Chameleon – A Pedagogical AI-Driven Tool for Doctors and Medical Students

Al-Hussein Abutaleb

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Master of Science in Artificial Intelligence

of the

University of Aberdeen.

Department of Computing Science, 2020

This page is left intentionally blank.

“The process may seem strange and yet it is very true. I did not so much gain the knowledge of things by the words, as words by the experience I had of things.”

— **Plutarch**

Declaration

No portion of the work contained in this document has been submitted in support of an application for a degree or qualification of this or any other university or other institution of learning. All verbatim extracts have been distinguished by quotation marks, and all sources of information have been specifically acknowledged.

Signed: ***Al-Hussein Abutaleb***

Date: 2020

Acknowledgements

First and foremost, I would like to express my never-ending gratitude to my parents who supported me unconditionally in every way. I am sorry for all the heartache, know that it is shared. I love you then, now, and forever. Special thank you to my brothers and my super sister, Amal. To my aunt Saborah Al-Hassan who thinks I'm an undiscovered Einstein. I am forever grateful for your love, kindness, and your unwavering belief in me when I didn't. To my dear friend Sylvia Finzi for your incredible understanding, generosity, love, and for still being my friend after all these years! To my dear friend and former supervisor in a past life, Paul Dobson. Thank you from the bottom of my heart. Your generosity, kindness, guidance, and open arms meant the world to me. For my dearest friends Amy Masson, Abdulaziz Kimawi, Abdi-Aziz Suleiman, Kevin Kennedy-Ryan II, and Jo Bridger. A special thank you to my friend the Docker-Master Abdulrahman Jama whose advice on how to resolve bugs in in the backend code and frontend React code was very helpful, indeed.

A special thank you to the volunteers whose input and feedback was instrumental in completing this project: Dr. Tim Gardner, Dr. Andrew Hunter, Dr. Sophie Pitts, Dr. Charles Moore, my father Dr. Nasrullah Abutaleb, my colleagues and friends from the AI MSc programme Jessica McGowan and Jasmin Uhlhorn.

And to my supervisor Dr. Bruno Yun a thank you that's laden with cake and flowers. Your guidance, support, and insights were the cane that never broke when I leaned on it. Thank you!

Abstract

Objective Clinical Structured Examinations (OSCE) are an effective and widely used tool to assess multiple competencies in medical pedagogy. A primary component of OSCEs is their reliance on the use of Standardised Patient (SP). A role often played by a professional actor mimicking the symptoms of an agreed upon condition with the examiners. This allows for fairer and better controlled assessment of students' competencies. Problematically, OSCEs are often costly and require significant resources to organise. This is partly due to the costs associated with hiring the actors. Moreover, in their current format, they mostly serve as a tool to assess existing skills in students rather than improve upon them; they are not a teaching tool. Medical students rarely encounter SPs to practice before the OSCE. In addition, the current and probable future COVID19 restrictions may constitute a serious hinderance to the assessment and improvement of clinical interviewing skills of physicians, both practicing and students alike. We argue that developing an AI driven conversational agent to serve the role of a virtual SP is a creative and cost-effective way to mitigate some of the limitations surrounding OSCEs. Chameleon[1] is a conversational agent that uses Google Cloud Platform's DialogFlow API to simulate a patient presenting with lower back pain. This is an end-to-end solution that includes: 1) Chameleon conversational flow design containing over 180 intents with over 2637 training sentences. 2) Frontend of the web platform built using Reactjs and Bootstrap. 3) Backend of the platform written in Python/Django and deployed using on. 4) A basic version of Chameleon containing placeholder responses that is purposefully designed for future expansion into other clinical presentations.

Our design of the Chameleon's intent space and the frontend is informed by two live-testing rounds practicing physicians. A final user evaluation round is conducted; all 33 medical schools in the UK were contacted via their representative medical students' societies (over 230 societies). Our survey shows that 75% of participants would use the software again, 91.6% can envisage Chameleon replacing the role of the SP after significant improvements, with 83% of respondents able to reach a diagnosis of the condition Chameleon suffers from. Users' rated Chameleon's ability to understand written text input as Good or better, 58%. Furthermore, our web design proved to be encouragingly intuitive, reliable, responsive, and easy to use for participants with clinical training.

[1] Can be visited here: https://chameleonpatients.co.uk/

Overall, this work presents a technology demonstrator that addresses an existing need in the medical education. And the results suggest promising potential for future expansion of Chameleon.

Table of Contents

List of Tables

List of Figures

Setting the scene

An informal introduction

Perhaps the earliest form of learning is modelling those who know better. Lion cubs learn hunting from their elders and humans learn how to perceive life, hate, love and speak from their parents. A long time ago, Man held a wooden board and scribbled on it. Perhaps the scribbling was with chalk, blood, or small bit of coal. Then the scribbling evolved into forms of expression such as language. Then Man learned how writing can disseminate knowledge via books. Then, radio. The television. The computer. And recently the internet. In addition to many other purposes, technology has always had a dialectical relationship with education.

This work uses a commercial conversational agent development platform to do the following:

- Build a conversational agent that can play the role of a patient.
- In doing so, it creates a new space for pedagogical tools that are not yet fully explored.

Why Chameleon?

Chameleon is a reptile of many colours and backgrounds. It moulds its appearance according to its survival needs. We this conversational agent to eventually be able to wear different clinical presentations and different patient profiles in a manner not unlike the way real Chameleons wear the colours of their surroundings.

Motivations

1. This work offers medical students the chance of practice their clinical interviewing skills on-demand and not just in an OSCE setting.
2. It also offers them the chance to practice for an OSCE beforehand.
3. May reduce the significant costs associated with hiring professional actors to play the role of Standardised Patients (SP) if commercialised.

Contributions of this work

1. Created a full end-to-end solution that satisfies motivation #1 for a backpain clinical presentation.
2. With minor modifications related to imposing a time limit and feedback, it can fulfil motivation #2.
3. It offers a roadmap for this agent to expand into a variety of clinical presentations other than backpain.

Introduction

Introduction to the OSCE

Global life expectancy at the turn of the 20th century is estimated to have been 30 years on average. Today, over a century later, average global life expectancy is 71.5 years (Roser et al., 2013). The improvements and breakthroughs achieved in medicine were, and still, a decisive factor contributing to not only increasing life expectancy but also ensuring that people live healthier for longer. Accompanying this success was a huge increase in the knowledge base of medicine. In part, due to new analytical and technological tools that were used to discover pathogens and identify associated symptoms. This huge expansive growth in medical knowledge over a relatively short period of time necessitated continuous improvements and modernisation to the pedagogy of medicine.

Prior to 1975 in the United States, medical students were assessed using a range of Multiple-Choice Questions (MCQ), written essays, and unstructured clinical assessments. The clinical assessments entailed the student examining several patients whilst being monitored by two senior physicians. The overall goal is to assess the student's ability to elicit clinical history, competency in carrying out physical examinations, and on their ability to interpret results of both history and examination. A criticism of this methodology is that it was prone to become a test for the students' factual knowledge evident by the disparity between students' performance in written versus clinical tests, and with the significant variance in the marks given by different examiners for an identical student's performance (Harden et al., 1975; Harden, 1988; Hubbard et al., 1965; Wilson et al., 1969).

In 1975, Harden and collaborators proposed a new strategy to examine medical students and address the weaknesses cited above. The strategy revolves around the students physically examining patients over several stations where in each station a different competency is evaluated (Original structure of the OSCE published in 1975. Note that at every station (boxes preceding questions) the student is invited to demonstrate an aspect of their clinical knowledge. For example, the student starts by extracting clinical history from a patient with breathlessness whilst being observed and scored by an examiner; following this the student is asked to answer a set of written questions (usually in the form of MCQs). Recreated Harden et al., 1975

). The proposed Objective Structured Clinical Examination (OSCEs) entails the student rotating over several stations where the student is assessed on specific clinical task,

ordinarily observed and marked by one or two examiners who follow a structured scoring list (Newble, 2004). Today, many variants for the OSCE exist and its use has extended to healthcare fields beyond medicine.

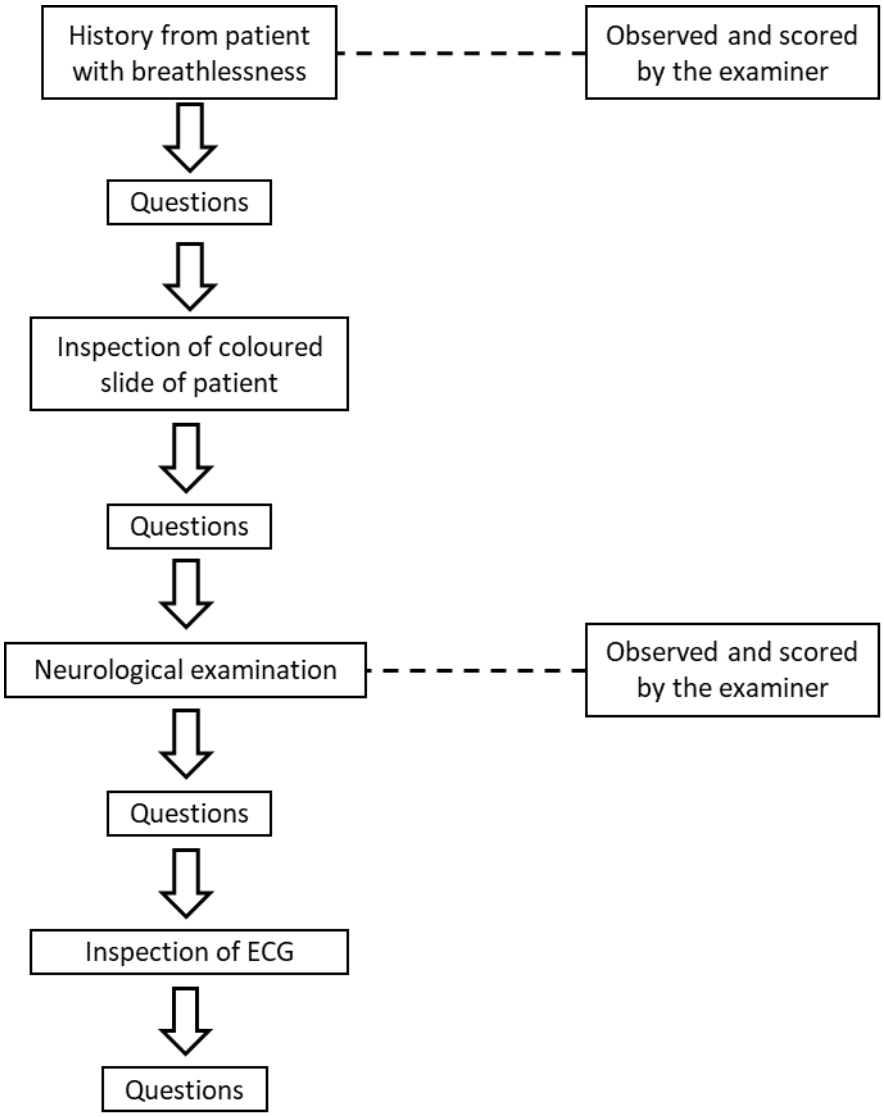


Figure 1 Original structure of the OSCE published in 1975. Note that at every station (boxes preceding questions) the student is invited to demonstrate an aspect of their clinical knowledge. For example, the student starts by extracting clinical history from a patient with breathlessness whilst being observed and scored by an examiner; following this the student is asked to answer a set of written questions (usually in the form of MCQs). Recreated Harden et al., 1975

Shortcomings of OSCE

A common critique against the OSCE is the significant cost entailed. This is partly due to the expense in structuring several stations for students to rotate on, associated admin and clinical working hours, senior physicians' time who prepare, mark, and conduct the exam, any

consumables used, and the cost of hiring actors to play the role of patients. The Standardised Patient (SP) is an individual, who may be an actor or a healthcare professional, who will play the role of a patient with a predetermined clinical presentation. The use of SPs is considered the gold standard in OSCEs. Occasionally, SPs are also responsible for scoring and assessing students' performance (Battles et al., 2004). Problematically, the cost of hiring actors and training them is not insignificant[2]. Brown et al. from the University of Aberdeen estimated that a two days OSCE for 185 final year medical students, costs the university £20,239 to train and hire actors to play the role of SPs. This constituted over 22.7% of the total cost of the exam (Brown et al., 2015). Furthermore, the practice of hiring paid actors or volunteers to play the role of the SP adds a degree of safety to student-patient interactions in a clinical setting. Mistakes are better made in a simulation, whether real or virtual, than on a patient expecting the highest standard of care (Battles et al., 2004). Despite the apparent high cost of hiring SPs, their use continues today due to them becoming an integral part of the assessment process; indeed, SPs are considered to be the highest contributing factor to the outcome of the OSCE (Harden, 1988).

Our survey of the field suggests that there exist other critiques on the ability of OSCE to assess clinical reasoning, the test's relative rigidity which follows its structured format, the validity of the test; is it measuring what we think it is measuring? And the apparent correlation between the reliability of the test and resources involved (higher number of stations increase reliability of assessment). However, the resources and costs entailed in running a valid and reliable 'high-stakes' OSCE seem to be the biggest issue of contention (Barman, 2005; Gupta et al., 2010; Park et al., 2015).

Using the OSCE as a pedagogical tool

Ronald Harden, the father of the clinical OSCE, states in Harden (1988) that the purpose of the exam is to assess students based on what they do in a clinical setting rather on what they know. Indeed, we note here that the intention behind the original OSCE was to assess the clinical skills of the student objectively and fairly, and not to serve as pedagogical tool. Learning success presumably takes place in the weeks, and months before the OSCE and not during the examination. However, recognising that OSCEs can be used as a pedagogical instrument, modifications to the original OSCE have been made to add such an element to the experience of the students. This is often done by: 1) the student receiving feedback immediately after the exam. 2) Allowing the students to observe expected performance before the exam, and 3) to

[2] Occasionally, volunteers are recruited. While the cost is certainly smaller, the costs around training and prepping the volunteers remain. We here assume that SPs are played by paid actors.

remain at the station to observe the performance of the next student. 4) Asking the student to play the role of the Standardised Patient (SP)[3] (Harden et al., 2015).

We argue that if the students had on-demand access to the factor contributing the most to the learning success, namely history taking and interaction with the SP, prior to the exam, their clinical skills for those competencies will improve. And thereby improving students' performance as doctors once they graduate.

Assessment of student's performance

Error! Reference source not found. depicts three clinical competencies that the OSCE a ssesses; we restricted the figure to competencies that can, or will, be addressed within the context of this project. Namely, those that can be addressed via an interview with the SP and are also amenable to computer simulation. The three examples are paraphrased and reformatted from Harden (1988). At each station, the associated standardised task is presented to the student. The examiners expect key performance metrics to be met. Scoring the students, whether by the SP or by a senior physician, is often done via a standardised checklist to ensure objectivity. Appendix A contains a detailed table of the original competencies assessed by OSCEs.

Figure 2 Examples of clinical competencies that can be assessed by the OSCE and are also amenable to computer simulation. These examples were paraphrase and reformatted from Harden (1988). There are also other examples amenable to computer simulation, for example, communication with medical staff, that we will not consider in this project. We restrict our focus the core role the SP plays.

[3] Introduced as pedagogical tool in medical education in the 1960s, SPs are impartial specialists (actors or volunteers) trained to portray the symptoms, behaviours, concerns of a real patient. They are also often responsible for scoring students' performance.

Structure of history taking

Taking clinical history from the patient is perhaps the activity any doctor, particularly internists[4], would spend the most time performing throughout their career. It is also the activity that contributes the most to the outcome of the OSCE examination (Longmore et al., 2014). It is a cornerstone that sets in motion a sequence of either sound or erroneous clinical decisions. The latter often leading to preventable suffering and increasing costs due to the increased time and resources that must be allocated to care for the patient post misdiagnosis. Moreover, diagnostic errors can lead to a litigation against the hospital or the physician in question. Indeed, it is conservatively estimated that medical errors are the third leading cause of death in the United States, arguably most of those deaths were preventable (Makary and Daniel, 2016). The costs of medical errors in generals are also conservatively estimated to be in the range $73.5 billion to $98 billion dollars annually (Andel et al., 2012). Furthermore, physicians' concern about making an error, and thus being sued, is not an insignificant behaviour modifier in a clinical context. Doctors are less willing take calculated risks in treatment even when the net outcome is in favour of the patient's wellbeing; Defensive Medicine has its own problematic clinical implications (Carrier et al., 2010; Catino, 2011; Rimmer, 2017). A secondary outcome of this project will be the significant cost savings that will result from deploying such a system successfully (Brown et al., 2015; Lee et al, 2018).

The Oxford Handbook of Clinical Medicine recommends that over 80% the diagnosis should be inferred from interviewing the patient alone (Longmore et al., 2014).

In a standard clinical interview with a patient, doctors are expected to investigate five aspects of the patient's clinical history:

1) History of the Presenting Complaint – this is where the doctor asks questions that are directly related to the disclosed reason of the patient visiting the clinic. For example, a patient complaining of recurring headaches could be asked about when the headaches started, what makes them better or worse? If there was an accident that caused them?…etc
2) Family History – here the clinician attempts to establish if there's a familial basis for the presentation. Thus, the doctor could ask if anyone in the family had suffered a similar problem or ask whether the patient's parents are still alive or not and establish cause of death if not.

[4] A doctor who specialises in Internal Medicine

3) Past Medical History: This line of questioning attempts to establish whether the current clinical presentation has anything to do with past medical problems the patient has suffered from. Questions such as whether the patient has ever been admitted to a hospital, any major surgeries, any past or current major medical conditions and so on.
4) Drug History – This is mainly questions surrounding any current prescribed/over-the-counter drugs the patient is taking. Allergies the patient suffers from are often asked in this segment of the interview.
5) Social History – Here the doctor tries to connect with the patient as a person and get to know them holistically. Questions around their place of living, their relationships, children/dependants, their work, whether they need special care at home, hobbies, exercise, smoking/alcohol habits, presence of pets, etc are asked. In addition to connecting with the patient, the doctor is usually looking for clues regarding the presentation that the patient themselves may not be aware of.
6) Functional Enquiry – this is an optional part of the interview and is occasionally ignored by experienced physicians (Longmore et al., 2014). It explores the presentation from a systems perspective. So, questions related to the cardiorespiratory system are asked, then respiratory, gastrointestinal, gastrourinary and neurological systems. Over two thirds of this part of the enquiry is modelled in our solution. However, intents that address this are embedded into other segments of the interview.[5]

The structure of clinical history will be revisited in later chapters.

[5] Intents are the meaning behind a user's utterance or question. For example, the questions 'are they alive?' and 'are they well?' can be classified under one intent – patent's living status.

Literature review

Chatbot Mediated Learning (CML)

Chatbot-Mediated Learning (CML) has been growing steadily in popularity over the last decade; however Winkler and Söllner's (2018) systematic review of chatbots pedagogical use suggests that CML technology is still the early Introduction Stage of the technology S-curve (Sood, 2010). Nonetheless, there are numerous pedagogical chatbot examples in the literature. For example, Colace and collaborators developed an e-tutor that can identify students' questions and select the most appropriate answer using ontological knowledge representation of the domain (Colace et al, 2018). Furthermore, in the mental health domain, where an experienced counsellor will often play the role of a patient when training less experienced counsellors; Demasi and collaborators introduced CrisisBot to virtualise the training. The goal was to reduce the risk to callers to a crisis hotline by allowing counsellors to train first with CrisisBot. Thus, saving the cost and time it will take to hire experienced counsellors (Demasi et al., 2020).

The use of Virtual Patient (VP) has also been widely adopted as part of medical education, however, often VP systems are usually classically designed software where the user interacts with the programme solely via buttons, dropdown menus, and mouse clicks as opposed to conversing with the patient as if it was a person in a clinic (Poulton and Balasubramaniam, 2011). In this work, the term Virtual Patient does not equal Standardised Patient.

With regards to training medical students, one of the earlier attempts was put forward by Kenny et al. (2007) where they introduced a virtual patient capable of displaying human-like character traits such as moodiness when asked the same question several times. Kazi et al. (2012) build an AIML based chatbot to answer queries from medical students, however, it did so not as a patient but as knowledge resource. Vash et al. (2007) used a virtual but non-speaking patient to run a randomised controlled[6] study on 4th year medial students. Learning outcomes were significantly better in students who used the system. Moreover, student wrote clinical questions in their own words and the system responded with the relevant clinical information. It also allowed the students to perform physical exams virtually.

More recently, El Zini and collaborators used deep learning to design a chatbot with a 3D avatar capable of interacting with medical students. Their work draws heavily on various deep

[6] Students in the control group were taught and assessed using classical educational material. The authors do admit that this is not a strictly RCS study.

learning techniques to build the entirety of the platform (El Zini et al., 2019). It is our observation that while the literature is abound with examples of virtual patients, example of simulated SP or VP for the purpose of clinical history training are scarce.

Chatbot development platforms

Our initial direction was to build the chatbot from ground zero using a machine learning framework such as TensorFlow. However, following discussion with Dr. Yun and members of the faculty of the Computing Science Department, a decision was made to use a specialised platform to develop the software and reorient the work to produce a full end-to end solution. Developing the chatbot from zero would have required a specialised large dataset, a significant amount of time, and will pull the focus of this work from the implications this bot may have on the experiential learning of medical students to the low-level technical details of implementing the chatbot. Furthermore, there has been an increasing number of sophisticated and commercially successful chatbot development platforms recently with a wide range of option to choose from, making the goal of producing a full solution achievable within the time frame (Pérez-Soler et al., 2021).

Our search for a chatbot development platform to use was guided by four factors: 1) Ease of use and intuitiveness[7]. 2) Price and cost for the duration of the project. 3) Ease of expanding the software in any future work by someone else other than the principal architect, myself. 4) Existing and future cloud based accessible services that the platform can use, for example Google Cloud offers robust and reasonable priced cloud storage service.

From the first and fourth criteria, we only focused on platforms that are follow utterance/intent/entity/parameter or an action based model and are also part of a reliable cloud computing suite.

Below is a brief description the main platforms that were under consideration.

Microsoft (MS) Bot Framework

Microsoft's QnA Maker

Part of Microsoft's Cognitive Services offerings which itself is part of MS's Azure cloud computing suite. Provides excellent user-friendly knowledge base integration to automatically create question and answers pairs to be reviewed by the user. Problematically, it does not

[7] This is subjective to the principal architect, me.

offer essential features of conversational agents such as intent detections. It is more suitable in a business context for a Q&A FAQ focused chatbot and not as a pedagogical tool.

Microsoft's Language Understanding Intelligent Service (LUIS)

A cloud-based API also offered by MS Cognitive Services as part of the Microsoft Bot Framework. LUIS' intention detection engine and custom entities make it an attractive option for building out conversational agent. Furthermore, LUIS automatically improves its performance with regular use (Ray and Mathew, 2018). Luis currently offers support for 50 languages. LUIS supports C#, Node.js., Go, Java and Python for development.

Amazon's Lex

Pay-as-you-go conversational AI development platform that is run by Amazon AWS. Features built-in automatic speech to text engine along a machine learning powered intent classification engine. Lex can be configured to function as Q&A FAQ engine like MS' QnA. Lex currently only supports six languages.

IBM Watson

Offered as part of the IBM Cloud; IBM's offering of cloud computing services. Watson was partly trained on the text of Wikipedia of 2011. The standard free offering is too limited to be used for this project. It offers only five 'skills' or actions that the chatbot executes upon identifying an intent. The Plus version costs $140[8], a price that is well above the personal funds allocated towards this project. In comparison with other platforms, IBM Watson is considerably more expensive. It currently supports thirteen languages (Summa Linguae Technologies, 2021).

Facebook's wit.ai

Open source chatbot framework acquired by Facebook in 2015. Currently, it supports 132 languages. Reports of a difficult training process dissuaded us from using it (discover.bot, 2019).

Google's DialogFlow

Part of the Google Cloud ecosystem, DialogFlow is also a cloud-based API that is capable of building complex conversational agents. Currently, Google offers two versions of DialogFlow. The Essentials (ES) version was released after Google bought the start-up api.ai in 2015. And the CX version which offers visual state-based flow to control the conversational flow.

[8] https://www.ibm.com/cloud/watson-assistant/pricing/

DialogFlow uses rule based and machine learning methods to detect intents. And its capabilities in terms of accuracy, speed and overall robustness are comparable to IBM's Watson (Thorat and Jadhav, 2020).

There is a significant overlap between the platforms just discussed. For example, all platforms support voice and text channels. All platforms offer a web platform and a visual UI for the design of the bot that are at similar levels of usability. Our survey suggests that there is no major difference in terms functionality, usability, and possible strength of the chatbot to be built. We settled on choosing Google's DialogFlow partly due its intuitive state-based visualisation of the conversational flow and, perhaps more crucially, the significant credit that Google offers to new users at no cost. An additional factor that influenced our final decision is the readability of the documentation. Dialogflow CX is well documented and the GitHub repository offeres plenty of code for developers to use.

Choice of DialogFlow CX over the ES version

- Relatively poorer speech processing in ES, ES does not offer the Advanced NLU option which in our personal experience offers a significant improvement.
- Choosing CX ensures that future upgrades of our bot could include high performance speech NLP processing

Table 1 (next page) compares some the main factors that were under consideration when surveyed available chatbot platforms.

Platform	Supported Rest API or SDK	Integration	Language supported	Cost	Notes
Amazon Lex	Python, Javascript, Java, Ruby. C#, PHP, Objective C	Facebook, Kik, Slack, Twilio SMS	6	$0.00075 per API request. Assuming a conversation with 40 requests makes the chat cost = $0.03 = £0.022[9]	10,000 text requests/month free for the first year
Google's DialogFlow CX	Python, Node.js, Java	Dialogflow Messenger, Facebook, LINE	56	$0.005 per API request[10]	Over $900 free credit =$600 standard free trial +$300 free trial of Google Cloud. The CX version offers an Advanced NLU option not available in the ES offering.
IBM's Watson	Curl, Java, Node, Python, Go, .NET, Ruby, Swift, Unity	Facebook, Slack, WordPress, Voice Agent	13	Flat fee of $140 per month	1 month free trial. The free version is too limited for the purposes of this project
Microsoft's Bot Framework	C#, Node.js	Facebook, Skype, MS Teams, Slack	13	$0.0005 per API request[11]	
Facebook's wit.ai	Node.js, Python, Ruby, Go, Unity	Facebook	132	Free	Very easy to integrate with Facebook. Problems with training the NLP engine reported.
SAP's Conversational AI	Python, Node.js, Ruby, PHP, Swift, Java,	Skype, Facebook, Slack,	19	£258/1000 chats/month = £0.258 chat/month	No free credit

Table 1 Side by side comparison of the main Chatbot design platforms and frameworks explored.

[9] Based on exchange rates on 26th August 2021
[10] Google announced a change in pricing effective from 1st September 2021: https://cloud.google.com/dialogflow/pricing#cx-agent-2021-09
[11] https://azure.microsoft.com/en-us/pricing/details/bot-services/

Choice of the clinical presentation of the bot

After initial consultations in the early stages of the project a decision was made to model Mixed Anxiety Depressive Disorder (MADD). MADD is a combination of symptoms indicating depression (a mood disorder) and anxiety (a neurotic and stress related disorder) both co-presenting severely enough to be recognised psychiatrically, however neither the depression nor the anxiety is distinctly predominant in the presentation (Spijker et al., 2010). Furthermore, Abd-Alrazaqe et al., (2020)'s review suggests that Depression is the second most modelled condition in healthcare chatbots suggesting available examples and code bases to use.

Our interest in MADD is because it could easily be mistaken for several other conditions. For example, shortness of breath, a symptom of panic attacks which are common in MADD diagnosed patients, could be interpreted as an indication of Pulmonary Embolism (PE), a serious and life-threatening condition. It follows that doctors would use contextual information about the patient to make the correct diagnosis. For example, patient's appearance, eye contact, body language, or personal hygiene. A young and physically healthy young person is less likely to have a PE than a smoking non-active overweight person who's 75 years old. Furthermore, symptoms of MADD are often associated with significant functional impairment adding an element of rising costs to delayed treatment due to misdiagnosis (Barlow and Campbell, 2000). The rationale here is that symptoms of anxiety and/or depression will often partially present somatically. Inexperienced doctors, or medical students, could easily misdiagnose the root causes of the conditions and order a multitude of unnecessary tests which compounds the problem as it prolongs patient's suffering and exerts further expenses to healthcare bodies. Thus, offering a educational tool that train doctors to be better a commonly misdiagnosed condition may generate more interest and engagement from the users.

However, following the first round of live-testing (next chapter), it became clear that correctly modelling the probable conversational flow of a patient who suffers from a mental illness would add significant complexity that may not be suitably addressed within the time frame of this project. A decision was made to follow a classical approach.

Our choice for back pain as the symptom to model was primarily driven by 1) the simplicity of the presentation in comparison with MADD. 2) This is a symptom that is often assessed in OSCEs[12]. 3) Low back pain is by far one of the most common clinical presentations in any population and is the leading cause of disability globally for people from all age groups and

[12] https://geekymedics.com/back-pain-history-taking/

both genders (Bener et al., 2013; Hartvigsen et al., 2018; Ramdas and Jella, 2018). Therefore, there's a high likelihood that healthcare professionals testing the system would have met a patient presenting with back pain in their practice. Rendering the feedback more grounded in real-life experiences of doctors.

Building Chameleon – a user informed approach

This chapter introduce the approach and iterative processes used to build Chameleon, our conversational agent. This includes two informal rounds of live testing with working doctors and two AI MSc colleagues.

Preamble to the procedure followed when building Chameleon.

Evaluation directed Chatbot development

The overarching theme of this chapter is that once the most basic version of the chatbot is deployed on Facebook, most of the subsequent development work is informed by the results of two live testing rounds conducted in this chapter. Indeed, Hobert (2019) suggests there's no standardised procedure for evaluating pedagogical conversational agents. Hobert, however, proposes four major steps in building and assessing pedagogical chatbots. 1) Wizard of Oz testing[13] – a human plays the role of the bot. This tests subject's openness toward the new technology. 2) Implementation of the NLP/NLU models that will run the chatbot. 3) Live testing with subjects in a laboratory setting. 4) Field testing with a large set of users. We make no attempt in this work to apply steps 1 and 2. In this chapter we focus on step 3. Step 4, is discussed in the evaluation and results chapters.

Intents

All intents in the basic placeholder version of Chameleon were harvested from the Oxford Clinical Handbook of Internal Medicine (Longmore et al., 2014) and from the first round of live testing with volunteers (stage two). The second round of live testing added intents only to the backpain version of the conversational agent.

Intents naming convention

Conventionally, intents are named [predicted intent].[action]. This is suitable in a business use case where the bot is expected to perform a service for the customer. For example, book a flight from London to New York. *Book.Flight* would be a suitable and intuitive name for the intent that detects the user's request. However, in Chameleon, the roles are largely reversed. The bot is not attempting to execute any actions of note on behalf of the user. The primary goal of the bot is to carry the conversation to help students practice their clinical interviewing

[13] These experiments are not unlike the human part of a Turing test (Turing and Haugeland,, 1950).

skills. The user is the one expected to ask the questions not the bot. Therefore, we deviate from the recommended naming convention by Google (Google Cloud Tech, 2021). Most intents are named to show in which section of the clinical interview they are expected to be triggered and to clearly describe the predicted intent of the user. For example, when taking past medical history, the physician may ask whether the patient has had any operations before. The intent to address this question is named: Past Medical History – Surgeries.

Training sentences

Where possible the following procedure was followed to create the training corpus for each intent, we here use the intents "Past Medical History - Depression" as a demonstrator:

1. A simple form of the expected question in the present tense is added to the corpus:

   Are you depressed?

2. The question is rephrased to a past tense[14] and added to the corpus:

   Have you been depressed?

   Were you depressed?

3. New auxiliary words are added to the question that do not change it semantically:

   Have you ever suffered from depression?[15]

4. One or several preambles are added:

   Is it ok if I ask if you are depressed?

   Do you mind if I ask if you suffer from depression?

   Alright if I ask whether you are depressed or not?

5. A synonym of one or multiple words that do not change the meaning are used in place of the core trigger word(s):

   Do you feel hopeless and sad often?

   Do you suffer from constant low mood?

   Have you felt hopeless and sad often?

6. Where appropriate, the word 'diagnoses, or any synonym[16] of it, is used in the question:

[14] Technically, this is a present perfect progressive tense indicating an event that began in the past and continues in the present. Where appropriate, we also use all past tense forms. For example, Have you been depressed? were you depressed? Did you get depressed? Etc.

[15] Or have you ever been diagnosed with depression?

[16] For example, "display symptoms of", "shown symptoms of", "tested positively for ", "were you ever told that…"

Were you ever diagnosed with depression?

7. A preamble that captures time is added:

   During the past month, did you have little interest or no pleasure in doing things?

   Over the past few months, did you have little interest or no pleasure in doing things?

8. Occasionally, two more of the above forms would be combined. For example, the user could ask:

   Have you felt low, depressed, and sad during the last month or so?

9. A complete rephrasing of the question without the main trigger word of 'depression'

   Has your mood been low often recently?

10. With more complex intents such as "Social History - Home - Movement in the House"; completely different questions may be asked that have very similar meaning. For example, if the doctor wants to know if the patient can move freely and unaided inside the house, they could ask any of the following:

    Can you move unassisted at home?

