

Received March 3, 2020, accepted March 31, 2020, date of publication April 16, 2020, date of current version May 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988252

Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot

ERIC HSIAO-KUANG WU¹, (Member, IEEE), CHUN-HAN LIN¹, YU-YEN OU²,
CHEN-ZHONG LIU³, WEI-KAI WANG¹, AND CHI-YUN CHAO¹

¹Department of Computer Science and Information Engineering, National Central University, Taoyuan City 32843, Taiwan

²Department of Computer Science and Engineering, Yuan Ze University, Taoyuan City 32003, Taiwan

³Graduate Institute of Network Learning Technology, National Central University, Taoyuan City 32843, Taiwan

Corresponding author: Eric Hsiao-Kuang Wu (hsiao@csie.ncu.edu.tw)

This work was supported by the ministry of science, Taiwan, under Grant 108-2221-E-008-034.

ABSTRACT E-Learning has become more and more popular in recent years with the advance of new technologies. Using their mobile devices, people can expand their knowledge anytime and anywhere. E-Learning also makes it possible for people to manage their learning progression freely and follow their own learning style. However, studies show that E-Learning can cause the user to experience feelings of isolation and detachment due to the lack of human-like interactions in most E-Learning platforms. These feelings could reduce the user's motivation to learn. In this paper, we explore and evaluate how well current chatbot technologies assist users' learning on E-Learning platforms and how these technologies could possibly reduce problems such as feelings of isolation and detachment. For evaluation, we specifically designed a chatbot to be an E-Learning assistant. The NLP core of our chatbot is based on two different models: a retrieval-based model and a QANet model. We designed this two-model hybrid chatbot to be used alongside an E-Learning platform. The core response context of our chatbot is not only designed with course materials in mind but also everyday conversation and chitchat, which make it feel more like a human companion. Experiment and questionnaire evaluation results show that chatbots could be helpful in learning and could potentially reduce E-Learning users' feelings of isolation and detachment. Our chatbot also performed better than the teacher counselling service in the E-Learning platform on which the chatbot is based.

INDEX TERMS E-Learning, chatbot, isolation, detachment, retrieval-based model, QANet.

I. INTRODUCTION

A. DEVELOPMENTS AND PROBLEMS OF E-LEARNING

With the advancement of modern technologies, people's lives are more convenient than ever before. Education and ways of learning have also begun to change with the development of new technologies that improve learning efficiency and effectiveness. One such example of these new technologies is the Summit Learning Platform [1], an online learning tool that helps students set and track learning goals at their own pace.

E-Learning is gaining popularity and is widely used in today's modern education. In [2], many higher learning institutions, such as universities, are using E-Learning platforms because younger students are familiar with these new

technologies and are potentially attracted by new types of learning methods.

The approaches in [3] and [4] design a new platform and system with new technologies involved, e.g., Semantic Web, the Hinting Computer E-Learning System, to enhance learning.

Small and portable devices, e.g., smartphones and tablets, nowadays pack enough computing power to perform any kind of application. Plus, these smart portable devices are more widespread than ever before. This helps E-Learning technologies shift toward smart devices and become more accessible to more users. Docebo's E-Learning trend whitepaper [5] shows that people rely on their smartphones for everyday tasks and carry them at all times. The whitepaper also shows that mobile learning markets are experiencing strong growth, signifying that E-Learning is shifting toward mobile platforms and becoming the new trend. In addition to smart devices, VR, AR and Intelligent Assistants are the recent new

The associate editor coordinating the review of this manuscript and approving it for publication was Chongsheng Zhang.

trends in E-Learning developments [6]. These new technologies create new ways for us to explore and learn.

E-Learning has a great amount of benefits over traditional classroom learning. The approach in [7] informs us that students can learn and study at anytime, anywhere as long as they can access the E-Learning platform, which is now almost certain since mobile learning is growing in popularity. In contrast, traditional learning methods in school classrooms are limited by the fact that students can only learn from the teacher at a certain place and time. Additionally, with E-Learning platforms, such as the Summit Learning Platform [1], students can manage their learning progression and courses freely to suit their personal learning style, which improves learning efficiency and engagement. In traditional classroom learning, it is impossible to design a course that suits every student's learning progression and style.

