Chameleon - A Framework for Developing Conversational Agents for Medical Training Purposes

Demonstration Track

Al-Hussein Abutaleb
University of Aberdeen
Aberdeen
alhussein.abutaleb@gmail.com

Bruno Yun
Department of Computing Science
University of Aberdeen
bruno.yun@abdn.ac.uk

ABSTRACT

Objective Clinical Structured Examination (OSCE) is used to assess multiple competencies in medical pedagogy such as efficiently eliciting relevant clinical history from a patient. Clinical interviewing is done in a question and answer fashion, making it amenable to computer simulation. We introduce Chameleon, a framework to create virtual patients in an OSCE setting using the conversational agent platform Dialogflow CX. Our framework consists of a generic chatbot that is capable of answering most questions (in a classic non-specific clinical interview) and which can be expanded to capture any clinical presentation, e.g. a patient with backpain.

KEYWORDS

Chatbot; OSCE; Education; Social Simulation

ACM Reference Format:
Al-Hussein Abutaleb and Bruno Yun. 2021. Chameleon - A Framework for Developing Conversational Agents for Medical Training Purposes: Demonstration Track. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 VIRTUAL STANDARDISED PATIENTS

Introduced in the 70's by Harden et al., the Objective Structured Clinical Examination (OSCE) is the universally adopted framework for assessing various clinical competencies in medical schools globally. The OSCE resolved problematic issues around objectivity, disparity of outcome, relevance, and fairness of previous assessment methods in medical pedagogy [6–8, 12, 16]. A key component of a "modern" OSCE is the interaction that takes place between the student and the Standardised Patient (SP); often a trained actor who plays the role of a patient with a predetermined clinical presentation. The use of SPs is considered the gold standard in OSCEs and adds an element of safety; mistakes are better made with an actor than with a patient [1]. The conversational flow between the SP and the student is structured, i.e. the student tries to ask the appropriate questions; the answers of which hint at what might be causing the patient's symptoms [11]. In this work, we follow the guidelines of taking clinical history discussed in [11], i.e. we consider a 6-steps conversational flow:

(1) Presenting Complaint (PC): The student asks the SP about the complaint, the intensity of symptoms, onset, triggers, prior incidences, amongst others.

(2) Family History: The student tries to establish if there's a genetic component to the PC. An example question is: "Has anyone in your family suffered from similar symptoms?".

(3) Past Medical History: This student attempts to establish whether the current clinical presentation is related with past medical problems the patient has suffered from.

(4) Drugs and Allergies History: The SP is asked about current drugs they are taking and any allergies they suffer from.

(5) Social and Lifestyle History: The SP is asked about their personal and professional life to discern clues about the cause of the patient's symptoms. For example, a patient who works as a construction worker complaining of persistent shortness of breath may warrant investigating whether they have been exposed to any dust or chemicals recently. Whereas a 16 years old active teenager who just finished high-school complaining from the same symptom is more likely to be suffering from panic attacks.

(6) Ideas, Concerns and Expectations: The interview concludes with the student asking the patient about their concerns, ideas, and expectations [14].

This 6-steps approach to history taking is used with real patients and information gained from the clinical interview will contribute 80% toward any diagnosis the doctor makes in practice [11].

With the on-going coronavirus pandemic and associated social distancing measures, the acceptance of blended learning in many universities, the high costs of training SPs, and the under-utilised power of OSCEs as a pedagogical tool a window of opportunity presents itself to use modern AI tools to fill an existing need in medical education. A need that is more pressing with global disruption to education caused by COVID19 [4, 5, 10]. Although there are existing virtual SP chatbots to train student in and outside of an OSCE context [2, 9, 15], they required a lot of handcrafted conversational data and must be trained from scratch for each different clinical presentation. In this paper, we present the first framework for developing varied conversational agents that simulate the role of the SP based on Google's Dialogflow CX. This novel framework is based on (1) a co-created generic medical chatbot and (2) the associated online platform to specialise it into any clinical presentation, hence we named this framework: *Chameleon*.

2 CHAMELEON: A GENERIC CHASSIS

Chameleon's reasoning engine is powered by Google Dialogflow CX and will match the student's textual question (e.g. "Do you have diabetes?") to the meaning behind it (its intent). For example, the intent for the aforementioned question would be "enquire.diabetes".