    Do you use a walking stick in your house?

    Is there a carer who help you get around in your house?

    Do you need to use the rails on the walls or the stairs to move about in your home?

    On average, how long does it take from you to walk from you front door to your bedroom?

In all intents, we attempted to maintain a minimum of 10 training sentences per intent as per DialogFlow CX's documentation (Google Cloud, n.d - a). Occasionally, the 10 sentences minimum requirement is relaxed either due to the nature of the intent[17], or due to an oversight that was only detected post deployment[18].

[17] For example, the intent "Are you well groomed?" has only 8 training sentences. This is an intent that may aid the doctor in accessing sub communication related the patient's appearance.

[18] Post deployment, we discovered several intents with less than 10 training sentences. All intents in both models are listed in detail in Appendix E

Design of Chameleon's responses

As discussed earlier, training sentences are needed for intent detection to cover a probable space of utterances. Thus, when writing the responses of the bot, we have ensured that the agent is verbose and clearly expresses the desired response. In part, this is to aid the debugging process. For example, in response to the detection of the intent "Drug History – Recreational Drug use", the agent responds with: "No, I have never used recreation drugs" instead of just "No". The trigger for this intent could be one of many triggers (e.g. do use recreational drugs, do you use LSD, do you eat mushrooms, do you smoke weed? Have you experimented with illicit substances...etc)

Design approach

Following our selection of GCP DialogFlow, we begin designing the conversational agent. Our approach follows the guidelines below:

1) Prioritise a design that allows for future expansion to cover multiple patient profiles with a variety of conditions. This necessitated that:

   a. The creation of a generic version where responses are dummy and nonspecific. For example, if the intent of 'Past Medical History – diabetes' is triggered, the response in this basic version would be: '[patient declares diabetes status]' as opposed to 'Yes, I have type II diabetes' or 'No, I am not diabetic'.
   Modifying this response in the future to reflect an actual symptom can be done easily via the DialogFlow CX console or via DialogFlow API[19]. This is done to allow for future modifications to reflect different symptoms, temperaments, and phrasing.

   b. This in turn meant that the generic version of Chameleon will only contain one layer of intents that form the backbone of the clinical enquiry. These represent the most common and broad questions doctor may ask patients in general and not specifically about a certain disease. For example, most patients, regardless of the nature of their clinical presentation will be asked about any drugs they are taking and if they have any allergies.

[19] In this work, we edited all agent responses manually using the DialogFlow CX console.

This basic intent space forms the basis from which any future adaptation of Chameleon to reflect any set of symptoms should start (Figure 3). We here emphasise that these basic intents represent possibilities of what could be asked and not what will definitively be asked by the physician. For example, one of these basic intents is to investigate whether the patient has had any heart disease. However, an 18-year-old physically active patient presenting with a sprained ankle is less likely to be asked this question in practice (and have this intent triggered).

Furthermore, predicting which intents could be triggered in a clinical interview will also depend on the patient's response. For example, this generic version that encompasses all the intents from the standard Oxford Clinical Handbook (Longmore et al., 2014) does not contain questions directly related to backpain, the condition we model in the deployed version of Chameleon. Therefore, to further expand Chameleon's design to capture an actual condition, backpain, a new second layer of intents is added. These are intents that are unique to this clinical presentation. For example, doctors are not expected to ask in detail about recent travel history of the patient when the presenting complaint is lower back pain following a gym visit. However, they will if they suspected the patient also recently went hiking for twenty days in the Amazon. Figure 4 and Figure 5 show a visual explanation of the intent space expansion just discussed.

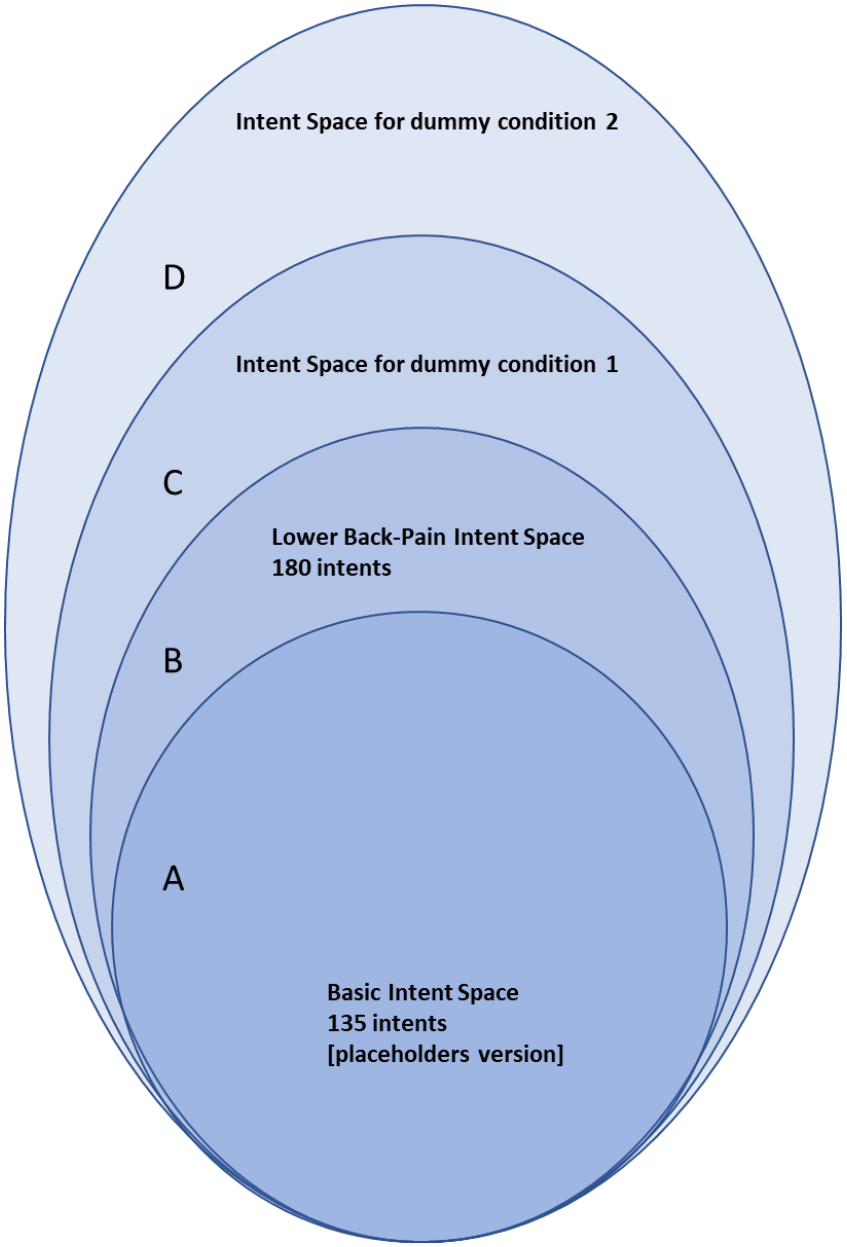


Figure 3 **A** - basic intent space comprising the placeholder version of Chameleon where all response are placeholders. **B** - After removing all the placeholder responses and replacing them with responses that reflect the clinical presentation – backpain, the intent space expands to accommodate the intents that are specific to the presentation. For example, a new intent is added to see if the patient is cautious and protective of their back[20]. **C** - The same principle applies to any future expansion of Chameleon. For example, if the patient presents with symptoms that the doctors suspects are indicative of a very rare disease, the overlap with the basic intent space may be smaller. This is due to the basic intents space mostly being designed to address most common clinical presentations in the population. **D** - However, it may also be the case that a new clinical presentation will share symptoms with all above intent spaces. All these considerations will have to start from the basic conversational structure of the basic placeholder version, however.

[20] A list of all intents and their responses is provided in Appendix E

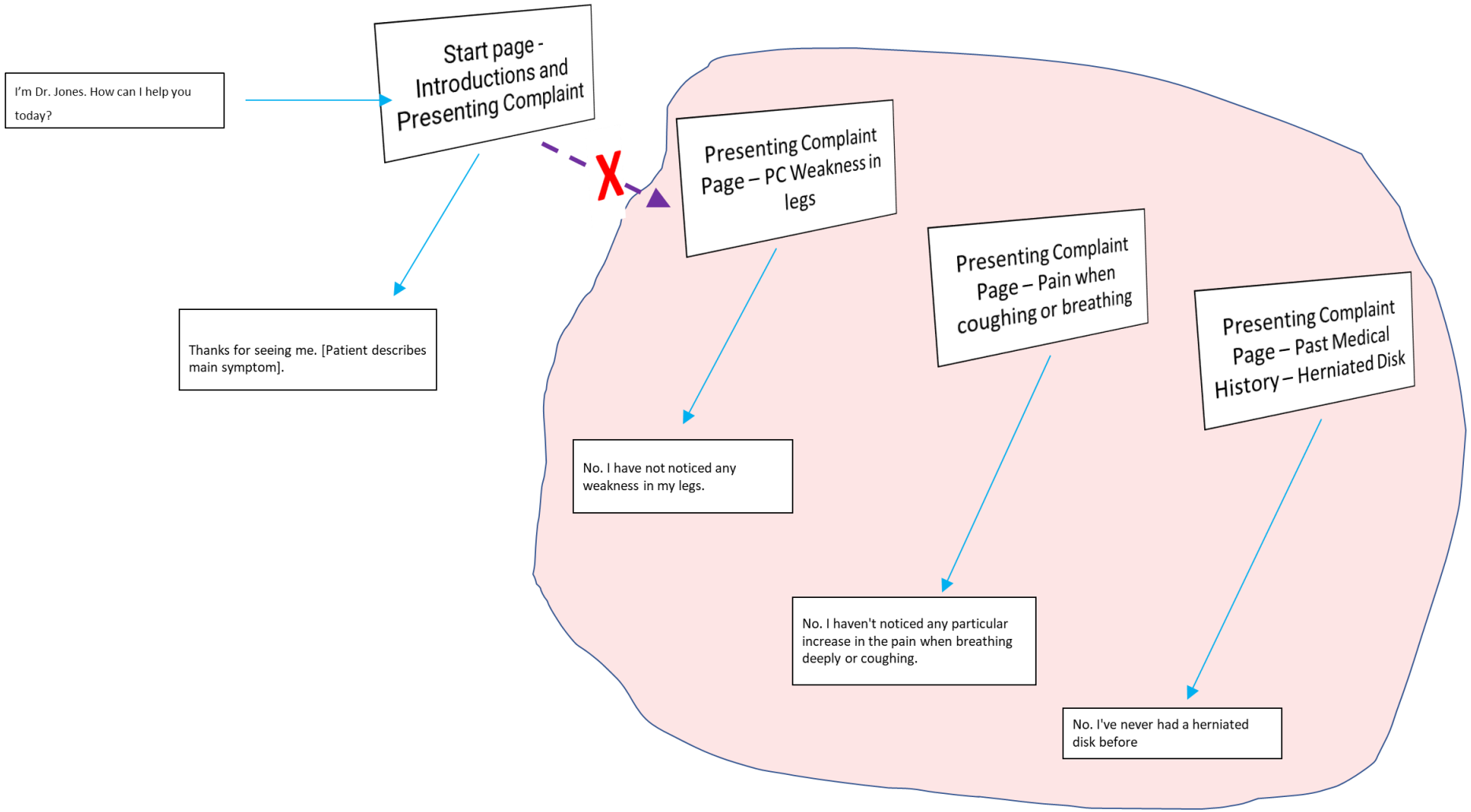


Figure 4 In the basic version of Chameleon, expected patient responses are replaced with temporary placeholders such as 'Thanks for seeing me. [Patient describes main symptom]'. These do not open a follow up question directly related to patient's answer.

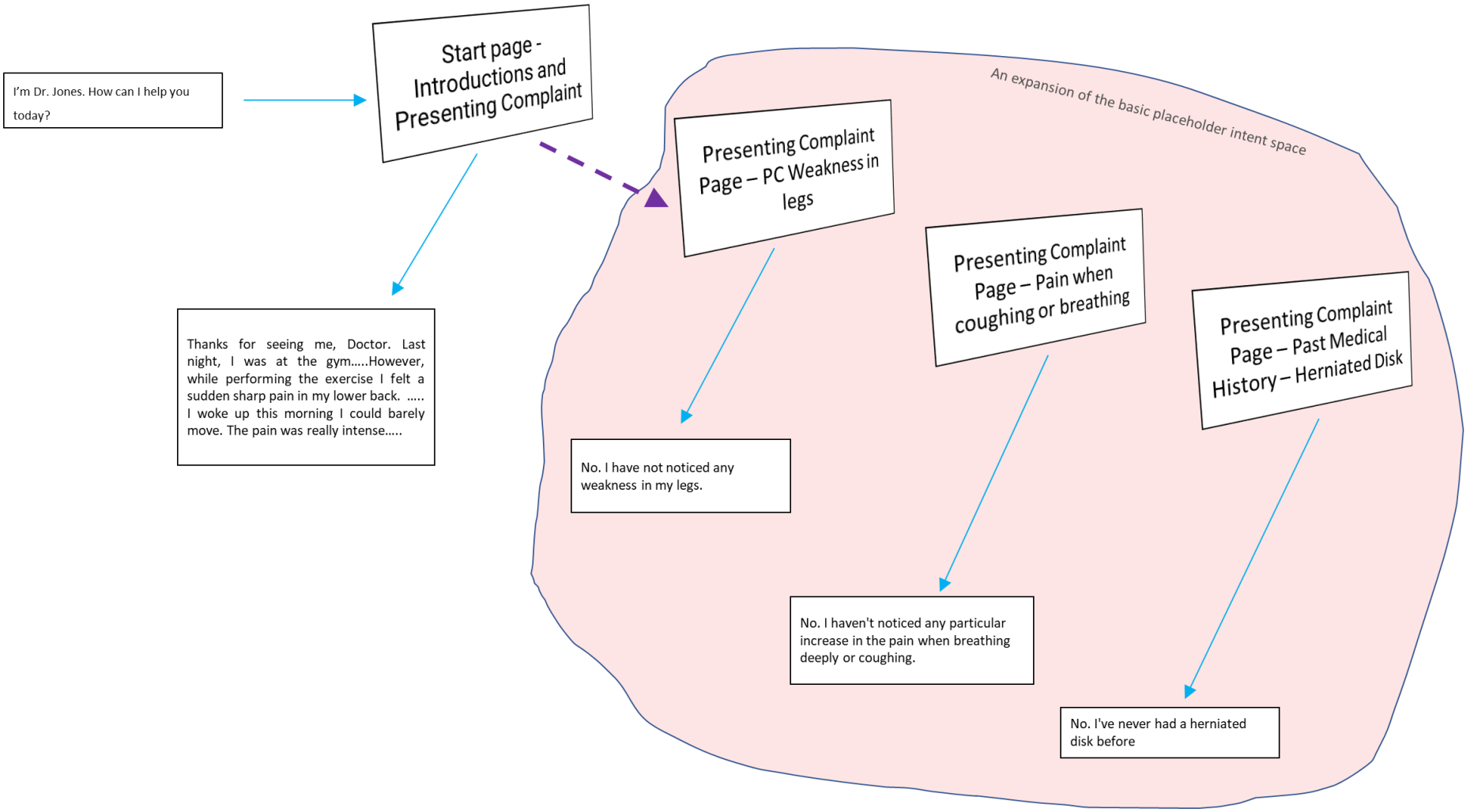


Figure 5 In the deployed version of Chameleon, the responses of the bot are tailored according to the condition it is mimicking, backpain, and thus a new line of questioning opens for the user with the necessary associated expansion in the intent space for the bot to be able to process the new questions.

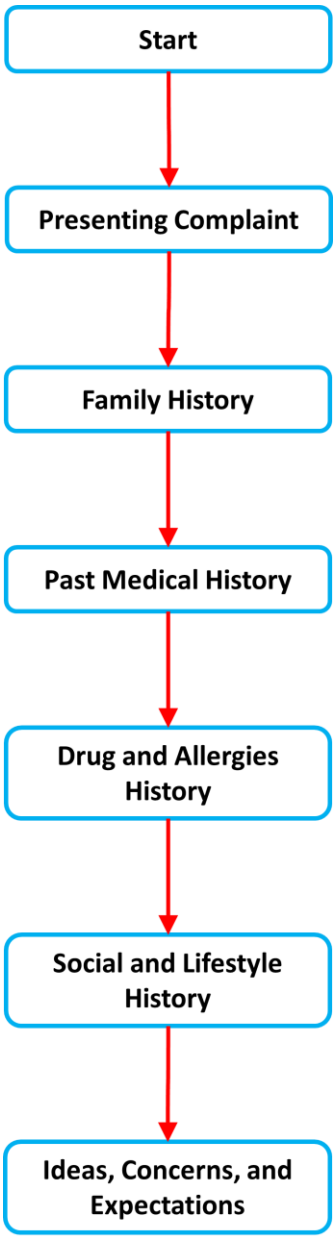
2) Intuitiveness from the user's perspective. The internal structure of the conversation must follow the algorithmic clinical interviewing methods doctors are trained to follow discussed in the Introduction. Briefly, in all clinical interviewing situations, whether in an OSCE or in a real clinical situation[21], the conversation begins by:
    a. The doctor introducing themselves and offering assistance to the patient with regards to the patient's motive for visiting them. The patient responds by explaining what their complaint is.
    b. This in turn triggers a unique set of questions by the doctor to get the patient to elaborate on symptoms. For example, when did it start? What were you doing at the time you noticed it for the first time? Does the pain move anywhere? Is this the first time?…
    c. Following the patient's elaboration on their symptoms, the physician then attempts to determine whether their complaint has possible genetic or heritable basis. Questions such as: Has anyone in the family complained from something similar? Are you parents healthy? Are there any conditions that run in the family?...
    d. The conversation then moves to the doctor asking their patient about their overall health and past medical history. For example, were you healthy as a child? Have you ever been admitted to a hospital? Have you undergone any surgeries in the past? Do you have any long-term medical conditions? Are you diabetic?...
    e. Which in turn leads to questioning the patient about any medications they are currently taking, type of drugs, efficacy, dosage, side effects, allergies, etc.
    f. Following that, the conversational moves to the doctor attempting to get a more a holistic view of the patient, their background, their living space, who they live with, and their wellbeing. The doctor also aims uncover important clues to the real causes of their complaint that the patient themselves may not be aware of. Questions regarding their social life, recreational habits, alcohol and tobacco use, and their lifestyle are asked. For example, the responses to some of these questions could suggest the presence of a mental health condition the patient may not have shared. In turn, this would put the symptoms in a new light due to the psychosomatic factor common in many presentations (Sayar et al., 2003). Or if the patient complains from constant wheezing, the doctor may ask kind of jobs have

[21] This particular flow applies more to an outpatient clinic setting.

they held in the past? To identify whether this is caused by dust or exposure to chemicals.

g. Finally, the interview often concludes with the doctor asking the patient about their Concerns, Ideas, and Expectations (ICE).

The figure below depicts the algorithmic conversational flow from start to finish.



Stage One – Building a working model – Facebook Deployment

The basic intents are created from the Oxford Clinical Handbook with a conversational flow of six pages[22]. Here, all intents' responses are written using temporary placeholders (*I have a diabetes* versus *[patient discloses whether they are diabetic or not]*). Our primary goal here is quickly to develop a basic version of that bot that can correctly understand users' clinical questions:

1. **Start page** – 8 intents: Handles initial introductions. Automatically, triggers the Presenting Complaint page once the patient informs the doctor of what the dummy complaint is.
2. **Presenting Complaint page** – 10 intents.
3. **Family History page** – 16 intents.
4. **Past Medical History page** – 6 intents.
5. **Drug and History Page** – 7 intents.
6. **Social History page** – 20 intents.

In line with stated aim for this stage, to maintain the priority to build a basic bot quickly, we utilise DialogFlow Facebook integration feature to deploy the bot. A page titled Chameleon is created on Facebook[23] and directly linked to DialogFlow via Facebook's API (Figure 6). Now, users that are given special access and can interact with the bot directly via Facebook Messenger (Google Cloud, n.d - b)[24]. The approach of using Messenger-like platform's APIs to develop chatbots is not uncommon and has been gaining momentum recently (Hobert and Meyer von Wolff, 2019; Smutny and Schreiberova, 2020; Wiratunga et al., 2020). Crucially, deploying via Facebook messenger takes around 2-3 hours only.

Transition between those pages can be triggered by the utterances listed in Table 2. Thus, any physician who is asking questions about Past Medical History and wishes to move the conversation on to Social History should write a sentence that is semantically close to "I'm going to ask you questions about your Social History". This triggers the state change, and the bot now can process questions relating to Social history. This rudimentary solution was implemented to facilitate our goal completing the basic structure as soon as possible to begin the evaluation.

[22] A page in DialogFlow is the state of the bot. Each of page contains several intents relating to the purpose of that page. For example, Past Medical History page will contain intents that fulfil any enquiries relating to this subject.
[23] https://www.facebook.com/ChameleonStandardisedPatient
[24] Users must be added as Facebook developers to the page first. Any attempt by the reader to access the bot via Facebook will not work without being added as a developer to the project first. Which we are happy to do given enough notice.

Intent	Training sentences	Response
Trigger – Family History	11	Changes the state to the Past Family History page
Trigger – Past Medical History	13	Changes the state to the Past Medical History page
Trigger – Drug History	9	Changes the state to the Drug History page
Trigger – Social History	14	Changes the state to the Social History page

Table 2 We use the Route Groups DialogFlow feature to insert these intents in all pages. The feature allows for any of these intents to be triggered from any state in the conversation. For example, if the user is on the Drug History page and wish to go back to the Family History page, they can trigger the "Trigger – Family History" by uttering "ok, let's talk about your family history" or a sentence like it.

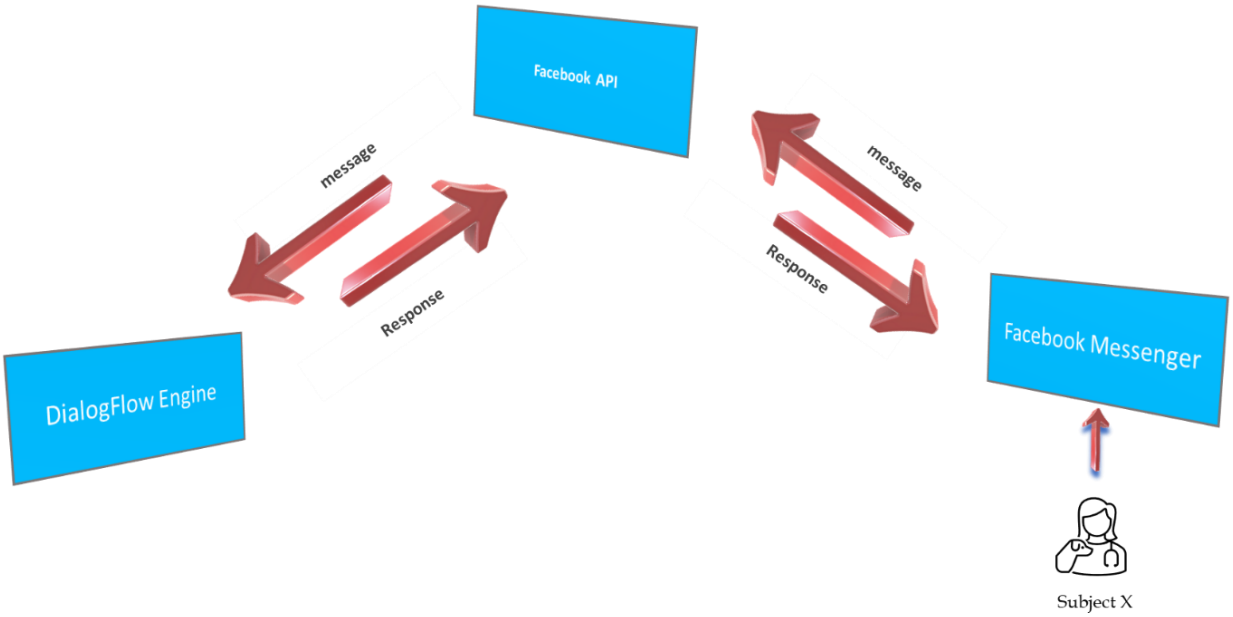


Figure 6 The complete system as it stands following Chameleon's integration with Facebook API. This allows us to start evaluation Chameleon's linguistic performance without having to develop a complete end-to-end solution

Stage Two – Live testing with volunteer domain specialists – First round

Following the conclusion of the procedure of stage one, seven interviews with four volunteer doctors working in the Aberdeen area and two colleagues from the AI MSc course were conducted[25].

In the tables below, Error Rate is defined as:

[25] All persons in this round and the next one have kindly come forward voluntarily. None were proactively recruited.

$$Error\ Rate^{26} = \frac{Instances\ of\ misinterpretation\ or\ no-match}{Total\ number\ of\ user\ utterances\ in\ the\ session}$$

And DialogFlow or Chameleon breakdown is defined as:

> Any point in the conversation in which the bot is unable to recognise any utterance. This includes utterances that match known training sentences verbatim.

The interviews were *not* conducted in a scientifically controlled manner. At times, users were given pointers on which intents to trigger in their own words. This was done to 1) test specific intents, 2) generate training sentences, and 3) help the users who might not know what to ask for next[27]. We therefore do not use the Error Rate defined above to quantify Chameleon's performance. Finally, the stated length of the live test does not include the personal introductions and explaining end goals of the project and how their assistance will contribute to this work.

The interviews broadly followed the format below:

1. Personal introductions
2. The subject is introduced to the project and the technology behind it using nontechnical language.
3. A mock-up of what we envisaged the future UI of Chameleon is going to look like is shown to the subject (Figure 9 in the next chapter). The subject is then asked if they have any comments on it.
4. The subject is invited to create a Facebook Developer Account. Following that, the subject is added as a developer to the project.
5. The live test begins.
6. A video recording of the test is kept and a log of the conversation with Chameleon is saved on Chameleon's Facebook page for review later. This is done with users' permission.

Below is a summary of each interview and associated results.

Interviews 1 and 2 – Subject A

Subject: Junior Doctor

The experiment

[26] This is a strict definition since it considers incidences of Chameleon finding no match due the absence of an intent not just due to an internal limitation of it. For example, if it was asked about its blood type, this would result in a no-match default response since the intent to handle this does not exists. This is still considered an error.

[27] The reader is reminded that at this stage Chameleon produces only placeholder responses.

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
13/07/2021	1	8	50%		33 minutes
	2	36	50%		
16/07/2021	1	27	27%	DialogFlow broke down twice mid interactions and the test was cut short	11 minutes
	2	7	40%		

Feedback

Intents	**1. Used user's unrecognised and misinterpreted utterances to expand training sentences for each respective intent.** **2. Separated tablets and over-the-counter drugs into two distinct intents** **3. Created an intent for recreational drug use**
Bugs	Identified a bug that causes DialogFlow to completely breakdown where the chatbot is no longer able to recognise any utterance including ones that are identical to training sentences. Reasons of the bug remained unknown throughout the project. It is our view this is maybe a bug DialogFlow CX integration without side APIs.

Outcome

1. Several intents were modified and created. We also expanded the training sentences for the intents the bot was unable to recognise using the subject's phrasing.
2. The subject pointed out that using MADD as condition to model will be difficult and will require more time and resources than is available. The subject correctly pointed out that doctors will often rely on contextual information to assess the likelihood that symptoms have an underlying psychosomatic root cause. Cues from eye contact, body language, the patient's appearance and personal hygiene, the patient's socioeconomic and cultural background are all factors that significantly affect the doctor's perception of the patient's complaint.[28] Following this discussion with the subject, we decided to use backpain instead of Mixed Anxiety Depressive Disorder as the condition to use with the final version of Chameleon, as was discussed previously.

[28] For example, a young patient with history of mental illness, unemployed, wearing unkempt clothing, and single, reporting shortness of breath will be assessed differently than a 70 years old patient also reporting shortness of breath who is overweight regardless of any present body language or appearance related information.

Interview 3 – Subject B

Subject: AI MSc colleague

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
16/07/2021	1	25	29%	Identified an issue with " Social History - Spouse's Health" intent	N/A

Feedback

None. The subject was neither a doctor nor a native English speaker.

Outcome

1. Confirmed that some of the minor issues were resolved following modifications made after Subject A interview.
2. Identified issues where the bot mistook several unrelated utterances to be related to the health of the patient's spouse.
3. Generated few extra training sentences.

Interview 4 – Subject C

Subject: Junior Doctor

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
18/07/2021	1	47	%36	Identified several new intents to be added	12 minutes

Feedback

Intents	**No new intents added**
Bugs	1. DialogFlow kept mistakenly triggering the intent 'Family History - Parents – Diabetes'. At one point, previously correctly identified intents were now misidentified as a diabetes in parents question. This was not a breakdown as defined above. There was no need to restart the conversation with Chameleon 2. Chameleon was unable to recognise questions related to the presence of heart disease in relatives despite an intent existing to address this possibility.

Outcome

1. Added new intents for: Patient's mobility at home, presence of carers at home, regular use of prescribed drugs[29], and for previous diagnosis of cancer in family and in the patient.
2. Expanded Family History page significantly to correct Chameleon's behaviour with regards to the "Family History - Parents – Diabetes". This included significantly increasing the number of training sentences of existing intents.
3. Used the user's unrecognised and misinterpreted utterances to expand training sentences for each respective intent.

[29] As opposed to over-the-counter/tablets

Interview 5 – Subject D

Subject: AI MSc colleague who had prior research experience with DialogFlow ES

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
20/07/2021	1	26	%27	N/A	16 minutes

Feedback

Intents	No new intents added
Bugs	1. Heart disease and diabetes in parents issues from earlier test remain.

Outcome

1. Expanded the training sentences for some of the misinterpreted intents.

Interview 6 – Subject E

Subject: Junior Doctor

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
21/07/2021	1	72	%53	%53	59 minutes

Feedback

Intents	**1. Added numerous non-existent intents the user pointed out, for example, recent travel history which becomes important if the patient is suspected of having an infectious disease.** **2. Suggested significantly expanding Alcohol and Smoking intents to detect possible signs of alcoholism, and a desire to quit both smoking and/or drinking.** **3. Suggested adding a line of questioning to explore past and current jobs held by the patient to uncover possible exposure to dust, chemicals, stress, etc.**

Outcome

1. Over twenty-two new intents added:
    a. Seven new intents added to expand Tobacco smoking habits, intentions to quit, past attempts to quit, type of smoking, use of nicotine replacements.
    b. Four intents added to expand upon alcohol drinking, previous attempts to quit, possible indicators of alcoholism. Most of this was drawn from the CAGE questionnaire (Bush et al., 1987).
    c. Added four intents to capture recent travel history to Latin America, the tropics, the far east, and abroad in general.
    d. Added three intents related to the patient's employment history and what job is held by their spouse.
    e. Added four new intents to determine the health and status of siblings.
    f. Two intents to describe the patient's living space and whether the patient's cares for anyone at home (for example, if the patient cares for their parents or children).
2. The subject confirmed insights provided by subject A that modelling conditions that have significant psychosomatic aspect will be challenging given the time and resources available.
3. This was by far the most informative and significant interview throughout this project.

Interview 7 – Subject F

Subject: Junior Doctor

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
26/07/2021	1	59	%40	Suggested adding a new section to the format of the interview	19 minutes

Feedback

Intents	**1. Suggested adding a page for intents to cover Ideas, Concerns, and Expectations (ICE[30]) to conclude the interview (Towle et al., 2006). This addresses possible concerns, worries, and hopes the patient may have as is the case in clinical practice. It also centres the patient and makes them part of their own treatment plan.** **2. Copied intent related to use of asthma inhaler into Drug and Allergies History from Past Medical History page. Furthermore, this was done to several intents as well where there is a possibility the user may attempt to trigger it in the 'wrong' section.**
Bugs	1. Chameleon gets confused if more than one question is combined in the same utterance. For example, asking about previous diagnosis of cancer in the patient versus asking about previous 'diagnosis of cancer, heart disease, or diabetes'. 2. Intent related to patient's activity at the time of symptom does not work.

Outcome

1. Five intents are added to model the ICE stage of the interview.
2. A trigger intent for ICE is added
3. Suggested adding vaping to the types of smoking detectable by Chameleon as doctors are observing a significant increase in its use as safer alternative to smoking.
4. Suggested changing Drug History to Drugs and Allergies History, and to change Social History to Social and Lifestyle History. This was done on the frontend page but not from DialogFlow side.
5. Expanded training sentences for several intents in the Social and Lifestyle History.

[30] https://geekymedics.com/ice/

The chatbot is now comprised of 135 intents and 7 pages. However, all responses of the chatbot are still placeholder generic responses. Next, we discuss Stage Three of the development where the placeholder responses are substituted with actual symptoms and a subsequent short round of final interviews.

Figure 7 (next page) shows the procedure followed in this round. The growth in the Chameleon logo is indicative of the growth in Chameleon's intents and training corpus however is not true to scale.