However, E-Learning does have several defects and limits. Some studies are focused on improving the learning efficiency of E-Learning. The approach in [8] proposes a multi-agent system to improve the learning process and provide more efficient knowledge acquisition. The approach in [9] uses a gamification method in their E-Learning environment. By implementing elements from gaming, e.g., points, competition, and levels, this system can increase students' motivation to learn. Some studies are focused on user feelings and experiences. The studies of [10] and [11] show that some E-Learning platforms, e.g., VOD streaming or open online courses, designed for individual learning may cause students to experience feelings of isolation and detachment. These feelings often occur due to lack of interaction with other students or educators. Also, the user interface (UI) of these platforms is only designed with learning functionality in mind and is not sensitive to student responses. These feelings may reduce the motivation to learn, which we will discuss in Related Work.

B. DEVELOPMENTS IN ARTIFICIAL INTELLIGENCE AND CHATBOTS

Artificial Intelligence (AI) has been a hot topic in recent years and more applications can now be achieved with AI-powered computers. AlphaGo [13] is an example of AI advancement that amazed the world by defeating the best human GO player. This was previously regarded as an impossible feat for computers. In [14], Benedict du Boulay discusses using AI in the education field to assist in learning. In [15], Carlos Ramos *et al.* talk about an ambient intelligence system powered by AI that could improve our everyday life. Google is also taking its AI-powered personal assistant [16] to a whole new level in the 2018 Google I/O [17]. Google announced Google Duplex [18] as Google assistant's new feature. Google Duplex could help users book or reserve services via telephone. This system will talk to a human clerk or receptionist to help the user complete a booking or reservation tasks just as a human assistant would. These AI advancements show that early rule-based dialogue or conversational systems like Eliza and Alice mentioned in [19]

could be greatly improved by machine learning. Apple's Siri [20], Amazon's Alexa [21] and Microsoft XiaoIce [22] are also examples of AI-powered conversation systems and personal assistants. There are implementations of Artificial Intelligence and Machine Learning Model in other fields as well. Authors in [23] designed a new brain-computer interface (BCI) called iAUI (intelligent adaptive user interface). In iAUI, they implement a recurrent quantum neural network (RQNN) to help filter out noise from the EEG signals in a real-world environment, thus improving signal quality.

Apart from the personal assistants or conversational systems mentioned above, there are more general conversation systems called chatbots. In comparison to those personal assistants, most chatbots are more function-centric and provide specific services. For example, Forbes Asia Bot [24] provides flash news and news subscription services for people to receive the latest news from their messaging app, and, TaxiGo Bot [25] helps users call a taxi. These chatbots are convenient and easy to use since they use human-like conversation and are based on messaging apps with which we are already familiar.

In the education field, there are some examples of chatbot implementation or tutoring systems. Got It Study [26] introduces a photo-analyzing and math-solving bot based on their online learning and helping services. The bot can help users solve math problems and provide step-by-step instructions. Deakin Genie [27] is a Smart Personalized Digital Assistant Chatbot created by Deakin University. This bot utilizes AI to help students manage their courses and learning, create schedules, set reminders and inform students about any course-related information. Jill Watson [28] is also a teaching assistant built by Prof. Ashok K. Goel. Using the power of the IBM Watson [29] AI platform, it helps students solve problems on learning. AutoTutor [30] is an intelligent tutoring system that holds conversations in a natural way, just like a human. AutoTutor uses a virtual teacher or tutor to guide and instruct students while learning on the E-Learning platform. It also incorporates strategies from human tutors and uses voice and visual representation (3D character) to create the feeling that a real tutor or teacher is learning alongside with students.

