

Development of Conversational Artificial Intelligence for Pandemic Healthcare Query Support

Wai Lok Woo^{1*}, Bin Gao², Raid Rafi Omar Al-Nima³, Wing-Kuen Ling⁴

¹Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, NE18ST, United Kingdom

²School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

³Technical Engineering College, Northern Technical University, Mosul 41002, Iraq

⁴Faculty of Information Engineering, Guangdong University of Technology, Guangzhou, China

Abstract

The paper proposes and describes the development of conversational artificial intelligence (AI) agent to support hospital healthcare and COVID-19 queries. The conversational AI agent is called "Akira" and it is developed using deep neural network and natural language processing. It is capable of reading the inputs from the user, understanding the input and identifying the intention, and outputting messages towards the user, and these steps are iterated until the user prompts to exit or the programme is terminated. A deep learning model has been trained, and Akira could converse with the user ranging from the conversation over 7 topics related to COVID-19, common cold and flu, mental health, sexual health, abortions, allergens, drugs and medicine. The paper also describes the importance of designing an interactive human-user interface when dealing with conversational agent. In addition, the context of ethical issues and security concerns when designing the agent has been taken into consideration and discussed. The conversational agent is demonstrated to answer queries from a pool of 57 participants.

Key Words: *Artificial intelligent agents; Healthcare; Query system; Machine learning; Automation machines*

***Corresponding Author:** Wai Lok Woo, Department of Computer and Information Sciences, Northumbria University, England, NE1 8ST, United Kingdom, Tel: +44 191 2273824; E-mail: wailok.woo@northumbria.ac.uk

Received Date: July 20, 2020, **Accepted Date:** September 28, 2020, **Published Date:** October 30, 2020

Citation: Woo WL, Gao B, Al-Nima RRO, et al. Development of Conversational Artificial Intelligence for Pandemic Healthcare Query Support. *Int J Auto AI Mach Learn*. 2020;1(1):54-79.



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to non-commercial purposes.

1. Introduction

Conversational AI agent (also commonly known as chatbot) has multiple benefits towards the business and towards the user's personal life as it starts to slowly grow on to the user and client of the businesses. With the help of the recent development on artificial intelligence, it has helped to increase its own usability as it becomes more advance and intelligent [1]. For business, conversational AI agents are used to emphasis on their companies' images and brands, it helps to build a better image of the company as it represents the company itself. The most important factor is that it automates process that require little human interaction or basic task, which then save the resources and money for the company and use it somewhere else. Besides, conversational agent only offers information that are needed, and it does not overwhelm the user with unnecessary information [2].

ELIZA [3] is the first conversational AI agent to be ever created in the history of computing science. ELIZA uses a person-centred approach [4,5] to engage with the users and the approach is called the Rogerian psychotherapist. PARRY [6] is a conversational agent created by a psychiatrist. The persona for PARRY is that he imitates or simulates a paranoid mental patient. PARRY is similar to ELIZA but with an attitude. The Artificial Linguistic Internet Computer Entity or A.L.I.C.E. [7] is another conversational agent which consists of two components i.e. chatbot engine and language knowledge model. A.L.I.C.E. stores the conversation patterns in a file called the Artificial Intelligence Mark up (AIML). Google assistant [8] is an interactive artificial intelligence chatbot in recent years, among the others which stands side by side with it are Amazon's Alexa, Apple's Siri, Microsoft's Cortana and IBM's Watson. All huge enterprises have made their own version of conversational AI agents which have their own unique system each of their own. Woebot [9] is a conversational AI agent developed by a psychologist and is designed to deliver cognitive behavioural therapy (CBT) towards the user, mainly for users that are suffering from mental illness. The interaction between the users and Woebot begins with a general inquiry with context, and receive inputs from users that are either emojis or words or both, that represent their current mood.

Open-domain chatbots such as MILABOT [10], Xiaoice [11], Gunrock [12] and Mitsuku [13] display human-like attributes, but rely on complex frameworks, such as dialogue managers with knowledge-based, retrieval-based, or rule-based systems. End-to-end neural network approaches [14,15], on the other hand, offer the simplicity of a single learned model. Despite much research, open-domain chatbots still have weaknesses that prevent them from being generally useful as they often respond to open-ended input in ways that do not make sense, or with replies that are vague and generic. Closed-domain chatbots, on the other hand, respond to keywords or intents to accomplish specific tasks. The human computer interaction community has long been interested in the interaction benefits offered by conversational interfaces. The general consensus is that conversational interfaces offer several advantages over traditional WIMP (Windows, Icons, Menus, and Pointers) interfaces. First, conversational interfaces offer a natural and familiar way for users to express themselves, which in turn improves the usability of a system. Second, such interfaces are flexible and can accommodate diverse user requests without requiring users to follow a fixed path. Third, such interfaces are often personified and their anthropomorphic features could help attract user attention and gain user trust. Researchers have built various AI agents to elicit information from a user through a one-on-one, text-based conversation. Smart mobile healthcare assistant [16] has been developed to improve patient doctor interaction and to promote the self-management of chronic diseases by the patients themselves. They can leverage clinical decision

support to require interpretation of data obtained from the user, and diagnosis or recommending appropriate treatment using probabilistic classification and case-based reasoning [17]. Suitability for a mental health intervention was investigated with a simple online conversational agent as a human surrogate for a health professional [18]. The intervention of alcohol drinking habits assessment of young adults was provided. In order for AI to function as a digital companion, there is a need to take into consideration of emotional aspects similar to personality such as friends and family, beyond rational aspects such as information provision and reasoning [19]. The major problems faced by the generic conversational agents is that, due to the presence of a corpus, it assumes that its knowledge will come only from the training dataset. Furthermore, generating AIML from a corpus cannot guarantee a coherent chat because of a fear of repetitive or out-of-context statements, which will worsen the chat experience. These two issues indicate a major defect of conversational agent that often results in a lack of intelligence and a lack of variation in terms of message generation due to the pattern-matching nature of AIML.

This paper proposes a simple conversational AI agent known as Akira which is capable of replying to users when a health-related question is asked. From healthcare viewpoints, conversational AI holds high potential to provide immediate access to care [20]. For the vast majority of routine medical conditions, it is noticed that people do not have to go to a doctor. The healthcare sector has a staffing shortage problem while patients are looking to access care struggle to get through to the physicians they want to talk to, for a host of reasons. This results in frustration for patients and lost revenue for healthcare providers. In addition, conversational AI can get user to the right caregiver and care setting more accurately and in a shorter time. As with the use of conversational AI in automating processes, the accuracy and convenience of the tool leads to greater adoption over time through machine learning and natural language processing tools, potentially leading to tackling more complex care over time. However, conversational AI can also cannibalize revenue. For the same reason that telehealth adoption is not growing faster, the use of conversational AI will also struggle, especially among the vast majority of health systems still on fee-for-service reimbursement models. This has implications for who gets to control the primary interface with a healthcare consumer at the time of need. In this paper, the proposed conversational AI agent Akira is designed during the pandemic Corona Virus Disease 2019 (COVID-19) to be general health information conversational agent. Akira has a personality to ensure that the conversation is engaging and does not sound dry and dreadful and it helps to lighten up the mood if a serious question were to be asked and the answer might be too heavy for them to handle. The design of the conversational agent uses the supervised learning-based algorithm with the help of deep neural network (DNN) to work together instead to produce a response instead of just relying on the natural language processing toolkits. These toolkits play an important role as it provides the functionality to stem and lemmatize the input sentence and to tokenise the words. A survey of consisting of 57 participants has been conducted to collect responses. These participants are recruited whose age is 18-50 years old and consisting of 24 male and 33 females. Each participant will list out five questions to ask a medical practitioner.

The paper is organized as follows: Section 2 discusses the design and development of the conversational AI agent. The design incorporates various natural language processing steps, conversation flow and persona of the agent, and proposes a deep learning model. Section 3 presents the results and analysis of the conversational AI agent on healthcare queries. Issues on user experiences, response system, ethics consideration, and security concerns are discussed in Section 4. Finally, Section 5 draws the conclusions and future direction of the work.

2. Design of Conversational AI Agent

Pandemics have unique characteristics that make them amenable to tailored interventions deliverable via conversational agents. In particular, pandemics differ from other natural disasters in three ways. Firstly, individual actions can significantly worsen outcomes in a pandemic, given that a single person may infect many others depending on their behaviour. Secondly, the fear of infecting others makes infectious diseases more insidious through disease-related stigma. Thus, people can feel personally responsible for unpleasant outcomes during a pandemic and may hide symptoms from others. Thirdly, the physical gatherings to connect with others in difficult times are supposed to be avoided during a pandemic, worsening the risk for future mental health problems. Given these, conversational agents therefore have unique affordances which may mitigate short- and long- term disease burden during infectious disease pandemics. In the designed conversational agent, natural language processing (NLP) will be a part of the artificial intelligence that communicates between the computers and the users using the natural languages. The objective of the NLP is to read what the user wrote, decipher the sentences or words, extract the intents of the users and understand the user's intents in a manner that is processable by the AI.