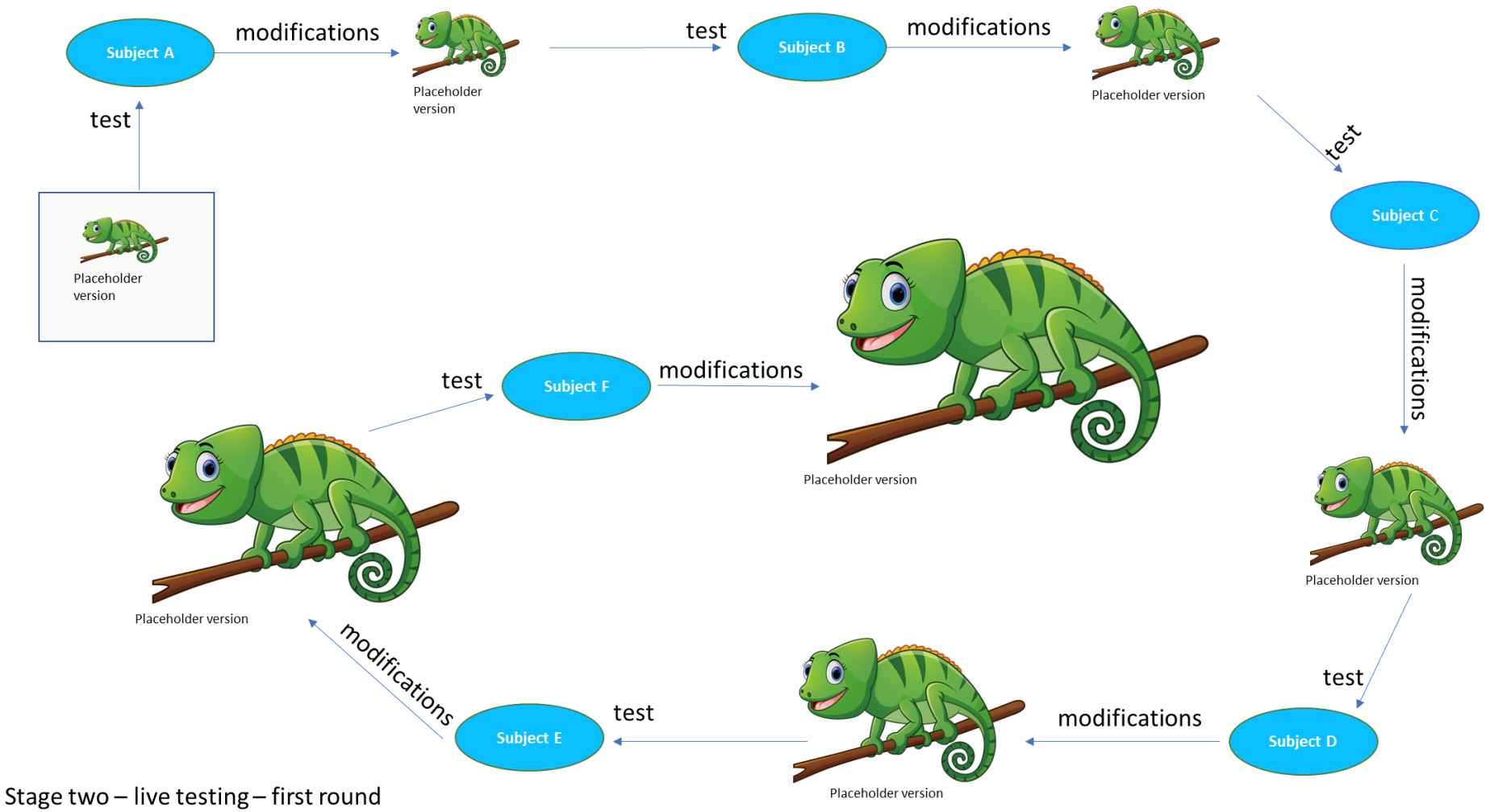


Figure 7 Starting from the basic placeholder version developed from the discussion in stage one, the live testing begins with Subject A, followed by modifications and testing with Subject B. This iteration is repeated until we reach Subject F, with Chameleon, still with placeholder response, significantly bigger and more capable. Chameleon now is comprised of 135 intents.

Stage three – Live testing with volunteer domain specialists – Second round

Following the discussion with subjects A and E, our earlier design decision to model MADD was dropped in favour of simulating a patient with back pain. We here make use of GeekyMedics, a very well-known and widely used online resource by medical students (Harvey, 2020), to harvest common symptoms and clinical questions specific to back pain history taking. We expand Chameleon by 1) adding 45 new intents mostly related to back pain. 2) Replacing all placeholder responses to responses that can conceivably come from a human.

The following two interviews followed a similar format to the earlier except for that now subjects were introduced the live web platform and its functionality first. Moreover, no introduction to the goals of the project was given.

They were conducted with Subject C from the previous live testing round and Subject G who has not used the system before. The biggest two differences between the previous round and this one is that 1) Chameleon now responds with natural language that may conceivably come from a patient. Furthermore, 2) the test is being conducted on a website (discussed in the next chapter) rather than on Facebook. This was a significant visual difference and subject C offered useful comments on the overall design of the website.

Finally, the overall tone of these interviews was slightly stricter than earlier, however not to the extent that it be can classified as controlled, particularly with Subject G due to a personal relationship with the subject.

Interview 1 – Subject C

Subject: Junior Doctor

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
13/08/2021	1	37	%40	Provided excellent comments on frontend design. The user also pointed out that Chameleon's tendency to misinterpret utterances as questions relating to parent's diabetes has been corrected.	14 minutes

Feedback

Intents	1. **Pointed out that there are intents relating to symptoms of backpain which were not programmed, for example, intents to see the subject suffered from unusual bowel/bladder dysfunction.**
Bugs	1. Some existing intents performed badly; utterances of the user were used to expand the training corpus of those intents. For example, the intent 'Past Medical History - Heart Disease' was not triggered when the user uttered 'Any problems with your heart?' or when asking 'Any heart disease?' 2. Discovered a bug in the main chat window of the front end. Any new messages cause the chat window to expand below whilst the scroll bar automatically scrolls up. This forces the user to manually scroll down the whole bowser window after each message sent to Chameleon. Minor remnants of this bug are still observable on the deployed version unfortunately.

Outcome

1. When asked for feedback regarding the frontend design:
    a. Suggested giving the patient an avatar to help the user contextualise the patient into context. We used an image created using a StyleGan2 on: https://thispersondoesnotexist.com/ (Karras et al., 2020).
    b. The subject commended the minimalist approach to designing the frontend. And when asked, advised against adding a CSS element on the right of the chat

box to contain 'How to use' instructions (Figure 13 – next chapter). Subject C's justification was that the layout of the frontend will be intuitive to anyone with clinical training.

c. When asked, the subject indicated that the placement of the navbar below the buttons on the left hand of the screen is clear and easily observable. Since the lowest button, Ideas, Concerns, and Expectations (ICE), is the last one usually clicked on in the interview. We asked the user this to ensure that the link to the survey is easily visible (Figure 13 – next chapter).

2. Expanded the training corpus of several existing intents. For example, in addition to the Heart Disease intent mentioned above, new training sentences were added to 'Social History - Exercise and Physical Activity', 'PC - location - Where do you feel it?', 'Family History - Parents – Hypertension', and 'ICE - Plan for action (patient's opinion)'.
3. Added an intent for 'weakness in the legs' following lower back injury, and intents for dysfunction in bladder or bowel. The user pointed out all of these can present as symptoms when a patient presents with lower back issues.

Interview 2 – Subject G

Subject: Senior Consultant, non-native English speaker

The experiment

Date of the test	Conversation	User Utterances	Error rate	Notes	Length
14/08/2021	1	22	%45	Subject pointed the importance of including depression and anxiety as possible causes for lower backpain	24 minutes

Feedback

Intents	**1. The user suggested adding intents for reduced sensation in the legs (could be indicative of a neuromuscular basis of the injury), prior use of diagnostic imaging of the spine, determining whether the subject's exercises history is relevant to the injury[31], and for determining if the subject is currently suffering from mental illness, and, interestingly, an intent to see if the subject is someone who is cautious about using their back.**

[31] No intent added to address this due to an oversight.

Outcome

1. Five new intents added to cover possible questions relating back pain injury specifically. Of note, an intent was added to ask the subject if they've had a herniated disk in the past and one to see if the patient is conscious about protecting their back when lifting objects.
2. Expanded training corpus for several intents further.
3. The user correctly pointed out that mental-health related factors such as somatization, stress, depression, and anxiety can cause low back pain (Bener et al., 2013). The Subject emphasised the importance of the bot being able to answer questions related to depression/anxiety specifically.

Figure 8 (next page) shows the iterative procedure followed in this second round. The growth in the Chameleon logo is indicative of the growth in Chameleon's intents and training corpus, however, it is not true to scale. Chameleon now is composed of 180 intents and 2638 training sentences. On average the size of the training corpus per intent is 14.7 sentences.

Concurrent with Chameleon iterative live testing was the web development work to build the website which the users will be using in the final evaluation round. This is discussed over the next chapter.

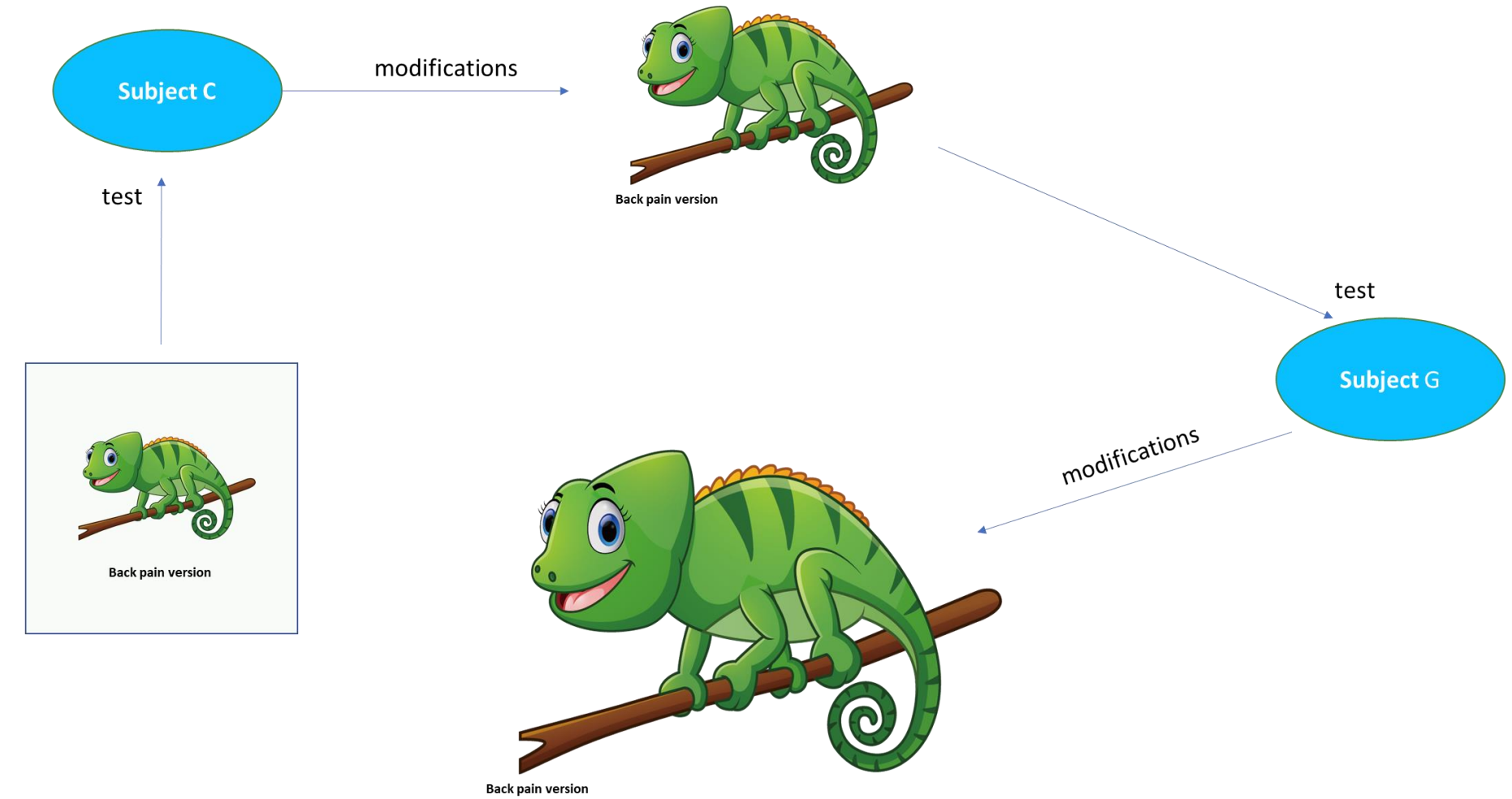


Figure 8 Starting from the finished placeholder version developed from stage two – first live testing round, the live testing begins with Subject C, followed by modifications and testing with Subject G. All testing conducted here was done using the website developed in the next chapter. Chameleon now is comprised of 180 intent and 2638 training sentences in total at an average of 14.7 sentences per intent.

Chapter Building the web platform

Declaration

We used numerous online educational resources and Google's extensive documentation of DialogFlow's API to build the website. Where possible a list of the sources we used to learn and implement the code used in this chapter was constantly updated. This list in Appendix H.

Design requirements for the web platform

Following our survey of the literature we first establish the following broad and refined design requirements for the website:

1 It needs to be easily expandable in the future. This follows from our design approach for building Chameleon.
2 The web solution needs to be guided by what doctors find intuitive in terms of presentation and content.
3 Reliable and Responsive.
4 Able to handle several users simultaneously.

The objectives stated above were significantly watered down from our original goals. Constraints of time and significant challenges during web development meant that we had to refine the objectives to what is stated above. For example, the original intention was for the frontend to give the users the ability to order simulated lab tests such as a full blood count or liver function test. This, and several other, feature(s) did not make it to the final deployed version. All discussion here pertains to the deployed version of the website.

We have included an early mock-up design of the frontend based on earlier optimistic design objectives in Figure 9 Early mock-up design of how we imagined the frontend of the web platform from which users can interact with Chameleon. On the right is a panel from which the user can order simulated lab tests and perform virtual physical examination..

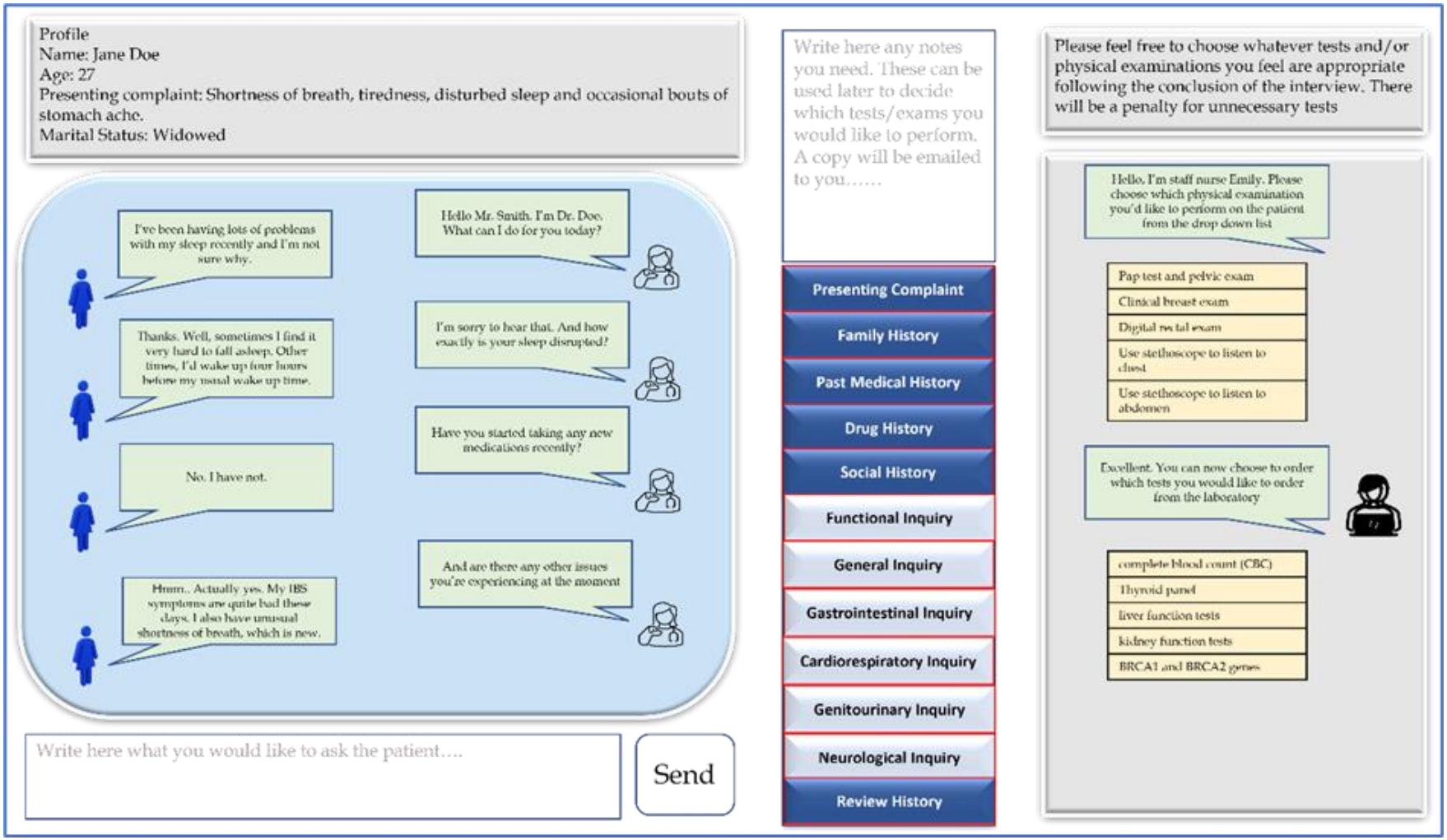


Figure 9 Early mock-up design of how we imagined the frontend of the web platform from which users can interact with Chameleon. On the right is a panel from which the user can order simulated lab tests and perform virtual physical examination. Despite many elements in this design not making it to the final deployed version, it still strongly influenced the way approached building the frontend.

Choice of tools

Backend

An initial attempt at building the backend relied on following two MOOCs where the backend was written in Nodejs (Bergant, 2016 and 2018). Unfortunately, despite significant efforts our code did not function as intended[32]. Our second attempt relied on transforming some of the approaches followed in these courses using Python/Django to run the backend. We used Google's publicly available python library to communicate with DialogFlow CX API[33] as a steppingstone.

Frontend

Our initial exposure to ReactJs in (Bergant, 2018) strongly favoured its use over other popular web frame works such as Angular or VueJs. ReactJs is repeatedly recommended as the tool to use when fast development time is required and when clarity of design is a priority (Reis, 2020). Furthermore, as a matter of personal opinion, we found it intuitive and easy to learn.

Deployment – Frontend tools used:

1. A domain: https://chameleonpatients.co.uk/ is purchased for the future website.
2. The react app is built using VS Code[34].
3. A public Github repository is used to host the frontend codebase: https://github.com/AlHusseinA/ChameleonFrontendv2
4. Used Postman API[35] platform to test API end points throughout development
5. Following the completion of a basic functioning frontend, we use Vercel[36] hosting service to deploy the codebase. This is an automated process where the code is pushed from the local host to Github which is linked to Vercel.

Deployment – Backend tools used:

1. A domain: https://mybackendchameleon.com/ is purchased to redirect traffic to the backend.
2. A Django project is created in PyCharm.

[32] Code not included.
[33] https://github.com/googleapis/python-dialogflow-cx and https://googleapis.dev/python/dialogflow-cx/latest/index.html
[34] Procedure followed here is the standard in React development: https://reactjs.org/docs/create-a-new-react-app.html
[35] https://www.postman.com/
[36] https://vercel.com/

3. A Github repository is used to host the backend codebase (this is a different repository than the one used for the frontend codebase): https://github.com/AlHusseinA/chameleon-backend-django

6. Used Postman API platform to test API end points throughout development

4. Following the completion of a basic working backend on the local host, we use DigitalOcean hosting service to deploy. This is an automated process where the code is pushed from the local host to Github which is linked to DigitalOcean.

Frontend design

CSS Layout

Taking inspiration from the preliminary mock-up design (Figure 9) we use the Classic Holy Grail Layout from (Kravets, 2020) to create the CSS layout which will host the future React components (Figure 10). Here, the intention is to place the chat window and buttons pane inside the Main Content. The Left Sidebar, Right Sidebar and the Footer were removed during development and were not included in the final iteration of the code.

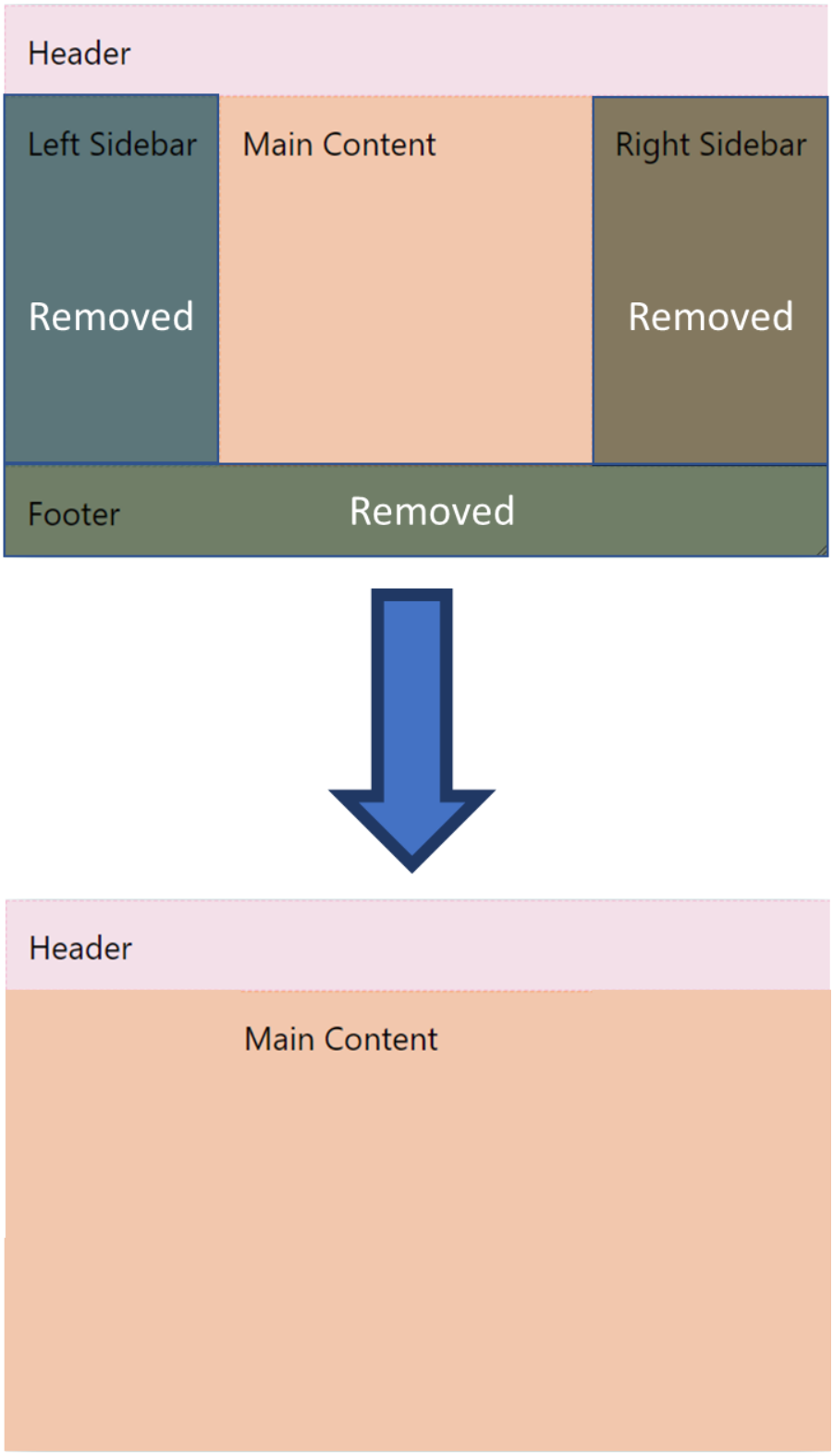


Figure 10 Basic design of the frontend. This formed the backbone of the frontend development. (above) The basic 1-line CSS layout that we started out with. (bottom) The shape of the final CSS layout. Left Sidebar, Right-Sidebar, and the Footer were removed. Main is where react Chat.js component will operate. Header is where react Header.js will operate.

React Components

Our final design is composed of 10 React components[37], Table 3 and Figure 11 present the components used in detail (next page).

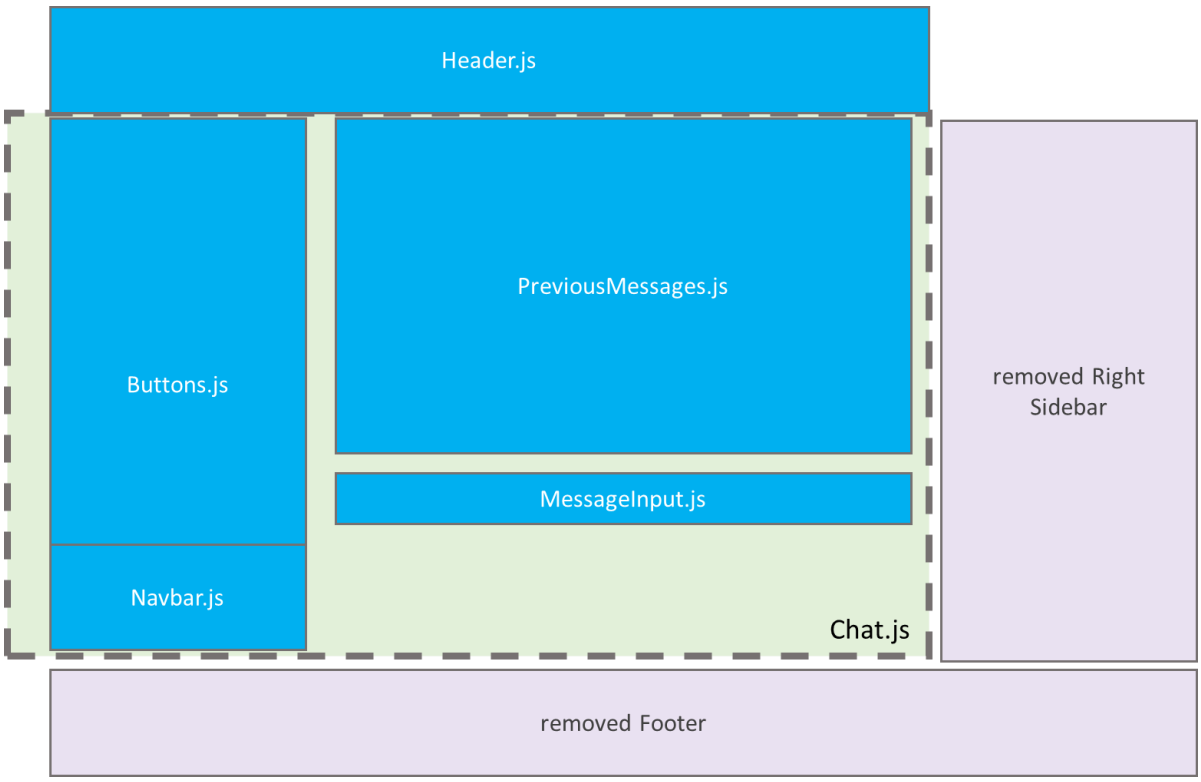


Figure 11 Final layout of the react components. Here we only show the components on the "\chat" page placed on top of the Header and Main CSS layouts (Left sidebar also removed, not shown). The components About.js, and FAQs.js are not shown. The area shaded in green is what is controlled by the Chat.js component which in turn is placed inside the main CSS layout.

[37] Briefly, a component in react is like a JavaScript function or class (when using ES6) that receives data in the form of a single 'props' object from other components. The component returns React element. It also can render HTML/CSS on the browser and manipulate React's own virtual Document Object Model (DOM) of the page. Itself an abstraction of the html DOM. React components are often declared in a way that enables reusability. For more information the reader is invited to visit: https://reactjs.org/docs/components-and-props.html

Component	Receives	Returns	Details
App.js	-----	The whole dynamic website	All components are descendants of App.js
About.js	-----	Returns a bootstrap container class showing background information about the project from Ethics Form C.	More information about the project for interested users.
Buttons.js	conversationUuid	The patient's biography box + buttons and associated logic + the Navbar component Figure 13.	1. Logic for the buttons on the left-hand side of the page (within the main CSS layout). Each click sends a hardcoded text message with conversationUuid to the backend. 2. Rendering of the patient's biographical information and photo. 3. Rendering of the Navbar.js component.
Chat.js	-----	Returns a bootstrap container that houses two bootstrap col classes. The first contains Buttons.js, and the second contains PreviousMessages.js and MessageInput.js	1. Renders and updates the input text box, message history box, and the buttons/links pane on the left-hand side. 2. Creates conversationUuid and passes it on to MessageInput.js and PreviousMessages.js. This is to attach the same session id to all communication between the frontend and the backend including fetching previously exchanged messages.
FAQs.js	-----	A bootstrap container which in turn contains list-group class	Displays organised and formatted FAQs for the users
Header.js	-----	Returns an html tag containing the main title of the page and Chameleon's icon on the top right corner	This component operates inside the header CSS layout discussed above Figure 10 Figure 11
Landing.js	-----	Returns a bootstrap container class to display user consent form.	1. Renders a formatted user consent form in bootstrap. 2. Logic for a button that the user must click on to agree to participate in the evaluation at the bottom.

MessageInput.js	conversationUuid	1. Returns the input text box and the associated submit button 2. Receives conversationUuid from the PreviousMessages component. This is such that react/backend knows that user's messages and fetched previous messagers belong to the same user (same session id as far as DialogFlow is concerned)	1. Logic for receiving input from the user and sending it to the backend with the conversationUuid using axios. 2. Logic for the text input and associated Submit button
Navbar.js	-----	A navbar of three links	1. Forwards the user to the questionnaire if they click on the respective link. 2. Renders the "/about" page when the respective link is clicked 3. Renders the "/faqs" page when the respective link is clicked
PreviousMessages.js	conversationUuid	1. Bootstrap card class that contains all exchanged messages between user and Chameleon. Messages are coloured differently based on id.	1. React hook useSWR remotely fetches the message history from the Django backend and continuously updates the rendering of the message history every 1000 ms. 2. Handles the logic of alternatively colouring the displayed messages 3. Handles the logic of continuously scrolling down the latest message into view inside the previous messages box. Otherwise, the user would have to scroll down manually to see that last message once the exchange is bigger than the size of the box.

Table 3 Details of all React components used on the website. The user is invited to examine Figure 11 and Figure 12 for a graphical, and perhaps more intuitive, architecture of the system.

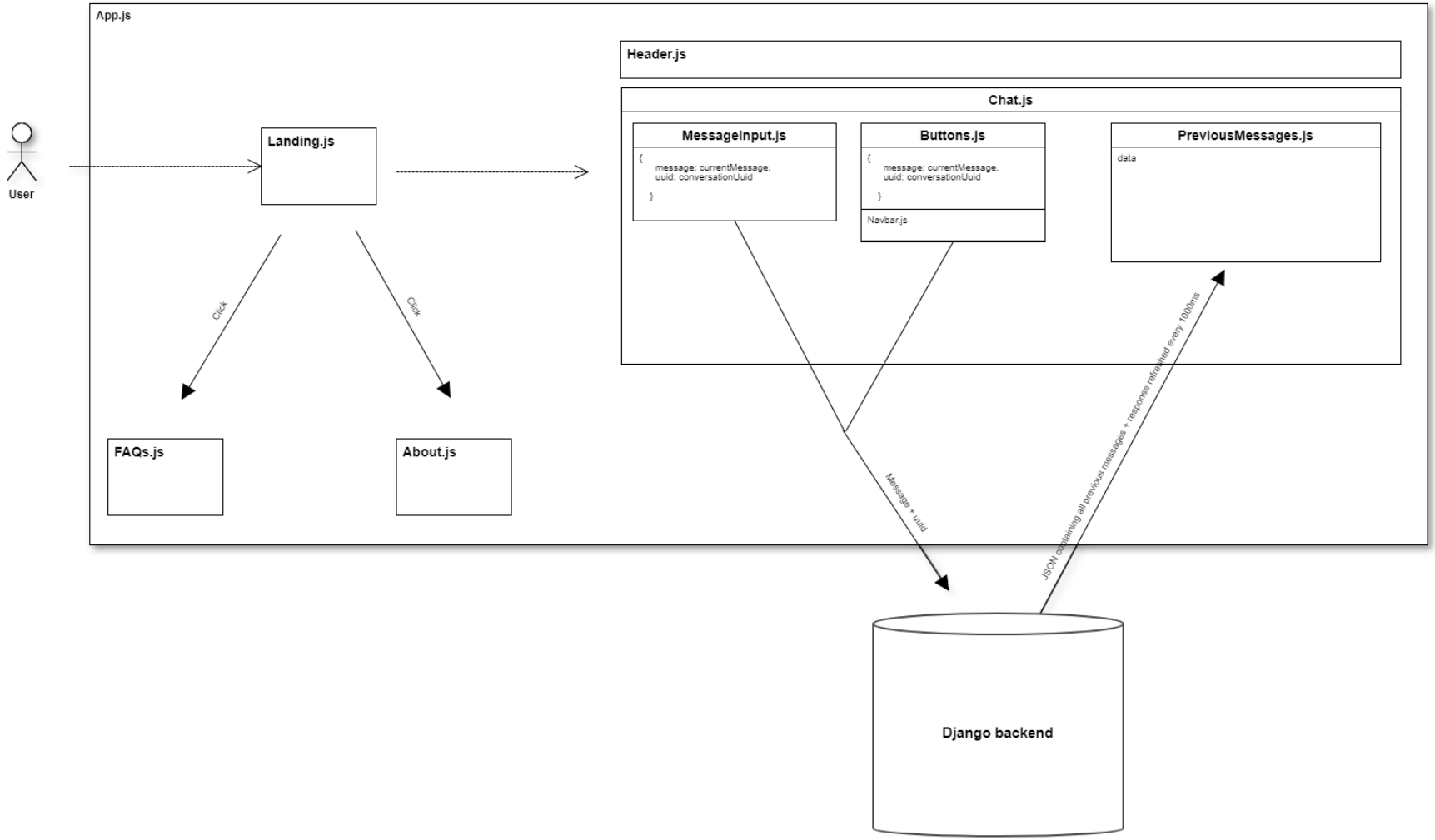


Figure 12 The user is first shown the Landing page rendered by the component Landing.js. Once the user consents to taking part in the evaluation, the main "/chat" page is rendered by App.js, the children components of Chat.js and Header.js. The components MessageInput.js, Buttons.js, Navbar.js and PreviousMessages.js are rendered within the Chat.js component. The components FAQ.js and About.js are only rendered if their respective links are clicked on by the user in the Navbar.js Component (shown in Figure 11).