C. PROPOSED GOAL AND ORGANIZATION OF THE PAPER

By understanding the development of E-Learning and the advancement of chatbots mentioned above, we want to explore and evaluate how a chatbot could perform as a learning assistant in an E-Learning environment. What are the advantages that a chatbot could bring to students? What are the constraints of using a chatbot as a learning assistant? We also want to see whether using a chatbot could reduce E-Learning issues such as feelings of isolation and detachment since texting with a chatbot is similar to many of our everyday interactions, i.e., texting. We designed a hybrid-model chatbot with the use of both retrieval-based and QANet models [31]. A retrieval-based model is aimed at solving specialized learning problems related to E-Learning courses,

TABLE 1. Translated example conversations with our chatbot.

User Input	Could you tell me what is the Pythagorean theorem?
Answer	The Pythagorean theorem is also known as Pythagoras' theorem. It states that the square of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the other two sides. Here is a tutorial video: ...
User Input	Do you have any tips to improve my concentration when studying?
Answer	First, delete all games and social media apps. Trust me! It really works :). Second, pay attention in class and correct the mistakes in your test. Third, go to sleep early and stop staying up late. You can't concentrate on learning if you are exhausted. Don't be like my friend who plays games all night, hits his bedroom door, and then falls asleep.
User Input	Can you suggest a good anime to watch?
Answer	Of course, here are some good anime: My Hero Academy S3, Hisone to Masotan and Golden Kamuy are some of the best anime this season.
User Input	What do you love to eat?
Answer	I love to eat cheesecake and drink bubble milk tea. What do you like to eat?

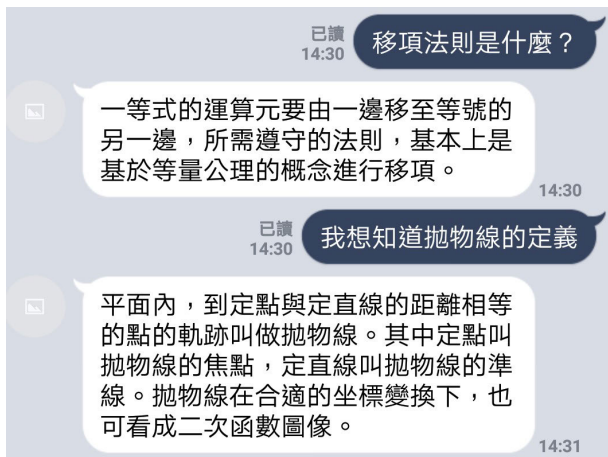


FIGURE 1. An example conversation (in Chinese) with our hybrid chatbot in a messaging app.

and a QANet model [31] is for general learning problems and casual chitchat or conversation with students that mimics a real learning mate or companion. This helps our chatbot respond more naturally and potentially reduces feelings of isolation and detachment that some E-Learning users experience. Fig 1 and Table 1 show some example conversations with our chatbot, which can answer math or general learning questions as well as engage in chitchat. Our hybrid chatbot design is focused on improving both learning efficiency and user experiences. Unlike most systems from other E-Learning studies which are implemented into E-Learning platforms, our chatbot is independent of the E-Learning platform. In other words, our chatbot is more versatile and can be adapted to different E-Learning platforms' learning courses or services without changing the design of the E-Learning platform itself. Our chatbot could also be used outside of E-Learning as a simple assistant or human-like companion chatbot.

Our goal is to build a hybrid-model chatbot learning assistant and explore its potential to improve user experience in E-Learning and reduce users' feelings of isolation and

detachment while increasing their learning motivation (the same benefits provided by learning in a conventional school setting). We also want to understand the constraints of using a chatbot as a learning assistant. We will compare our chatbot with a teacher counselling service from the E-Learning platform on which our chatbot is based to see how well our chatbot performs.

The rest of the paper is organized as follows: section II discusses related works about the advantages of chatbots, E-Learning platforms our system is based on, isolation and detachment issues and research solutions of these issues. The benefits and shortcomings of some machine-learning models used in chatbots will be discussed in section II. In section III, we will first discuss the neural network model and the QANet model used in our system. Then, we will describe our system's architecture and user scenarios. Finally, we will present the training details and testing datasets used for the QANet model. In section IV, the experiment and questionnaire design will be shown and the results will be discussed. Lastly, in section V, we will provide the conclusion of the performance, advantages and constraints of our designed chatbot and discuss future improvements and works.