2.1. Sentence Segmentation and Tokenization

Sentence segmentation [21] is the part of the NLP where the process of processing a large group of text into sentences and by using the punctuation as a mark to know when the sentences stop. Most of the time, the set of punctuation rules are defined. By doing that, it makes processing a lot easier to process and handle for the computers. However, it is noted different languages have different types of punctuation rules, so the rules might apply to some but not all which defines the boundaries. This is language specified which has to be dealt with using separate language models.

After segmenting the sentences, the NLP processes the words in the sentences into tokens for each word. It can tokenize each word by locating the word boundaries such as white spaces or blank spaces, the point where one words ends and the other begins [22]. Tokenizing is important because it helps to process each word for more accuracy and lemmatizing the words later in the stages of trying to understand the intention and predict the intention of the user.

The purpose of segmenting and tokenizing the sentence is to enable different weights on the tokenized words in the sentence where it could determine the outcome of the response by responding towards the heavier weight token. e.g. ELIZA used this technique to determine the outcome of the conversation.

2.2. Stemming and Lemmatization

Stemming and Lemmatization plays an important role in a conversational AI agent as both functions are crucial to identify the common base or the root of the words. The stemming uses a crude heuristic process that removes the suffixes of the words [23], while lemmatization performs properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is the lemma [24]. When the token that is confronted is the word saw, stemming it returns just as, where else lemmatization would attempt to return either see or saw depending on whether the use of the token was as a verb or a noun. The most common algorithm for stemming is Porter's algorithm

[25]. Porter's algorithm has been proven to be very effective, how it works is that it consists of 5 phases of word reductions, applied sequentially. Within each phase there are various conventions to select rules, such as selecting the rule from each rule group that applies to the longest suffix. In the first phase, this convention is used with the following rule group in Table 1.

Table 1:

Examples of stemming.

Rule	Example
SSES → SS	caresses → caress
IES → I	ponies → poni
SS → SS	caress → caress
S →	cats → cat

The lemmatizer does a full morphological analysis to accurately identified the lemma for each word. However, it only provides a modest amount of benefits towards retrieval. In order to perform Lemmatization, it is necessary to have a detailed dictionary which the algorithm can look through to link the words back to its base form, where else stemming does not. In our conversational agent design, we have used the natural language toolkit (NLTK) and spaCy.

2.3. Conversation Flow

Having a conversation with a conversational agent is easy but keeping the conversation within the context is the tricky part as some information can't be gathered in one go, it needs to be break down and analyse by the bot to understand what the user's desired answers is. The agent could be programmed to guide the flow of the conversation at the same time keeping the context by asking specific questions towards the users. Users tends to adapt their language and their behaviour to the agent answers e.g. words mentioned by bots in their answers were copied by users in their next turn questions [28]. Triggers are important as well in the conversation as it keeps the conversation going and help to guide the flow of the conversation into a different topic. It also helps to guide the user when the users are lost. Conversation flow is important as it determines the differences between a good chatbot or a bad chatbot besides the algorithm used. As the humans go, a good conversation between 2 people could eventually lead to a better chemistry or a connection towards that person, so the objective of the chatbot is to allow that connection or that chemistry to happen so the conversation could be more engaging and does not feel dry and boring. It is also important to allow the agent to have a persona, by allowing it to have that, it helps to build the conversation better and allow it to be more engaging.

The design of the conversation flow can be quite challenging as it requires the conversation to flow naturally and at the same time, the response has to sound humane [27]. The design requirement is that of a generic healthcare conversational agent and it is not meant to be used for personalisation. It provides a general information regarding the question that will be asked, e.g. "I feeling feverish", "I'm coughing a lot lately". So, the agent will respond by giving a very generic answers by saying, e.g. "Oh no, it seems like you might have a fever.", "You might have got a cold.". These responses would be prompt by the agent and some of the responses will redirect the users towards websites that have proved to be credited and a good source if they would like to know more about it. Besides, the response of the agent has been prompted to sound humanlike, sometimes it would make sounds that only a human would, sounds like e.g. "Hmm", "Oof", "Huh", etc. Figure 1

shows part of conversation flow to allow the user to ask more question and be able to engage with the user for a longer time period.

There are a different kind of responses to negative input such as “I’m feeling sick”, in response the agent would reply with “Oof, could you describe the symptoms to me?”, this is only the beginning of the conversation, the complexity increases as more tests are undertaken. Improvement on the type of response that will be inputted after the agent ask “So, how can I help you today?”, user could input sentences like, “I think I have a flu”, “I have a fever”, “I feel cold”, etc. New intents are designed to accommodate to the new input that the user might give towards the question asked by the agent.

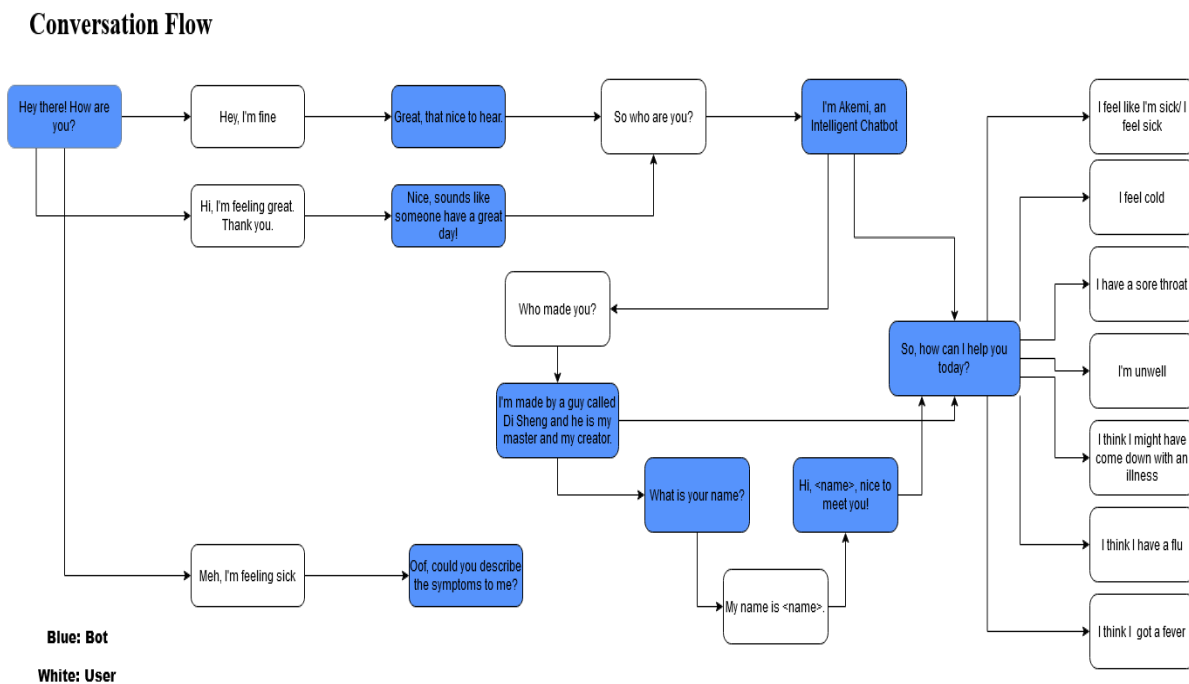


Figure 1: Part of the conversation flow of the human and agent interaction.

There are a different kind of responses to negative input such as “I’m feeling sick”, in response the agent would reply with “Oof, could you describe the symptoms to me?”, this is only the beginning of the conversation, the complexity increases as more tests are undertaken. Improvement on the type of response that will be inputted after the agent ask “So, how can I help you today?”, user could input sentences like, “I think I have a flu”, “I have a fever”, “I feel cold”, etc. New intents are designed to accommodate to the new input that the user might give towards the question asked by the agent.

2.4. Design of the Agent Persona

In order to help make the conversational agent as user friendly as possible and at the same time more engaging towards the users, a persona has been developed. The persona helps agent to promote its identity and at the same time, it helps boost the user perception when using it [28]. A name is given to the proposed conversational agent called Akira. Akira is a Japanese name that

stands for intelligence and bright, which suits Akira as her personality was designed to make her sounds like this way.

Figure 2 shows a simple profile of Akira that was created to help develop the persona and implement it into the Akira's responses in the conversation at the same time the conversation flow. It is featured on a mobile application. The design helps to shape the conversation by allowing Akira response and sometimes she would respond in a 3rd person form. This provides a framework for the nature of Akira's response to user input, for example how agreeable, extroverted or neurotic the conversational agent should be. At the same time, Akira helps the user by providing the correct information as well as it is part of her goals as shown in the profile. She deals with problems that most user faces including limited information about health, poor dealing with problems, and lack of information in mental health. Akira learns about these problems and it would slowly improve in time after interacting with more users.