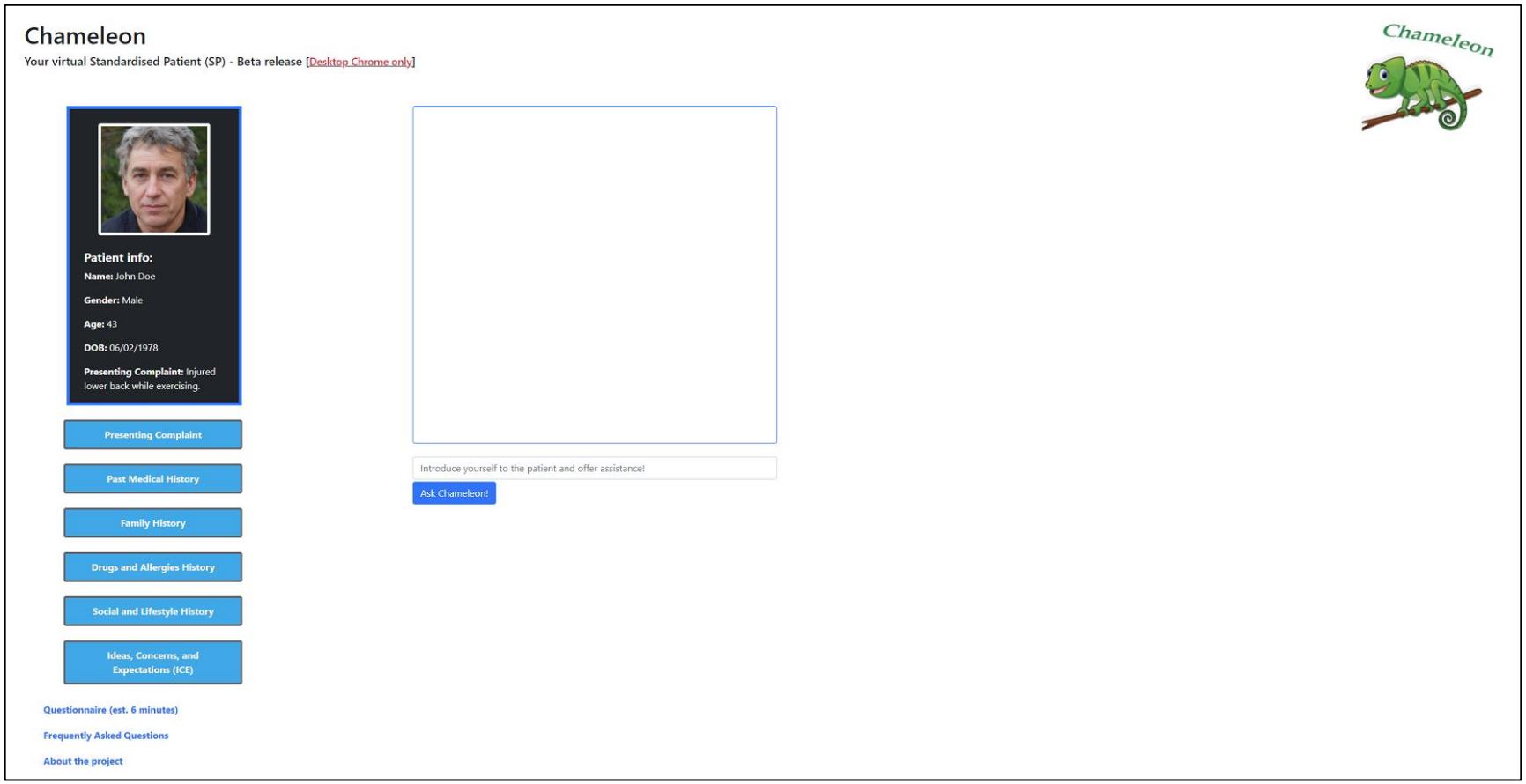


Figure 13 The final deployed frontend. The space on the right of the main chat window is where our early mock-up envisaged the placement of lab testing/physical exams functionality. A decision during the development was made to place an element containing "How to" instructions[38]. However, Subject C who tested the website dissuaded us. The justification was that anyone with clinical training will find the website intuitive to use. This decision is clearly reflected in the evaluation results, as we will see in the Results chapter, where a clear majority of subjects assessed the website to be intuitive and easy to use.

[38] Code not included

Frontend known bugs

1. The whole page keeps slightly scrolling down with every input. This was only discovered after deployment. This is not the same bug reported in the second live testing round but a consequence of it nonetheless.
2. On the landing page, there's a divider line beneath the button. It is unnecessary.
3. The buttons are invisible on the Safari browser.
4. The Previous Messages box does not render correctly on Firefox
5. On the '/chat' page, if refreshed there is flickering, and the scroll bar of the main page moves downward a little.
6. Elements on the chat page don't resize correctly when browser size is changed.
7. Does not work elegantly on mobile apps.

Backend design

A Django project is created in a local virtual environment[39]. Following that a Django app 'api' is created within the project. We make use of the infrastructure provided by Django and Django Cookiecutter library and place our logic as follows:

service.py

Houses the core logic of the backend. The three functions below are taken from Google's dialogflow repository[40] with minor modifications.

Function	Details
get_client( )	1. Creates a session with the necessary credentials between the client and dialogflow's api (agent)
build_intent_query( )	1. Builds the query (regardless of whether it was text or event) according to Google's format
detect_inent_texts( )	1. This will call build_intent_query( ) and passes the result to dialogflow. 2. The response is stored in 'returned_message'
call_dialogflow( )	1. Returns a call to detect_inent_texts( )

Table 4 Core logic of the backend

views.py

Views main function is to house classes that generate and process the http responses and requests. All communication between Google's DialgoFlow API and the React frontend takes place through the IntentionView, EventView, and PreviousMessagesView classes.

Class	Details
IntentionView	1. Receives json object containing a message (user's question to Chameleon) and the uuid. The json object is passed to dialogflow and response is received by calling call_dialogflow( ). 2. Adds the message received, and the id of the sender/receiver[41] to the data base via the Conversation object in model.py.

[39] https://docs.djangoproject.com/en/3.2/intro/tutorial01/
[40] https://github.com/googleapis/python-dialogflow-cx/blob/HEAD/samples/snippets/detect_intent_texts.py
[41] 'Me' for the user and 'Chameleon' for Chameleon

PreviousMessagesView	1. Will constantly be retrieving all messages stored in the PostgreSQL db. 2. This is done via a second end point in url.py dedicated to retrieving the archive. 3. PreviousMessages.js React component receives a data object via the useSWR hook discussed in Table 3. In turn, PreviousMessages.js will use the id ('Chameleon' or 'Me') of the messages to determine the background colour they're displayed in.
EventView	This is legacy code that we wrote around the same time we designed the first iteration of the frontend. The original function of this class is to send an 'event' query to DialogFlow as opposed to a text query (as we do in the final deployed version). This is kept in place and not deleted to enable future iterations to model click events queries.

Table 5 Views.py - classes that handle the logic for API communication

models.py

Controls the data flow to and from the postgres database. It contains one Conversation class with attributes for uuid, created, and messages. Each will house information regarding the message stored.

urls.py

Defines three API end points. One for communication between DialogFlow and the backend, and two for communication between the PreviousMessages.js and MessageInput.js components ReactJs-side and the Django-python backend.

Class	Details
Endpoint for dialogflow	1. path("api/", include("config.api_router"))
Endpoint for MessageInput.js	1. path('intentions/', IntentionView.as_view()) 2. React-side: const BASE_URL = process.env.REACT_APP_API_ENDPOINT const url = `${BASE_URL}/intentions/`
Endpoint for PreviousMessages.js	1. path('messages/<conversation_uuid>', PreviousMessagesView.as_view()) 2. React-side: const BASE_URL = process.env.REACT_APP_API_ENDPOINT const url = `${BASE_URL}/messages/${conversationUuid}`

Table 6 API end points in url.py

plucky.json

JSON object containing the necessary authentication code. Please see Appendix K for a detailed procedure on how to obtain the credentials.

Deployment

DigitalOcean was our first choice for a hosting platform due to 1) It being able to handle dozens of users concurrently[42]. 2) Affordability, it is priced at $5 per month for the 1GB RAM VM 'Droplet' option. 3) Offers plenty of educational material that include Django deployments.

Figure 14 and Figure 15, offer a visual representation of the backend and the entirety of the end-to-end solution, respectively.

The reader is invited to Appendix K for the complete deployment procedure.

[42] No firm estimate on maximum number of users per unit time from DigitalOcean docs. However, estimates on the internet range from 50-200 users simultaneously for the type of Droplet we used.

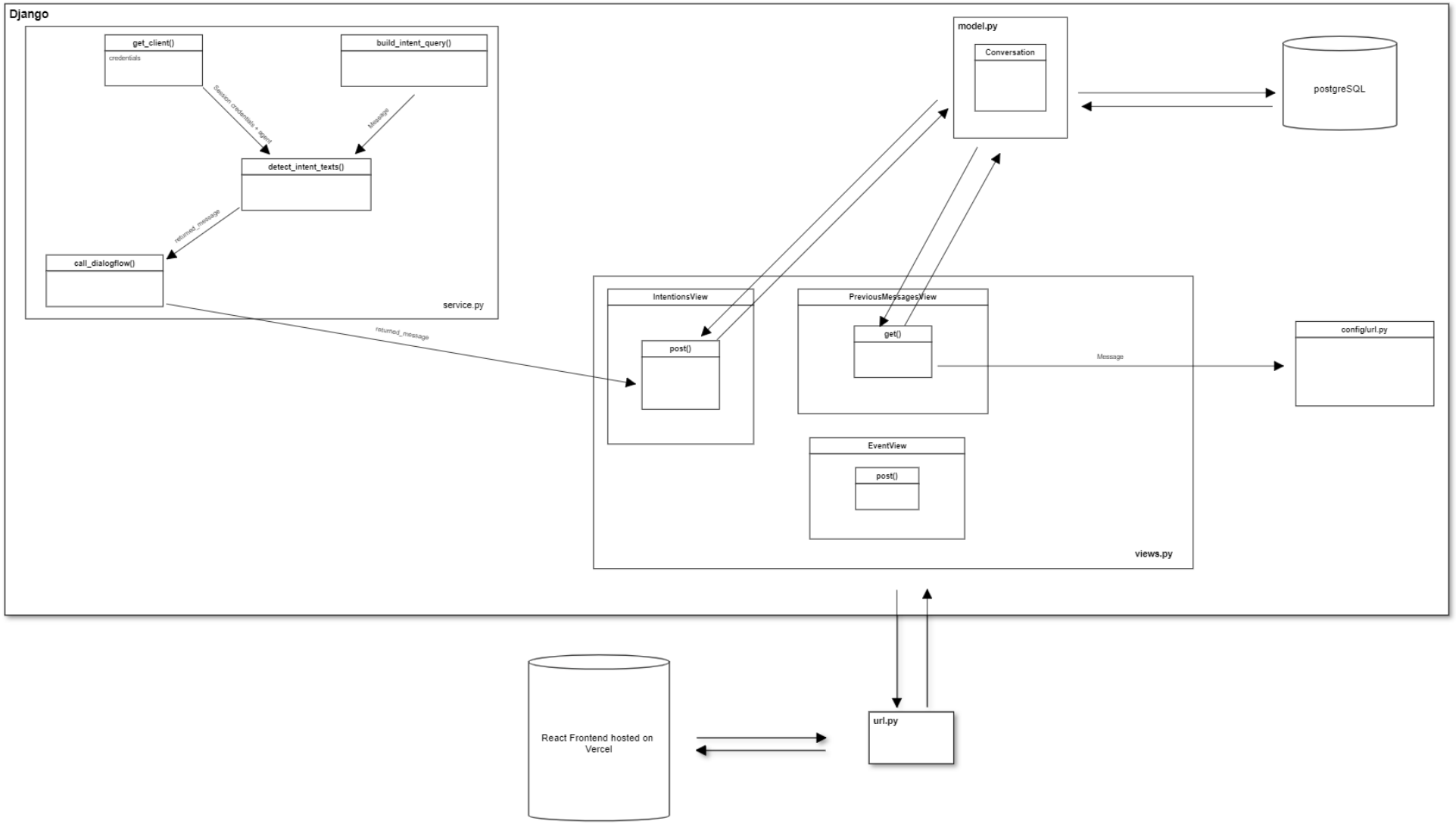


Figure 14 url.py defines three API endpoints. 1) When a question is typed in the text box created by the MessageInput.js component, the text is wrapped and a uuid is attached to it. It is then posted using axios to the IntentionView class via its respective API end point in url.py. 2) Following that all exchanged messages are constantly being saved in the database. To display the message history on the screen, PreviousMessages.js component requests the whole archive from the PreviousMessagesView class via its respective endpoint. PreviousMessages.js then formats the received object of all previous messages such that messages from the user are green and those from Chameleon are blue. 3) Once the received question from the user has been formatted and authenticated in service.py, it is then sent to DialogFlow for the response to be received on the third defined endpoint in url.py.

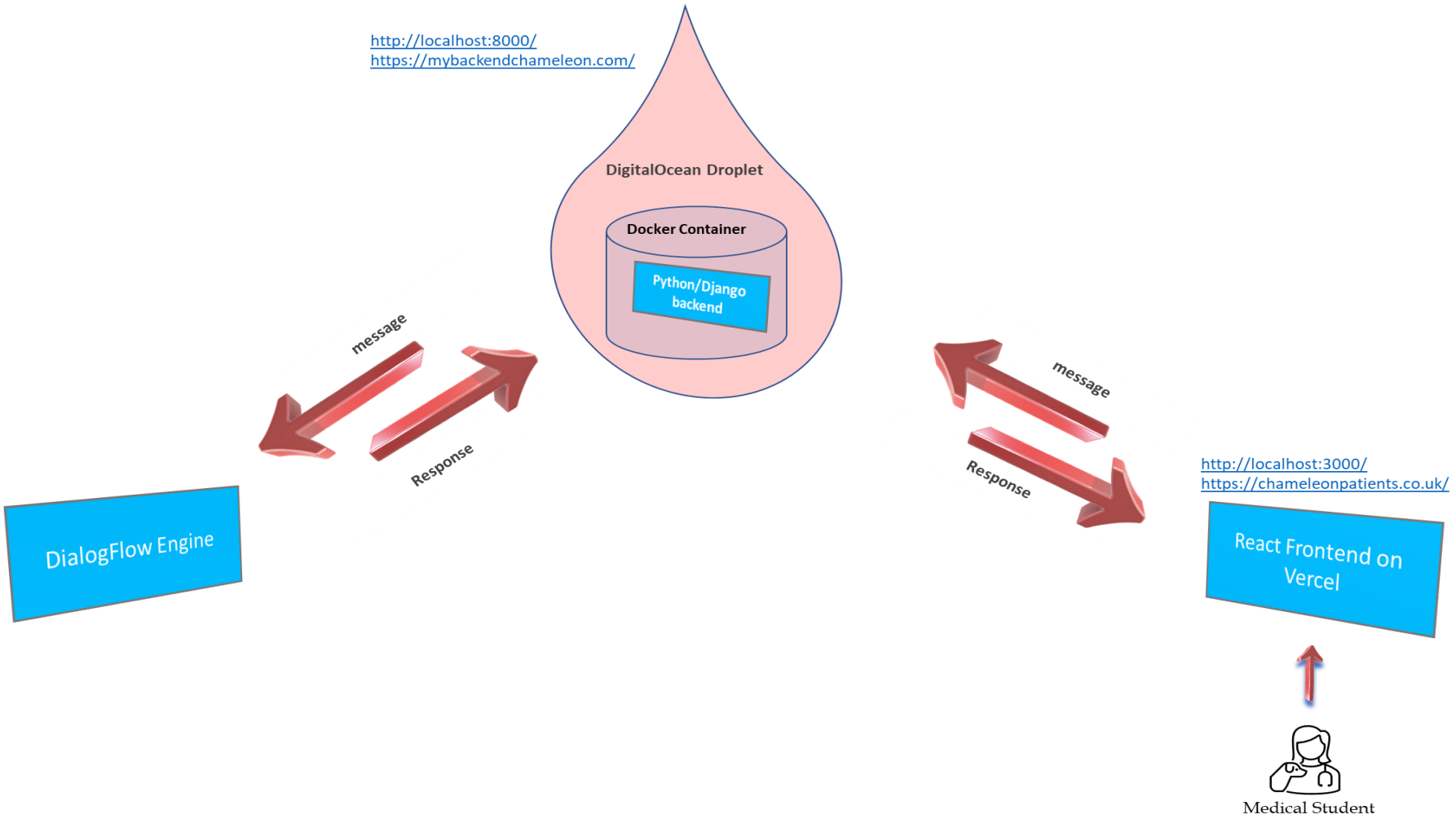


Figure 15 The complete end-to-end solution. The user types in their questions on the react frontend, the query gets posted to an IntentionsView object in the backend which in turn passes the query to the Dialogflow engine via its respective endpoint using functions in service.py. The response is then passed back to the frontend the same way it was sent.

Evaluation

Design of the survey

Our survey was designed to address user perception with respect to three key metrics 1) overall performance of Chameleon. 2) Overall performance of the website used to interact with Chameleon. 3) Technology adoption and acceptance. 4) Complete experience of the user.

In our earlier two rounds of live-testing, the shortest exploratory performance live test took 12 minutes. The longest took over 55 minutes. This is a significant investment in time from any user with no tangible and immediate return.

Assuming Chameleon performs perfectly, we estimate from the two tests in the second round of testing, that a doctor will reach a diagnosis with high confidence in no more than 10 minutes and thus conclude the best-case scenario interaction with Chameleon[43]. This imposes a constraint on how detailed our user evaluation survey is going to be since the response rate of participants has a negative linear correlation with the length of the survey (Fan and Yan, 2010; Hoerger, 2010). Survey length was a central factor on how detailed our survey would be.

In total, and according to MS Forms own algorithm, it will take an average user 6 minutes to complete the survey. However, three volunteers[44], one native speaker of English and two non-native speakers, found that on average it took them 4.6 minutes which roughly confirms Microsoft's estimation. A 6-minutes investment from the user was deemed reasonable.

The curious reader is invited to Appendix I for the complete survey used and discussed in this chapter.

Structure of the survey

A total of 26 questions were formulated. Three were open ended and the rest were close ended questions. A mixture of yes/no, Likert scale, and two rating scale questions were used. When designing this questionnaire, the following was considered:

- Minor time commitment from the user.

[43] This does also assume that Chameleon doesn't breakdown as observed in live testing, that the user wants to find a diagnosis and thus will ask the necessary question, the user is a doctor, etc.
[44] Colleagues from the AI MSc course.

- The questions address the main objectives of the survey.
- Clear and easy to answer.

Participants' Background

We ask two question to establish whether the participants have a clinical background. This is done to establish a correlation between those who have such a background and their perception of the way the website is designed. Our prediction is that doctors and 4th or 5th year medical students will, indeed, find it intuitive to use the website.

Question	Format
Are you a 4th or 5th year medical student?	Yes/No
What is your speciality in healthcare?	Single-answer multiple choice (Physician/Dentist/Lab Specialist/Nurse/Other)

Table 7 Questions on users' clinical background

Chameleon's overall performance as a Standardised patient

Robustness of Chameleon as a software

On several occasions during the live testing rounds Chameleon suffered from complete breakdown where it was completely unable to detect any utterance including utterances that match instances in the training corpus. Quantitatively establishing Chameleon's breakdown will shed light on the current robustness of the solution. Furthermore, while the term 'breakdown' was explicitly defined for the user, users may have perceived that the system broke just by continually repeating and rephrasing an utterance for which no intent exists. At the time of deployment, there was no mechanism of delineating which users experienced real breakdown, and which did not.

Overall performance of Chameleon as a conversational agent

Questions here address performance with respect to Chameleon's ability to understand language, respond correctly, the naturalness of its prescribed responses, diagnostic outcome of the interaction, suggestions on what could have made reaching a diagnosis easier, possible

conversational flow outside the backpain focus, and any ethical concerns the users wish to express.

Technology adoption and acceptance

One question explicitly that asks the users if they can conceive of Chameleon eventually replacing the role played by SPs. This aspect is also implicitly explored via other questions (next chapter).

Table 8 (below) Questions that address Chameleon's overall performance as a Standardised patient.

Question	Format
Did Chameleon break? Was there a point in the interaction where no matter what you asked the chatbot was unable to understand your input?	Yes/No
Overall how do you rate Chameleon with respect to: **1. Correctly understanding what you asked** **2. Its ability to to[45] mimic a real patient** **3. Naturalness of Chameleon's responses**	Matrix - 5 points Likert scale
Were you able to reach a diagnosis? **1. Yes. With high confidence.** **2. Yes. With moderate confidence.** **3. Yes. With low confidence.** **4. No**	Single-answer multiple choice:

[45] Typo in original survey.

If you did not reach a high confidence diagnosis, which of the following would have made reaching a high confidence diagnosis easier for this presentation? 1. **Options to perform a virtual exam on the affected area (using an animated avatar of the patient's body)** 2. **Options to order medical imaging procedures and view the results (using AI generated images of the affected area)** 3. **Options to order lab tests (using randomly generated but realistic readings)**	Multiple-answers multiple choice:
Overall, how do you rate Chameleon's performance as a virtual standardised patient?	Rating scale questions (Net Promoter Score)
Given your interaction with the chatbot Chameleon, do you think it could eventually replace the use of Standardised Patients in OSCEs? 1. **Yes, immediately.** 2. **Yes. After significant improvements to the current version.** 3. **Neutral.** 4. **No. Not with the performance of the version just tested.** 5. **No. Not even after significant improvements.**	Single-answer multiple choice:

Did you try to ask the virtual patient questions that are not directly related to the symptoms? For example, under Social and Lifestyle History you could have asked it about its recent travel history	Yes/No
(optional) What were the topics you asked about that were not directly related to the symptoms?	Open ended
In your subjective opinion, which of these statements do you agree most or least with: **1. I think I would like to use Chameleon again** **2. I think I would like to use Chameleon with a different clinical presentation**	Matrix - 5 points Likert scale
(Optional) Do you have any concerns surrounding the ethics of using Chameleon as a pedagogical tool?	Open ended

Performance of the website

Most of the questions were taken and modified from the SUS-A questionnaire (Brooke,1996). These questions assess the performance of the web development work conducted in the previous chapter, the intuitiveness of the design for doctors, ease of use, responsiveness and reliability of the design. In total, 8 questions were asked, all were 5-point Likert scale questions.

Dedicating a third of the questionnaire to evaluate the website is due to the significant impact it places on users' perception of the performance of the core content of a website, Chameleon in this case. Indeed, Thielsch and Blotenberg's, (2013) work suggests usability and content of

a website are the two most significant factor when it comes to users' perceptions and intention to revisit the website.

Question	Format
With regards to the website you used to interact with Chameleon. How do you rate the following: 1. **Ease of use** 2. **Intuitiveness of the User Interface (UI)** 3. **Responsiveness (is it fast?)** 4. **Reliability (does it work correctly?)**	Matrix - 5 points Likert scale
In your subjective opinion, which of these statements do you agree most or least with: 1. **I found the design of the buttons, message history, and patient biographical information well integrated.** 2. **I thought there was too much inconsistency in the website** 3. **I believe most doctors and medical students with clinical training would find the website easy to use** 4. **I found the website cumbersome to use**	Matrix - 5 points Likert scale

Table 9 Questions related to the performance of the website

Complete user experience

One ranking scale question that asks the following:

Considering your complete experience with our software solution (website and chatbot), how likely would you be to recommend its use to a colleague for training or educational purposes?[46]

Openness to the technology and user suggestions/feedback

Two questions Table 10. One that offers the users to choose multiple answers as to where future development work should focus on. The other question is an open ended and optional question where users are free to provide whatever feedback they feel is appropriate. Our open-ended question is deliberately worded with a preamble that contextualises the importance of the participation of the user for the purpose of advancing Chameleon's capabilities. Furthermore, the question removes some of the bias inherent in the earlier multiple-choice questions that ask for users' opinions with regards to future development, or what could have made reaching the diagnosis easier (Table 8 – last question) (Reja et al., 2003).

Question	Format
What do you think future releases of Chameleon should focus on? - **The website** - **The prescribed responses of Chameleon** - **The internal algorithms responsible of identifying the meaning behind your questions**	Multiple-answer multiple choice question

[46] This question was taken from a software evaluation template that can be found on: https://www.questionpro.com/survey-templates/software-evaluation/

User feedback is instrumental in pulling our development focus to where it is most needed. **Any suggestions or feedback on how to improve Chameleon?**	Optional and open ended

Table 10 Questions on openness to the technology and user suggestions/feedback

Recruitment of participants

The recruitment drive was focused on medical students. All thirty-three medical schools in the UK were approached. A full list of the schools was created from the UK Medical Schools Council and from Wikipedia[47]. The following procedure was followed to create a database of contact details:

1. On each school's website, emails for all relevant medical students' societies were collected. This includes societies for sub-specialities such as orthopaedics or emergency medicine, it also includes some medical societies that represent minority students such as Muslim, Christian and LGBT medical societies.
2. Where possible, Facebook links for the pages of those societies were also collected.

A table of the full details of this database is included in Appendix F.

Following the collection of the contact details, we implement the following procedure to establish contact with the societies:

1. A draft standard email is written and emailed to 233 medical societies across the UK. A copy of the email used is included in Appendix G. Recipients were explicitly informed of the expected 6 minutes time commitments to fill the survey.
2. 45 societies were also contacted via Facebook or Instagram using the same draft email. Not all of those were contacted via emails.

Our procedure ensures that members from all medical schools in the UK were invited to take part whether via email or social media.

[47] https://www.medschools.ac.uk/studying-medicine/medical-schools and https://en.wikipedia.org/wiki/List_of_medical_schools_in_the_United_Kingdom

Society	Membership
Galenicals (University of Bristol Medical Students Society)	1969
University of Cambridge MedSoc	3468
Imperial College School of Medicine Students' Union	5245
Liverpool Medical Students' Society - LMSS	1669
Plymouth Cardiology Society (PCS)	312
UCL Medical Society	3455
University of Aberdeen GP Society	402
Glasgow Medical Research Society	1433
Glasgow MedTech Society	81
Cardiff University Medsoc	2046

Table 11 The medical societies that responded positively to the survey. In all cases, the response was a promise to forward the details of the survey/website to their membership which totals to an unreliable estimate of 20,647 based on the size of their Facebook pages.

Response

In total, only 10 students' societies responded positively. All 10 agreed to forward the email communication they received from us to their members. The survey was online from 00:00 18th August 2021 to 12:00 noon 24th August 2021. Only twelve participants responded and filled the survey .

Table 11 includes the names of the societies that responded positively. The membership number is not an accurate reflection of their membership, however. This is the number of members they have on their Facebook pages. The total number of presumed students contacted is 20,647.

Results

Background of participants

All 12 participants were either practicing physicians or 4th/5th medical students. 67% reported they are currently a student.

Overall performance of Chameleon

Question	Results and Insights
Overall how do you rate Chameleon with respect to: 1. **Correctly understanding what you asked** 2. **Its ability to mimic a real patient** 3. **Naturalness of Chameleon's responses**	➢ 67% (8) of participants reported that Chameleon did suffer from a breakdown. ➢ In response to questions of how the users perceived Chameleon's performance. A significant minority 41% reported Fair when asked if Chameleon understood what they asked. A majority, 58%, reported Good or better to the same question. ➢ In total, 75% (9) reported Good or better with regards to the naturalness of the prescribed answers. 58% Good, 8.3% Very good, and 8.3% Very good. ➢ And 66% of users rated Chameleon's performance as a in mimicking the role of the SP as Good or better. These results suggest that while Chameleon displayed acceptable but not exceptional performance when it came to correctly understanding users' queries, when it did understand the queries, it was convincing. Figure 16.
Were you able to reach a diagnosis?	83% (10) said they were able to reach a diagnosis. One subject with high confidence, seven with moderate confidence and two with low confidence. This result suggests that some measure of experiential learning has taken place via the interaction with Chameleon. Reaching a diagnosis, including for simple presentation, requires the user to objectively assess the information gained from the interaction with Chameleon (Figure 17).

Table 12 Overall performance of Chameleon - part 1

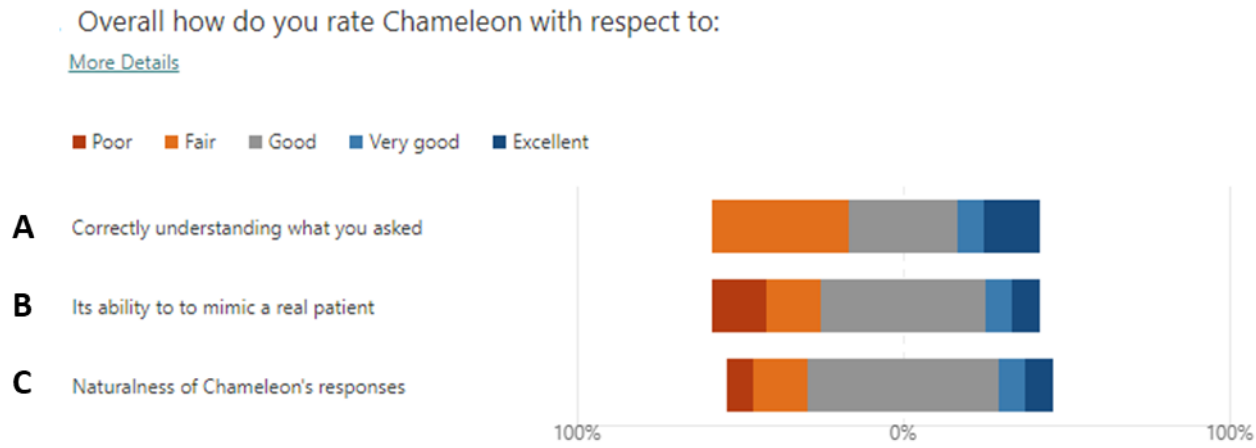


Figure 16 In response to the A, a sizable minority, 41.7%, indicated Fair suggesting less than ideal performance. In response to B, 66% (8) of participants reported Good or better. In C, 75% (9) also reported Good or better. B and C suggest that while the bot may have not shown exceptional performance with regards to understanding users' inputs, it was convincing when it did so.

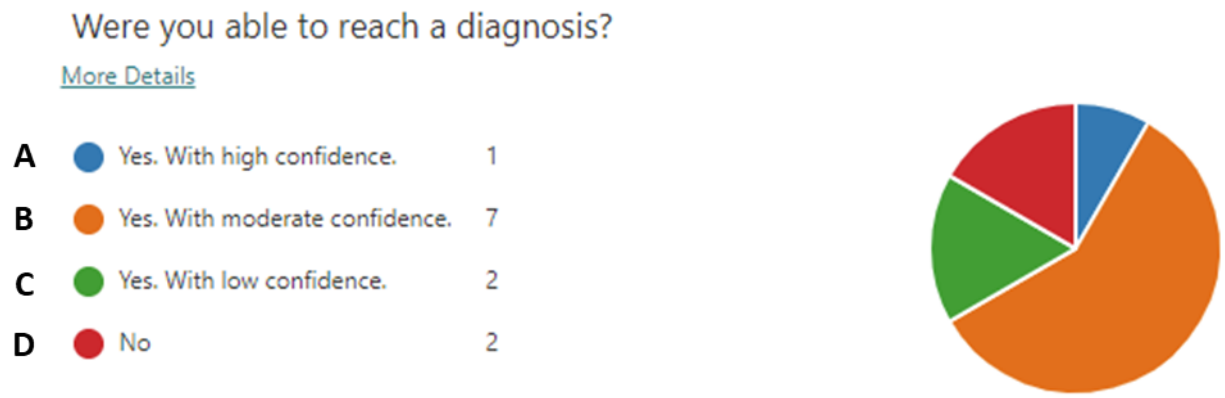


Figure 17 Overall, 10 subjects reached a diagnosis with low or higher confidence. Only two subjects were not able to reach a diagnosis. Suggesting that the role of the SP was successful in achieving some measure of learning by encouraging the user to objectively asses the interaction to reach a diagnosis.