II. RELATED WORK

A. E-LEARNING ISSUES AND SOLUTIONS

In the studies of [10], [11], and [12], the authors discuss the benefits and problems of E-Learning. They also provide approaches for addressing the issues of isolation and feelings of detachment that some students experience.

E-Learning courses are more student-focused, while a traditional classroom is more teacher-focused. The benefit of a student-focused platform or course design is that students have full control of their learning progression. The learning time and learning environment can also be adjusted freely by students. This makes learning more efficient. However, some students may experience feelings of isolation and detachment while taking E-Learning courses. The reason this may happen is because students are separated from other students

and teachers in E-Learning platforms. There is no human interaction between the learner and other students or teachers while learning. Most operations of E-Learning platforms are based on mouse or touch input which gives students the sense of controlling machines, not interacting with fellow humans. According to [12], students who experience feelings of isolation and detachment may lose their motivation to learn and potentially drop out from their E-Learning courses. The approach in [32] also confirms that students taking courses without any face-to-face contact with other students or tutors could lead to feelings of frustration. Ultimately, the student may quit due to loneliness and a lack of direction and help.

In [10], Montebello create a web 2.0 platform to let users share, collaborate and source information, ideas and knowledge. This creates indirect interaction between users. In [11], Al-Samarraie *et al.* re-designed the UI of their E-Learning platform. Their platform can dynamically change the layout and course content to suit the users' need. It also provides the user with learning suggestions and tips. Users can have a fresh experience every time they use this E-Learning platform, which helps reduce feelings of isolation and detachment. Both [10] and [11] use methods to address isolation and detachment issues, but these methods still lack direct human interaction, e.g., face-to-face conversation or other conversation in the form of audio, video or text. We think having a conversation is the best way to reduce isolation and detachment since this behavior occurs naturally in a traditional classroom, where students can create interpersonal connections with other students or teachers. Mourad *et al.* in [12] combine a course management system and a video conference system to create a virtual classroom for students to discuss and share ideas. The virtual classroom concert performs very well at reducing feelings of isolation and detachment, but there is a downside. One benefit of E-Learning is the option to take courses and learn freely at any time; video conferencing will limit this benefit since not everyone learns at the same time.

B. ADVANTAGES OF CHATBOTS AND BASIS OF DEVELOPMENT

Chatbots are easy to use and convenient. The way we interact with chatbots is similar to how we interact with each other. Having a conversation with a chatbot is just like having a conversation with our friends through devices like computers, smartphones or other technology, so it is pretty easy to get started. There are some advantages of using a chatbot when compared to human services, such as telephone or online customer services. A chatbot service is more efficient and can respond to customers 24/7, which means it is not limited to normal working hours, unlike humans. Also, a chatbot service can handle multiple customer requests at once which is impossible for a single human. Utilizing a chatbot to provide customer service could drastically reduce human labor costs and the stress of human services, improving customer experiences. Many companies have started implementing services such as customer QA and online booking. Social media platforms are also developing chatbot integration on existing

messaging platforms, such as Facebook Messenger [33], Telegram [34]. Some tech companies are also releasing bot-building platforms and services, such as MS Bot Builder [35] and Chatfuel [36].