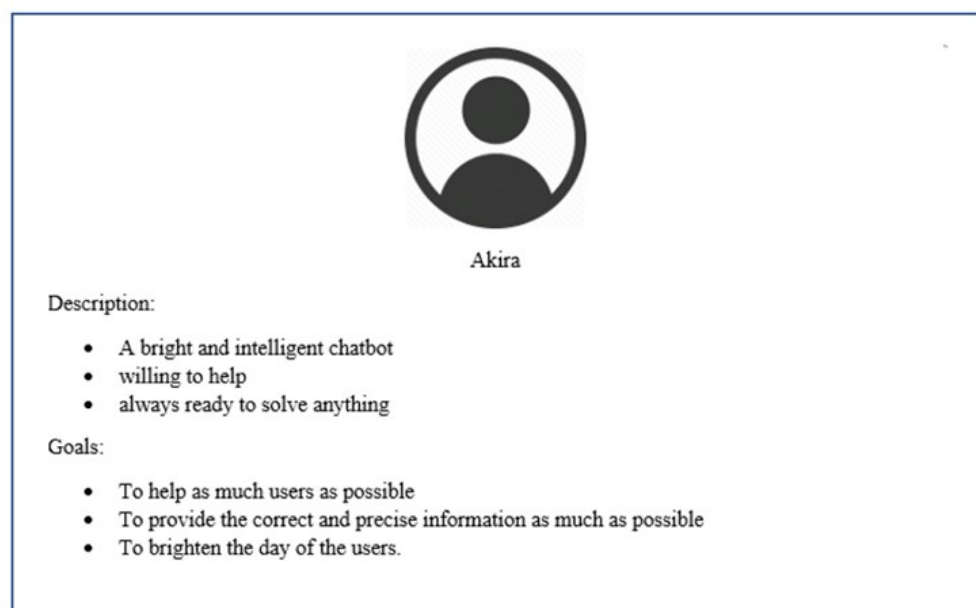


Figure 2: Profile of conversation AI agent.

2.5. Machine Learning

The conversational agent in rule-based method is based on what it says on the rules. These rules are created manually so its act as a pattern for the input request, it would work on a "if-then" logic, an action or a response would be triggered by it. The development of a chatbots started from analyzing the input response based on the decomposition rules which are triggered by the keywords in the responses [3] An example provided by the conversation agent "ELIZA" is that when the user asks, "It seems that you like me". ELIZA can only recognize 2 words from it which are "you" and "me", but it does not know what the sentence "It seems that" and the word "like" means. Based on the rule of decomposition, "0 YOU 0 ME" (where 0 stands for indefinite number of words) and the rule of reassembly "WHAT MAKES YOU THINK I 3 YOU" (where 3 stands for the third component of the subject decomposition, here is "like"), ELIZA will then response with, "What makes you think I like you" [29]. It allows users to believe that they are chatting with a human but in truth they are not. A rule-based chatbot can process faster as it does not rely on supervised learning, therefore allows the quick response towards the user's questions and

requests. It does not require the agent to go through training as well because it is not relying on supervised learning algorithms which requires it to be taught beforehand. However, even if it could be fast and does not require training, it still cannot answer questions which are outside of its rules or complicated questions in general. The conversational agent tends to lack the ability to imitate the human-like behaviour.

In this development, supervised learning-based neural network approach is used. A neural network learns to perform tasks by analysing training examples, usually the examples are pre-labelled in advance. In this case, for the conversational agent, it would be the intents and the labels inside the intents that are being pre-labelled, or hand labelled before training it in the neural network [7]. Akira model is based on Deep Neural Network (DNN) where it consists of 4 layers of neural network as shown in Equation (1):

$$x_i^{(l+1)} = f^{(l+1)}\left(\sum_j w_{ij}^{(l)} x_j^{(l)} + w_{i0}^{(l)}\right) \quad (1)$$

where l denotes the layer number e.g., $l=1,2,3$, and $x_j^{(l)}$ is the j -th input at the l -th layer, $w_{ij}^{(l)}$ is the weight of the l -th layer connecting the j -th node to i -th node in the next layer, $w_{i0}^{(l)}$ is the bias term of the l -th layer connecting to i -th node and $f^{(l)}(\cdot)$ is the activation function of the l -th layer. Two types of activation function are used given by the rectified linear units (ReLU) and the softmax function as shown in Equations (2) and (3), respectively, as follows:

$$\text{ReLU:} \quad f_{\text{ReLU}}(x_j^{(l)}) = \max(0, x_j^{(l)}) \quad (2)$$

$$\text{Softmax:} \quad f_{\text{softmax}}(x_j^{(l)}) = \frac{\exp(x_j^{(l)})}{\sum_i \exp(x_i^{(l)})} \quad (3)$$

For hidden layers, ReLU is used i.e., $f^{(2)}(\cdot) = f^{(3)}(\cdot) = f_{\text{ReLU}}(\cdot)$ while for the output layer, the softmax is used i.e., $f^{(4)}(\cdot) = f_{\text{softmax}}(\cdot)$. The input to the DNN i.e., $x_j^{(1)}$, is the bags-of-words (BoW) [30] which are computed from the tokenized words after the stemming and tokenization. BoW may not be enough to make the conversational agent intelligent as it is still incapable of understanding complex input and produces unreliable results and questionable outputs where else, if BoW is used with DNN, this allows the agent to be capable of understanding the input and be able to output a more reliable and precise result. Alternative to BoW is to use Word Embedding. However, if the dataset is small and context is domain specific, BoW may work better than Word Embedding. Context is very domain specific and one may not find corresponding vector from pre-trained word embedding models. Thus, BoW may be preferred in such situation. The DNN works by going into the intent files and getting the tag, responses and patterns from the intent's files and from there, it trains the model. In order for it to increase its accuracy and decrease the rate of errors, the batch normalization and dropout of 20% have been used. The proposed architecture is as follows:

- 1st layer: input layer with number of nodes equals to the dictionary dimension in the BoW
- 2nd layer: hidden layer with ReLU activation function, batch normalization, dropout rate set to 20%

- 3rd layer: hidden layer with ReLU activation function, batch normalization, dropout rate set to 20%
- 4th layer: output layer has 44 nodes corresponding to the total number of intents, softmax activation function

Thus, the DNN is developed as the intents learning model. In the 2nd layer, the number of neurons is twice the number of the input nodes. This is to enable a nonlinear mapping (i.e., features embedding) on the input to the high dimensional space. Hence, the outputs of 2nd layer are the nonlinear features of the BoW (or Word Embedding). The 3rd layer and the 4th layer have the same number of nodes but with different activation functions. Thus, the 3rd layer performs a nonlinear mapping from the high dimensional features space to the classifier space in the 4th layer, where the softmax (multi-class) classifier is utilised. The presence of the 3rd layer is necessary to realise a nonlinear classifier if we treat the outputs from the 2nd layer as features. The DNN is trained using the adaptive learning rate optimization algorithm (ADAM) [31]. The algorithm leverages the power of adaptive learning rates methods to find individual learning rates for each parameter. It also has advantages of Adagrad [32], which works really well in settings with sparse gradients, but struggles in non-convex optimization of neural networks, and RMSprop [33], which tackles to resolve some of the problems of Adagrad and works really well in on-line settings. Adam can be looked at as a combination of RMSprop and Stochastic Gradient Descent with momentum. It uses the squared gradients to scale the learning rate like RMSprop and it takes advantage of momentum by using moving average of the gradient instead of gradient itself like stochastic gradient descent with momentum. Cross-entropy with L_2 norm regularization is used as the loss function to optimize the DNN as shown in Equation (4):

$$L = \frac{-1}{K} \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^K \left[d_i(n) \log x_i^{(4)}(n) + (1 - d_i(n)) \log (1 - x_i^{(4)}(n)) \right] + \frac{\lambda}{2N} \|w\|^2 \quad (4)$$

where $d_i(n)$ are the i -th intent of the n -th training sample, $x_i^{(4)}(n)$ is the corresponding i -th output of the DNN, w is the vector concatenation of all the weights in the DNN, and λ is the regularization parameter which controls the DNN from overfitting the data points. The regularization parameter can be determined using the Bayesian evidence framework or Variational Bayesian approaches [34-38]. Overfitting the DNN on the training dataset is minimized by the use of the weights regularization and cross validation framework [39,40]. Training of the DNN has been conducted over a dataset of more than 1,000 intent instances. The dataset is split into 80:20 for training and testing phases. In addition, a 10-fold cross-validation has been adopted [41,42]. Various ablation studies have been performed to determine the optimal performance of the DNN. Batch sizes between 8 to 32 and allowing the DNN to train for 2000 epoch have been explored. Table 2 shows the test accuracy of various configurations in terms of the number of neurons in the 2nd layer and 3rd layer. These are the results after 10-fold cross-validation and test with unseen dataset. The best performance is attained with 128 neurons in the 2nd layer and 64 neurons in the 3rd layer. The table also clearly show that larger networks i.e., DNN with more than 128 neurons in the 2nd layer and more than 64 neurons in the 3rd layer tend to overfit although the overfit has been somehow ameliorated by a certain extent due to the dropout and regularization. However, once the networks grow overly large, then even the dropout is not useful. This can be seen in the lower right hand corner of the table where the test accuracy hovers in the region between 60%-69%.

Table 2:

Test accuracy under various configuration of the DNN.