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

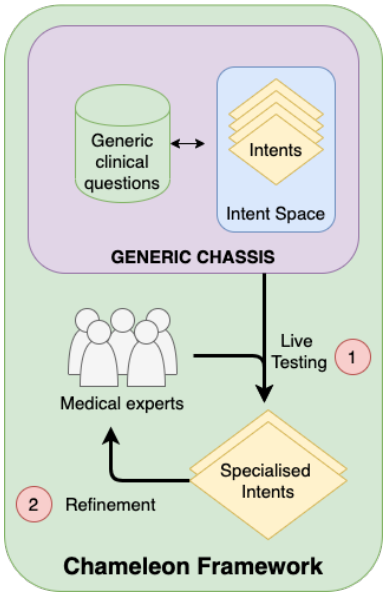

Figure 1: The Chameleon development framework.

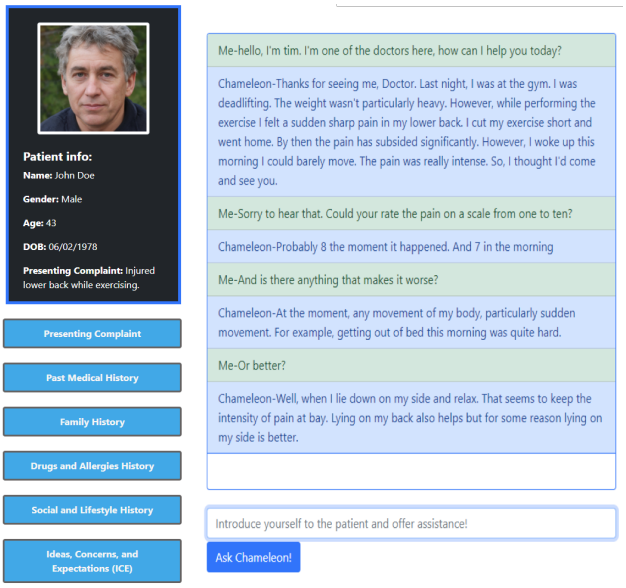

Figure 2: Screenshot of the Chameleon online interface.

An intent can have more than one associated question since semantically identical questions can be phrased differently.

We co-created, via interviews with medical experts from the Scottish NHS, a question dataset containing 2638 common and generic clinical questions [11]. This dataset was used as the training corpus for the generic chatbot. Then, for each question in that dataset, we extracted its intent which resulted in a set of 180 intents which we call the *intent space*. We designed the intent space to capture the meaning of questions that are generic and which could be asked during a variety of clinical presentations.

This generic chassis allows for Chameleon to be easily expanded to simulate a specific medical condition such as back pain or even mental illnesses. This can be done by collecting new specific questions (and related intents) for that particular clinical presentation during successive live testing with experts on our online platform (see Figure 1). Chameleon, our proposed framework to build a conversational agent, is distinguished by: (1) the short development time to create new SPs, (2) its capability to play the role of the SP, and (3) its ability to shape shift to reflect any medical conditions.

To illustrate the feasibility our approach, a demo version that simulates a specialised SP with lower back pain has been deployed[1] and a demo video is available at https://youtu.be/QWdDZYU8JKg. Figure 2 shows the interface of the demo which is composed of three main parts: (top-left) A short presentation of the SP including a computer generated picture (consistent with gender and age) and the presenting complaint, (right) the chat log, and (bottom-left) a navigation menu to move between the 6 steps of the conversational flow. Note that this interface was co-created with medical experts and informed by user requirements. The structured interaction with the SP through the navigation menu was chosen for educational purposes and using the bot requires familiarity with clinical interviewing basics. The prototype was developed using ReactJs, Django, Python, and deployed on DigitialOcean.

Due to the difficulties in recruiting medical students and staffs, only 15 participants took part in the evaluation of the specialised SP with back pain. Among them, 58% rated Chameleon's understanding of questions as "Good or better" and 66% rated its performance in mimicking a real SP as "Good or better". While the performance of the specialised SP could still be improved, this proof of concept shows the feasibility of creating chatbots from Chameleon. At the end of the interview, 83% of users reached a diagnosis with varying degrees of confidence, suggesting critical engagement with the SP and experiential learning taking place. Users perception of the web design was overall "Very Good" to "Excellent" and 92% of users expressed willingness to use the SP for training in the future.