Our chatbot is developed and built on the E-Learning platform of [37], [38] and [39]. This platform features VOD course streaming for K-12 students, learning records, learning notes, live handwriting note functionalities, and more. It also provides a custom messaging app for use among students, parents, teachers and counsellors for conversation and counsel. This allows students to seek help through conversation with teachers or counsellors while learning. Aside from texting in the messaging app, the E-Learning platform also provides a telephone service for students to speak to teachers. These are great student services that can help reduce feelings of isolation and detachment. However, just like in [12], the services are limited by the time and resources the teacher or counsellor has available since they can only provide these services at certain times to a limited number of students. We designed our chatbot with these learning-assisted services in mind and to also act as a personal companion with no time limitations. It can handle multiple student interactions at once as well. Although our chatbot is superior to human teachers and counsellors in multiple ways, we have chosen to integrate the services of teachers and counsellors from [37], [38] and [39] into our chatbot. Reference [40] suggests that human operators are able to appropriately manage unexpected situations and enrich the content of social interaction between robots and humans. Even though a chatbot can answer a lot of questions, the design corpus of a chatbot does have limitations. To further improve the user experience of our chatbot, we designed a way for students to switch a conversation from the chatbot to a conversation with a human teacher or counsellor inside the same dialog. In the evaluation section, we will compare the effectiveness of our chatbot to the teacher services of [37], [38] and [39] in the same E-Learning platform.

We also considered the work of [41] when designing our chatbot's response corpus and dialogs. Authors of [41] investigate the conversations between multiple chatterbots and human users. They found that asking chatbots "are" and "where" questions resulted in higher response satisfaction levels while other interrogative-style input questions such as "why" do not. By understanding the work of [41], we could enhance our corpus to answer more questions of "are" and "where," but in our targeting domain such as education, there are more questions in the interrogative-style input. This factor could affect our chatbot's satisfaction results, and we have taken extra care when designing the response corpus for these types of questions.

C. MODELS USED IN CHATBOT

Current commercial chatbots typically use two types of models to create responses [42]: a retrieval-based model and a generative-based model.

A retrieval-based model [43] is a common commercial chatbot choice since it is easier to develop and maintain, e.g., Pizza Hut's pizza ordering bot [44] and the CNN News bot [45]. This model is essentially a matching process between the input question and an output answer. We can use cosine similarity or a trained CNN, LSTM model to calculate the matching probability between a question and an answer. All bot responses are maintained and saved in databases. When an input question comes in, this model will calculate the matching probability of the input question and all responses in the database. The higher the matching probability is, the more probable that the response is right. Google's automated email reply suggestion system [46] is powered by the same type of retrieval-based model. This model is great for a QA system that needs precise and accurate responses since responses are maintained in databases. The quality of responses can also be ensured. However, this model suffers from low coverage in responses like [47] because natural language is complex. It cannot answer a question or input sentence that is not in the coverage of pre-defined responses. Building a database that contains all kinds of topics, answers and responses is incredibly time consuming, hard to maintain and difficult to cover all possible responses in natural language.

Unlike a retrieval-based model, a generative-based model [48] does not store pre-defined responses in a database. It creates a response by the model itself so it can use it in applications of open or wide coverage responses like chitchatting. The most used generative-based machine learning model is sequence-to-sequence models. Sequence-to-sequence models, known as seq2seq, consist of one encoder and one decoder that are based on a neural network model. In the training state, the encoder receives word embedding vectors from the question string as its input, and the decoder receives word embedding vectors from the response string as output. By pairing the encoder's vector result to the decoder's input, we can train the model to create responses by its trained decoder. However, this model is hard to build compared to the retrieval-based model since it needs a huge selected dataset to train the model. In [49], Hu *et al.* built a tone-aware chatbot for customer care or services with a generative-based model. The conversation datasets they used were collected from Twitter and 1.5 million conversations were selected in training. Although the training results in [49] show that its chatbot can produce great responses, the helpfulness of the responses still fall behind that of a human agent. The quality of the responses produced by the model is also harder to control since the quality is related to the selection of training datasets.