		Number of neurons in 3rd layer							
		16	32	64	96	128	160	192	256
Number of neurons in 2nd layer	16	62.1	65.2	71.5	72.2	72.7	79.2	73.7	70.3
	32	66.2	66.4	70.8	71.8	71.4	77.2	69.7	69.2
	64	64.1	68.2	75.2	73.9	77.2	77.2	68.7	66.5
	96	67.2	68.9	83.8	86.7	74.7	72.2	69.7	64.7
	128	72.8	77.3	90.6	85.6	82.4	71.5	70.3	68.4
	160	71.9	71.1	87.1	86.4	69.9	69.2	68.2	61.8
	192	68.7	69.7	77.2	77.3	65.9	65.6	62.1	60.9
	256	63.5	67.8	70.2	67.8	65.3	62.9	62.8	59.1

The DNN has also been compared with other statistical models, including multinomial logistic regression [43], linear support vector machine (SVM) [44], SVM with Gaussian kernel [45] and Extreme Learning Machine (ELM) [46]. Table 3 shows the evaluation on the test data using metrics such as accuracy, sensitivity, specificity, and area under receiver operating curve (ROC). The table strongly indicated that DNN is more superior than other models in all of metrics.

Table 3:

Performance metrics of models.

MODELS	ACCURACY (%)	SENSITIVITY	SPECIFICITY	ROC AREA
Multinomial logistic regression	72.1	0.711	0.682	0.795
Linear SVM	78.8	0.897	0.574	0.815
SVM (Gaussian kernel)	86.1	0.825	0.828	0.837
Extreme Learning Machine	82.5	0.795	0.811	0.811
DNN	90.6	0.878	0.899	0.913

3. Results and Analysis

In order to obtain as many intents as possible to predict what the user might input or ask when interacting with Akira, a survey has been conducted to determine what type of questions a user would ask on a normal visit to a doctor. A group of participants were recruited, and each participant will list down a minimum of 5 questions. Only the ones that consented towards the privacy issues were collected, if not, they are destroyed and remove from the database. Besides, some of the responses were void because some participants did not take the survey seriously and provided an invalid response towards the question asked, so the results need to go through another filter to check for more or any invalid responses. Table 4 shows the intents and examples of patterns of sentences captured during the questionnaires. In total, there are 44 intents which are spread over the following domain questions: (i) COVID-19, (ii) disease (cold and flu), (iii) mental health, (iv) sexual health, (v) abortions, (vi) allergens, (vii) drugs and medicines, (viii) others. The proportion of these questions are shown in Figure 3. These were obtained from a survey of consisting of 57 participants in the UK whose age is 18-50 years old and consisting of 24 male and 33 females. These participants were recruited in March 2020 for a period of 4 weeks, identified

themselves between moderate to good health, and had previous experience in using conversational AI or willing to use it.

After inputting new intents according to the participant's response, we need to determine the response for Akira. Since providing correct or precise information is part of the requirements, so research has been carried out to ensure that all responses given by Akira has to be precise or accurate in order to not misinform the users and cause misunderstanding between Akira and the users. WebMD [47] or UK National Health Service (NHS) [48] are used as references for Akira's response. Both are reliable health website which is widely approved by others as well and Akira will refer towards these websites if the user wants to know more about it. Questions about mental health has been quite prevalent during the pandemic. Therefore, some forms of support are still provided towards the users as mental health is complex and it requires a professional to properly diagnose the symptoms or the condition the user's might be in. Therefore, Akira would only provide support such as talking to someone or giving a supportive message towards the user.

Table 4:
Intents and examples of patterns used in Akira.

Number	Intents	Examples of patterns
1	greeting	"Hey", "Is anyone there?", "Hello", "Hi"
2	reply_greetings_good	"I feel great", "I'm fine. Thank you", "I feel amazing", "I'm feeling wonderful", "I'm alright", "I'm ok", "Im good", "Im great"
3	reply_greetings_bad	"I dont feel fine", "I'm not ok", "I dont feel great", "Not really"
4	goodbye	"Bye", "See you later", "See ya", "Nice chatting with you", "Till next time", "Have a nice day"
5	thanks	"Thanks", "Thank you", "That's helpful", "Awesome, thanks", "Thanks for helping me"
6	understood	"I see", "I understand now", "Huh, ok"
7	greetings_to_bot	"How are you?", "How do you do?", "What's up?"
8	creator	"Who made you?", "Who created you?"
9	name	"what is your name", "what should I call you", "Who are you?"
10	usability	"What do you do?", "How does it work?"
11	question	"Can I ask you a question?", "I would like to ask you a question"
12	ask_more_question	"I would like to ask you another question?", "Can I ask you one more question?"
13	diseases_question	"I feel sick", "I feel unwell", "I don't feel fine"
14	disease_cold	"I feel cold", "My body hurts", "I'm sweating a lot", "I feel like I have a fever", "I have a mild fever"
15	disease_cold_symptoms	"Yes, I do", "I can't smell and taste as well", "My nose is always stuffy", "I sneeze a lot"
16	disease_cold_symptoms_2	"My throat feel sore", "I can't smell anymore", "My nose is always stuffy", "I sneezed a lot", "difficulty breathing"
17	disease_flu_symptoms	"Yes, I'm experiencing high fever", "My fever is longer than usual", "My body hurts a lot"
18	mental_health	"I feel sad", "I feel down", "depression", "Can't sleep", "feel like killing myself", "insomnia", "anxiety", "binge eating", "Bipolar disorder", "eating disorders", "OCD", "I hurt myself", "self-harm"
19	safe_sex	"How to practise safe sex?", "Do I have to wash my vagina before sex?", "Is it safe to have sex during my period?", "How do I

Number	Intents	Examples of patterns
20	contraception	practice safe sex?" "Are there anyways to prevent pregnancy?", "How to prevent from getting pregnant?", "How to avoid getting pregnant?", "How to practice safe sex", "Does condoms prevent pregnancy?", "I'm allergic to latex, what should I do?"
21	pregnant	"How do I know I'm pregnant?", "My period is late, am I getting pregnant?", "I feel nausea these days", "I have morning sickness and I feel like vomiting", "I do not take my contraception constantly, what are the chances of getting pregnant?"
22	abortion	"Where can I get an abortion?", "Is it ethical to have an abortion"
23	abortion_more_info	"Am I legal to get an abortion", "Does it hurt to have an abortion?", "When is it too late for an abortion?", "Is it safe to have an abortion?"
24	STI	"My penis is itchy. What should I do?", "My vagina smells really bad", "Why does my vagina smells bad", "Why is it painful when I pee", "My discharge smell really bad", "There are spots on my penis", "It hurts when I ejaculate"
25	consent	"How do I ask for consent?", "Is consent important when doing sexual intercourse", "What happens if he or she didn't give her consent?"
26	victim	"I have been violated", "I have been hurt", "I feel victimised", "I feel abused"
27	emergency	"I need help", "Help me", "911", "999"
28	allergens_mild	"I have rashes", "itchy nose", "runny nose", "red eyes", "itchy eyes", "wheezing"
29	allergens_severe	"I have swollen face", "My body hurts a lot", "skin feels very itchy", "severely itchy"
30	drugs	"Is it legal to take weed?", "Is it dangerous to take drugs?", "How often can I take marijuana?"
31	drugs_penalties	"What are the laws for drugs?", "What happens if I were to be found with drugs?", "Is it legal to take drugs", "Is it legal to sell drugs?"
32	medicine	"Can I still purchase medicine?", "What medicine can I take?"
33	medicine_painkiller	"Is it safe to take paracetamol/aspirin/painkiller/medication?", "What are the side effects of taking paracetamol/aspirin/painkiller/medication?"
34	drugs_penalties	"What are the laws for drugs?", "What happens if I were to be found with drugs?", "Is it legal to take drugs", "Is it legal to sell drugs?"
35	corona_info	"What is a coronavirus?", "What is COVID-19?", "What is SARS-CoV-2?"
36	symptoms_corona	"What are symptoms for coronavirus?", "What are symptoms for new coronavirus?", "What are symptoms for COVID-19?"
37	get_infected_corona	"I got COVID-19, what should I do?", "I got infected with the new coronavirus, what should I do?"
38	spread_corona	"How does coronavirus spread?", "How does the COVID-19 spread to others?", "How does the COVID-19 transmits to others?"
39	symptoms_less_infect	"Can a person who shows no symptoms of COVID-19 infect others?", "Can COVID-19 be caught from someone who shows no symptoms?", "Can I catch COVID-19 from someone who seem healthy?"