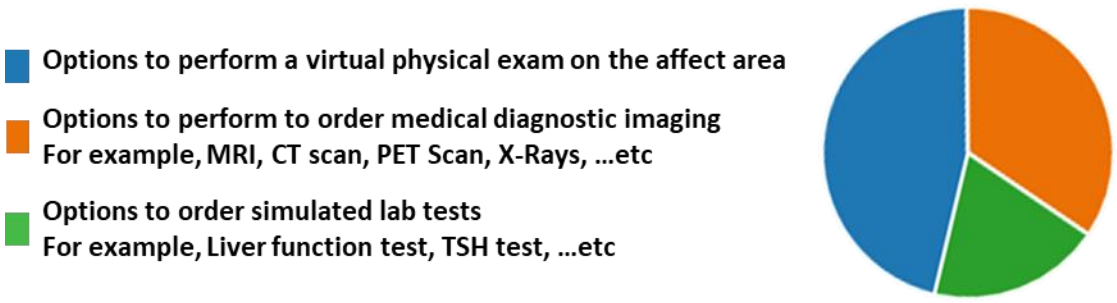


Figure 18 When offered the choice of what would make them able to reach a diagnosis easier, users selected the option of being able to perform virtual physical exams with the highest frequency, followed by the option to order medical imaging studies, followed by being able to order simulated lab tests. The results are suggestive of a level of acceptance and openness toward the use of conversational agent as a pedagogical tool.

Question	Results and Insights
If you did not reach a high confidence diagnosis, which of the following would have made reaching a high confidence diagnosis easier for this presentation? 1. **Options to perform a virtual exam on the affected area (using an animated avatar of the patient's body)** 2. **Options to order medical imaging procedures and view the results (using AI generated images of the affected area)** 3. **Options to order lab tests (using randomly generated but realistic readings)**	This question indirectly assesses users' openness toward the use of this technology and its future iterations in the future in teaching clinical medicine. Respondents will entertain the notion that Chameleon has some real-world applications, even if not at the state it is currently in, if they can see themselves and others using it. It also implicitly primes the respondents to answer the important open-ended question in Table 16. The question inherently restricts users in expressiveness, this can be resolved with answering the final question. Moreover, this approach is also applied to the penultimate question. The highest frequency (12) was for including an option to allow users to perform a virtual exam on the affected area, followed by options to order medical diagnostic imaging virtually (9). And (5) for ordering simulated lab tests. All three options were in the original design, and perhaps too ambitious, objectives set for Chameleon very early in the project. The results are suggestive of the validity and usefulness of those objectives to medical doctors and clinicians. Furthermore, we estimate that if the Chameleon presented with symptoms of a condition that requires some lab analysis of body fluids; users would potentially have favoured the option that allows them to order simulated lab test. For example, if the patient presented with symptoms suggestive of hypothyroidism; the doctor would need to order a Thyroid-stimulating hormone (TSH) blood test to confirm the diagnosis. Users are likely to have favoured future focus on the first two options as they are the most relevant to Chameleon's current presentation. Being exposed to a condition that can benefit from medical imagining and physical examination, backpain, primed the subject to favour, generalise, those options. Figure 18

Overall, how do you rate Chameleon's performance as a virtual standardised patient?	No users were promoters, 4 were passive, and 8 were detractors. NPS score of -67. Taken into the context of the previous questions, this result, while certainly disappointing, is certainly likely to change given future modifications the users suggested.
Given your interaction with the chatbot Chameleon, do you think it could eventually replace the use of Standardised Patients in OSCEs? (Figure 19)	➢ 67% (8) of users answered either 'No. Not with the performance of the version just tested' or 'No. Not even after significant improvements.' ➢ 33% (4) of users answered 'Yes. After significant improvements to the current version' (Figure 19) ➢ These results are inline with the above rating scale question (8 detractors), however, also indicate of a level of openness should the necessary modifications were deployed. When considering the 'Yes, after improvements' option and the 'No. Not with current performance' option to be identical in the subtextual meaning of what is being asked from the users; it is in fact 91.6% of participants who envisage Chameleon, with significant improvements, replacing the use of human SPs in OSCEs.

Table 13 Overall performance of Chameleon - part 2

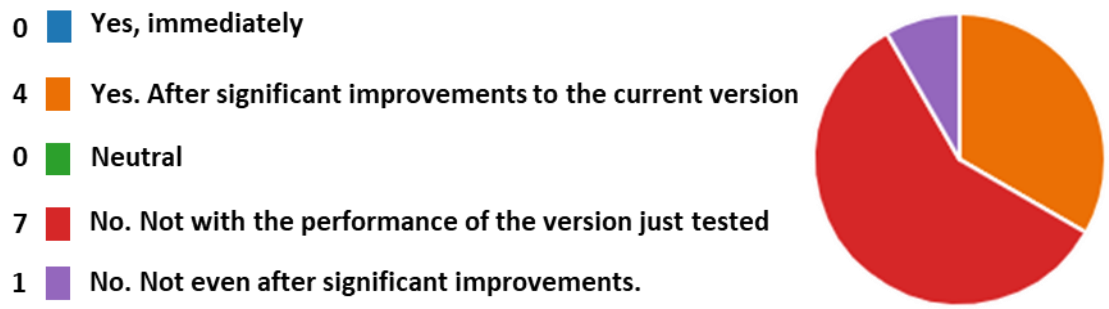


Figure 19 In response to the question 'Given your interaction with the chatbot Chameleon, do you think it could eventually replace the use of Standardised Patients in OSCEs?', 91.6% of participants indicated they can envisage Chameleon replacing the use of human SPs in the future given significant changes. Question B (orange) and D (red) are considered equal in subtextual meaning here.

Question	Results and Insights
In your subjective opinion, which of these statements do you agree most or least with: **1. I think I would like to use Chameleon again** **2. I think I would like to use Chameleon with a different clinical presentation** **(Figure 20)**	➢ 75% (9) of users either Strongly agreed or Agreed that they would like to try Chameleon again. ➢ 75% (9) of users also either Strongly agreed or Agreed they would like to see a different clinical presentation in Chameleon. ➢ The participants here are invited to think subjectively about their experience of interacting with Chameleon and report possible future use depending on that experience. This is another question that indirectly attempts to gauge the level of openness to this technology. Despite the result of the earlier rating scale question (Table 13) being suggestive that most users would not recommend the use of the system to others (8 detractors, no promoters); a clear majority here when used different phrasing state that they themselves would use the system again without modifications (there's no way to tell for certain if the same 9 users who Agreed/Strongly agreed with the first statement are the same 9 users who also Agreed/Strongly agreed with the second statement). (Figure 20)
What were the topics you asked about that were not directly related to the symptoms	42% (5) of users attempted to explore the conversational capabilities of Chameleon. However, only two subjects explicitly mention how they attempted to do so. One asked Chameleon to explain what 'deadlifting' is[48]. The other user asked Chameleon about its diet.

[48] John Doe, the imaginary SP, presents to the clinic with low back pain after a deadlifting session in the gym the previous night.

Do you have any concerns surrounding the ethics of using Chameleon as a pedagogical tool?	Three users answered this question in detail: 1. First participant raised a valid concern of whether Chameleon would be able to detect and 'effectively challenge any offensive or poorly-phrased statements' entered by the doctor as a normal SP would. The user pointed out that in real life, repeated use of such statements may normalise them and, if not challenged, doctors may believe that such statements are acceptable. The user did not offer examples.
	2. The second participant suggested that 'works well as an additional resource for students to use as practice', however the user opined that it should not be used as a replacement for a human SP, whether actor or real patient in an OSCE setting.
	3. Importantly, the third participant asked about Chameleon's catering for users with learning disabilities. The user specifically mentioned 'Dyslexia, eyesight, typing, colour scheme'. Furthermore, the user pointed out that assessing students in a written format for a clinical exam is not representative of a real OSCE since a student could lose marks for issues that are not directly related to the interview. No examples were given.

Table 14 Overall performance of Chameleon - part 3

Figure 20 75% (9) of users indicate that they would like to use Chameleon again, and/or use it again with a different clinical presentation. Users' willingness to further dedicate time to experiment with this solution is suggestive of their receptivity and openness to it.

Performance of the website

Interestingly, while a small majority of users rated the performance of Chameleon as a conversational agent as 'good' or above. A very clear majority rated the website overall as Very good or Excellent.

Question	Results and Insights
With regards to the website you used to interact with Chameleon. How do you rate the following: **1. Ease of use** **2. Intuitiveness of the User Interface (UI)** **3. Responsiveness (is it fast?)** **4. Reliability (does it work correctly?)**	- 58% (7) users rated the website as either Excellent or very good in terms of Ease of use. 83% (12) of users rated it Good or above in terms of Ease of use. - 67% (8) of users rated the Intuitiveness of the UI as Excellent. 92% (11) of users considered the Intuitiveness of the UI as Good or above. This particular result is in line with Subject C's recommendation during the second round of life testing. They suggested that the design as it stood at the time was more than intuitive and that most clinicians would find it easy to use and interpret. - 92% (11) of users considered the website to Excellent or Very good when asked about Responsiveness of the site. - 67% (8) of users rated the Reliability of the website as either Excellent or Very good. 25% (3) of users rated it as Fair or Poor. - Excellent ratings for responsiveness and reliability of the website are partly due to the backend being hosted on VM on DigitialOcean.

In your subjective opinion, which of these statements do you agree most or least with: 1. **I found the design of the buttons, message history, and patient biographical information well integrated.** 2. **I thought there was too much inconsistency in the website** 3. **I believe most doctors and medical students with clinical training would find the website easy to use** 4. **I found the website cumbersome to use**	- 83% (10) of users either Strongly agreed or Agreed that the rendering of the Chat.js React component containing previous messages box, text input box, buttons, navbar, patient information was well integrated. No users Strongly disagreed with this statement. - 83% (10) of users either Strongly disagreed or Disagreed with the statement suggesting inconsistent design of the website. - 92% (11) of users either Strongly agreed or Agreed suggesting that doctors or medical students with clinical training would find the website easy to use. - 50% (6) of users either Strongly disagreed or Disagreed with the statement suggesting that the website was cumbersome to use. 17% (2) of users Strongly agreed or Agreed that it was.

Table 15 Overall performance of the website

(this space is intentionally left blank)

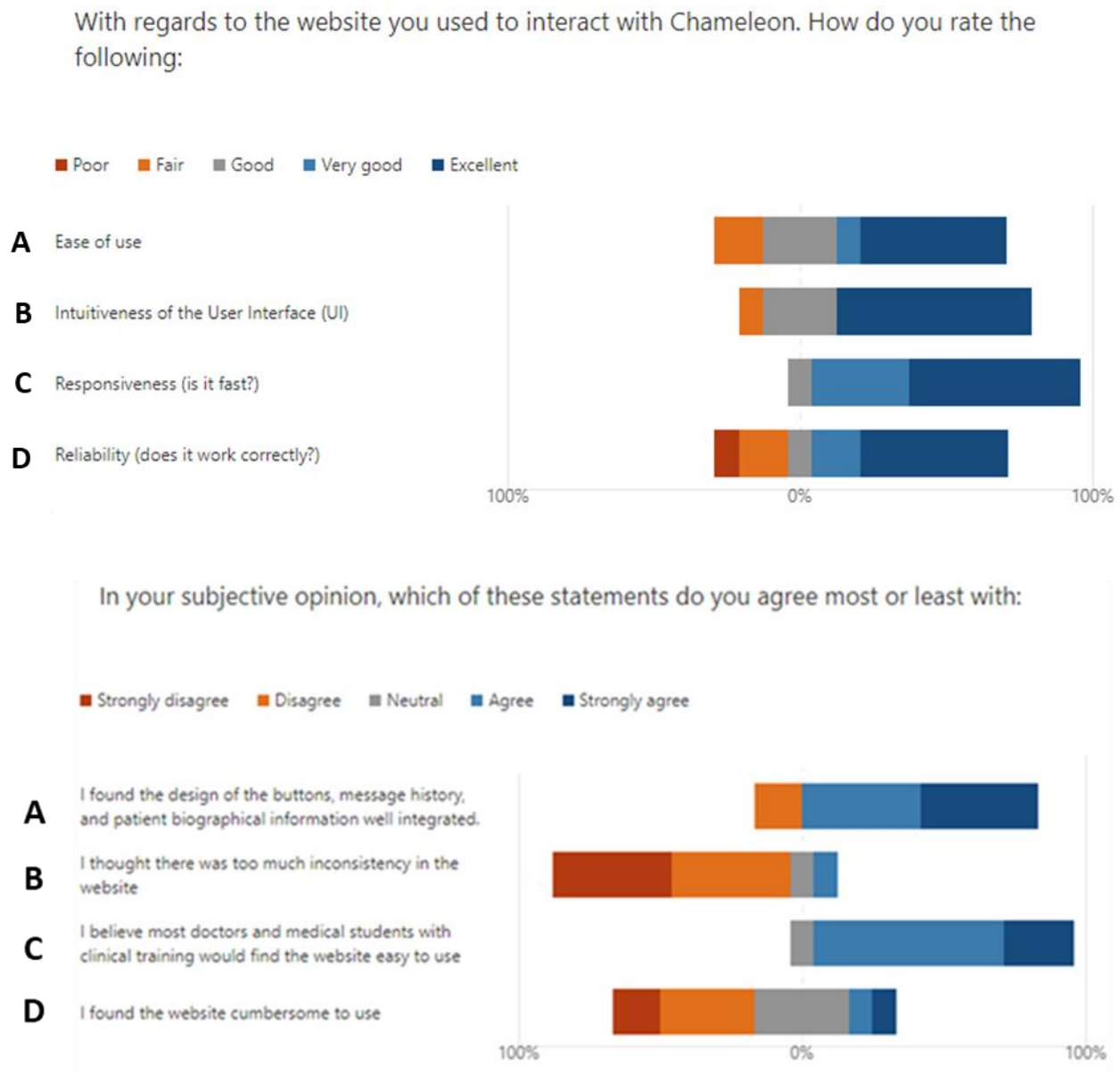


Figure 21 Most participants rated the website in terms of functionality and presentation as either Very good or Excellent. The decision to host the backend on VM with dedicated hardware resources may have contributed to users perception of the website being fast and reliable. Furthermore, early design mock-up and input from volunteers in the live testing rounds may have resulted in the users perceiving the website as easy to use, intuitive, consistent, and practical.

Complete user experience, feedback, and suggestion

Question	Results and Insights
What do you think future releases of Chameleon should focus on? - **The website** - **The prescribed responses of Chameleon** - **The internal algorithms responsible of identifying the meaning behind your questions**	- All users indicated that they think any future work on Chameleon should focus on the 'The internal algorithms responsible of identifying the meaning behind your questions'. For 50% of users, this was the *sole* concern. - 50%(6) added they would also like to see improvements to the prescribed responses. - 25% (3) users suggested that future improvements should also focus on the website. - The results here are suggestive of 1) Intent space was not large enough to accommodate all possible scenarios and questions users could ask with regards to this specific clinical presentation (backpain). It is worth noting here that users cannot to tell the difference between Chameleon not understanding what they said due to an absent intent to address their query or due to a weak training corpus for the intent that should have been triggered. 2) Despite most users in Table 12 rating the naturalness of the prescribed response as Good or better, 50% of them here would like to see future work to focus on improving them in addition to better NLU. This is in part due to most of the users, in fact, rated the naturalness of the responses as Good. Only 1 user rated them as Very Good, and 1 user rated them as Excellent.
Considering your complete experience with our software solution (website and chatbot), how likely would you be to recommend its use to a colleague for training or educational purposes?	Only one Promoter user, one Passive and 10 Detractors. NPS -75 suggesting most users would not recommend using Chameleon to someone they know.

User feedback is instrumental in pulling our development focus to where it is most needed. **Any suggestions or feedback on how to improve Chameleon?**	- 50% (6) of users responded with detailed answers. - Making use of a freely available sentiment analysis tool[49], we show that 3 users responded positively, two negatively, and one user who was neutral. Details in Table 17 (next page). - The relatively high response rate of this optional question may be explained by the wording of the question as it contextualises the effect that the feedback/suggestion will have on the system the participants just tested. It may further explain why half of the users answered it in a detailed manner (average word length = 135 words per answer), they may have perceived that their responses matter more than other optional questions, such as the ethical concerns question which only 3 participants answered. An alternative explanation is that the rigorous scientific training and curiosity of doctors introduces a bias towards action that advances science, which in this case is a textual contribution. Furthermore, we can clearly see the usefulness of this question mitigating the bias introduced by the earlier questions with regards to future releases and the question that asked what could have made the diagnosis easier (Reja et al., 2003). Users notably suggested future directions of future focus not mentioned in the earlier question.

Table 16 Future focus and complete user experience

[49] https://monkeylearn.com/sentiment-analysis-online/

ID	Results and Insights	Sentiment
1	- After significant development, this user believes that Chameleon can be 'very helpful' to doctors. - The user pointed out that some of Chameleon's responses sound 'very medical'. The user used the example of Chameleon using the word 'subsided' when describing pain after taking pain killers, instead of 'got a bit better' which sounds more like what a real patient would say.	Positive – 97.8% confidence - 49 words
2	- The user criticised the way Chameleon operates. They pointed out that using prescribed answers limits the naturalness of the conversational flow and limits the doctor should they want to ask ICE/Social History related question outside the ICE/Social History pages. - The user pointed out that triggering the same intent repeatedly yields the same prescribed answer from Chameleon. There's a distinct lack of variety to Chameleon's response. - Moreover, the user perceived the division of the interview into sub-segments as unrealistic and 'never going to happen in real life'. - The user suggested a measure to track user's frustration, respect and have the chatbot adjust its responses accordingly. - The user was critical of the fact that the site is not entirely compatible with the browser Firefox (and Safari), mobile devices, and pointed out the bug cited in the web development chapter (namely, that the whole page scrolls down after every response) is 'annoying' .	Negative – 99.7% confidence – 288 words
3	- Suggested adding an intent to address the queries: o 'have you taken any painkillers'. o The type of exercise that the SP was doing - This user found Chameleon at the Family History section of the enquiry struggling but otherwise thought it was 'very cool to see in practice'	Positive – 93.5% confidence – 45 words

4	- Pointed out that Chameleon 'does not perform well' when the queries contain spelling mistakes. The user mentioned that consequently their level of 'enjoyment' was reduced as they 'had to edit my answers'[50] - The user expressed that they think our approach to providing a tool for practice is good and is more useful to younger doctors.	Positive – 70% confidence – 77 words
5	- Mentioned Chameleon 'did not let me introduce myself to the patient' - Stated that in their opinion the patient's answers felt natural - Expressed surprise that Chameleon could not answer the basic question 'what are your expectations?' in the ICE section of the interview - Pointed out that the bot was not able to answer a query regarding asthma medication - Suggested adding a button 'finish consultation' whereby the user is presented with a text box to enter their diagnosis to submit to the programme. And for the programme to indicate whether this diagnosis was correct or not and explain the reasoning. And for Chameleon to point out to the user questions that they should have asked but did not. - Suggested adding an intent to allow Chameleon to give its consent for the consultation to go ahead in the Start page. - Pointed out that Chameleon did not answer correctly when asked 'how the pain/the problem is affecting their life' - The user likes that Chameleon was able to answer detailed questions related to Social and Lifestyle History - Suggested adding an 'OSCE practice' mode option to 'practice scenarios within a time limit of 5 minutes'	Negative – 60% confidence – 317 words
6	- Asked questions about the pain that Chameleon did not recognise - Asked further questions about pain and Chameleon misinterpreted the questions to do with weakness instead. - Asked Chameleon about sensation in perineal area[51], Chameleon misinterpreted the question to be about the legs.	Neutral 77.2% confidence – 36 words

Table 17 Insights from participants' detailed feedback and suggestions

[50] Google's Dialogflow CX does not automatically correct spelling or grammatical mistakes in the Europe region.
[51] Area of the pelvic floor between the thighs and is the most inferior part of the pelvis.

Discussion

This work built an end-to-end solution for a conversational agent. Chameleon, a pedagogical tool for medical students, aims to fulfil an existing need in medical education and practice. Such a need is further made the more urgent with the recent COVID19 pandemic imposing restriction on the closeness between medical students and Standard Patients (SP) in an OSCE setting. We argued that OSCEs currently are broadly used as an assessment tool, not as a pedagogical instrument. We made the case that using technology could expand the application of OSCE to non-assessment settings in medical education. This is partly due to the cost and significant overhead involved in organising an OSCE; factors that are much nullified in a virtualised OSCE.

We introduced Chameleon as a technology demonstrator for a conversational agent that can play the role of the SP. It thus serves three main goals: 1) Allows the users to hone their clinical interviewing skills. 2) Creates a basic infrastructure of future expansions into different clinical presentations and/or different patient profiles. 3) As a proof-of-principle demonstrator, it opens new possibilities for practitioners and educators alike to explore the application of AI technology in the field of medical pedagogy.

Chameleon was built over three iterative steps of design – test – modify. A working basic version with generic placeholder responses was deployed on Facebook early in the development work. A first round of live-testing was conducted with volunteer junior doctors and two colleagues from the AI MSc course in the University of Aberdeen. In total, seven interviews were held. The resulting Chameleon was made up of 7 pages (states) and 135 intents. We then proceed and extend Chameleon's capabilities to reflect the symptoms of a real-life condition that is often presented to medical students in OSCEs, backpain. A second round of life-testing is conducted with only two subjects. Following modifications, Chameleon is now comprised of 180 intents.

Concurrent with Chameleon's development and testing, we design and deploy a full-stack web solution. A frontend console developed using Reactjs framework based on early mock-up designs. A backend developed using Django and Google's publicly available Dialogflow API functions. Our backend codebase is deployed on

DigitalOcean due to its affordability with respect to its ability to serve dozens of users at the same time. Once deployment is successfully completed, we contacted 233 medical societies covering all 33 medical schools in the UK. Ten societies responded positively which converted to 12 participants filling our survey with regards to Chameleon's performance.

Reflections on the results of the survey and the limitations of Chameleon

- The newly released[52] Dialogflow CX still suffers from minor production bugs. Some of these are noticeable in the console[53]. Others are related to reliability and robustness. Crucially, the breakdown behaviour observed in some of the live-testing seems to have occurred with users. Over two thirds of users reported that they have observed this behaviour. However, unlike the live-testing rounds where our intervention confirms that this behaviour is indeed a temporary breakdown in Dialogflow intent detection, no such confirmation can be stated for users' reports. A possible explanation is some of the users were simply, and repeatedly, asking questions that are significantly outside the training corpus for whatever intent they were trying to trigger. Furthermore, one of the users correctly pointed out that spelling mistakes can cause the DialogFlow engine to miss-classify or not recognise text queries. DialogFlow ES automatically corrects for spelling mistakes since 2018[54]. However, in the CX offering this option has not been deployed to the europe-west2 region which hosts our bot[55].

- While 80% of users stated that they were able to reach a diagnosis, this result is not indicative of what Chameleon purports to do. The clinical presentation used, backpain following a gym visit for a healthy 43 years old, is too simple and easy to resolve. Future studies will need to modify the presentation so that it is 1) multi-faceted presentation: for example, the patient is suffering from back pain that is constant and independent of physical activity, in addition the patient is Type II

[52] Dialogflow CX was released on December 15 2020
[53] For example, when clicking on a page, then clicking on 'view all' to see all intents under that page. The user is presented with a list of 10 intents with the option to extend the view to 25, 50, or 100. If the user selects 50, then selects an intent from the list to view in detail, the option of viewing 50 is forgotten and the view is reverted to 10. It can slightly be cumbersome when there are many intents on the page.
[54] Google Developers' Blog
[55] https://stackoverflow.com/questions/68194202/auto-correct-spelling-option-not-available and https://cloud.google.com/dialogflow/cx/docs/concept/region#limits

diabetic and depressed (not reported). 2) Explore the role of environmental factors on health: for example, while the patient is being seen by the doctor, he/she could remember to inform the doctor of previously unreported recurring headaches. A possible route of investigation would be to ask the patient of any new stressors in their life, the type of work they do, their sleep and diet, new baby, etc.

- Overall performance of Chameleon as a conversational agent is Good but not excellent. Below are factors that may have contributed to this performance:
    - Our design contains only 180 intents. To the extent possible, the intents do address the most common expected questions in a clinical consultation, however, this is based on either medical textbooks or conversations with only five doctors, not a representative sample by any mean. Moreover, none of the doctors we interviewed was a medical student. We estimate that the junior doctors graduated four to five years ago[56]. Any changes in the way medical students are taught to approach and start the consultation would not have been captured in the training corpus or intents coded. This is of note since it was our observation that all the junior doctors who tested the bot with no exception started the conversation using broadly the same sentence: "Hello, I'm X. I'm a doctor here. How can I help you today?". Suggestive of similar training background. Furthermore, Human language is complex and doctors will tend to express themselves differently either due to slightly different training background, different accents and cultures or to be more approachable to the patient.
    - Our approach to designing the conversational flow neglected the use of conversational entities and parameters. Entities are the types of words/categories that DialogFlow can extract from users' utterances. However, they would have to be manually annotated in the training corpus beforehand. For example, the words 'pain', 'headache', 'vomiting' can all be manually annotated under an entity called 'Symptoms'. Parameters are the value of the specific entity. In the previous example, in the sentence: 'Do you

[56] Based on their current positions. No explicit questions about their graduation dates were asked.

have a headache?' headache would be extracted as a parameter of the entity type Symptoms. The benefit of this approach is that it allows for conditional intents, that is intents that are triggered once a set of parameters are true. For example, if John Doe replies to the question 'Do you have a headache?' with a 'yes', thus, internally setting the parameter 'entity = headache'. One possible expected question would be 'is it a migraine?'[57], but to trigger this intent the condition entity == headache must be true first (Adamopoulou and Moussiades, 2020). In our approach we coded all intents unconditionally, they can be triggered at any point if the bot's state is on the respective page and this does significantly fasten the development process of the bot, however, it does limit the naturalness of the twists and turns of the conversational flow. Another consequence of this is that Chameleon, as the design currently stands, is unable to remember facts told to by the user. For example, if the user introduces themselves by saying "Hello, I'm Lisa. I'm a doctor here. How can I help you?". Throughout the interaction Chameleon will not remember the doctor's name. This is partly due to the System Entity options offered by Google are still incomplete in the CX version of Dialogflow[58].

- 27 intents were found to have a training corpus of less than the minimum recommended by Google of 10 sentences. It is possible that this may have contributed to DialogFlow hallucinating utterances and consequently misinterpreting intents (Reiter, 2018).
- Our second round was composed of only two interviews. This was for the backpain version of Chameleon. In our view, a larger second round with more domain specialist would have improved Chameleon's performance at deployment. Given enough resources and time, we would recommend that future research recruits user via Amazon Mechanical Turk or any similar service to generate intents and expand training corpus.

[57] Or 'is it a pulsating headache?', 'where in your head does it hurt?', 'did you try any pain killer?', in this line of questioning these won't be asked unless John Doe replies in the affirmative for the question 'do you have a headache?'
[58] In the ES version, a System Entity for names can automatically recognise virtually any name uttered by the user. This option does not exist yet in the CX version at the time of writing of this paragraph.

- Chameleon's evaluated performance is in line with reported performance of chatbots and conversational agents used in medicine as ranging from mixed to positive amongst domain specialist including for ones dedicated for student training purposes (Milne-Ives et al., 2020).
- Importantly, the lack of rigorous statistical analysis of the evaluation data does limit the insights that could be gained from data and renders our insights qualitative in nature.

Ethical Considerations on the future use of Chameleon

Artificial Intelligence technology is, in many ways, reflective of existing biases that exist in society; the data which AI/ML models rely on is a harvest from the deeply flawed human existence. In clinical settings, we find such biases can have severe consequences that could mean a patient losing their life. Deploying a software that mimics ill humans, and in doing it is mimicking aspects of human suffering, is, in our view, an endeavour that must be subjected to rigorous and sensitive ethical approval. Who decides what racial identity the bot will adopt? If future iterations of Chameleon do succeed in allowing the users to perform virtual examinations, such as inspecting a generated skin legion, on what basis is the racial background of the avatar is selected? Does the user select 'Afro-Caribbean' from a dropdown menu? Existing biases in society suggest that it is likely that non-Caucasian Chameleon will rarely be selected, exacerbating existing biases in doctors' assessments of black and brown patients.

The same considerations above apply to gender. Biases that exist in society around not taking the suffering of women seriously may also be propagated here in the hypothetical scenario where Chameleon is widely adopted. Do doctors get to see Chameleon as a woman suffering from depression, low libido, and endometriosis? Will doctors attribute symptoms of Myalgic encephalomyelitis in women to mental illness as a first diagnosis before exploring other possibilities more so then in men (Clarke, 1999)?

References

Abd-Alrazaq, A., Safi, Z., Alajlani, M., Warren, J., Househ, M. and Denecke, K., 2020. Technical metrics used to evaluate health care chatbots: scoping review. *Journal of medical Internet research*, *22*(6), p.e18301.

Adamopoulou, E. and Moussiades, L., 2020, June. An overview of chatbot technology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 373-383). Springer, Cham.

Andel, C., Davidow, S.L., Hollander, M. and Moreno, D.A., 2012. The economics of health care quality and medical errors. *Journal of health care finance*, *39*(1), p.39.

Barlow, D.H. and Campbell, L.A., 2000. Mixed anxiety-depression and its implications for models of mood and anxiety disorders. *Comprehensive Psychiatry*, *41*(2), pp.55-60.

Barman, A., 2005. Critiques on the objective structured clinical examination. *Annals-Academy of Medicine Singapore*, *34*(8), p.478.

Battles, J.B., Wilkinson, S.L. and Lee, S.J., 2004. Using standardised patients in an objective structured clinical examination as a patient safety tool. *BMJ Quality & Safety*, *13*(suppl 1), pp.i46-i50.

Baumgart, D.C. and Sandborn, W.J., 2012. Crohn's disease. *The Lancet*, *380*(9853), pp.1590-1605.

Bener, A., Verjee, M., Dafeeah, E.E., Falah, O., Al-Juhaishi, T., Schlogl, J., Sedeeq, A. and Khan, S., 2013. Psychological factors: anxiety, depression, and somatization symptoms in low back pain patients. *Journal of pain research*, *6*, p.95.

Bergant, Jana. ChatBots: Messenger ChatBot - DialogFlow and nodejs, 2016, Udemy.com, viewed on 20 August 2021, < https://www.udemy.com/course/chatbots/> [59]

Bergant, jana. Create Chatbot for Website with React and Node.js, 2018, Udemy.com, viewed on 20 August 2021, < https://www.udemy.com/course/chatbot-for-website-with-react-and-nodejs/>

Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189* (194), pp.4-7.

Brown, C., Ross, S., Cleland, J. and Walsh, K., 2015. Money makes the (medical assessment) world go round: The cost of components of a summative final year Objective Structured Clinical Examination (OSCE). *Medical teacher*, *37*(7), pp.653-659.

Bush, B., Shaw, S., Cleary, P., Delbanco, T.L. and Aronson, M.D., 1987. Screening for alcohol abuse using the CAGE questionnaire. *The American journal of medicine*, *82*(2), pp.231-235.

Carrier, E.R., Reschovsky, J.D., Mello, M.M., Mayrell, R.C. and Katz, D., 2010. Physicians' fears of malpractice lawsuits are not assuaged by tort reforms. *Health affairs*, *29*(9), pp.1585-1592.

Catino, M., 2011. Why do Doctors practice defensive medicine? The side-effects of medical litigation. *Safety science monitor*, *15*(1), pp.1-12.

[59] The procedure for determining the correct publication date can be found here: https://webapps.stackexchange.com/questions/142885/how-to-find-udemy-course-creation-date

Clarke, J. N. (1999). Chronic fatigue syndrome: gender differences in the search for legitimacy.

Colace, F., De Santo, M., Lombardi, M., Pascale, F., Pietrosanto, A., & Lemma, S. (2018). Chatbot for e-learning: A case of study. *International Journal of Mechanical Engineering and Robotics Research*, *7*(5), 528-533

Demasi, O., Li, Y. and Yu, Z., 2020, November. A Multi-Persona Chatbot for Hotline Counselor Training. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings* (pp. 3623-3636).

discover.bot. 2019, 16 June. CHATBOT DEVELOPMENT – Building Bots with Wit.ai. Viewed on 26 August 2021. < https://discover.bot/bot-talk/guide-to-bot-buiding-frameworks/wit-ai/>

El Zini, J., Rizk, Y., Awad, M. and Antoun, J., 2019, July. Towards a deep learning question-answering specialized chatbot for objective structured clinical examinations. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-9). IEEE.

Fan, W. and Yan, Z., 2010. Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior*, *26*(2), pp.132-139.

Gaitonde, D.Y., 2012. Hypothyroidism: an update. *South African Family Practice*, *54*(5), pp.384-390.