Due to the issues outlined above, we turned toward the QANet model [31]. QANet [31] was proposed by the Google Brain team and CMU. QANet was originally designed for reading comprehension and question answering. It uses the information learned from an article to answer questions about the article. Basically, the response from the model is a piece of content from the article. The advantages of the QANet

model over a traditional RNN-based generative model is that training time of the QANet model is significantly less, and the responses of the model can be easily modified without re-training since it retrieves the response from an input article. This model is perfect for our chatbot because our target users are K-12 students. The quality of response content is important and needs to be designed carefully. It is hard to collect huge amounts of conversation data that is also appropriate for our target users. With QANet, we do not need to worry too much about the content of training datasets and can focus on improving the QANet model and design of our own conversation responses. Although QANet is not designed for chitchat or casual conversation, we see great potential for QANet to be used in this kind of everyday conversation. We implemented it into our chatbot to show it is possible to use QANet as a conversational model. [49] shows that talking style plays an important role in human conversation, so we aimed to build our QANet model and response content to be as joyful to chat with as possible. We evaluated EM/F1 scores with SQuAD standardize datasets from [50] and [51], placing performance test of response context design and training time of our implemented QANet model in section IV.

Haptik [52] proposed a novel idea that combines both a graph-based model (modified retrieval-based model) and a generative-based model to create a hybrid system as a personal assistant. In each conversation, a user's question input will first be processed by the graph-based model and carry out name entities recognition. If the graph-based model is unable to answer the user's question (out of domain, mixed domain, spelling error), the input question will be sent to and handled by the generative-based model. The generative-based model is trained with conversation data from user chats between human assistants or their graph-based model. The main purpose of their generative-based model is to create better chatting experiences for users without responses like, "Sorry, we cannot answer your question," and "I don't understand your question," which are common answers when the retrieval-based model fails to match a response to the question. If the generative-based model is still unable to answer the question, an output threshold mechanism determines whether the user's question should be redirected to a real human assistant.

Haptik's proposed system [52] is a great concept that uses the generative-based model to complement the retrieval-based model since it is limited to a designed domain range. However, the retrieval-based model still generates better and more maintainable answers.

Our proposed hybrid system has a similar concept as [52], but the models used in our system are focused on and designed for different purposes. We use a retrieval-based model as the E-Learning assistant to answer student questions with specialized, accurate, professional and maintainable E-Learning course-related responses created by us with the help of K-12 teachers. Unlike a human learning tutor, our chatbot is more of a learning assistant and companion, since we designed our bot to not only answer E-Learning questions,

but to also perform functions like weather forecasts, study recommendations, learning progression records and course notification, just like a personal assistant, e.g., Apple's Siri [20]. The retrieval-based model focuses on improving the learning efficiency of E-Learning and the other everyday tasks of a personal assistant. The QANet model is mainly used and trained for general answers and everyday chitchat which make it feel just like having a classmate or friend at school. This diminishes the impression of texting with a bot and further reduces feelings of isolation and detachment. In paper [53], Jean Chagas Van *et al.* improved their robot design to suit the needs of a robot companion. These robot companions are designed to interact with target users who have little time to be with other humans. Since human interaction can have a significantly positive affect on an individual's feelings, we designed our QANet to focus on chatting with users and acting as a learning companion to reduce feelings of isolation and detachment.

We designed and built these models with different purposes to suit their individual properties and achieve the best results from each model. However, in order to get the best result, we must correctly classify a user's question into the appropriate model and match it with appropriate responses. This is the challenge of our hybrid-model system. We discuss our system design and user-question classification in section III and evaluate the performance of our hybrid chatbot in section IV.

III. DESIGN AND IMPLEMENTATION

In this section, we discuss background knowledge of the machine-learning model we used. Then, we present our system's architecture and workflow. Finally, we discuss the training data, preprocessing, and parameters used in our model.

A. BACKGROUND

GloVe, an abbreviation of "Global Vector," is an unsupervised learning algorithm for obtaining vector representations of words [54]. In more colloquial terms, GloVe performs translations, converting the word into a vector that the computer can understand. Interestingly, when GloVe is translating, it can also understand concepts within the words. Here is a classic example of GloVe: When we think of England, we may think of its capital, London. GloVe has the ability to do this after training. What is even better is that its training time is very short, meaning that the practical value is very high.