Number	Intents	Examples of patterns
40	corona_vaccine	"Is there any vaccine that are being developed right now for the corona virus?", "Are there going to be vaccine for it soon?", "Can I prevent myself from getting Covid-19?","Can I vaccinate myself from this disease?"
41	pet_corona	"Can my pet catch the new coronavirus?", "Can my dogs catch the new COVID-19?", "Can my pet catch the COVID-19?"
42	corona_travel	"How do I do my grocery shopping?", "How do I go out and buy the necessity?", "How can I protect myself when I go out?"
43	Helpline_corona (context coronavirus)	"What/Which number to call?", "Numbers to call", "How to get help?", "Suggest help"
44	Helpline_general	"What/Which number to call?", "Numbers to call", "How to get help?", "Suggest help"

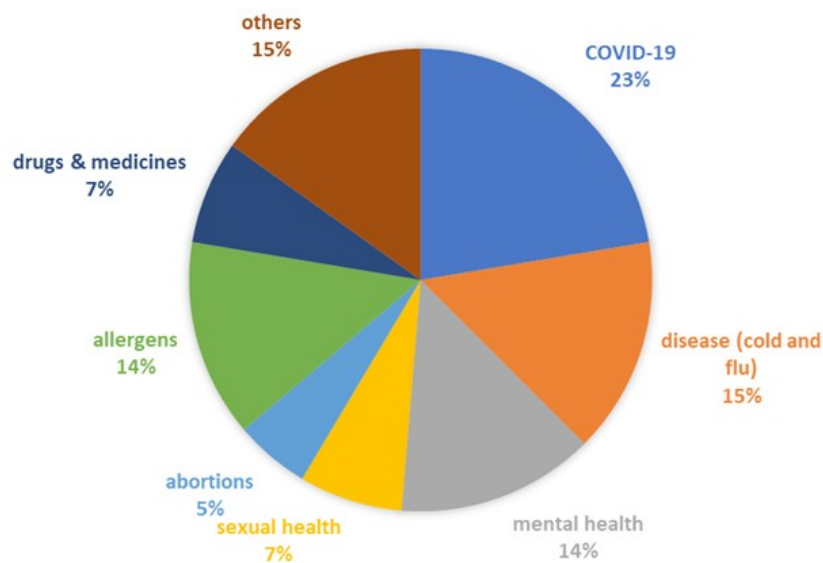


Figure 3: Domain questions asked in the questionnaires.

As shown in the Figure 3, there are questions directed towards sexual health, which is fairly normal and important as well to take care of personal hygiene and understand the circumstances towards the actions that was taken. Akira has been programmed with some intents that are related to sexual health such as sexual transmitted diseases, hygiene, menstrual cycle, safe sex and pregnancy.

To investigate the performance of Akira, we will proceed in several test. In the first test, we show the case where Akira is not trained by the DNN. Hence, Akira is not able to understand what the user was inputting or anything that the user inputted, and it always pass to the else statement in the chat () function. The else function handles the errors or intentions where Akira could not understand, or if it too complex that Akira cannot identify the intent, or if the intent is not coded in the intent files. If Akira receives intents that are unrelated then it will prompt the message saying "Sorry, Akira doesn't quite understand what you just typed. Try asking a more specific question or

other questions then maybe Akira could understand. Sorry for the inconvenience.” Figure 4 shows a snippet of the code developed using Python.

```
def chat():
    print("Start talking with Akira! She won't bite! :) (type exit to stop)")
    while True:
        inp = input("You: ")
        if inp.lower() == "exit":
            break

        results = model.predict([bag_of_words(inp, words)])[0]
        results_index = numpy.argmax(results)
        tag = labels[results_index]

        if results[results_index] > 0.7:
            for tg in data["intents"]:
                if tg['tag'] == tag:
                    if 'context_set' in tg:
                        context = tg['context_set']
                    if not ('context_set'== context):
                        responses = tg['responses']
                        print(random.choice(responses))
            else:
                print("Sorry, Akira doesn't quite understand what you just typed, try asking a more specified question or another questions then maybe Akira could understand. Sorry for the inconvenience cause.")
```

Figure 4: A snippet of the chat function code.

Figure 5 shows what happens when the Akira is not trained beforehand and the response were all saying that she could not understand what the user is saying. This is because there is nothing to match or compare it with since there is not any module that Akira can use it to predict with. After Akira is trained properly with the training dataset, it goes through another test run to check if Akira could read the intentions and understand what the user meant. The second test is successful as it can read the intent of the user and reply with the user’s desired answers. Figure 6 shows the example of Akira responding the user’s desired answers.

Akira: Start talking with Akira! She won't bite! :) (type exit to stop)

You: Test

Akira: Sorry, Akira doesn't quite understand what you just typed. Try asking a more specific question or another questions then maybe Akira could understand. Sorry for the inconvenience.

You: hello

Akira: Sorry, Akira doesn't quite understand what you just typed. Try asking a more specific question or another questions then maybe Akira could understand. Sorry for the inconvenience.

Figure 5: Example of Akira with non-trained AI.

```

Akira: Start talking with Akira! She won't bite! :) (type exit to stop)
You: Hey
Akira: Hi, you alright?

You: im great
Akira: That great to hear. Nice.

You: how are you?
Akira: Akira feels great, thanks for asking.

You: Who made you
Akira: I'm made by a guy called X-Men and he is my master and my creator.

```

Figure 6: *Example of working output.*

Several more tests have been conducted to test whether Akira can handle whatever input that she was given and be able to produce a response for it, not just that, a test on its usability and a scenario test case to see how Akira would response to that scenario, and would it be able to reach its final task. The third test is the grey-box testing where the inputs are chosen by the tester and it is to see whether it would still function the way it is. The goal of this test is to test the error handling capability and make sure Akira will return the error handling message for intents that she could not understand and the first test was successful as it was not able to understand the input of an URL and the input of a mashed-up words which does not make any sense.

Figure 7 shows that Akira is in fact able to prompt the error messages towards intents or input that does not make sense to her. During the grey-box testing, Akira is not able to prompt the error handling messages when it is provided with an intent where she could not understand. Since she could not understand the intent, she decides to take the closest intent and use the response from that specific intent. In this case it is the intent "safe sex". The input that is fed into Akira is the URL "<https://hackr.io/blog/types-of-software-testing>" and it returns as a response that is from the intent for "safe sex" which is rather unexpected as it is not near the type of input that the user would respond. So, in order confirm whether it is replying that specific intention is because it consists the word "test" in one of the pattern in the intent or that if it does not understand any intent it would prompt a response from that intent to the user. Before the fourth test is conducted, a few tweaks are added into the intent file to prevent Akira from getting into that intent file if it is not related. The tweaks that are added is that one of the patterns in the "safe sex" intent that consists the word "test" has been removed temporarily and to make sure it goes smoothly on this run and to make the test consistent, the same input was used. This time, a different output is prompted and the response belongs to the intent "disease_cold_symptoms_2". This is shown in Figure 8. None of the words in the patterns tag contain the word "testing" in it.

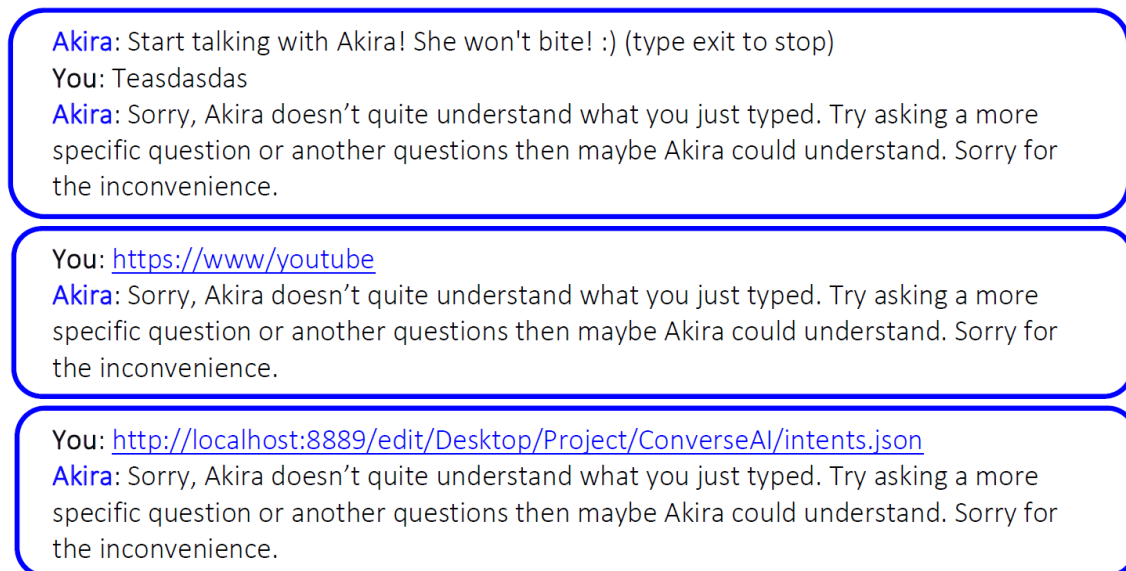


Figure 7: Akira's response to the unidentified queries.

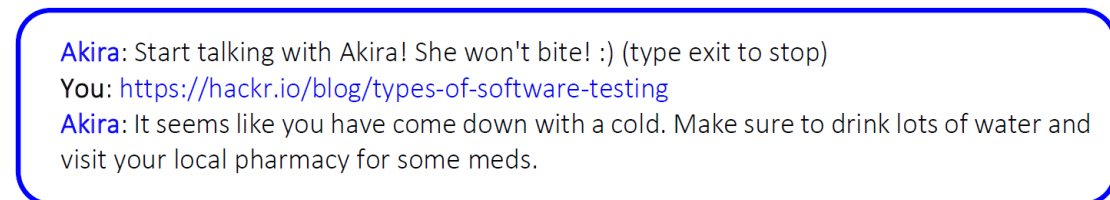


Figure 8: Akira's response to the URL.