3 CONCLUSIONS AND FUTURE WORK

Attempts to virtualise standardised patients for medical education, and in an OSCE in particular, are few due to the cost needed to collect medical conversational data. Existing chatbots are mostly trained for specific illnesses and follow a bottom-up development approach. For instance, Kenny et al. (2007) built 'Justin' to help with training doctors in the mental health domain and their speech recognition engine was handcrafted from scratch [3, 13] while El Zini et al. used deep-learning to build a virtual SP with training data collected manually. As noted above, both approaches required significant investment in time and resources. This is the first attempt that leverages the advantages offered by modern chatbot development platforms to address this gap. The framework presented herein circumvents previous limitations by (1) using a generic and reliable chatbot chassis which significantly fasten development time, (2) accommodates future expansions easily, and (3) having its development and deployment driven by user requirement.

[1] https://chameleonpatients.co.uk/

REFERENCES

[1] JB Battles, SL Wilkinson, and SJ Lee. 2004. Using standardised patients in an objective structured clinical examination as a patient safety tool. *BMJ Quality & Safety* 13, suppl 1 (2004), i46–i50.

[2] Julia El Zini, Yara Rizk, Mariette Awad, and Jumana Antoun. 2019. Towards a deep learning question-answering specialized chatbot for objective structured clinical examinations. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–9.

[3] Jonathan Gratch, Jeff Rickel, Elisabeth André, Justine Cassell, Eric Petajan, and Norman Badler. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent systems* 17, 4 (2002), 54–63.

[4] Andrew K Hall, Markku T Nousiainen, Paolo Campisi, J Damon Dagnone, Jason R Frank, Karen I Kroeker, Stacey Brzezina, Eve Purdy, and Anna Oswald. 2020. Training disrupted: Practical tips for supporting competency-based medical education during the COVID-19 pandemic. *Medical teacher* 42, 7 (2020), 756–761.

[5] Peter Hannon, Katie Lappe, Claire Griffin, and Danielle Roussel. 2020. Objective Structured Clinical Examination: From Exam Room to Zoom Breakout Room. *Medical education* (2020).

[6] Ronald M Harden. 1988. What is an OSCE? *Medical teacher* 10, 1 (1988), 19–22.

[7] Ronald M Harden, Mary Stevenson, W Wilson Downie, and GM Wilson. 1975. Assessment of clinical competence using objective structured examination. *Br Med J* 1, 5955 (1975), 447–451.

[8] John P Hubbard, Edithe J Levit, Charles F Schumacher, and Truman G Schnabel Jr. 1965. An objective evaluation of clinical competence: new technics used by the National Board of Medical examiners. *New England Journal of Medicine* 272, 25 (1965), 1321–1328.

[9] Patrick Kenny, Thomas D Parsons, Jonathan Gratch, Anton Leuski, and Albert A Rizzo. 2007. Virtual patients for clinical therapist skills training. In *International Workshop on Intelligent Virtual Agents*. Springer, 197–210.

[10] Zhen Chang Liang, Shirley Beng Suat Ooi, and Wilson Wang. 2020. Pandemics and their impact on medical training: lessons from Singapore. *Academic Medicine* (2020).

[11] Murray Longmore, Ian Wilkinson, Andrew Baldwin, and Elizabeth Wallin. 2014. *Oxford Handbook of Clinical Medicine-Mini Edition*. OUP Oxford.

[12] David Newble. 2004. Techniques for measuring clinical competence: objective structured clinical examinations. *Medical education* 38, 2 (2004), 199–203.

[13] William R Swartout, Jonathan Gratch, Randall W Hill Jr, Eduard Hovy, Stacy Marsella, Jeff Rickel, and David Traum. 2006. Toward virtual humans. *AI Magazine* 27, 2 (2006), 96–96.

[14] Peter Tate. 2005. Ideas, concerns and expectations. *Medicine* 33, 2 (2005), 26–27.

[15] Javad H Vash, Masud Yunesian, Mohammad Shariati, Amir Keshvari, and Iraj Harirchi. 2007. Virtual patients in undergraduate surgery education: a randomized controlled study. *ANZ journal of surgery* 77, 1-2 (2007), 54–59.

[16] GM Wilson, Rosemary Lever, R McG Harden, JIS Robertson, and J MacRitchie. 1969. Examination of clinical examiners. *The Lancet* 293, 7584 (1969), 37–40.