In the GloVe training process, let us explain the concept of the above algorithm once again in spoken language. We now randomly extract two words from an article. During the training process, the context of these two words is imported into GloVe to determine the value of their vectors. Each word will have a vector value in a pre-defined N-dimensional vector space (we use 300 as the number of N in our training) and the same word will have the same vector value. During the training process, if the context of these two words is determined to be similar, then GloVe will adjust their vector values to bring them closer to each other in the vector space and

vice versa. Based on this concept, it can be inferred that the trained and completed GloVe model will have closer vector values in vector space for words having similar contexts. For example, France, Britain and Germany will have close vector values. Words with abstract concepts such as happy, joyful and cheerful will also have similar properties. After training is complete, we can compare two random words with their vector values to determine whether the contexts or concepts of these two words are similar.

In the preprocessing stage, we must segment Chinese sentences into words in order to turn them into word vectors. Chinese is much more complicated than English. Words are not separated by a blank space in Chinese, so we need a specialized process to segment sentences into words. We use Jieba Chinese text segmentation [55], an open source word segmentation module, with the Traditional Chinese dictionary in our system to help process user input.

In [31], Carnegie Mellon University and the Google Brain team proposed a reading comprehension and question answering model called QANet. Unlike other question answering models, instead of using recurrent neural networks (RNN), QANet uses convolution neural networks (CNN) and a self-attention mechanism. RNN-based models are often slow in training and inferring results due to their sequential nature. On a SQuAD dataset and evaluation [50], a machine reading comprehension test proposed by Stanford University [51], QANet performs similarly to other RNN-based models but with training and inference speeds up to 13 times and 9 times faster, respectively. QANet also currently holds the best performance model of SQuAD1.1 datasets with a 84.454 ExactMatch (EM) score and a 90.490 F1 score.¹ This proves that QANet is not only fast but also performs as well as other models. Another QANet quality that prompted us to implement this model in our chatbot is that it does not generate answers or responses directly from the model like most RNN-based models. QANet selects a section from articles of the reading comprehension test as the answer to a question. This means we give the QANet model an article that contains the content we want the chatbot to use for an answer. Based on the question, it will select an appropriate section (a few words or sentences) from the article we gave as an answer. This property is especially important to us. We not only want our model to be trained with as much data as possible to improve its performance, but we also want responses and answers that are appropriate and suitable for our target users, K-12 students. It would be very time consuming to preprocess all the training data to suit our needs and to train with an RNN-based model. Also, response results of an RNN-based model are far more difficult to control than QANet. With QANet, we can train it with as much data as we can and not to worry about the content of training data. We can focus more time on the design of article content for our chatbot's responses.

¹The performance of QANet model is referenced as of July 12th, 2018.

One problem with implementing QANet as our chatbot’s everyday conversation model is that QANet is designed for reading comprehension tests, not chatting. To address this issue, we have to design a way to provide QANet with the correct article, which is full of everyday conversation, to answer a user’s input. This is important since the length of chatting article is limited by the length of articles used in training. We use the same classification technique used in our retrieval-based model to choose the right article for the user’s input, then let the QANet model process the article and select an answer to respond. To the best of our knowledge, we are the first² to implement a QANet model in a chatbot designed for everyday conversation. The working flow of our hybrid chatbot is illustrated in System Architecture and Work Flow section below.

B. SYSTEM ARCHITECTURE AND WORKFLOW

First, we will discuss our system’s hardware layout and design. Then, we will walk through the workflow of our hybrid model system.

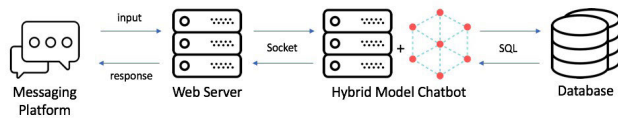


FIGURE 2. Hardware architecture of our proposed hybrid chatbot.

Fig 2 shows our proposed system’s hardware architecture.³ The design is described as follows: A web server is used to handle the social messaging platform’s API requirements and load balancing. After receiving the message, the web server will transfer it to our hybrid-model chatbot via socket. The chatbot queries the appropriate response and sends it back to the web server and social media platform. Meanwhile, the chatbot server will save the dialogue (user’s message and chatbot response) into the database for future improvement. The database also contains the response material: retrieval-based and QANet model.