While conducting the fourth test, another problem is encountered during the middle of the test. It was that Akira would break when the user enters nothing as in blank spaces created from the space button and just send it towards Akira just like that. The programme would display error messages when trying to find out the intent for it and associating it with a context. In order to work around this, Akira would identify this and respond with the following error handling message "Akira can't respond to you, if you didn't type anything, so type something!". Besides that, an addition of an if-else statement is added towards the chat () function to pass the input towards the else block if the input is not blank.

The objective of the fourth test is to ensure Akira produces the correct responses or the user's desired response when Akira is given an input. In order to test this, a cross reference has to be made between the intent file and the input data that has to be inputted. In other word, the input has to be one of the patterns in the intents that are located in the intent files, As long the input shares similar words distribution with the patterns that exist in the intent file and that the DNN has correctly identified the respective intent, then Akira will be able to issue the intended response.

Figure 9 shows that multiple intents have been tested for Akira and all of the responses were all correct. Therefore, Akira is able to correctly identify the user's intents and able to provide the correct response for each query. The fifth test examines the case where the users have a chat with Akira with a specific agenda in the mind, and the user would slowly prompt towards that question

if Akira is able to keep up with the conversation with the user. In this scenario, the user is trying to ask question regarding coronavirus, so Akira would need to be able to tackle the question that is asked and provide the user with a desired response. Figure 10 shows the conversation between the user and Akira.

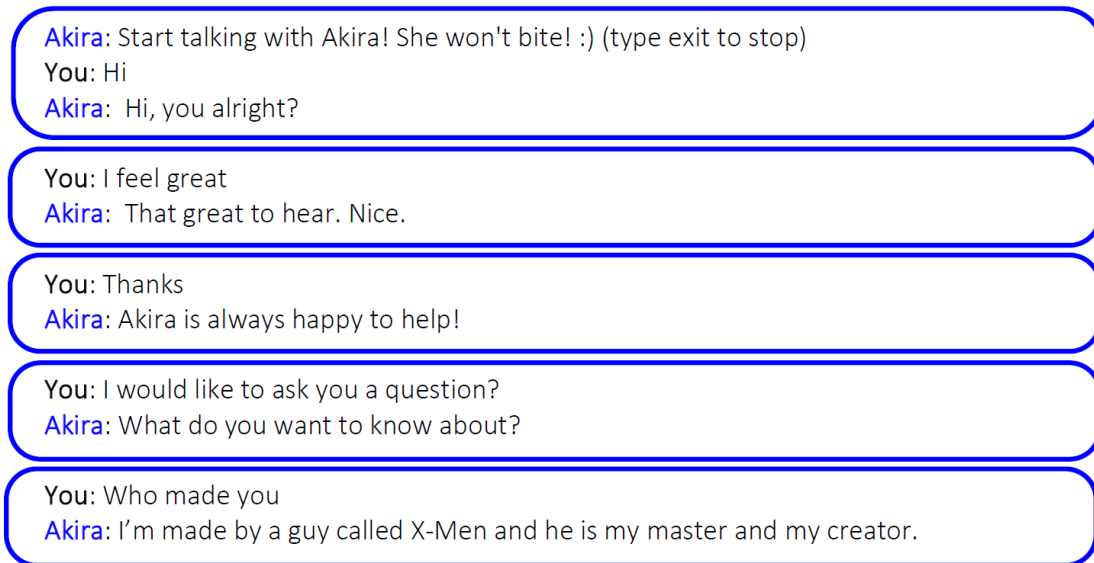
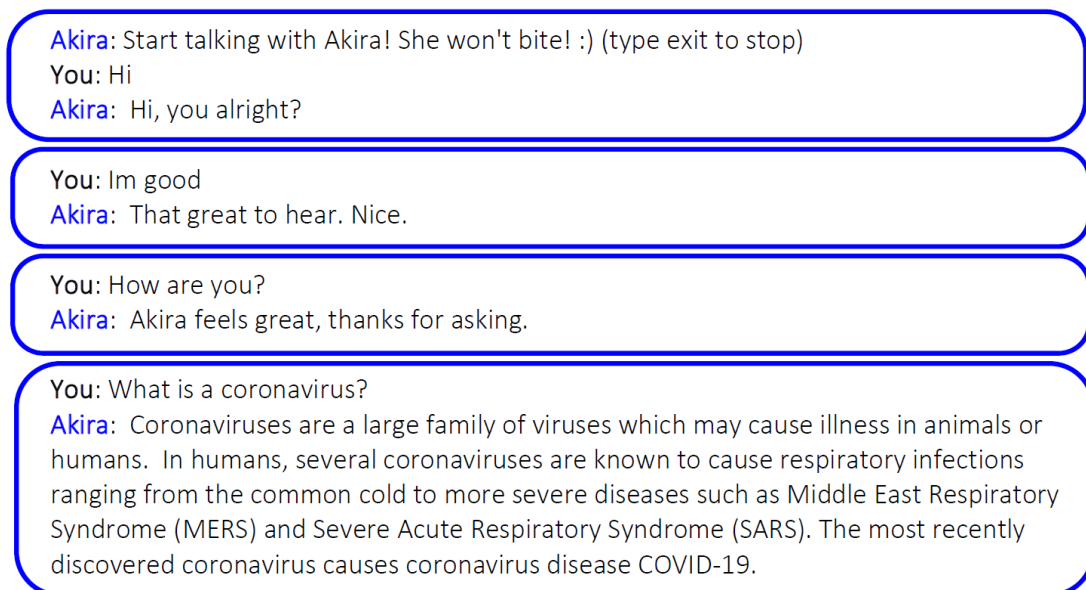


Figure 9: Akira's response to multiple intents.



You: What are the symptoms for the new coronavirus?

Akira: The symptoms take about a maximum of 3-14 days to show, but the most common symptoms of COVID-19 are fever, dry cough, and tiredness. Some patients may have aches and pains, nasal congestion, sore throat or diarrhoea. These symptoms are usually mild and begin gradually. Some people become infected but only have very mild symptoms. Most people (about 80%) recover from the disease without needing hospital treatment. Around 1 out of every 5 people who gets COVID-19 becomes seriously ill and develops difficulty breathing. Older people, and those with underlying medical problems like high blood pressure, heart and lung problems, diabetes, or cancer, are at higher risk of developing serious illness. However, anyone can catch COVID-19 and become seriously ill. Even people with very mild symptoms of COVID-19 can transmit the virus. People of all ages who experience fever, cough and difficulty breathing should seek medical attention.

You: im experiencing some fever.

Akira: Ohhhh dear ... it seems you might have a flu, you should visit your local pharmacy for medical attention immediately.

You: Should I take paracetamol?

Akira: Please refer to the website (<https://www.for-example-pharmacy-websites>) for further information. Please talk to a doctor or pharmacist before making any changes to your daily medications.

You: Can I ask more questions?

Akira: What do you want to know about?

You: How does coronavirus spread?

Akira: People can catch COVID-19 from others who have the virus. The disease spreads primarily from person to person through small droplets from the nose or mouth, which are expelled when a person with COVID-19 coughs, sneezes, or speaks. These droplets are relatively heavy, do not travel far and quickly sink to the ground. People can catch COVID-19 if they breathe in these droplets from a person infected with the virus.

You: I feel depressed to hear this.

Akira: I'm sorry to hear that, dude. I can recommend you to the website for more information on well-being care (<https://www.nhs.uk/conditions/stress-anxiety-depression/mental-health-helplines/>)

You: thanks

Akira: Akira is always happy to help!