Google Cloud (n.d - a). Intents – Training phrases. Viewed on 23 August 2021 <https://cloud.google.com/dialogflow/cx/docs/concept/intent?authuser=1#tp>

Google Cloud (n.d - b). Messenger from Facebook. Viewed on 23 August 2021 <https://cloud.google.com/dialogflow/es/docs/integrations/facebook>

Google Cloud Tech (2021) 7 best practices for Dialogflow CX. [online video]. Available at <https://www.youtube.com/watch?v=OMurp_9B1Z4>

Gupta, P., Dewan, P. and Singh, T., 2010. Objective structured clinical examination (OSCE) revisited. *Indian pediatrics*, *47*(11), pp.911-920.

Harden, R.M., 1988. What is an OSCE?. *Medical teacher*, *10*(1), pp.19-22.

Harden, R.M., Lilley, P. and Patricio, M., 2015. *The Definitive Guide to the OSCE: The Objective Structured Clinical Examination as a performance assessment*. Elsevier Health Sciences.

Harden, R.M., Stevenson, M., Downie, W.W. and Wilson, G.M., 1975. Assessment of clinical competence using objective structured examination. *Br Med J*, *1*(5955), pp.447-451.

Hartvigsen, J., Hancock, M.J., Kongsted, A., Louw, Q., Ferreira, M.L., Genevay, S., Hoy, D., Karppinen, J., Pransky, G., Sieper, J. and Smeets, R.J., 2018. What low back pain is and why we need to pay attention. *The Lancet*, *391*(10137), pp.2356-2367.

Harvey, A., 2020. Ten years of Geeky Medics. *BMJ*, *369*.

Hobert, S. and Meyer von Wolff, R., 2019. Say hello to your new automated tutor–a structured literature review on pedagogical conversational agents. TRACK 4: LERN- UND WISSENSMANAGEMENT (E-LEARNING AND KNOWLEDGE MANAGEMENT).

Hobert, S., 2019. How are you, chatbot? evaluating chatbots in educational settings–results of a literature review. *DELFI 2019*.

Hoerger, M., 2010. Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, behavior, and social networking*, *13*(6), pp.697-700.

Hubbard, J.P., Levit, E.J., Schumacher, C.F. and Schnabel Jr, T.G., 1965. An objective evaluation of clinical competence: new technics used by the National Board of Medical examiners. *New England Journal of Medicine*, *272*(25), pp.1321-1328.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T., 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8110-8119).

Kazi, H., Chowdhry, B.S. and Memon, Z., 2012. MedChatBot: an UMLS based chatbot for medical students. *International Journal of Computer Applications*, *55*(17).

Kravets, Una. 2020. 1-Line Layouts. Viewed on 21 August 2021, <https://1linelayouts.glitch.me/>

Lee, C.B., Madrazo, L., Khan, U., Thangarasa, T., McConnell, M. and Khamisa, K., 2018. A student-initiated objective structured clinical examination as a sustainable cost-effective learning experience. *Medical education online*, *23*(1), p.1440111.

Longmore, M., Wilkinson, I., Baldwin, A. and Wallin, E., 2014. *Oxford Handbook of Clinical Medicine-Mini Edition*. OUP Oxford.

Makary, M.A. and Daniel, M., 2016. Medical error—the third leading cause of death in the US. *Bmj*, *353*.

Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M.H., de Pennington, N., Mole, G., Normando, E. and Meinert, E., 2020. The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, *22*(10), p.e20346.

Newble, D., 2004. Techniques for measuring clinical competence: objective structured clinical examinations. *Medical education*, *38*(2), pp.199-203.

Park, W.B., Kang, S.H., Myung, S.J. and Lee, Y.S., 2015. Does objective structured clinical examinations score reflect the clinical reasoning ability of medical students?. *The American journal of the medical sciences*, *350*(1), pp.64-67.

Pérez-Soler, S., Juárez-Puerta, S., Guerra, E. and de Lara, J., 2021. Choosing a chatbot development tool. *IEEE Software*.

Poulton, T. and Balasubramaniam, C., 2011. Virtual patients: a year of change. *Medical teacher*, *33*(11), pp.933-937.

Ramdas, J. and Jella, V., 2018. Prevalence and risk factors of low back pain. Portal Regional da BVS, *Informação e Conhecimento para a Saúde*, ID: sea-194146.

Ray, A. and Mathew, R., 2018, December. Review of Cloud-Based Natural Language Processing Services and Tools for Chatbots. In *International conference on Computer Networks, Big data and IoT* (pp. 156-162). Springer, Cham.

Reis, João. (2020) . ANGULAR VS REACT: A COMPARISON OF BOTH FRAMEWORKS . Imaginary Cloud. Viewed 23 August 2021. <https://www.imaginarycloud.com/blog/angular-vs-react/>

Reiter, Ehud. 2018. Hallucination in Neural NLG. 12 November 2018. Viewed on 25/08/2021. < https://ehudreiter.com/2018/11/12/hallucination-in-neural-nlg/>

Reja, U., Manfreda, K.L., Hlebec, V. and Vehovar, V., 2003. Open-ended vs. close-ended questions in web questionnaires. Developments in applied statistics, 19(1), pp.159-177.

Rimmer, A., 2017. Can I avoid complaints by practising defensively?. *BMJ*, *359*.

Roser, M., Ortiz-Ospina, E. and Ritchie, H., 2013. Life expectancy. *Our World in Data*.

Sayar, K., Kirmayer, L.J. and Taillefer, S.S., 2003. Predictors of somatic symptoms in depressive disorder. *General Hospital Psychiatry*, *25*(2), pp.108-114.

Smutny, P. and Schreiberova, P., 2020. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, p.103862.

Sood, A., 2010. Technology S-Curve. *Wiley International Encyclopedia of Marketing*.

Spijker, J., Batelaan, N., de Graaf, R. and Cuijpers, P., 2010. Who is MADD? Mixed anxiety depressive disorder in the general population. *Journal of affective disorders*, *121*(1-2), pp.180-183.

Summa Linguae Technologies. (2021). Building Chatbots with Zero Experience: Lex vs Dialogflow vs Watson vs Rasa [White paper]. <https://summalinguae.com/wp-content/uploads/2021/03/Which-Chatbot-Platforms-are-Easiest-to-Build-and-Train-Summa-Linguae-Technologies.pdf >

Thielsch, M.T., Blotenberg, I. and Jaron, R., 2014. User evaluation of websites: From first impression to recommendation. *Interacting with Computers*, *26*(1), pp.89-102.

Thorat, S.A. and Jadhav, V., 2020, April. A review on implementation issues of rule-based chatbot systems. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

Towle, A., Godolphin, W., Grams, G. and LaMarre, A., 2006. Putting informed and shared decision making into practice. *Health Expectations*, *9*(4), pp.321-332.

Turing, A.M. and Haugeland, J., 1950. *Computing machinery and intelligence* (pp. 29-56). Cambridge, MA: MIT Press.

Vash, J.H., Yunesian, M., Shariati, M., Keshvari, A. and Harirchi, I., 2007. Virtual patients in undergraduate surgery education: a randomized controlled study. *ANZ journal of surgery*, *77*(1-2), pp.54-59.

Wilson, G.M., Lever, R., Harden, R.M., Robertson, J.I.S. and MacRitchie, J., 1969. Examination of clinical examiners. *The Lancet*, *293*(7584), pp.37-40.

Winkler, R. and Söllner, M., 2018. Unleashing the potential of chatbots in education: A state-of-the-art analysis.

Wiratunga, N., Cooper, K., Wijekoon, A., Palihawadana, C., Mendham, V., Reiter, E. and Martin, K., 2020. FitChat: Conversational Artificial Intelligence Interventions for Encouraging Physical Activity in Older Adults. *arXiv preprint arXiv:2004.14067*.

Appendix A – Original competencies scored by OSCEs

The table below lists out the sixteen example competencies that the OSCE is designed to assess. Over the next few pages we will revisit some of those competencies when discussing the role of chatbot mediated learning in medicine.

No.	Clinical skill or competency being assessed	Examples
1	History taking from a patient with a problem.	Patient presents with shortness of breath and insomnia.
2	History taking for the purpose of reaching a conclusive diagnosis.	Patient presents with constant fatigue, dizziness, fainting (symptoms of bradycardia[60]), poor concentration, weight gain and Depression. The student reaches the conclusion that the patient is likely to have an underactive thyroid (Gaitonde, 2012).
3	Demonstrating/educating the patient on management of condition under examination.	Patient is diabetic and is taught how to inject herself with insulin .
4	General advice to the patient	Patient's BMI is over 39[61], and he is advised on lifestyle changes to lose weight. For example, reducing processed foods, focus on lean proteins and vegetables and to walk to work if possible.
5	Explaining to the patient any forthcoming tests and procedures.	Patient is suspect to have Crohn disease. A colonoscopy and a biopsy are necessary to confirm the diagnosis (Baumgart and Sandborn, 2012). The procedure is explained to the patient.
6	Communication with other members of hospital/clinic staff	Student briefs a nurse with regards to a patient with suicidal ideations, or a patient who is terminally ill.
7	Communication with next of kin of the patient	Informing a mother that her daughter, a minor, is diagnosed with Leukaemia
8	Physical examination of a system or part of the body	Examination of the shoulder joint, suspected rotator cuff tendonitis.
9	Physical examination to follow up a problem	Patient has congestive heart failure, or any other chronic progressive illness such as Multiple Sclerosis (MS)

[60] Slower than normal heart rate
[61] NHS classifies any BMI above 30 as 'obese' and at a higher risk of obesity-associated health complications: https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/

10	Physical examination to assert the correctness or falseness of a previously made diagnosis.	Patient previously suspected of having thyrotoxicosis[62]. However, the symptoms are quit varied and may overlap with other conditions. Student orders a thyroid stimulating hormone (TSH) test [NOTE TSH is not a physical examination. Revisit this paper: https://www.bmj.com/content/349/bmj.g5128]
11	A diagnostic procedure.	Patient is suspected of having an enlarged prostate. The student conducts a digital rectal exam (DRE)
12	Written communication.	The patient is suspected of having ulcerative colitis. The student writes a referral letter.
13	Interpretation of clinical findings	The student can interpret lab results of patient's blood test, read the patient chart and records
14	Management	The student writes a prescription
15	Clinical critical thinking	The student reviews and critiques a published article or a pharmaceutical white paper/advertisement
16	Problem solving skills	---------

Example of competencies assessed by OSCE as suggested by Harden, 1988. Some of the examples have been paraphrased, others were taken verbatim.

[62] Excess thyroid hormone in the body

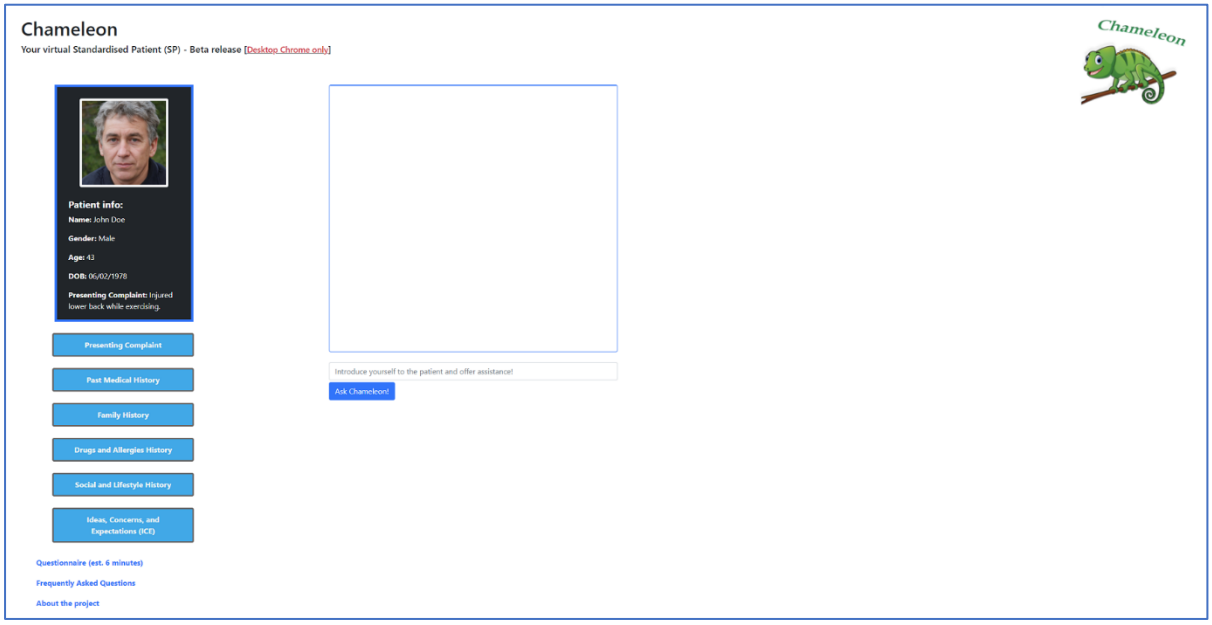
Appendix B – Icon of Chameleon

Source: https://www.pngwing.com/en/free-png-vitjn

License: Non commercial use, DMCA

Appendix C – Deployed Frontend

This mock-up was based on very optimistic and ambitious design objectives (below) that were set in June 2021. However, significant challenges in developing the solution within the available time limit meant that a lot of the ambitious objective were ignored. For reference, the final design of frontend is below:



The initial design objectives:

1. The frontend should offer doctors the option of performing virtual physical exams (doctors selects appropriate exam and a video/animation is played with the affected area clearly shown)
2. It should also offer the option of ordering a virtual lab/diagnostic test. For example, where necessary the doctor can order a full blood count or measure pressure. This is envisaged as the relevant numbers presented on screen when ordered.
3. The enquiry should include the Functional part of the clinical interview cited in the Oxfor Clinical Handbook of Internal Medicine
4. All actions and utterances by the user are marked. If user forgets to ask important questions that deemed relevant to reaching ac correct diagnosis, they are marked down:
    a. An NLG report is emailed to the user following the conclusion of the enquiry
    b. The report points out mistakes by omission or commission, unnecessary tests or physical examinations the student ordered.
    c. A list of resources is supplied to help the student fill the presumed gaps of knowledge which led to the mistakes.
5. That system can display multiple profiles of patients and diseases.

Appendix D – FAQs of the website

Chameleon's Frequently Asked Questions (exact text on https://chameleonpatients.co.uk/faqs)

Frequently Asked Questions

What is Chameleon?

Chameleon is an AI powered conversational agent (ie chatbot). It is the software which the website you just used communicates with.

Why is it called Chameleon?

Because our future plans will enable Chameleon to mimic the symptoms of a variety of clinical presentations in many different patient profiles.

What Artificial Intelligence technology does Chameleon use?

Chameleon uses a Google Cloud Platform (GCP) service called DialogFlow. DialogFlow uses advanced Natural Language Understanding models to detect intents (see below) of the user.

What are intents?

In the context of chatbot design, an intent is the meaning behind the user's linguistic utterance. For example, the sentences "Do you have high blood pressure?" and "Were you ever diagnosed with hypertension?", though different syntactically, are near identical semantically. They both trigger the same intent of 'hypertension'. Both sentences let Chameleon know that you are asking about the exact same thing. They both will trigger the same response.

How many intents are programmed in Chameleon?

There over 200 intents programmed in Chameleon

What is a Standardised Patient (SP)?

In medical education, it is common that medical schools will hire professional actors to play the role of patients with a specific set of symptoms. This process is standardised. Thus, several actors could be examined by different students, but all will present the same way if told to (for example, to ensure fairness in marking). Occasionally, the role of the standardised patient is played by other healthcare professional. The practice of hiring SPs is, however, quite costly.

What's all this supposed to be doing?

Chameleon digitises the role that a human actor plays. It is preprogramed to present the symptoms as a human would. At this stage, this work serves as a proof of principle and therefore not all design objectives are fulfilled at this point.

What is your objective from developing this software?

Our main objective is to build an AI-driven pedagogical tool for medical students and working doctors. Chameleon allows the users to practice asking questions algorithmically and logically as they would do in real life clinical practice. It does so at very little financial cost.

Why is the interface designed this way?

We followed a minimalist approach. The interaction with the agent is the main show and purpose behind using this tool. We have plans in the future to add a pane on the right to enable the users to order simulated lab tests, physical examinations, or medical imaging studies.

Why are there no instructions on how to use the interface?

Following consultation with several doctors, we made a design decision to keep the focus on the chat window and the buttons' panel on the left-hand side of the screen. We believe medical students who have received some training in clinical interviewing will find it intuitive.

How long will the questionnaire take?

We estimate it will take around 5-6 minutes for native English speakers

Are you collecting any private data?

No. We are not.

Will my data be shared with anyone?

No. All the data generated from the questionnaire will be used for research purposes. It will be kept on a secure university of Aberdeen account until November 2021 only.

Do you have any future plans for Chameleon?

At the moment, Chameleon is designed for research purposes. However, we have plans to expand the range of capabilities that it can offer.

Is John Doe's picture real?

No. The picture you see was generated using an AI technique called SyleGAN.
Visit: https://thispersondoesnotexist.com/ if you are interested.

Appendix E – All intents and responses in deployed Chameleon

Intents for back pain version

180 unique intents (intents can be coded into more than one page/flow but not twice on the same page) + 6 Trigger intents (not shown in table) = 186 total

2638 training sentences

On average

Intents that are coloured in green only exist in the deployed version.

Intents in Red are ones with a training corpus of less than 19 sentences.

	Section	Intent	Training sentences	Response	Notes
Start – 13 intents, 153 sentences					
1	Start Flow	Default Welcome Intent			
2		Introductions and Presenting Complaint - no name	46	Thanks for seeing me, Doctor. Last night, I was at the gym. I was deadlifting. The weight wasn't particularly heavy. However, while performing the exercise I felt a sudden sharp pain in my lower back. I cut my exercise short and went home. By then the pain has subsided significantly. However, I woke up this morning I could barely move. The pain was really intense. So, I thought I'd come and see you.	
3		How Are You	13	I am 44 in 6 months	
				I am 43 years old	
				I'm the early afternoon of my life, Doc. I'm 43.	

4		What is Your Name	11	My name is John Doe I'm John I'm John Doe	
5		Expression of Sympathy	17	You are too kind, thanks Thanks.	
6		Nice to Meet You – Salutations	7	Nice to meet you, too!	
7		Ok ok I see	7	.....	
8		How old are you?	7	Aside from this back problem, I'm doing alright thanks.	
9		Confirmation of patient's name and DOB	12	Yes, that's me. I am John Doe born on that date Yes. that's correct. Yes. that is my name and date of birth	
10		Confirmation of patient's name and DOB - explicit	8	I'm John Doe. Born on February 6th 1978.	
11		What are you wearing	8	I'm wearing a dark navy blue shirt and jeans.	
12		Are you well groomed?	8	I shave daily and I try to present in the best way I can to the world.	
13		Tell me about yourself	9	Well, my name is John Doe. I'm a 42 years old playwriter. I work from home as a freelancer. I am married to Jane Doe, my wonderful partner. We have three daughters.	
Presenting complaint – 41 intents, 758 sentences					

1		PC - Elaborate On Symptoms	20	As I said, I went to the gym yesterday after work. I warmed up and then started deadlifting. In my second set, probably on the 5th or 6th rep. I was lowering the bar slowly and when it was around my knee level I felt a very intense and sharp pain in my lower. The pain was quite intense, I cut my exercise short and went home. It was fine in the evening after I took some paracetamol. But this morning I found it very hard to get up. It my back was very painful. I could barely move.	
2		PC - Elaborate on Symptoms - Time	32	Well, it really started yesterday when I went to the gym to deadlift after work. So I guess around 20 hours.	
				I would say from the time of injury, so a little bit less than a day	
				The injury only happened yesterday	
3		PC - Describe Intensity Of Symptom in Words	26	Well, at the moment it happened, I was lowering the bar slowly to the ground. I felt a very sharp hot pain in my lower back. It was quite intense, actually. I let out a shriek that everyone heard. It wasn't the most dignified scene.....as I said, the pain was mostly manageable by the evening. It's just in the morning when I woke up it was quite intense. More of a pressure feeling than sharp, it comes and goes but gets worse when I move.	
4		PC - Describe Intensity of Symptoms Scale 1 to 10	24	Probably 8 the moment it happened. And 7 in the morning	
5		PC - Possible causes and triggers	23	Honestly, I think I got excited. I haven't been to the gym in a while and my form may have not been the best. I think I didn't pay attention to keeping a straight back when I was lowering the bar.	

6	Presenting Complaint	PC - Things that exacerbate conditions	17	At the moment, any movement of my body, particularly sudden movement. For example, getting out of bed this morning was quite hard.	
7		PC - Things that make it better	18	Well, when I lie down on my side and relax. That seems to keep the intensity of pain at bay. Lying on my back also helps but for some reason lying on my side is better.	
8		PC - Previous consultations/visits to doctor	21	No. You're the first doctor I see. It only happened yesterday. You're the first doctor I see.	
9		PC - Order of Symptoms	12	I wouldn't see there was an order. The pain is the biggest thing at the moment. There's no order to it as there isn't anything else of significance happening.	
10		PC - Activity at time of first symptom	14	I was deadlifting in the gym. I was lowering the bar when it happened. I would say the bar was a little bit lower than my kneecaps. I was lowering the bar when performing a normal stance deadlift. The bar was about knee height	
11		PC - location - Where do you feel it?	19	That whole area [John points at an area between and around L4 and L2.	
12		PC - Description of pain in one word	12	At moment of injury, Sharp. This morning: achy	
13		PC - Pain movement	19	I haven't noticed that the pain moves per se. It's quite localised to that area. But it does get worse when I move.	
14		PC - Do you see your GP regularly	11	I do see my GP when I feel there's a need to. But I haven't done so in months.	
15		PC - What kind of pain was it?	11	I would say, at the moment of the injury it was quite sharp. Somewhat calm in the evening. But very achy in the morning with sharp pain if I move suddenly.	

16		PC - Pain Radiation	24	No. I haven't noticed that it radiates or spreads to anywhere.	
17		PC - Waking up because of the pain	16	Not last night. I think I slept alright. It's just when I woke up where it was unbearable.	
18		PC - Is the pain constant or does it come and go?	11	I would say, it comes and goes...	
19		PC - Concurrent Symptoms	23	No. I don't think I noticed any other symptoms happening at the same time.	
20		PC - Bladder Dysfunction	21	No. Not really. I think my bladder is ok. I can empty it completely and don't think I have lost control.	
21		PC - Bowel movement	15	No, not really. I did not notice any changes to my bowel movement.	
22		PC - Things that make better or worse or staying	17	Hmm....I'd say probably sudden jerky movements like when I tried to get up from my bed this morning make it worse. Lying down on my side makes it better.....and maybe just lying on my back with a small pillow underneath stabilises it.	
23		PC - Weakness in legs	13	No. I have not noticed any weakness in my legs.	
24		PC - Can you walk?	18	Well, I needed to lean on the rail of the stairs last night after the gym and today when I went down from my bed room. But on a flat surface I think my walking is ok. Nothing unusual, really.	
25		PC - Sensation in the legs	23	No. I think my sensation in my legs is completely normal.	
26		PC - Similar Past Symptoms	15	No. I don't think I've ever experienced anything like this before. My back has been ok so far.	
27		PC - Back Protection measures - bending over	9	If I bend over to lift something heavy, I'll make sure I'm in the squat position with my knees bent and my back straight.	
28		PC - Pain when coughing or breathing	25	No. I haven't noticed any particular increase in the pain when breathing deeply or coughing.	

29		PC - Sexual Function	16	Ehem...to be completely honest, yesterday we thought it might reduce the pain. We didn't have full penetrative intercourse but my partner simulated me to orgasm. It helped with the pain only for few minutes, I'm afraid. But yes, I think my sexual function is healthy.	
30		Functional Enquiry - General - Fever	22	No. I don't think I developed a fever after the injury.	Placed in Presenting Complaint page
31		Functional Enquiry - Cardiorespiratory Enquiry - Chest Pain	11	No. Don't think so. No pain in the chest area after I hurt my back.	
32		Functional Enquiry - Cardiorespiratory Enquiry - Orthopnoea (breathlessness when lying down)	23	No. When I lied down on my bed last night there were no problems breathing.	
33		Functional Enquiry - Cardiorespiratory Enquiry - Oedema	12	No, I've noticed swollen legs before.	
34		Functional Enquiry - Cardiorespiratory Enquiry - Palpitations	17	No. I haven't felt any heart palpitations following the injury.	
35		Functional Enquiry - Cardiorespiratory Enquiry - Wheezing	18	No. I think my breath is fine. I did not notice any wheeziness.	
36		Functional Enquiry - Gastrointestinal Enquiry - Abdominal Pain	16	No. I have not noticed that my torso hurts except for my lower back.	
37		Functional Enquiry - Gastrointestinal Enquiry - Tenesmus	20	Well, to be honest, I have pooed this morning and I did notice that there's some left inside.	

38		Functional Enquiry - Genitourinary Enquiry - Incontinence (stress or urge)	33	No, I have noticed any leaking or incontinence.	
39		Functional Enquiry - Genitourinary Enquiry - Dysuria (painful micturition)	15	No. I can urinate without pain.	
40		Functional Enquiry - Genitourinary Enquiry - Urinary Abnormalities (Colour/Haematuria (streaks or pink urine)	29	No. I have not noticed any blood when urinating.	
41		Functional Enquiry - Genitourinary Enquiry - Hesitancy (Difficulty starting to urinate)	17	No. I think I can maintain a steady flow easily when urinating. It's not difficult to release either.	
Family History – 33 intents, 387 sentences					
1		Family History - Grandparents - Heart Diseases	13	Yes. My Grandfather from my father side passed away from a stroke. I don't think the others have heart problems, however.	
2		Family History - Parents - Hypertension	19	My mother has high blood pressure. But my dad is ok.	
3		Family History - Father - Status	8	He is alive and well.	
4		Family History - Father - Cause of Death	12	Thankfully, he is still with us. He is not dead. He's alive and well.	
5		Family History - Mother - Cause of Death	13	My mother is not dead! She is still with us. Alive and well.	
6		Family History - Mother - Status	8	Yes, she's a live and well.	
7		Family History - Parents - Diabetes	9	Yes. My mother has diabetes type II. My dad is alright though.	
8		Family History - Parents - Heart Disease	11	No, I don't think either of them has heart disease.	
9		Family History - family genetic diseases	18	No. I don't think there are any genetic disease in our family. Not to the best of my knowledge.	

10	Family History	Family History - Parents - Status	9	Yes. They are both alive, thankfully.	
11		Family History - lumped conditions	7	My mother has high blood pressure. I think I have a cousin who had stomach cancer for a while, but she's ok now. Also, my granddad from my father's side passed away from a stroke. But everyone else is fine as far as I know....	
12		Family History - Stroke	7	Only my grandfather from my father side of the family. He passed away sadly from a stroke	
13		Family History - Grandparents - Diabetes	8	No. None of them has diabetes to the best of my knowledge.	
14		Family History - General - High Blood Pressure	17	Yes. I think my Grandmother from mother side has high blood pressure.	
15		Family History - General - Heart Disease	13	Yes. My grandfather passed away from a stroke.	
16		Family History - General - Cancer	13	One of my first cousins sadly has stomach cancer. She's alright now. I think they removed most of her stomach.	
17		Family History - Parents - Cancer	12	No, I don't think either of them had it.	
18		Family History - Parents- Epilepsy	11	No, neither are epileptic.	
19		Family History - General - Epilepsy	11	I don't think anyone in our family is epileptic.	
20		Family history - Grandparents - Status	13	They're all alive and well except for my granddad from my father's side. He passed away from a stroke.	
21		Family History - Parents - Cause of Death	12	Thankfully, they're both still alive.	
22		Family History - Siblings (do you have)	9	Yes. I have younger sister.	
23		Family History - Siblings - Health	19	My sister is quite healthy. She's only 33.	
24		Family History - Children (do you have any - birth)	18	Yes. I have three daughters.	
25		Family History - Children - Health	14	Yes. Thankfully, all three of them are doing wonderfully health-wise.	

26		Family History - Rheumatic Diseases	8	I don't think so. Until yesterday, I don't think I've had any musculoskeletal or tendon issues, nothing that I can remember at least.	
27		Family History - Parents - Rheumatic Diseases	13	I think my mom has arthritis but I don't think my dad has any issues.	
28		Family History - Parents - Osteoporosis	12	Not to the best of my knowledge	
29		Family History - Family members with back problems	13	No. I don't think I know anyone in the family with a similar lower back complaint.	
30		Family History - General - Mental Health	13	I think that my aunt has been sadly depressed for a long time. I think she's the only one with a mental health condition.	
31		Family History - Aunts (exist)	8	Yes. I have one aunt and two uncles.	
32		Family History - Uncles (exist)	8	I have one aunt and two uncles.	
33		Family History - General - Diabetes	8	Yes. My mother has diabetes Type II.	
Past Medical History – 27 intents, 394 sentences					
1		Past Medical History - Epilepsy	12	No. I've never had any seizures, never been diagnosed with epilepsy.	
2		Past Medical History - Long term conditions	13	Thankfully, I don't have any long-term or chronic conditions.	
3		Past Medical History - Osteoporosis	13	No. I don't think I've ever had or been diagnosed with osteoporosis.	
4		Past Medical History - MIJTHREADS	14	I haven't had any serious medical issues at all. I've been quite lucky, I guess.	
5		Past Medical History - Diabetes	11	No. I don't have diabetes.	
6		Past Medical History - Bacterial infections	10	Ehem....Yes. I picked up Chlamydia when I was in uni. But it was fine. I regularly get tested so I discovered it quite quickly.	
7		Past Medical History - Respiratory problems	14	Yes. I used to use an inhaler as a child. My asthma was often trigged by animals and my mothers perfumes. But I'm fine now. I use it the inhaler probably once a year or so. Very rarely.	

8	Past Medical History	Past Medical History - Depression	19	No. Thankfully, I don't think I'm depressed. I have a rich family life and rewarding career. Broadly speaking, I really do feel content.	
9		Past Medical History - Current Medical Conditions	11	No. My health has been generally excellent. I've been blessed.	
10		Past Medical History - Past Medical Imaging	11	No. I haven't been through any medical imaging studies in the past for my back.	
11		Past Medical History - History of health in childhood/teenage	15	I'd say I had quite a healthy childhood aside from the occasional asthma attacks I mentioned.	
12		Past Medical History - Major Past Illnesses	11	Aside from breaking my arm when I was younger, I don't think I've had any major illnesses or diseases.	
13		Past Medical History - Sleep - Insomnia	27	My sleep has been ok, I don't think I have insomnia.	
14		Past Medical History - Type of Diabetes	15	I don't have diabetes.	
15		Past Medical History - Surgeries	13	Yes, I broke my arm when I was a teenager. They had to insert a trauma nail to fix the bone. They removed it after few months. It was a long time ago	
16		Past Medical History - Heart Disease	15	No. My cardiovascular has been generally excellent. I haven't experienced anything worrying.	
17		Past Medical History - Viral infections	11	No. I don't think I have picked up any serious viruses aside from the usual cold or flu. It's not COVID19 though, I have been tested for that.	
18		Past Medical History - Herniated Disk	10	No. I've never had a herniated disk before	
19		Past Medical History - Hospital stays	13	Only once. I broke my left arm playing football when I was a teenager. It was a long time ago	
20		Past Medical History - Current Blood pressure	11	I believe my latest blood pressure reading was 130/90. Pretty normal I think.	
21		Past Medical History - High Blood Pressure	13	It's in the normal range. 130/90 I think.	

22		Functional Enquiry - General - Recent Weight Loss	18	I put on 2 kgs during the first lockdown. I've generally kept active. So the increase is small.	
23		Functional Enquiry - General - Night Sweats	22	No. No night sweats thankfully.	
24		Functional Enquiry - General - Any lumps	13	I don't think so. I haven't noticed any lumps anywhere over my body.	Placed in Past Medical History page
25		Functional Enquiry - General - Fatigue	34	Not really. I haven't noticed any unexplained fatigue recently.	
26		Functional Enquiry - General - Appetite	14	The injury happened only last night. But I did eat well at dinner and this morning at breakfast. I haven't noticed anything out of the ordinary.	
27		Functional Enquiry - General - Rash	11	No. I haven't noticed any unexplained rashes on my body.	
Drug History – 19 intents, 219 sentences					
1	Drug History	Drug History - Over-the-counter	13	Just the usual vitamins and supplements. Sometimes I get paracetamol or ibuprofen like yesterday.	
2		Drug History - Improvement from drugs	8	Yes. The paracetamol and the ibuprofen help keep the back pain manageable.	
3		Drug History - Insulin	10	I don't self-administer any insulin. I'm not diabetic.	
4		Drug History - Inhalers	10	As I mentioned, I used to use inhalers regularly as a child. These days, I use it probably once a year. Very rarely.	
5		Drug History - Antidepressants - Dosage and Frequency	8	I don't take any antidepressants.	

