strongly agree" to each statement.

4.2 Preliminary usability test on NLP voice interaction

We also carried out in October 2017 a preliminary usability test of the voice and natural language interaction with 8 medical students from two different colleges of medicine: 3 fourth-year students from the University of Insubria, Varese (Italy) and 5 third-year students from the Gulf Medical University, in Ajman (United Arab Emirates, UAE).

The objectives of the test were to collect their feedback about the voice-based interaction modality with respect to the mouse-based interaction modality. In particular, our aim was to investigate whether the new system provided suitable solutions to the typical issues of a voice-based interface system, which are *discoverability* (how does the user know what to say?) and *learnability* (how easy is it for users to accomplish tasks without previous training?) [17, 18].

We conducted some experiments with the medical students to assess if they were able to complete a simulation using voice and to find out the questions to ask.

The tests were organised in three steps:

- *welcome phase*, to explain the reason of the participants' involvement and to make a demo of the two versions of the simulator;
- *user testing phase*, where each participant was asked to interact with the voice-based version of the VPL and complete at least one visit. Each session lasted for about 45 minutes. During this activity, an observer took note of the participant's behaviour, comments and difficulties;

- *focus group*, to ask students questions concerning their experience using the simulator: opinions, problems, satisfaction, user friendliness, recommendations and suggestions for improvements.

The main results of the test and the following focus group are summarised below.

Students were generally able to complete a simulation using voice; they rarely felt stuck because they knew what to say and when, since they were trained to follow the scheme of the medical interview. They appreciated that the system showed the textual alternatives only after various failed attempts to interact with voice.

In comparing the previous interaction modality with the new voice-based one, students preferred the new one, because it is more realistic.

In general, the reaction of students was very positive: they considered the system interesting and enjoyable; they appreciated the proposed training approach because it allowed them to learn how to ask questions in the right way and would have liked to use such a tool in their curriculum. The interface was judged clear and clean, there were no doubts about how to start the simulator, conduct the visit, ask for tests, and exit.

In spite of the overall positive opinions, participants encountered some difficulties during the test and proposed a number of suggestions for improvements. Some problems were connected to *non-matching events*: this was because the system did not have answers to some questions useful for a complete anamnesis of the patient (e.g. the question “do you have children?” is not one of the pre-designed ones, so there is not a specific video for that) or did not correctly transcribe the speech input (the spoken sentence was not correctly converted by the speech-to-text system due to a poor pronunciation or ambient noise). The ideal solution to solve these issues would be to extend the set of questions the system is able to understand and to use the simulator in a quiet environment. Participants also encountered few cases of *wrong matching*: the matcher sometimes selected a wrong canned text with no relation with a speech input. To reduce this problem, we used canned texts that contain as many words with unique sense as possible. Additional tests have shown that this restriction is tolerable from a user experience perspective and the system achieved high acceptance.

In the future we plan to further assess the system, by involving more students and using additional tests: for instance an A/B testing to compare the voice-based interaction modality with the mouse-based one and some standard user-experience questionnaires, such as SUS (System Usability Scale) or SASSI questionnaire, which is specific to assess voice interaction design [19].

4.3 Ongoing challenges and next steps.

VPL has been appreciated in all the contexts we have used and proposed it. Nevertheless, we are in a process of continuous development to provide medical students from around the world a tool that can be as close as possible to an authentic experience with a real patient. By staying in close and continuous contact with the tutors we are able to gather in-the-field feedback that we can use to improve future releases of the tool.

In particular, future challenges will include providing a larger number of cases to allow students to train with hundreds of different cases and patient personalities, and to expand and diversify the use of this tool to the training of nurses, pharmacists, caregivers and other clinical professionals.

Another challenge will be to improve the discoverability of topics with a better in-game help system (there is a very basic version at present) and a wider range of topics that can be managed by the NLP engine.

Moreover, we are planning to further improve the richness of the dialogue in the exchanges, to embrace still more relational styles and to provide different outcomes.

We are also currently testing on small groups some custom-made AI algorithms to improve the user experience of the students and encourage the community-based knowledge creation. While the students interact with the simulator, the system learns what are the most common interview paths and what are the learning objectives generated as a result of those path, also measuring the time taken to generate them (intended as number of questions and not the runtime of the simulation). Later, the list is compared to the expected learning objectives of the case and the questions are ranked by efficacy and popularity.

As a result, students can find in the simulator optimized medical interview paths, based on of the collective community-generated best practices.

References

1. S. Barry Issenberg, William C. Mcgaghie, Emil R. Petrusa, David Lee Gordon, and Ross J. Scalese. Features and uses of high-fidelity medical simulations that lead to effective learning: a beme systematic review. *Medical Teacher*, 27(1):10{28, 2005.
2. Margaret Bearman, Debra Nestel, and Pamela Andreatta. *Oxford Textbook of Medical Education*. Oxford University Press, 2013.
3. De Ascaniis S., Cantoni L., and Sutinen E. and Talling R. A lifelike experience to train user requirements elicitation skills. In *Design, User Experience, and Usability: Understanding Users and Contexts. DUXU 2017. Lecture Notes in Computer Science*, vol 10290. Springer, Cham, 2017.
4. Diana F Wood. Problem based learning. *BMJ*, 326(7384):328{330, 2003.
5. Margaret Bearman, Debra Nestel, and Pamela Andreatta. *Oxford Textbook of Medical Education*. Oxford University Press, 2013.
6. Malone HR, Syed ON, Downes MS et al (2010) Simulation in neurosurgery: a review of computer-based simulation environments and their surgical applications. *Neurosurgery* 67:1105–1116.
7. Bashir G (2010) Technology and medicine: the evolution of virtual reality simulation in laparoscopic training. *Med Teach* 32: 558–561.
8. Neequaye SK, Aggarwal R, Van Herzeele I et al (2007) Endovascular skills training and assessment. *J Vasc Surg* 46: 1055–1064.
9. Cosima Jahnke, Albrecht Elsasser, Gudrun Heinrichs, Rudiger Klar, Christoph Bode, and Thomas K. Nordt. Neue wege in der kardiologischen aus- und weiterbildung. *Medizinische Klinik*, 101(5):365{372, May 2006.
10. <https://www.kynectiv.com>

11. <http://openlabyrinth.ca>
12. <http://websp.lime.ki.se>
13. Jerey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532{1543, 2014.
14. John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. From paraphrase database to compositional paraphrase model and back. *TACL*, 3:345{358, 2015.
15. <https://cloud.google.com/speech/>
16. Minsky, Marvin, and Seymour A. Papert. *Perceptrons: an introduction to computational geometry*. MIT press, 2017.
17. Eric Corbett and Astrid Weber. What can I say?: addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceeding MobileHCI '16 Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Italy — September 06 - 09, 2016*, pages 72–82, 2016.
18. Laura Klein. *Design for voice interfaces, Building products that talk*. O' Reilly Media, 2016.
19. Lewis, J. R. (2016) Standardized Questionnaires for Voice Interaction Design, April 2016, Vol 1, Issue 1, *The Journal of AVI, Voice Interaction Design*