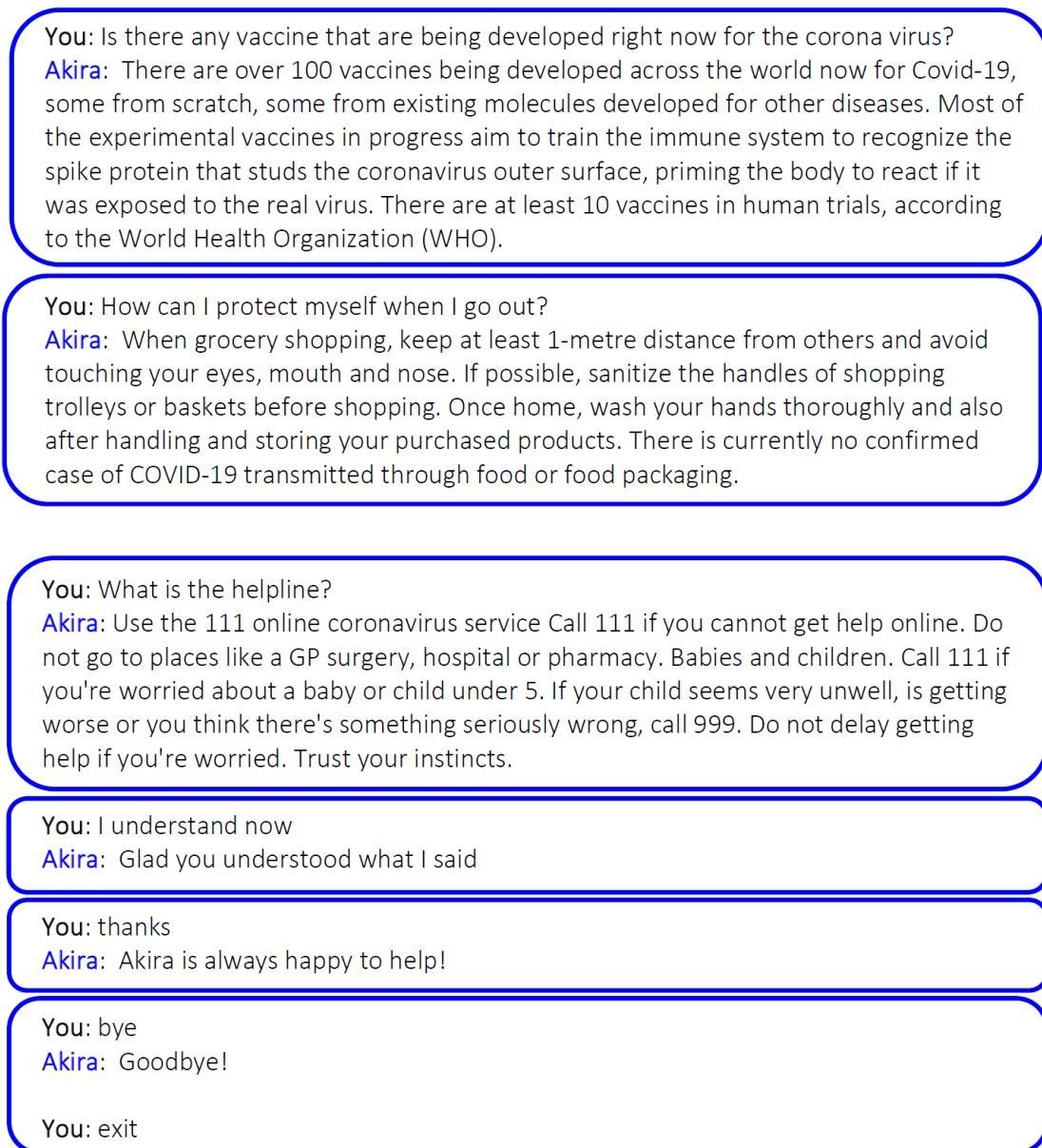


Figure 10: A snippet conversation of a user with Akira.

As shown in Figure 10, Akira is able to converse with the user over several topics ranging from the current coronavirus pandemic, flu symptom, and wellbeing, and in all cases the user is able to understand what Akira is responding, which allows the testing to meet its objective. The developed conversational AI agent is meant to support healthcare query during the pandemic and not meant as a casual chatter system. All the 57 participants were asked to score their satisfaction with system with five queries. User satisfaction is often subjective and in order to quantify this, a score is assigned corresponding to the experience. The score is ranged between 0 for none (which denotes no satisfaction), 1 for low, 2 for medium, and 3 for high (which denotes highly satisfied). Table 5 shows the score of user satisfaction after each query. The user satisfaction remains high for the first three queries. Notably, the average scores are 3, 2.91 and 2.82 for the first three consecutive queries. The user satisfaction drops to a score of 2.68 after the 4th query, and score of 2.56 after the 5th query. In terms of the overall user satisfaction with the developed conversational AI agent, out

of 57 participants, 82.5% rated with a high score, 15.4% rated medium, 1.4% rated low, and only 0.7% rated none. This is shown in the last row of Table 3. The mean of the average scores of the system is 2.74, which shows that users are highly satisfied.

Table 5:

Users satisfaction score for the first five queries.

	Range of satisfaction				Average score
	None ('0')	Low ('1')	Medium ('2')	High ('3')	
After 1st query	0	0	0	57	3.00
After 2nd query	0	0	5	52	2.91
After 3rd query	0	0	10	47	2.82
After 4th query	0	2	14	41	2.68
After 5th query	2	2	15	38	2.56
Average number of participants	0.4 (0.7%)	0.8 (1.4%)	8.6 (15.4%)	47 (82.5%)	

4. Discussions

Akira is capable of talking like a human with the help of the persona and the way its response to inputs but they are definitely plenty of room for improvement ranging from the AI model towards the way it responds towards user. First, the response given by Akira can be simpler and shorter as to not overwhelm the user with too much information as once, just enough information would be great and practical. However, Akira has the ability to understand the context of the conversation and reply based on the context. At the moment, she is incapable of understanding emotions which make her less intelligent and cannot fool the user into thinking that Akira is a bot and not a human. Since understanding emotions requires a deeper understanding in analysing the input and a much intelligent model to handle that input and be able to generate weights on each of the token and direct the conversation without any help from the intent's file.

4.1. User Audience and Experiences

As the user audience are meant to target people from as young as 18 to people as old as 50 which are capable of using a keyboard and a mouse but currently, there is only a small portion of data from participants from 18-20. The data could be rather unreliable or rather targeted towards people from that range of age, then that means it is no longer targeted towards users as young as below 18 or users older than 50. The survey has only been conducted with limited number of participants as a start and it is intended to obtain more inputs from more participants so that the results or the question or the intents will have more diversity and increase its users age range.

Akira is aiming to be a general health conversational agent, therefore it is good to increase diversity to tackle health problems that the users might have from all ages.

Akira could play an important role for users who are rather shy and introvert. This could be their go-to programme or app to ask question regarding their health or their body's condition as people tend to surf the web regarding the symptoms that they are experiencing and finding out causes before consulting a doctor for medical attention. A survey was conducted, and it has shown 72% people has searched for health-related information online which increases the importance for the user to use this while feeling comfortable and it does not intrude their personal space. At the same time, when people google their symptoms or health-related issues online, some information might be misleading, sometimes even false, so it is important to credit the source of information coming from. In this way, the information given by Akira will not be misleading, so users can form a trust and allow them to ask more question in the future. Besides, it is important to assure the users that it is important to stay calm and tell the users how much attention is need on their health issues that they are experiencing as overloading them by telling them that there is a chance of death could cause health anxiety and they would start buying health products which might not be good for them in some cases. So, it is important to assure the users with just enough information so it will not burn them out mentally. On the note of user friendliness, colours were added to Akira response so it would allow the users to read better and to distinguish between user's query and Akira's response.

4.2. Intents and Response System

The intents currently exist in Akira are relatively small and they are specific to support healthcare query during the pandemic COVID-19. It currently captures the common queries based on the undertaken survey. Akira needs more improvement if she wants to grow. For example, the response system could be improved as it heavily relies on the users input to be similar as the ones pre-existed in the intent files. Due to the fact that some of the response are very long, the user might not read it all and just loss interest after all due to the fact that there was so much to read [49]. Besides, it is possible to introduce options which the user can select to act as a guide to provide support to the users so that Akira could decrease the amount of information that needs to be enter and at the same time it creates user friendliness. In addition, it decreases the length of the response from Akira as user can freely choose if they want to know more about it or just the very basic version of it, which would make this a rather engaging conversation and allow the user to diverse into other topics to talk or discuss about as well.

Context is important and it allows the designer to control over what happens to the conversation. It defines the conversation flow and the next steps of the chat. Additionally, the context fallback is triggered when the query cannot be matched. The common type of context is situational. For example, intention #43 and #44 have the same patterns except that the former is only activated when the detected context of the conversation is related to coronavirus. Otherwise, intention #44 will be activated. Also, once the context is identified, the response will be clamped on the topic which is the class of intents associated with the context. For example, intents #1 to #10 are related to casual conversation, intents #13 to #17 related to common disease symptoms, intents #19 to #26 related to safe sex and abortions, intents #28 to #34 related to medicines and allergens, and intents #35 to #43 associated with the topic of coronavirus.

4.3. Ethical Issues

Issues regarding trust among the AI and privacy of data has been quite the topic of discussion among the recent years. One of the objectives of the conversational agent is to act humanlike, but by doing that it will cause some distrust among the users who interacts with it since the users believe it might be an actual human being behind the screen chatting with them, but in fact, it is just a computer programme. With the recent Google Duplex reveal, Google demonstrated a call between the chatbot and the person in the saloon. During the demonstration, the chatbot sounded human like by using “uhm” or the word “mm-hmm”, besides that, the ability of it to response quickly after the person has finished their sentence or has asked a question. Designer of the conversational agents will have to keep ethics in mind while they are designing the chatbot because the users has concern for their privacy and the data that are collected.

Chances of agent being corrupted by unethical users are minimal with rule-based agents since the responses are all programmed. With supervised learning-based agents which require large pool of dataset, the developer can exercise a limited control over the responses of the agent. The advantages for this are that the responses are provided by the users if there is a trusted group of users where the agents can be train with. If not, the agents can develop an inconsistent personality and designer has no knowledge of what it is being taught. At worst, it turns into a Hitler loving, racist, sexist, homophobic piece of nasty software that swears a lot. e.g. Microsoft’s Tay chatbot in 2016 [50,51].