6		Drug History - Drugs without prescription	9	No. I have not used any drugs with a prescription.	
7		Drug History - Thoughts on vaccines	13	I'm 100% pro vaccines, pro -logic. Not a fan of Andrew Wakefield.	
8		Drug History - Side effects from current drugs	14	I haven't noticed any side effects from the paracetamol.	
9		Drug History - Injections	9	No. I don't take any injections.	
10		Drug History - Allergies	20	No. No allergies from anything. As I mentioned, pets used cause asthma attacks for me when I was a child but not anymore.	
11		Drug History - Antidepressants	12	No. I don't take any antidepressants.	
12		Drug History - Supplements	12	I take my usual vitamins plus Ashwagandha and Lion's Mane tablets	
13		Drug History - herbal and alternative remedies	11	Sometimes I drink green tea. I think that's as close as I'm going to get to herbal treatments.	
14		Drug History - Dosage	10	I took two 500mg paracetamol tablets last night on my way back home from the gym. A couple more before bed. And 4 500mg tablets this morning.	
15		Drug History - Prescribed Medications	15	No. I don't take any regular prescribed drugs.	
16		Drug History - Oral contraceptives	11	As a man, I only used condoms.	
17		Drug History - Tablets	10	Yes. Paracetamol and ibuprofen for the pain	
18		Drug History - Vitamins	13	Yes. Each day, I take Vitamin D 4000UI, Omega 3 1200mg, Vitamin K 2600mcg, Magnesium 500mg, and a Vitamin B Complex pill.	
19		Drug History - List of medications	11	I don't have a list. As I mentioned my health is generally ok. It's only paracetamol and ibuprofen that I use for the pain.	
Social History – 47 intents, 584 sentences					
1		Social History - Work - Jobs in the past	10	I worked in marketing for a number of years before this. And before that I worked as a sales man for a paper company.	

2	Social History	Social History - Smoking - Giving up in the past	13	Well, I never smoked anything past my third year in uni. That is about 31 years, I guess.	
3		Social History - Marital Status	20	I have a partner, Jane.	
4		Social History - Work - Day job	11	I am a playwriter.	
5		Social History - Living Situation (alone or shared)	14	I live with my wife, Jane, our three daughters.	
6		Social History - Spouse's job	11	Jane is pianist. She's a pianist.	
7		Social History - Loneliness	10	No. I do not feel lonely.	
8		Social History - Home - Description	15	Sure. It's a 3-bed house. We have a small front garden. and a backyard where the kids play. All the bedrooms are upstairs. The eldest one Jo has her own room, the younger one shares another room. My partner and I have our own master bedroom. We have a small well aired and dry cellar. There's medium size sitting room when you first coming to the house. And the kitchen is quite ordinary, a bit smallish. A table for the girls to eat on, electric oven and plenty of cupboards.	
9		Social History - Smoking - History of Smoking	12	I smoked socially when I was in uni. It wasn't a habit. Just few cigs on a night out.	
10		Social History - Travel - Asia	13	No. We haven't travelled anywhere in the last 18 months.	
11		Social History - Smoking- Frequency/Quantity	13	I don't smoke at all.	
12		Social History - Alcohol - Drink in the morning	8	Oh no, thank god no.	
13		Social History - Alcohol	12	I only socially drink. Rarely. Perhaps, 2-4 glasses of red wine per week.	
14		Social History - Alcohol - Guilt	13	I really don't drink that much. Four red wine glasses per week at most. I have no feelings of guilt with regards to this amount.	

15		Social history - Alcohol - Bing drinking	13	No. I don't binge drink. Did it few times in uni but never past that.	
16		Social History - Frequency of dog walking	15	We walk Gold I would say twice a day at a minimum. Plus plenty of exercise in the backyard. We share the walking duties between me and my partner.	
17		Social History - Smoking - When did you start smoking?	10	I don't smoke. I smoked socially few and far between on nights out.	
18		Social History - Does the patient care for anyone?	10	Well, I look after my daughters.	
19		Social History - Home - Location of home	19	We live in Rosemount & Midstocket.	
20		Social History - Smoking - Do you smoke?	10	No, I don't smoke.	
21		Social History - Alcohol - Desire to reduce intake	12	I don't smoke.	
22		Social History - Home - Movement in the House	12	I only needed help this morning after I woke up but not before.	
23		Social History - Work - Exposure	10	No. I am a playwriter and I work from home.	
24		Social History - Stick use	14	Thankfully, I've never used a walking stick. But this morning I wished I had one....	
25		Social History - Travel - Latin America	15	No, we've never been to South America before.	
26		Social History - Travel - Africa	10	No. We haven't travelled anywhere in the last 18 months.	
27		Social History - Recreational Drug Use	14	No. I have never used recreational drugs.	
28		Social History - Things Patient Cannot Do Because of Illness	11	Well, I could barely move this morning. And if I hadn't taken the painkillers I probably would not have been able to make it here today. I do feel quite constrained in what I can do physically at the moment, there's no way I can go back to the gym anytime this week.	

29		Social History - Smoking - Desire to stop smoking now	13	I don't smoke.	
30		Social History - History of exercise	13	Well, I've been working out in the gym regularly since July 2017. Specifically, I started deadlifting in late 2016.	
31		Social History - Hobbies	15	Well, I quite like physical exercise. I go to the gym 3 times a week on average, I also practice Yoga daily. My other hobbies are just some gardening in the backyard and reading. The garden still looks like Fangorn Forest, to be honest...	
32		Social History - Alcohol - Frequency/Quantity	17	I would guess it's around 2-4 glasses of red wine per week on average. I'm not much of a drinker.	
33		Social History - Travel - Recent travel abroad	13	No. We haven't travelled anywhere in over 18 months.	
34		Social History - Travel - Tropics	10	No. We haven't travelled anywhere in the last 18 months.	
35		Social History - Spouse's Health	16	Yes. My partner's health is excellent thankfully. She's physically fitter than I am, actually.	
36		Social History - Work - Heavy Lifting	11	Not at all. I'm a playwriter. My work forces me to sit in front of the screen for the whole day.	
37		Social History - Work - Spouse'	21	My partner is a pianist.	
38		Social History - Smoking - Nicotine replacement	18	No. I don't use any nicotine replacements.	
39		Social History - Home - Establish Presence of Dependents at Home	11	Yes. We have three daughters. We all live together with my wife, Jane Doe.	
40		Social History - Smoking - Duration of smoking in the past	8	I think it was about 2 years. My second and third year in uni. But even then it was only few and far between. Never formed the habit.	
41		Social History - Closeness with Family	8	Yes. I'm quite close to my family and my relatives.	
42		Social History - Exercise and Physical Activity	18	Yes. I exercise regularly. Strength training 3 times per week and I keep a daily Yoga practice.	

43		Social History - Pets	16	Yes. We have a dog, Gold.	
44		Social History - Happiness	14	I would probably say yes. I'm quite content. I have a blissful life.	
45		Social History - Pet - Gold	12	Gold? He's a Golden Retriever. We had him since he was 12 weeks, he's 3 years old now. He is quite healthy, if this is what you are asking. We take him to the vet once every two months. He's been chipped and vaccinated.	
46		Social History - Home - Establish Presence of Stairs at Home	6	There are stairs. Thankfully we have a rail, so I had to lean on it this morning when I was getting down.	
47		Social History - Smoking - What do you smoke?	36	I don't smoke anything currently.	
Ideas, Concerns, and Expectations – 5 intents , 143 sentences					
1	Ideas, Concerns, and Expectations (ICE)	ICE - Hopes	19	I was hoping for some rehabilitation exercises, pain killers and to know if this is serious or not.	
2		ICE - Plan for action (patient's opinion)	30	Well, I think a range of appropriate rehabilitation exercises with painkillers should do it.	
3		ICE - Worst possible outcome	14	The worst thing that comes to mind is a herniated disk and nerve damage to my spine.	
4		ICE - Ideas (patient opinion about the symptoms)	51	To be honest, I think I pulled something in my back. But I'm not entirely sure and that's why I'm here.....	
5		ICE - Concerns (origins of conditions and anything specific)	29	The intensity of the pain this morning got me a little bit worried, that's all. I don't have any other concerns or questions. Thanks!	

Appendix F – Details of medical societies contacted

Key:

Yellow highlights – either sent an email or a facebook message (draft of email sent at the end of the table)

Orange Response – Unclear if invitation passed on to members

Blue Response – Replied either via email or on Facebook and suggested members will be sent the invitation

Light grey – Not contacted

No.	Medical School	University	Method of contact	Students' Societies contacted		Response
1	Aston Medical School	Aston University	Facebook + email: astonmedsoc@gmail.com	Aston Medical Society	30	
			Instagram	Aston Surgical Society	452	
2	Anglia Ruskin School of Medicine	Anglia Ruskin University	Facebook + email: medsoc.angliaruskin@gmail.com	Anglia Ruskin MedSoc	74	
3			Facebook + email: arusurgicalsociety@gmail.com	ARU Surgical Society	372	
4			Facebook + email: aruccts@gmail.com	ARU Cardiology & Cardiothoracic Society	143	
5			Facebook + email: arugpsoc@gmail.com	ARU GP Society	56	
6			Facebook + email: arunapschelmsford@gmail.com	ARU NAPS - Neonatology and	86	

				Paediatric Society		
7			Facebook + email: aruparasoc@gmail.com	ParaSoc - ARU Chelmsford	543	
8			Facebook	Anglia Ruskin MedSoc	74	
9	Barts and the London Schools of Medicine and Dentistry	Queen Mary (University of London)	Facebook + email: anatomy@bartslondon.com	BL Anatomy Society	1086	Responded. However, indicated that a new committee is coming in two weeks, and they will be deciding on whether to forward the details to members or not.
10			Facebook + email: welfare@bartslondon.com	Barts and The London Cardiology Society	989	
11			Facebook + email: blgpsociety@gmail.com	BL GP Society	689	
12			Facebook + email: medtechqmbl@gmail.com	Barts and The London MedTech Society	542	
13			Facebook + email: bartslondonmess@gmail.com	BL MESS	935	
14			Instagram + email: paediatrics@bartslondon.com	Paediatric Society	877	

15			Facebook + email: blorthopaedicsociety@gmail.com	Barts and The London Orthopaedics Society	618	
16			Facebook + email: blplasticsurgerysoc@yahoo.com	BL Plastic Surgery Society	513	
17			Facebook + email: blpems@bartslondon.com	Barts and The London Prehospital and Emergency Medicine Society	1231	
18			Facebook	Barts and The London School of Medicine and Dentistry	20408	
19	University of Birmingham College of Medical and Dental Sciences	University of Birmingham	Facebook + email: bims@uobmedsoc.com	BIMS - Birmingham Internal Medicine Society	1877	
20			Facebook + email: med-soc-president@adf.bham.ac.uk	Birmingham MedSoc	5210	
21	University of Bristol Medical School	University of Bristol	Facebook + email: secretary@galenicals.org.uk	Galenicals	1969	Promised forward to members with the caveat that all are on holiday and participation will be low.
22			Facebook + email: uob.anaesthetics.criticalcare@gmail.com	University of Bristol Society	374	

				for Anaesthetics and Critical Care		
23			Facebook + email: cardiologysociety.bristol@gmail.com	University of Bristol Cardiology Society	577	
24			email: ems.bristoluni@gmail.com	Bristol University Emergency Medicine Society (BUEMS)		
25			Facebook + email: bristolgpsoc@me.com	Bristol GPSOC	29	
26			email: bristolmedtech@gmail.com	Bristol MedTech		
27			Facebook + email: bristol.uni.paeds@gmail.com	BUMPS - Bristol University Medics' Paediatric Society	680	
28			email: uobsimulationsociety@gmail.com	Simulation Society		
29			Facebook + email: bristolsems@outlook.com , uobsems@outlook.com	University of Bristol Sports and Exercise Medicine	138	
30			Facebook + email: cc16349@bristol.ac.uk , brisorthosoc@gmail.com	Bristol University OrthoSoc	592	
31			email: uobwilderness.medicine@gmail.com	Bristol University Wilderness Medicine Society		

32	Brighton and Sussex Medical School	University of Brighton University of Sussex	Facebook	Brighton and Sussex Medical School	2810	
33			Facebook + email: wildernessbsms@gmail.com	BSMS Wilderness Medicine Society	508	
34			Instagram	BSMS SurgSoc - Official Brighton and Sussex Medical School Surgical Society	631	
35			Facebook + email: bsmspaedsoc@gmail.com	BSMS PaedSoc	357	
36			Facebook	BSMS MedTech	79	
37			Facebook	BSMS Research Society	349	
38			Facebook + email: cardiosoc@bsmsmedsoc.co.uk	BSMS CardioSoc	100	
39			Facebook + email: info.bsmsamecs@gmail.com	BSMS AMECS	494	
40			Facebook + email: comms@bsmsmedsoc.co.uk	BSMS MedSoc	966	
41			Facebook + email: ctss@bsmsmedsoc.co.uk	BSMS Cardiothoracic Surgical Society-CTSS	213	
42	University of Buckingham Medical School	University of Buckingham	Instagram + email: ubmssurgicalsociety@gmail.com	UBMS Surgical Society		
43			Instagram+ email: Buckinghampeerledteaching@outlook.com	Peer-Led Teaching Society		
44			Facebook + email: uob.gp.soc@gmail.com	University of Buckingham GP Society	946	

45			Instagram + email: buckinghamemergencysociety@gmail.com	Emergency Medicine Society	287	
46			Instagram+ email: buckinghamemergencysociety@gmail.com	General Medicine Society UBMS	173	
47	University of Cambridge School of Clinical Medicine	University of Cambridge	Facebook + email: cuscrs@gmail.com	Cambridge University Students' Clinical Research Society	2081	
48			Facebook + email: academic@clinsoc.co.uk , secretary@clinsoc.co.uk , president@clinsoc.co.uk	ClinSoc		
49			Facebook + email: cambridgepaedsoc@gmail.com	Cambridge University Paediatrics Society	1016	
50			Facebook + email: president@cambridgemedsoc.com	Cambridge MedSoc	3468	Positive response, promised to forward to members
51			Facebook + email: cambridgegpsoc@gmail.com	Cambridge GP Society	497	
52			Facebook + email: trinitymedsoc@gmail.com	Trinity College Medical Society	371	
53			Facebook + email: cusurgicalresearchsoc@gmail.com	Cambridge University Surgical Research Society	33	

54			Facebook + email: foundation.medtech@gmail.com	Cambridge MedTech Foundation	843	
55			Instagram + email: cumhs@cambridgesu.co.uk	CU Medical Humanities Society		
56	University of Exeter Medical School	University of Exeter	Facebook + email: uems-web@exeter.ac.uk	University of Exeter Medical School	5130	
57			Facebook + email: exess@groups.exeterguild.com	Exeter Surgical Society (ExeSS)	1790	
58			Facebook + email: exetermedsoc1@gmail.com	Exeter MedSoc	1666	
59	Hull York Medical School	University of Hull University of York	Facebook	Hull MedSoc	201	
60			Facebook + email: hull.surgicalsociety@hyms.ac.uk	SurgSoc Hull	388	
61			Facebook + email: pcp@hyms.ac.uk	HYMS Pre-Hospital Care Programme	1300	
62			Facebook	Hull York Medical School	2098	
63	Imperial College London Faculty of Medicine	Imperial College London	Facebook + email: icsm.secretary@imperial.ac.uk	Imperial College School of Medicine Students' Union	5245	Promised to forward the info to members
64			Facebook + email: surgical.soc@imperial.ac.uk	ICSM Surgical Society	4484	
65			Facebook + email: uk.amsa@gmail.com	Asian Medical Students' Association	1672	

				United Kingdom - AMSA UK		
66			Facebook + email: gp.careers@imperial.ac.uk	Imperial College London General Practice Society	555	
67			Facebook + email: icsmcoding@ic.ac.uk	ICSM Coding Society	825	
68			Facebook + email: muslimmedics@imperial.ac.uk	Muslim Medics	2436	
69			Facebook + email: medtech@imperial.ac.uk	MedTech Imperial	1183	
70			Facebook + email: medhums@imperial.ac.uk	ICSM Medical Humanities	405	
71			Facebook + email: phem@imperial.ac.uk	Imperial PHEM	629	
72			Facebook + email: sora@imperial.ac.uk	ICSM Society of Research and Academia	1200	
73	Keele University School of Medicine	University of Keele	Facebook + email: soc.medicaleducation@keele.ac.uk	Keele Medical Education Society	252	
74			Facebook + email: medicine@keele.ac.uk	School of Medicine, Keele University	562	
75			Facebook	Keele Medsoc	537	
76			Instagram + email: soc.womeninsurgery@keele.ac.uk	Women In Surgery Keele	354	
77			Instagram	Keele SEM society	56	
78			Facebook	Keele Surgical Society	835	
79			Facebook + email: soc.medicaleducation@keele.ac.uk	Keele Medical Education Society	252	

80			Instagram	GP Society Keele	367	
81	Kent and Medway Medical School	University of Kent Canterbury Christchurch University	Facebook + email: enquiries@kmms.ac.uk	Kent and Medway Medical School	529	
			Instagram + email: medsoc@ccsu.co.uk	KMMS MedSoc	237	
82	King's College London GKT School of Medical Education	King's College London (University of London)	Facebook + email: msa@kcl.ac.uk	GKT MSA - Medical Students' Association	4126	
83			Facebook + email: kclmerj@gmail.com	KCL Medical Education and Research Journal Society	464	
84			Facebook + email: kclmsgpsoc@gmail.com	KCLMS GP Society	777	
85			Facebook + email: extendedmedicinesociety@gmail.com	KCL EMDP	334	
86			Facebook + email: acmsatkcl@gmail.com	African Caribbean Medical Society at Kings College London	240	
87			Facebook + email: kclwms@gmail.com	KCL Wilderness Medicine Society	1611	
88			Facebook + email: kclwim@gmail.com	KCL Women in Medicine Society	256	
89			Facebook + email: kclsurgicalsociety@gmail.com	KCL Surgical Society	2597	

90	Lancaster University Medical School	University of Lancaster	Facebook + email: lancastermedsoc@gmail.com	Lancaster MedSoc	493	
91			Facebook + email: luwildernessmedicine@gmail.com	Lancaster University Wilderness and Emergency Medicine Society	171	
92			Facebook + email: lancsgpsoc@gmail.com	Lancaster GPSociety	155	
93			Facebook + email: bmemedicslancaster@gmail.com	BME Medics Lancaster	116	
94	University of Leeds School of Medicine	University of Leeds	Facebook + email: medsocleeds@gmail.com	Medsoc Leeds	1996	
95			Facebook + email: cuttingedgeleeds@hotmail.com	Cutting Edge Leeds	1175	
96			email: hitleeds1@gmail.com	Healthcare Innovation Team		
97			Facebook + email: general@msrcleeds.co.uk	Leeds MSRC	1224	
98	University of Leicester Medical School	University of Leicester	Facebook + email: lmspaedsoc@gmail.com	Leicester Paediatric Society	520	
99			Instagram + email: cardiosoclusuma@gmail.com	LUSUMA CardioSoc	574	
100			email: leicestermmf@gmail.com	MMF Revision Society		
101			Instagram + email: uolgpsociety@gmail.com	Leicester GP Soc	440	
102			Facebook	University of Leicester Rheumatology Society	110	

103	University of Nottingham - Lincoln Medical School	University of Lincoln	Facebook + email: anatomical@lincolnsu.com	Lincoln Anatomical Society	30	
104			Facebook	University of Lincoln Med Soc	125	
105			Facebook	Lincoln GP Society	55	
106			email: mzyss46@nottingham.ac.uk	SCRUBS Lincoln Cell		
107	University of Liverpool School of Medicine	University of Liverpool	Facebook + email: president@lmssonline.co.uk	Liverpool Medical Students' Society - LMSS	1669	Instructed to post on the wall (fb).
108			Facebook + email: pre-hospital@society.liverpoolguild.org	Liverpool Pre-Hospital and Emergency Medicine Society	1624	
109			email: womeninmedicine@society.liverpoolguild.org	Women in Medicine		
110			Facebook + email: bmemedics@society.liverpoolguild.org	BME Medics Liverpool	865	
112			Facebook	Liverpool Clinical Anatomy Society	113	
113			Facebook + email: livgpsoc@outlook.com	University of Liverpool GP Society	963	
114			Facebook + email: cardiovascularthoracic@society.liverpoolguild.org	Liverpool Cardiovascular and Thoracic Society	1115	
115			Facebook + email: laces@society.liverpoolguild.org	Liverpool Acute, Critical & Emergency	7585	

				Medicine Society - LACES		
116			Facebook + email: lifestylemedicine@society.liverpoolguild.org	University of Liverpool Lifestyle Medicine Society	333	
117			Facebook + email: meded@society.liverpoolguild.org	MedEd	3131	
118			Facebook + email: orthopaedics@society.liverpoolguild.org	Liverpool Orthopaedic and Trauma Society	396	
119	University of Manchester Medical School	University of Manchester	Facebook	MMRSSoc - Manchester Medical Research Student Society	1416	
120			Facebook + email: semsocuk@gmail.com	SEMSoc - The Sports and Exercise Medicine Society	1607	
121			Facebook + email: uomgps@gmail.com	General Practice (GP Soc)	1667	
122			Facebook + email: emsocmanchester@hotmail.com	Emergency Medicine Society	1972	
123			Facebook + email: manchestermeded@gmail.com	Manchester MedEd	873	
124			Facebook + email: mmrssoc@gmail.com	Medical Research Students	1416	
125			Facebook + email: womeninmedicine.manchester@gmail.com	Women in Medicine Society	592	
126			Facebook + email: president@manmedsoc.co.uk	ManMedSoc	3112	

127	Newcastle University School of Medical Education	University of Newcastle	Facebook + email: academicmedncl@gmail.com	Newcastle Academic Medicine Society	1133	
128			Facebook	Newcastle Anatomy Society	167	
129			Facebook + email: newcorthosoc@gmail.com	Newcastle University OrthoSoc	518	
130	University of Nottingham School of Medicine	University of Nottingham	Facebook + email: nottsmedsoc@gmail.com	Nottingham MedSoc	2125	
131			email: fomsfnottingham@gmail.com	GPSoc Nottingham		
132			email: medsocteaching@gmail.com	MedSoc Teaching		
133			email: nusems@outlook.com	NUSEMS (Sports Exercise Medicine)		
134			email: nutos@nottingham.ac.uk	Trauma and Orthopaedic Society		
134			email: uonlms@gmail.com	Lifestyle Medicine Society		
135			email: enquiries@wamsnottingham.com	WAMS (Widening Access to Medical School)		
136	Norwich Medical School	University of East Anglia	Facebook + email: ueamedsoc1@gmail.com	UEA MedSoc	1116	
137			Facebook + email: anatomysoc1.uea@gmail.com	UEA Anatomy Soc	197	
138			Facebook + email: ueagpsociety@gmail.com	UEA GP Society	466	
139			Instagram + email: uea-links@sja.org.uk	UEA first aid society		

140			Facebook + email: norwichpcp@gmail.com	Norwich Pre-Hospital Care Programme	1655	
141			Facebook + email: norwichsurgicalsociety@gmail.com	Norwich Undergraduate Surgical Society	352	
142			Facebook + email: ueawms@gmail.com	UEA Wilderness Medicine Society	203	
143			Facebook + email: norwichmeded@gmail.com	Norwich MedED	1722	
144			Facebook + email:			
145	University of Oxford Medical Sciences Division	University of Oxford	Facebook + email: president@oxfordmedsoc.com	Oxford MedSoc	1925	
146			Facebook + email: admin@oxfordmedicaleducation.com	Oxford Medical Education	12562	
147	Plymouth University Peninsula Schools of Medicine and Dentistry	University of Plymouth	Instagram + email: info.pumasoc@gmail.com	Plymouth University Medical Academic Society (PUMASoc)	140	
148			Facebook + email: plymouthgpsoc@gmail.com	Plymouth GP Soc	375	
149			Instagram + email: plymouthcardiology@gmail.com	Plymouth Cardiology Society (PCS)	312	Positive response, promised to forward to members
150			Instagram + email: nursocuop@gmail.com	NurSoc	269	
151			Facebook + email: committee.pms@gmail.com	Peninsula MedSoc	621	
152	University College London Medical School	University College London	Facebook + email: secretary@medicalsociety.org.uk	UCL Medical Society	3455	Promised to forward the

		(University of London)				info to members
153			Facebook + email: uclmedtechsociety@gmail.com , su.medtechsociety@ucl.ac.uk	MedTech UCL	2525	
154			Facebook + email: uclwildernessmedicine@gmail.com	UCL Wilderness Medicine Society	1088	
155			Facebook + email: su-anatomy.society@ucl.ac.uk	UCL Anatomy Society	1561	
156	University of Sheffield Medical School	University of Sheffield	Facebook + email: medsoc@sheffield.ac.uk , publicity@medsoc.net	Sheffield MedSoc	1819	
157			Facebook + email: actsoc@sheffield.ac.uk	Sheffield Acute Care and Trauma Society (ACTSoc)	726	
158			Facebook + email: gpsoc@sheffield.ac.uk	Sheffield GP Society	600	
159			Facebook	Sheffield Orthoplastic Society	168	
160			Facebook + email: sheffieldbamemedics@gmail.com	Sheffield BAME Medics Society	106	
161			Facebook + email: sportsmedicinesoc@sheffield.ac.uk	Sheffield Sports & Exercise Medicine Society	515	
162			Facebook + email: sams@sheffield.ac.uk	Sheffield Academic Medicine Society (SAMS)	591	
163	University of Southampton School of Medicine	University of Southampton	Facebook + email: medsoc@soton.ac.uk	Southampton University Medical Society	2394	

164			email: semsoc.soton@gmail.com	Southampton Sports & Exercise Medicine Society		
165			email: gpsoc@soton.ac.uk	Southampton University GP Soc		
166			Facebook + email: sotonmedtech@gmail.com	MedTech Southampton	245	
167			Facebook + email: mems.medsoc@gmail.com	UoS - Management and Entrepreneurship for Medics Society		
168			Facebook + email: orthosocietyuos@gmail.com	University of Southampton Orthopaedic Society	86	
169			Facebook	PHEMSoc Southampton	69	
170	University of Sunderland School of Medicine	University of Sunderland	Facebook	University of Sunderland Medical Society	77	
171			Facebook + email: medicaleducationsunderland@gmail.com	MESS - Medical Education Society Sunderland	8	
172			Facebook + email: sunderlandparamedicsociety@gmail.com	Sunderland Paramedic Society	248	

173	St George's, University of London	St. George's (University of London)	email: tandosgul@outlook.com	SGUL Trauma and Orthopaedics Society		
174			Facebook + email: emergency@su.sgul.ac.uk	St George's Emergency Medicine Society		
175			Instagram + email: mededsgul@gmail.com	St George's Medical Education Society		
176			email: integrativemedicine@su.sgul.ac.uk	St George's Lifestyle Medicine Society		
177			Facebook + email: mwfed@su.sgul.ac.uk	Medical Women's Federation SGUL	133	
178			Facebook + email: parasoc@su.sgul.ac.uk	St George's Paramedic Society	114	
179			Facebook + email: sgul.gpsociety@outlook.com	St George's General Practice Society	318	
180			email: physiosoc@su.sgul.ac.uk	SGUL Physio Soc		
181			Facebook + email:			
182	University of Warwick Medical School	University of Warwick	Facebook + email: WMSInspire@warwick.ac.uk, medsoc@warwicksu.com	Warwick MedSoc	1071	
183			Facebook + email: wmsgpsoc@gmail.com	Warwick Medical School GP Society	454	

184	University of Aberdeen School of Medicine and Dentistry	University of Aberdeen	email: aberdeenglobsurg@outlook.com	Aberdeen Global Surgery Society (AGSS)		
185			Facebook + email: ogstonmedia@gmail.com	Ogston Surgical Society	907	
186			Facebook + email: abdnplastics@gmail.com	Aberdeen University Plastic, Reconstructive, and Aesthetic Surgery Society	186	
187			Facebook + email: aberdeengpsoc@gmail.com	Aberdeen GP Society	402	Promised to forward the info to members
188			Facebook + email: abdnoutreach@outlook.com, aberdeenmedicoutreachsociety@hotmail.com	Aberdeen Medic Outreach Society	547	
189			Facebook + email: abdnmedtech@gmail.com	Aberdeen MedTech	40	
190			Facebook + email: aberdeensems@gmail.com	ASEMS - Aberdeen Sport & Exercise Medicine Society	505	
191			Facebook + email: assam@nsamr.ac.uk	Aberdeen Student Society for Academic Medicine (ASSAM)	359	
192			Facebook + email: auaces@gmail.com	Acute, Critical and Emergency Care Society ACES	1302	

193			Instagram + email: aberdeenBMS@outlook.com	Black Medics Scotland Aberdeen		
194			email: Macleod.im@outlook.com	Macleod Internal Medicine Society		
195			email: uoaorthosoc@outlook.com	The Orthopaedic Surgery and Rheumatology Society (THOR)		
196			Facebook + email: aberdeenmedsoc@gmail.com	University of Aberdeen Medical Society	2073	
197			email: wildmed@abdn.ac.uk	Wilderness and Expedition Medicine Society		
198			Facebook + email: aberdeenneuro@gmail.com	University of Aberdeen Neurological and Neurosurgical Society		
199			Facebook + email: uoacvs@outlook.com	University of Aberdeen Cardiovascular Society	633	
200			Facebook + email: aberdeenmwa@gmail.com	Aberdeen Medical Women's Association	240	
201			Facebook + email: abdnmedicalhumanitiessociety@gmail.com	Aberdeen Medical Humanities Society	205	

202	University of Dundee School of Medicine	University of Dundee	Facebook + email: ducas@dundee.ac.uk	Clinical Anatomy	445	
203			Facebook + email: dumes@dundee.ac.uk	DUMES Dundee University Medical Education Society		
204			Facebook + email: ogsoc@dundee.ac.uk	Obstetrics and Gynaecology	316	
205	The University of Edinburgh Medical School	University of Edinburgh	Facebook + email: cmvm.communications@ed.ac.uk	Edinburgh Medical School		
206			Facebook + email: lgbt.edinburghmedics@gmail.com	Edinburgh University LGBT+ Medics	258	
207			Facebook + email: wellmededinburgh@gmail.com	WellMed Edinburgh	436	
208			Facebook	Edinburgh BAME Medics	299	
209			Facebook + email: euorthosoc@gmail.com	Edinburgh University Trauma & Orthopaedics Society - EUTOS	839	
210			Facebook + email: paediatricssociety@gmail.com	Paediatrics Society	933	
211			Facebook + email: edneurosoc@gmail.com	Edinburgh University Neurological Society	1688	
212			Facebook + email: edmedgp@gmail.com	EUGPS - Edinburgh	467	

				University GP Society		
213			Facebook + email: eumehs@gmail.com	Edinburgh University Medical Ethics and Humanities Society	555	
214			Facebook + email: euems10@gmail.com	Emergency Medicine Society	789	
215			Facebook + email: eumededsoc@gmail.com	Medical Education Society	503	
216			Facebook + email: mwfstudents.edinburgh@gmail.com	Medical Women's Federation		
217			Facebook + email: ed.wilderness@gmail.com	Edinburgh University Wilderness Medics Society	409	
218	University of Glasgow School of Medicine	University of Glasgow	Facebook + email: glasgowuniversitycap@gmail.com	Clinical Assessment & Practice (CAP) Group	593	
219			Facebook + email: guclinicalanatomy@gmail.com	Clinical Anatomy Society	1130	
220			Facebook + email: gucendocrinology@gmail.com	Clinical Endocrinology and Diabetes Society	295	
221			Facebook + email: guemergencymedsoc@gmail.com	Emergency Medicine Society	1306	
222			Facebook + email: glasgowgastrosociety@gmail.com	Gastroenterology Society	146	

223			Facebook + email: medicinsightglasgow@gmail.com	Medic Insight Glasgow	1549	
224			email: uofgmiis@gmail.com	Medical Infection and Immunology Society	--	
225			Facebook + email: glasgowmedicalresearchsociety@gmail.com	Medical Research Society	1433	Promised to forward the info to members
226			Facebook + email: glasgowmwf@gmail.com	Medical Women's Federation	235	
227			Facebook + email: medchir@hotmail.com	Medico-Chirurgical Society	1706	
228			Facebook + email: glasgowuniversitymedtech@gmail.com	MedTech Society	81	Promised to forward the info to members
229			email: glasgomfs@outlook.com	Oral and Maxillofacial Surgery Society	--	
230			Facebook + email: glasgowunipalliativesociety@gmail.com	Palliative Care Society	118	
231			Facebook + email: GURUSociety2020@gmail.com	Renal and Urology Society	27	
232			Facebook + email: gurespmedsociety@gmail.com	Respiratory Medicine Society	104	
233			Facebook + email: shihedinburgh@gmail.com	Homelessness and Inclusion Health Society (HIHS)	261	

234			Facebook + email: euanaesthesia@gmail.com	Anaesthesia and Critical Care Society	632	
235	University of St Andrews School of Medicine	University of St. Andrews	Instagram + email: butemedsoc@st-andrews.ac.uk	Bute Medical		
236	Cardiff University School of Medicine	Cardiff University	Facebook + Instagram + email: medsoc@cardiff.ac.uk	Cardiff Medsoc	2046	Positive response, promised to forward details to members
237	Swansea University Medical School	Swansea University	Facebook + email: medsoc@swansea-societies.co.uk	Swansea Medsoc		
238						
239	Queen's University Belfast School of Medicine	Queen's University Belfast	Facebook + email: gp-soc@qub.ac.uk	QU-GPSociety	936	
240			Facebook + email: medicsformedics@qub.ac.uk	Medics for Medics Society		
241			Facebook + email: msfc@qub.ac.uk	Medical Students For Choice	582	
242			Facebook + email: amsa-ni@qub.ac.uk	Asian Medical Students' Association	366	
243			Facebook + email: acms@qub.ac.uk	African and Caribbean Medical Society	164	
244			Facebook + email: medicalethics@qub.ac.uk	Medical Ethics Society	227	
245			Email: friendsofMSF@qub.ac.uk	Friends of Médecins Sans Frontières		

246			Instagram + email: cmf@qub.ac.uk	Christian Medical Fellowship		
247			Facebook	Queens University Academic Medicine Society	499	
248			Email: bmsa@qub.ac.uk	BMSA (Belfast Medical Students Association)		
249	Edge Hill University Medical School	Edge Hill University	Facebook + email: edgehillmedicalschool@gmail.com	Edge Hill Medical Society	29	
250	London School of Hygiene & Tropical Medicine	London School of Hygiene & Tropical Medicine	Facebook + email: alumni@lshtm.ac.uk	London School of Hygiene & Tropical Medicine (LSHTM) Alumni Association	12808	
251	University of Central Lancashire School of Medicine	University of Central Lancashire	email: Uclanaicu@Gmail.com	Anaesthesia & ICU Society		
252			Facebook + email: uclancardiothoracic@gmail.com	UCLAN Cardiothoracic Society	116	
253			Instagram + email: uclancardiologysociety@gmail.com	UCLan Cardiology Society	339	
254			Facebook + email: uclancts@gmail.com	UCLAN Clinical Teaching Society	18	
255			Facebook + email: empulse1718@gmail.com	EMpulse UCLan	132	
256			Facebook + email: uclangpsociety@outlook.com	UCLan GP Society	110	

257			Facebook + email: paedsocuclan@gmail.com	UCLan Paediatric Society	67	
258			Facebook + email: uclansurgsoc@gmail.com	UCLan Surgical Society	351	
259	Brunel Medical School, Brunel University London	Brunel University London	Facebook	Brunel University Physician Associates	60	
260	Ulster University, School of Medicine	Ulster University	Facebook + email: bmsa125@gmail.com	The Ulster Medical Society	210	

Total number is 258, not 260.