4.4. Security Concerns

Conversational AI has security concerns as it is a very attractive target towards exploiters because they contain sensitive information about the users and for industrialised chatbots, they handle information such as credits/debit cards, SSN, bank accounts and others Sensitive PII (Personally Identifiable Information). Just like any new technologies, its benefits come with a huge responsibility and cyber-security risks. When the system is not well designed, there is a high chance that conversational agents can be compromised and become vulnerable to attacks. There are various ways to get into the system through the conversational agents, first is the spoofing when the exploiter illegally accesses and uses another user’s credentials. Besides, the data can also be tampering as well which means changing or modifying the data with malicious intents which could lead to victims entering sensitive information and lead to information disclosure. Information disclosure happens when the exploiters expose the stolen data into the deep web. Denial of service takes place when the exploiters send large number of packages towards the server to overload the system. Lastly, the exploiters can aim to gain the privilege levels in the system through the conversational agents, if they do, they could gain access to administration pages and steal database information. The types of threats that could be a concern towards the chatbot are spoofing, tampering, repudiation, information disclosure, denial of service, and elevation of privilege. It is important to mitigate or prevent the damages that could be done by the exploiters before it happens. Information that was expose or personal information that are stolen are usually threatened to be sold by the exploiters into the deep web [52].

5. Conclusion

This paper presents the initial development of an integrated and human-interactive conversational AI system called Akira. It is not designed as a diagnostic or clinical decision-making tool but an assistant to doctors to support healthcare queries during the pandemic COVID-19. A deep learning model has been trained, and Akira could converse with the user ranging from the conversation over 7 topics related to COVID-19, common cold and flu, mental health, sexual health, abortions, allergens, drugs and medicine. The paper has also presented the importance of designing an interactive human-user interface when dealing with conversational agent. It has discussed the context of ethical issues and security concerns when designing the agent. The conversational agent is demonstrated to answer queries from a pool of 57 participants. Further work is needed to improve the capability of the conversational agent. Firstly, there is need to include more diseases into our system. Secondly, a symptom synonym thesaurus related to both pandemic coronavirus and common flu should also be produced. In this way, questions with more understandable symptoms for the patients can be generated. Although the trained deep neural network has been trained using bag-of-words due to the relatively small size of the dataset, as future work, we will experiment with transformers and BERT-like language representation as the conversation agent for healthcare query support.

Conflict of interests: The authors declare that there is no conflict of interest.

References

1. Woo WL. Future trends in I&M: Human-machine co-creation in the rise of AI. *IEEE Instrum Meas Mag.* 2020;23:71-3.
2. <https://www.singlegrain.com/artificial-intelligence/why-chatbots-are-a-must-have-for-businesses-and-how-to-build-one/>
3. Weizenbaum J. ELIZA-a computer program for the study of natural language communication between man and machine. *Commun ACM.* 1966;9:36-45.
4. Bassett C. The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI Soc.* 2019;34:803-12.
5. <https://blog.ubisend.com/discover-chatbots/chatbot-eliza>
6. <https://arxiv.org/abs/1609.06479>
7. Wallace RS. The Anatomy of A.L.I.C.E. In: Epstein R, Roberts G, Beber G (eds) *Parsing the Turing test.* Springer, Dordrecht. 2009;pp.181-210.
8. <https://www.pocket-lint.com/apps/news/google/137722-what-is-google-assistant-how-does-it-work-and-which-devices-offer-it>
9. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Ment Health* 2017;4:e19.
10. <https://arxiv.org/abs/1709.02349>
11. Zhou L, Gao J, Li D, et al. The design and implementation of xiaoice, an empathetic social chatbot. *Comput Linguist.* 2020;46:53-93.
12. <https://arxiv.org/abs/1910.03042>
13. <https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7>

14. Shang L, Lu Z, Li H. Neural responding machine for short-text conversation. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China. 2015.
15. <https://arxiv.org/abs/1506.05869>
16. Hommersom A, Lucas PJF, Velikova M, et al. MoSHCA-my mobile and smart health care assistant. 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom), Lisbon, Portugal. 2013.
17. Koh BHD, Woo WL. Multi-view temporal ensemble for classification of non-stationary signals. IEEE Access. 2019;7:32482-91.
18. Elmasri D, Maeder A. A conversational agent for an online mental health intervention. International Conference on Brain and Health Informatics. 2016.
19. Kim BK, Roh J, Dong SY, et al. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. J Multimodal User Interfaces. 2016;10:1-17.
20. Ni L, Lu C, Liu N, et al. MANDY: Towards a smart primary care chatbot application. In: Chen J, Theeramunkong T, Supnithi T, Tang X (eds) Knowledge and Systems Sciences. KSS 2017. Communications in Computer and Information Science. Springer, Singapore. 2017.
21. Hildebrand J, Hemati W, Mehler H. Recognizing sentence-level logical document structures with the help of context-free grammars. Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France. 2020.
22. Hung BT. Domain-Specific versus general-purpose word representations in sentiment analysis for deep learning models. In: Satapathy S, Bhateja V, Nguyen B, Nguyen N, Le DN. (eds) Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing. Springer, Singapore. 2020.
23. Jabbar A, Iqbal S, Tamimy MI, et al. Empirical evaluation and study of text stemming algorithms. Artif Intell Rev. 2020.
24. Tarasiev A, Filippova M, Aksyonov K, et al. Using of Open-Source Technologies for the Design and Development of a Speech Processing System Based on Stemming Methods. In: Ivanov V, Kruglov A, Masyagin S, Sillitti A, Succi G. (eds) Open Source Systems. IFIP Advances in Information and Communication Technology, Springer, Cham. 2020.
25. Porter MF. An algorithm for suffix stripping. Program-Electron Lib. 1980;14:130-7.
26. Fryer LK, Coniam D, Carpenter R, et al. Bots for language learning now: Current and future directions. Lang Learn Technol. 2020;24:8-22.
27. Hill J, Ford WR, Farreras IG. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. Comput Hum Behav. 2015;49:245-50.
28. Fryer LK, Nakao K, Thompson A. Chatbot learning partners: Connecting learning experiences, interest and competence. Comput Hum Behav. 2019;93:279-89.
29. Peng Z, Ma X. A survey on construction and enhancement methods in service chatbots design. CCF Trans Pervasive Comp Interact. 2019;1:204-23.
30. Gidaris S, Bursuc A, Komodakis N, et al. Learning representations by predicting bags of visual words. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. 2020.
31. <https://arxiv.org/abs/1412.6980>

32. Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Technol.* 2011;12:2121-59.
33. Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks For Machine Learning. 2012;4:26-31.
34. Titterton DM. Bayesian methods for neural networks and related models. *Statist Sci.* 2004;19:128-39.
35. Sulistyono SB, Wu D, Woo WL, et al. Computational deep intelligence vision sensing for nutrient content estimation in agricultural automation. *IEEE Trans Autom Sci Eng.* 2018;15:1243-57.
36. Al-Tmeme A, Woo WL, Dlay SS, et al. Underdetermined convolutive source separation using GEM-MU with variational approximated optimum model order NMF2D. *IEEE-Acm T Audio Spe.* 2017;25:35-49.
37. Gao B, Woo WL, Dlay SS. Variational regularized 2-D nonnegative matrix factorization. *IEEE Trans Neural Netw Learn Syst.* 2012;23:703-16.
38. Gao B, Woo WL, Dlay SS. Single channel blind source separation using EMD-subband variable regularized sparse features. *IEEE Trans Audio Speech Lang Process.* 2011;19:961-76.
39. Luo Q, Gao B, Woo WL, et al. Temporal and spatial deep learning network for infrared thermal defect detection. *Ndt&E Int.* 2019;108:102164.
40. Fei Y, Ye P, Qiu F, et al. Crack characterization in ferromagnetic steels by pulsed eddy current technique based on GA-BP neural network model. *J Magn Magn Mater.* 2020;500:166412.
41. Mojarad S, Venturini B, Fulgenzi P, et al. Prediction of nodal metastasis and prognosis of breast cancer by ANN-based assessment of tumour size and p53, Ki-67 and steroid receptor expression. *Anticancer Res.* 2013;33:3925-33.
42. Mojarad SA, Dlay SS, Woo WL, et al. Cross validation evaluation for breast cancer prediction using multilayer perceptron neural networks. *Am J Eng Appl Sci.* 2011;4:576-85.
43. Dzhamtyrova R, Kalnishkan Y. Universal algorithms for multinomial logistic regression under kullback-leibler game. *Neurocomputing.* 2020;397:369-80.
44. Chang YW, Lin CJ. Feature ranking using linear SVM. *JMLR Workshop and Conference.* 2008;3:53-64.
45. Chamasemani FF, Singh YP. Multi-class support vector machine (SVM) classifiers -- an application in hypothyroid detection and classification. *Sixth International Conference on Bio-Inspired Computing: Theories and Applications, Penang, Malaysia.* 2011.
46. Huang GB, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern. Part B.* 2012;42:513-29.
47. <https://www.webmd.com/>
48. <https://www.nhs.uk/>
49. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit Med.* 2020;3:65.
50. Jameel T, Ali R, Toheed I. Ethics of artificial intelligence: research challenges and potential solutions. *3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan.* 2020.
51. Bostrom N, Eliezer Y. The ethics of artificial intelligence. In: Frankish K, Ramsey W(eds). *Cambridge Handbook of Artificial Intelligence.* Cambridge University Press. New York. 2014;pp.316-34.

52. Wei B, Hamad RA, Yang L, et al. A deep learning driven light-weight phishing detection sensor. *Sensors*. 2019;19:4258.