Total number contacted 249

Sources:

https://www.medschools.ac.uk/studying-medicine/medical-schools

https://en.wikipedia.org/wiki/List_of_medical_schools_in_the_United_Kingdom

Appendix G – Draft email/Fb message used

An invitation to test an AI-driven Standardised Patient

Hello [MEDSOC name],

I'm a researcher at Aberdeen University. We designed an AI-driven virtual Standardised Patient. And we would appreciate it if you could test the software and register your impression of it in a 6 minute survey.

Here's the link: https://chameleonpatients.co.uk/

Would it be too much to ask if you could forward this email to your members, please? Users' evaluation of the software will help us improve it and uncover previously unknown bugs.

Many thanks!

Al-Hussein Abutaleb

MSc candidate, University of Aberdeen

Appendix H – Documentations and Educational material used

- Google Dialogflow python github repository: https://github.com/googleapis/python-dialogflow-cx/blob/HEAD/samples/snippets/detect_intent_texts.py
- Google Dialogflow CX documentation – Python Client Library
  https://cloud.google.com/dialogflow/cx/docs/reference/library/python
  https://googleapis.dev/python/dialogflow-cx/latest/index.html
- Udemy - ChatBots: Messenger ChatBot - DialogFlow and nodejs
- Udemy - Create Chatbot for Website with React and Node.js
- Net Ninja – Full Modern React Tutorial
  https://www.youtube.com/watch?v=j942wKiXFu8&list=PL4cUxeGkcC9gZD-Tvwfod2gaISzfRiP9d
- Net Ninja – Django Tutorial
  https://www.youtube.com/watch?v=n-FTlQ7Djqc&list=PL4cUxeGkcC9ib4HsrXEYpQnTOTZE1x0uc

Appendix I – The Survey

Chameleon
User experience survey (6 minutes)

Please bear in mind that the chatbot and the website were designed with **doctors and medical students** (4th year and beyond) in mind.

Section 1

1.Are you a 4th or 5th year medical student?

- Yes
- No

2.What is your speciality in healthcare?

- Physician
- Dentist
- Lab specialist
- Nurse
-

3.Did Chameleon break? Was there a point in the interaction where no matter what you asked the chatbot was unable to understand your input?

- Yes
- No

4.Overall how do you rate Chameleon with respect to:

	Poor	Fair	Good	Very good	Excellent
Correctly understanding what you asked					
Its ability to to mimic a real patient					
Naturalness of Chameleon's responses					

5.Were you able to reach a diagnosis?

- Yes. With high confidence.
- Yes. With moderate confidence.
- Yes. With low confidence.
- No

6.If you did not reach a high confidence diagnosis, which of the following would have made reaching a high confidence diagnosis easier for this presentation?

- Options to perform a virtual exam on the affected area (using an animated avatar of the patient's body)
- Options to order medical imaging procedures and view the results (using AI generated images of the affected area)
- Options to order lab tests (using randomly generated but realistic readings)

7.Overall, how do you rate Chameleon's performance as a virtual standardised patient?

Extremely poor

Outstanding

8.Given your interaction with the chatbot Chameleon, do you think it could eventually replace the use of Standardised Patients in OSCEs?

- Yes, immediately.
- Yes. After significant improvements to the current version

- Neutral.
- No. Not with the performance of the version just tested
- No. Not even after significant improvements.

9.Did you try to ask the virtual patient questions that are not directly related to the symptoms?
For example, under Social and Lifestyle History you could have asked it about its recent travel history

- Yes
- No

10.(optional) What were the topics you asked about that were not directly related to the symptoms

11.With regards to the website you used to interact with Chameleon. How do you rate the following:

	Poor	Fair	Good	Very good	Excellent
Ease of use					
Intuitiveness of the User Interface (UI)					
Responsiveness (is it fast?)					
Reliability (does it work correctly?)					

12.In your subjective opinion, which of these statements do you agree most or least with:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think I would like to use Chameleon again	○	○	○	○	○
I think I would like to use Chameleon with a different clinical presentation	○	○	○	○	○
I found the design of the buttons, message history, and patient biographical information well integrated.	○	○	○	○	○
I thought there was too much inconsistency in the website	○	○	○	○	○
I believe most doctors and medical students with clinical training would find the website easy to use	○	○	○	○	○
I found the website cumbersome to use	○	○	○	○	○

13.What do you think future releases of Chameleon should focus on?

- ☐ The website
- ☐ The prescribed responses of Chameleon
- ☐ The internal algorithms responsible of identifying the meaning behind your questions

14.(Optional) Do you have any concerns surrounding the ethics of using Chameleon as a pedagogical tool?

15.Considering your *complete* experience with our software solution (website and chatbot), how likely would you be to recommend its use to a colleague for training or educational purposes?

○	○	○	○	○	○	○	○	○	○	◉

Not at all likely

Extremely likely

16.(optional) User feedback is instrumental in pulling our development focus to where it is most needed.
Any suggestions or feedback on how to improve Chameleon?

Appendix J – Ethical Approval Forms

PSEEB_Annex_A_Ethics_Checklist

RESEARCH ETHICS REVIEW CHECKLIST

The University Research Ethics & Governance Framework applies to all aspects of research undertaken within the University, including research undertaken by undergraduate and postgraduate students as a part of their dissertation, thesis or coursework. All academic staff, PGT, PGR students and UG students therefore should consider the ethical dimensions of their research using the self-assessment checklist below and, where necessary, seek ethical review and approval.

Ethical review must be sought for any project that answers **YES** to any one or more of the questions on this page:

Applicant Name: Al-Hussein Abutaleb	
Project Title: Chameleon – A virtual standardised patient	
	YES/NO
Does the project involve human participants?	**YES**
Does the project involve personal data? If YES, is the personal data from an existing dataset? If YES, (a) did the participants who originally provided the data give permission for it to be used in further research? (b) is the dataset publicly available?	**NO**
Does the project involve human remains?	**NO**
Does the project involve surveys or questionnaires?	**YES**
Will the participants of the project include staff or students of the University, or colleagues or clients in a work environment?	**YES**
Does the project involve children (under 18 years) or vulnerable adults?	**NO**
Does the project involve any clinical procedure or involve clinical populations?	**NO**
Could participants experience physical or psychological harm or discomfort?	**NO**
Does the project involve the collection of material that could be considered of a sensitive personal, medical or psychological nature, or is constrained by other data protection requirements?	**NO**
Does the project involve concealment or deception or deliberately misleading participants?	**NO**
Does the project involve face-to-face interviews or the collection, preservation or use of sound and/or video material involving human participants?	**NO**

Does the project involve collecting personal data from websites or from social media (e.g., Facebook, Twitter)?	**NO**
Is there any potential for conflict of interest between research funder, investigators and/or participants that may affect funding, dissemination or other research outcomes?	**NO**
Could the project lead to financial gain for funders, investigators or participants?	**NO**
Is the research likely to have any significant detrimental or lasting impact on the environment? *This includes the natural environment but also buildings and structures created by people, especially ones of historical or archaeological importance.*	**NO**
Does the project give rise to a realistic risk to the national security of any country?	**NO**
Does the project involve the collection of genetic resources (Nagoya Protocol)?	**NO**

CONFIRMATION BY APPLICANTS:

I confirm that I have discussed this checklist with my supervisor. *(For students only)* ☒

I understand that if during my project the answers to any of the above questions change, I must complete a new checklist and seek ethical review if necessary. ☒

FOR FUNDED RESEARCH ONLY

Do you have, or are you applying for, external funding?
YES/NO

If Yes, give the name of the funder.
…….....

Does the research funder require the project to undergo ethical review?
YES/NO

Do you need written evidence that your project has undergone ethical review?
YES/NO

If Yes, please fill out these details:

SURNAME/FAMILY NAME:…….....FIRST NAME:……

STUDENT ID NUMBER:……EMAIL ADDRESS:……..…….

DATE:……..

In the event an ethical review is required, please complete the Physical Sciences & Engineering research Ethics Review and Application Form (Annex B). If you are a student then you should do this in consultation with your dissertation supervisor or course co-ordinator. You may also need to attach other documents such as a Participant Information Sheet, a Consent Form or a schedule of interview questions.

Most reviews will be undertaken by circulation to appropriate reviewers within the University. The outcome of the review will be communicated to you by the Clerk of the Ethics Board as soon as practicable.

For further information please contact the Clerk to the Ethics Board in the first instance, copsethics@abdn.ac.uk.

PSEEB_Annex_B_Research_Ethics_Review_Form

Physical Sciences & Engineering Ethics Review and Approval Form

IMPORTANT NOTE: Research projects cannot begin until ethical approval has been granted.

Please complete the relevant sections of this form if, after filling out the relevant ethical review checklist (Annex A), you have identified a potential ethical issue. Please send the completed form and supporting documentation to copsethics@abdn.ac.uk.

Name: **ID number: *(for students only)***	Al-Hussein Abutaleb 52096834
School: **Department or discipline:**	The School of Natural and Computing Sciences Computing Science
Programme (e.g., PhD, MSc): *(for students only)*	MSc Artificial Intelligence
[63]Project Title: **Course Number and Name: *(for students only)***	**Chameleon – A virtual standardised patient** MSc Project in Artificial Intelligence (CS5917)

[63] Project = the **particular piece of work for which you are applying for ethical approval** (not your overall programme of research)

Names of other individuals involved in the research/project?	Al-Hussein Abutaleb Dr. Bruno Yun
Name and email address of main supervisor: *(for students only)*	Dr. Bruno Yun Bruno.yun@abdn.ac.uk
Application date:	08/08/2021

Please note:

1. Research involving NHS staff, patients, facilities and premises is subject to ethical review by the NHS North of Scotland Research Ethics Service. This includes research involving individuals when their status as NHS staff or patients is relevant to the research, even when a medical condition is not the subject of the research. Research involving adults who do not have the capacity to consent is also subject to these ethical review procedures.

2. Research involving animal and biological materials is subject to Home Office regulations. Forms and guidance can be obtained from the university's Research Governance Section (researchgovernance@abdn.ac.uk).

3. Research involving the collection of genetic resources (organisms, microorganisms, DNA, RNA, proteins, small molecules) from signatories to the Convention on Biodiversity/Nagoya Protocol requires a formal agreement to be in place before this research can begin. Contact your Business Development Officer for further guidance (https://www.abdn.ac.uk/staffnet/secure/research-grant-funding-2405.php#business-development-team-)

CHECKLIST

The purpose of this checklist is to make sure no information has been inadvertently left out and to allow reviewers to assess the application more quickly. **If you do not complete the checklist and attach a completed Annex A, the application will be returned to you.**

I confirm that if my project changes significantly then I will notify the Ethics Board. ☒

I have **attached** a completed checklist (Annex A). ☒

I confirm that I have discussed this application with my supervisor. ☒
(for students only)

I have completed the University's online ethics training. NO
(for Staff and PGR students only. NB: PGT students will undertake this training at the discretion of their Programme Coordinator.)

This project requires me to **travel outwith the UK** NO
If YES, please provide the following confirmation:-

- I will comply with the requirements of the University's Overseas Travel Policy, including obtaining permission to travel (where required by the policy), completion of a risk assessment and will obtain University travel insurance cover. ☐

Other Attachments (delete YES/NO as appropriate):

I have attached a Participant Information Sheet. YES

I have attached a Consent Form . YES

I have attached a schedule of questions for surveys and/or interviews. YES

Section 1: Research projects involving human participants (not NHS staff or patients)

If you answered `No' to Q1 and/or Q2 then omit Q 3 – 18 and proceed straight to Q19. In this case, please explain your answers to Q1 and/or Q2 in Section 7.

Recruitment Procedures

		Yes	No
1	Does your project involve human participants? This includes use of surveys, questionnaires, on-line surveys and tests, focus groups and workshops where human participants provide information or data to inform the research.	**YES**	
2	(a) Does your project involve human remains? (b) If so, does your work conform with the Historic Environment Scotland guidelines?		**NO**
3	Does your project involve people less than 18 years of age?		**NO**
4	Does your project involve people with learning or communication difficulties?		**NO**
5	Is your project likely to involve people involved in illegal activities?		**NO**
6	Does your project involve people belonging to a vulnerable group, other than those listed above?		**NO**
7	Does your project involve people with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues or clients?	**YES**	
8	Does your project involve people who do not have English as their first language?		**NO**
9	Does your project require the recording of audio or video of participants or of others not involved in the research?		**NO**
10	Do you plan to conceal your own identity during the course of your project?		**NO**

*Please explain in **Section 7** how you will recruit your participants. If you answered **'Yes'** to any of the above questions, please give details.*

*If you answered **'Yes'** to **Q1** then you must provide a Participant Information Sheet and a Consent Form. For web-based research, screenshots of the appropriate web pages suffice.*

If your project involves surveys or interviews then you must provide a schedule of questions.

*If you answered **'Yes'** to **Q3, Q4** or **Q6** then you may need to apply for disclosure through Disclosure Scotland.*

Consent Procedures

		Yes	No
11	Do you have set procedures that you intend to use for obtaining informed consent from all participants, including (where appropriate) parental consent for children?	**YES**	
12	Will you tell participants that their participation is voluntary?	**YES**	
13	Will you obtain written consent for participation, including for audio and/or video recording?	**YES**	
14	Will you tell participants that they may withdraw from the research at any time and for any reason?	**YES**	
15	Will you give potential participants a period of time to consider participation?	**YES**	
16	Does your project involve concealment or deliberately misleading participants?		**NO**

*Please explain in **Section 7** how you will obtain consent from participants. If you answered '**Yes**' to **Q16** or '**No**' to any of the other questions, please give details.*

Possible Harm to Participants

		Yes	No
17	Is there any realistic risk of any participants experiencing physical or psychological harm or discomfort?		**NO**
18	Is there any realistic risk of any participants experiencing a detriment to their interests as a result of participation?		**NO**

*If you answered '**Yes**' to either question, please explain in **Section 7** how this risk was assessed and how you propose to manage it.*

Section 2: Data protection, handling, security and storage

IMPORTANT NOTE:

The General Data Protection Regulation imposes a number of obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

If you are using personal data, you should consider whether your research requires a Data Protection Impact Assessment and complies with the University Data Protection policy.

If you are, you now need to see the Data Protection Checklist for Researchers for guidance.

If you then feel that a DPIA may be required or you need data protection advice, then you should contact the Data Protection Officer dpa@abdn.ac.uk.

Please provide the following confirmation:

I have read the above guidance and have met the relevant data protection obligations.

X **Please tick the box to confirm**

In addition, you should also check the requirements for a Data Management Plan (DMP) in the Research Data Management Policy and Guidance.

Once checked, please confirm the requirement by **ticking one of the following:**

X *No requirement for DMP*

☐ *DMP required and this is attached*

Please see here for guidance on creating a DMP. For further support, contact digitalresearch@abdn.ac.uk

		Yes	No
19	(a) Will any non-anonymised and/or personalised data be generated and/or used? (b) Will you use an existing dataset in your research? (c) If 'yes', do you have permission to do so?		**NO, NO**
20	Will any data be stored (temporarily or permanently) anywhere other than in University managed data storage or on University managed encrypted computers?	**YES**	
21	Will you gain access to sensitive[64] data about living individuals or organisations that is not already publicly available elsewhere? If 'Yes', will you gain the consent of the individuals concerned?		**NO**

[64] Sensitive data includes data that relates to racial or ethnic origin, political opinions, religious beliefs, trade union membership, physical or mental health, sexual life, actual and alleged offences.

22	Does your project require access to personal data about participants from other parties (e.g., teachers, employers), databanks or files? *If yes, please explain in Section 7 how you will gain the consent of these participants.*		**NO**
23	Does the project involve collecting personal data from websites or from social media (e.g., Facebook, Twitter)?		**NO**
24	Will the data be stored, collected or accessed from: - outside the UK? - outside the EU?		**NO,NO**
25	Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory or extremist? *If yes, see here for an explanation of the obligations of the researcher and the university under the Prevent duty.*		**NO**
26	Are there any contractual conditions attached to working with or storing the data? (E.g., an NHS Digital data sharing agreement.)		**NO**
27	Could working with this data damage the University's reputation? (E.g., bad press coverage, public protest.)		**NO**
28	Could working with this data cause an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)		**NO**

For further advice on Data Protection and GDPR, please refer to www.abdn.ac.uk/dataprotection.

Please provide details in **Section 7** of how you intend to ensure that data is stored securely and in line with the requirements of the Data Protection Act and funding bodies (if applicable). Please give specific consideration to whether any non-anonymised and/or personalised data will be generated and/or stored and what precautions you will put in place regarding access.

If you answered '**Yes'**, to any of the questions above, please give details in **Section 7**.

Section 3: Research involving possible harm to the environment

		Yes	No
29	Is the research likely to have any significant detrimental or lasting impact on the environment? *This includes the natural environment but also buildings and structures created by people, especially ones of historical or archaeological importance.*		**NO**

If you answered '**Yes'**, please explain in **Section 7** how this risk was assessed and how you propose to manage it. Say whether relevant guidelines exist in your discipline, and whether you intend to follow them.

Section 4: Research which may have an adverse impact on national security

		Yes	No
30	Does your project give rise to a realistic risk to the national security of any country?		**NO**

If you answered '**Yes'**, please give details in **Section 7**. Explain how this risk was assessed and how you propose to manage it.

Section 5: Funding and conflict of interest

		Yes	No
31	Is your project funded by the university or an outside organisation, or have you applied for funding?		**NO**
32	Is there any potential conflict of interest between research funder and researchers or participants and researchers which may potentially affect the research outcome or the dissemination of research findings?		**NO**
33	Might the project lead to financial gain to funders, investigators or participants?		**NO**

If you answered '**Yes'** to any question, please give details in **Section 7.** Explain any potential conflict of interest and how you propose to manage it.

Section 6: Collection of genetic resources

		Yes	No
34	Does the project involve the collection of genetic resources (organisms, microorganisms, DNA, RNA, proteins, small molecules) from signatories to the Convention on Biodiversity/Nagoya Protocol?		**NO**

If you answered '**Yes'**, then a relevant agreement must be in place before the research can begin. This agreement must provide prior informed consent with mutually agreed terms and the must be in keeping with the Convention on Biodiversity/Nagoya Protocol and be obtained via the national focal point of the provider country. Please explain in **Section 7** how you propose to arrange the agreement. Please indicate the need for confidentiality where appropriate.

Section 7: Additional Information

All questions must be answered fully in the space provided (however each box can be expanded as necessary). Incomplete or incorrectly completed forms will be returned to the applicant, delaying the process of obtaining ethics approval.

7.1	**Project description** *Please attach a project descriptor or summary document (where available)*	In this project, we will assess the ability of a chatbot, called Chameleon, to play the role of the patient in an objective structural clinical examination (OSCE). The participant will interact with Chameleon via a web interface (available at https://chameleonpatients.co.uk/chat) and fill a Google Form questionnaire to assess their perception of Chameleon's performance.
7.2	**Start date and duration**	Start date of the project: 24/05/2021 Start date of the questionnaire:16/08/2021 Duration of project: 96 days Duration of questionnaire:5 days
7.3	**Methodology**	- Users will be shown the participation sheet. - Users will complete the consent form - Following that they invited to interact with Chameleon the chatbot. - Following that, the users are invited to complete a short questionnaire
7.4	**Recruitment of participants**	- Social media groups of medical students' Society - Social media groups of working doctors/GPs
7.5	**Consent**	Yes, example of the consent form is attached to this application (sheet E)
7.6	**Harm to participants**	None
7.7	**Data storage**	The data will be temporarily stored online on a third-party repository (Google). Following the completion of the data collection phase, all cloud-based data will be permanently deleted. Data will be kept in local storage until the end of November 2021.
7.8	**Ethical considerations** *Concise statement of the ethical considerations raised*	• No minors under the age of 18 will be asked to participate. No one with any special needs or learning difficulties will knowingly be asked to participate. • The study does not seek the recruitment of anyone from any particular group. Any disproportionate

	by the project and how you intend to deal with them. Include details related to Sections 3-6 above, if applicable.	representation of any ethnic, religious or political group is likely coincidental. Furthermore, this work will not at any point ask for any personal information from participants including ethic/religious/political background etc. • The study does not involve deception, for research purposes, of any kind. • The study automatically assumes that participants have access to the internet. It thus excludes individuals who do not have the means to 1) access the internet 2) read and type in English. This is a by-product of the nature of the study. We are only targeting English speaking doctors and medical students. • All data collected will not be shared, sold, or voluntary delivered to anyone. Only the investigators, myself and Dr. Bruno Yun will have access to the data. • No identifiable information of any kind pertaining to the participants will be collected.

For any contractual or intellectual property questions, please contact the business development team in Research & Innovation (june.middleton@abdn.ac.uk).

Make sure you have completed the checklist on page 2.

FOR STAFF AND PGR STUDENTS: Please send your completed application form and any supporting documentation to copsethics@abdn.ac.uk.

FOR PGT STUDENTS: Please refer to the ethical review procedures outlined in your Project Guidelines or contact your Postgraduate Programme Coordinator for further advice.

PSEEB_Annex_C_Participant_Information_Sheet_template

[Address of School/Department]
[School/department phone number]
[School/department email address]
[A standard School or Department letterhead can be used.]

PARTICIPANT INFORMATION SHEET
Chameleon – A virtual standardised patient

Principal Investigator(s): Al-Hussein Abutaleb

Supervisor(s): Dr. Bruno Yun

I am Al-Hussein Abutaleb, an MSc AI student at the University of Aberdeen Department of Computing Science. I would like to invite you to consider participating in the research project "***Chameleon – A virtual standardised patient***". Below is some information about the project, to help you decide whether you would like to take part.

Participation in the research project is completely voluntary. You can withdraw from the project at any time, without having to give a reason.

AIMS

We created a pedagogical tool to help healthcare students and working healthcare professionals to improve their differential diagnosis skills by probing a chatbot that simulates the symptoms of a real patient.

The goal of this experiment it to collect conversational data to improve the performance of the developed chatbot and to collect feedback from participants about the chatbot through a questionnaire.

WHAT YOU WILL BE ASKED TO DO

We would like you to interact with Chameleon, the chatbot, in a manner like what you would do if it was a real patient. Afterwards, we will ask you to fill a short questionnaire. The questionnaire will focus on the performance of the software, your impressions, and any comments you would like to add. No personal information will be collected at any time.

We expect to deploy this questionnaire for a period of 3-5 days between Monday 16th of August and Friday 20th August. We estimate that it will take no longer than 5 minutes to complete. Both Chameleon web interface and the questionnaire can be accessed from any computer with an internet connection. The website has not been optimised for mobile devices, however.

RISKS

There should not be any risks to you from the experiment. You are free to experiment with Chameleon without answering the questionnaire. You can also withdraw at any time from both the questionnaire and the interaction with Chameleon.

DATA MANAGEMENT AND STORAGE

The data collected will be stored on Google Forms and deleted at the end of November 2021. I will not record any personal information such as names or email addresses, nor will I be asking any personal

information at any point. The information collected is purely about the performance of the tool, your perception of it, and general comments about performance.

To the best of our knowledge, the nature of the questions will not allow any retrospective identification of the users.

CONFIDENTIALITY AND ANONYMITY

The University's Privacy Notice for Research Participants is available here

Raw data and the identity of participants will not be released to anyone outside the research team. The data you provide will be analysed and may be used in publications, dissertations, reports, or presentations derived from the research project, but this will be done in such a way that your identity is not disclosed. We may use quotes from the comments, or create word clouds from the comments of all users in the final dissertation. No names or attribution of any kind will be made with any quotes used.

CONSENT

If you agree to take part in the research, you will be asked to indicate your consent by clicking the button below.

SPONSORS

There are no sponsors for this research.

Thank you for considering taking part in this research.

If you have any questions about this research please contact me:

Al-Hussein Abutaleb a.abutaleb.20@abdn.ac.uk

For any queries regarding ethical concerns you may contact the Convener of the Physical Sciences & Engineering Ethics Board at the University of Aberdeen:

Email: copsethics@abdn.ac.uk

This research project was approved by the Physical Sciences & Engineering Ethics Board on Monday 9th August 2021

PSEEB_Annex_E_Consent_form_template_online

Chameleon – A virtual Standardized Patient

Consent form for participation in the research project *(Chameleon – A virtual Standardized Patient).*

Please read the statements below and tick the final box to confirm you have read and understood the statements and upon doing so agree to participate in the project.

I confirm that the research project ***(Chameleon – A virtual Standardized Patient)*** has been explained to me. I have had the opportunity to ask questions about the project and have had these answered satisfactorily.

I consent to the material I contribute being used to generate insights for the research project (***(Chameleon – A virtual Standardized Patient***).

I understand that my participation in this research is voluntary and that I may withdraw from the research at any time (until the point of data analysis) without providing a reason.

I consent to allow the fully anonymised data to be used for future publications and other scholarly means of disseminating the findings from the research project.

I understand that the information/data acquired will be securely stored by researchers, but that appropriately anonymised data may in future be made available to others for research purposes. I understand that the University may publish appropriately anonymised data in its research repository for verification purposes and to make it accessible to researchers and other research users.

☐ I confirm that I have read and understood the above statements (check the box).

Appendix K – Detailed Local Deployment Procedure

Requirements

- Python 3.9
- Docker (with virtualisation enabled on your machine)
- VS Code + PyCharm (or any other programme you are comfortable with)
- Google account

Frontend

1. Unzip the backend and frontend code bases.
2. Navigate to the directory of the frontend folder. Run the terminal. Type the command code . to launch VScode.
3. The whole project should now be loaded on to VS Code.
4. From the menu, click 'Terminal' an select a 'New Terminal'
5. Run the command npm start run
6. A localhost:3000 should automatically open in your browser. You should now be running the frontend as it was deployed. However, any attempt to communicate with DialogFlow will not succeed until the backend is running.

Backend

1. Run PyCharm.
2. Select a New Project, navigate to where you unzipped the backend folder
3. Select the Python 3.9 interpreted and for the project to inherit all global libraries.
4. Go to Backend_chameleon_python/chatbot/chatbot/plucky.json file in PyCharm. Change the placeholders to *your correct authentication keys* (procedure below).
5. Go to Backend_chameleon_python/chatbot/config/settings/test.py and Backend_chameleon_python/chatbot/config/settings/local.py, manually insert your own DJANGO_SECRET_KEY
6. In the terminal, navigate inside the folder chatbot, then run the commands (you must have installed Docker first):

   docker-compose -f local.yml build

   docker-compose -f local.yml up

7. Docker will now create a container for the backend. It should be able to communicate with the front on end on your localhost:3000. However, no communication yet with DialogFlow has been established.

DialogFlow

1. Go to: https://cloud.google.com/dialogflow/cx/docs/basics
2. Click the blue button 'Get started for free'. No charge should be applied to your existing billing account. Google will automatically give you $300 credit.
3. Select appropriate Country and Organisation.
4. Google may ask you to confirm your identity via an SMS message.
5. Again, if your account does not have an associated payment method, Google will ask for your card details. Unfortunately, this is a must for the deployment to succeed even if on your local machine.
6. Once this is done, click on 'start my free trial'
7. Google Cloud creates a default project named 'My First Project'. You should be able to see it in the Console of Google Cloud Platform (you will be directed there automatically)
8. Go to: https://dialogflow.cloud.google.com/cx/projects
9. [IMPORTANT] In the window that appears, copy the ID of 'My First Project' to a safe place. And select this project.
10. Google Cloud should now present you with the option to enable DialogFlow API. Click 'Enable API'
11. You will not be able to proceed past this point if you have not entered your verified card details.
12. You should be presented with the option to either 1) create a new agent 2) Use pre-built agent
13. Select 'Create agent'
14. Enter the Display name for the agent, for example, you could just type 'chameleon' for simplicity.
15. Change location to 'europe-west2 (London, England, UK)'
16. Change Time Zone to '(GMT) Europe/London'
17. Leave the default language settings to 'en-English'
18. You will now be directed to the console of Dialogflow CX

19. On the left-hand side pane, below the tab ‘Manage’, click on the ‘+’ sign. Select ‘Uplaod’. Navigate to the model’s file you downloaded included in the submission and click on it.
20. On the left-hand side pane, you should see two new flows. 1) Functional Enquiry. 2) Standard. Ignore the Functional Enquiry.
21. Hover the mouse curser over ‘Standard’. Click on the three vertical points that appear. Select ‘Flow Settings’
22. Under the ‘NLU model’ dropdown list, select ‘Advanced NLU’, Wait until training is complete. It should not take more than two minutes.
23. Click Save.
24. Go back to the console. On the left-hand side pane, select ‘default start flow’. This is the default flow that automatically gets created when you created your agent. It cannot be deleted. We are going to use it to divert any traffic coming to it our ‘Standard’ flow.
25. Click on the ‘Start’ page icon in the page. There should be nothing else on the page.
26. Click ‘Default Welcome Intent’ under Routes.
27. Under ‘Condition’ select the ‘customize expression’ and type ‘true’ in the condition text box.
28. Under Fulfilment, delete everything under ‘Agent says’.
29. At the bottom of the panel, under ‘Transition’, select ‘Flow’ then from the drop down menu, select ‘Standard’. Problematically, Google does not allow you to delete the ‘Default Start Flow’ that gets automatically created when you import the downloaded files. So, what we just did is we created a condition that will always be true as long as the ‘Default Welcome Intent’ gets trigged.

    So, when you hook Chameleon to your local machine, always start the conversation with ‘Hi’ or ‘Hello’ *ONLY* and the proceed with the introducing yourself as a doctor. For example:

    Me: Hi

    Chameleon: --- (we deleted everything under fulfilment)

    Me: Hello, I’m Joe. I’m one of the doctors here. What can I do for you today?

    Chameleon: Thanks for seeing me, Doctor. Last night, I was at the gym. I was deadlifting. The weight wasn't particularly heavy. However, while performing the exercise I felt a sudden sharp pain in my lower back. I cut my exercise short and went home. By then the pain has subsided significantly. However, I woke up this morning I could barely move. The pain was really intense. So, I thought I'd come and see you.

Authentication codes

1. You must create a service account and attach it to the Google Cloud project you created. Follow the procedure detailed here: https://cloud.google.com/dialogflow/cx/docs/quick/setup
2. This section of the documentation details how you can obtain the necessary IDs to plug into the plucky.json object: https://cloud.google.com/dialogflow/cx/docs/quick/api#ids
3. You should have your:
    a. PROJECT_ID
    b. REGION_ID
    c. AGENT_ID
    d. Your google service account email
        i. Your service account will enable you to create a private key and a private key id
4. Plug in all the variables into plucky.json

A quick way to generate the necessary credentials:

1. On your Google Cloud Console, click on APIs & Services
2. Select Dialgoflow API from the list of available APIs(it should already be enabled)
3. On the top right corner, click on 'CREATE CREDENTIALS'
4. Select DialogFlow API when asked
5. Select the option 'Application Data'. This will create a service account for you.
6. To the question "Are you planning to use this API with Compute Engine, Kubernetes Engine, App Engine or Cloud Functions?" Select 'No, I'm not using them'
7. Enter your service account name and associated details
8. Click 'DONE'