Fig 3 shows our proposed system workflow. The workflow is described as follows: After receiving the user’s message, our system will first check whether it belongs to a specific topic of dialogue module defined in the retrieval-based model, and if so, it will enter the dialogue module of that specific topic. In the dialogue module, a program will collect information within the input message. If information in the user’s message is insufficient, the system will respond with a message to collect required information. If the user’s message is checked and no topic is matched, the input message will then be directed to the QANet model to process. We also provide a contact module for users to contact a real human teacher if needed due to limitations of our current build of

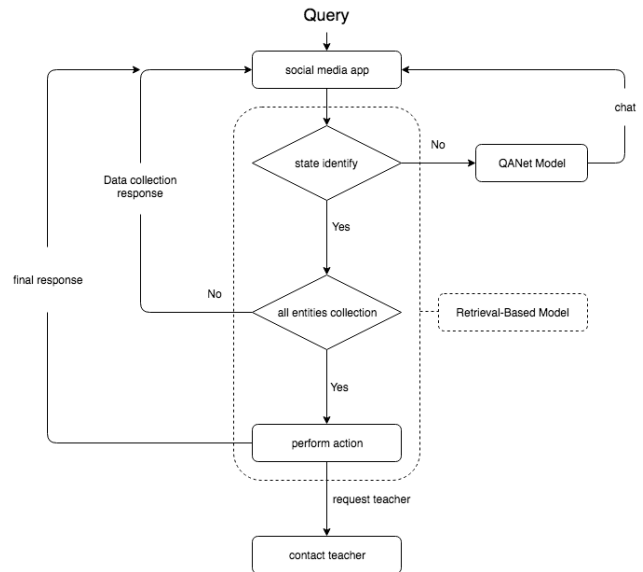


FIGURE 3. Workflow of our proposed hybrid chatbot.

dialogue modules. If a user asks the same question many times or asks a real teacher for help, our system will contact a teacher service to take over and reduce potentially negative experiences. The human teacher service used here is from the E-Learning platform on which our chatbot is based.

```

# main domain(topic) matching algorithm function
func mainMatch(self, input_sentence, threshold=0, root_node){
    let focused_rule be [], result_list be [], at_leaf_node be False

    #set root's children nodes as rules
    set focused_rule to [self.rules[root_node]]

    # start Loop from root node, end at leaf node
    while at_leaf_node is not True do
        set at_leaf_node to True

        # get word_vector_match result for each domain nodes
        # word_vector_match function will return max similarity value,
        # matched domain term, matched input word
        get [max_sim, domain_term, match_word] as result from word_vector_match function,
        for each rule in focused_rule

        append each result to result_list

        # get the highest max_sim domain node
        let top_domain_node be the domain node from result_list with the highest max_sim

        if self.rules[top_domain_node] has child node then
            set result_list to [], at_leaf_node to False

            #set best matched domain's children nodes as new rules
            set focused_rule to [self.rules[top_domain_node]]
        end if
    end while

    return result_list with the best matched domain
}

# word vector matching function
func word_vector_match(self, input_sentence, threshold=0){
    let max_sim be 0.0

    get maximum word_vector_similarity_value larger than threshold as max_sim,
    between each word in sentence and each domain_term in rule

    return max_sim, domain_term, match_word
}
    
```

FIGURE 4. Topic-matching pseudo algorithm.

Fig 4 shows the topic-matching algorithm used in the dialogue module of our retrieval-based model. This algorithm will match the user’s input to the best possible topic by word

²The claim here is based on our research as of July 12th, 2018.

³ Icons in Fig 2 are made by Good Ware, Alfredo Hernandez, Smashicons from <http://www.flaticon.com/>.

vector comparison and a confidence index threshold. With topic matching, we can design and implement different tasks and responses for different categories of user input.

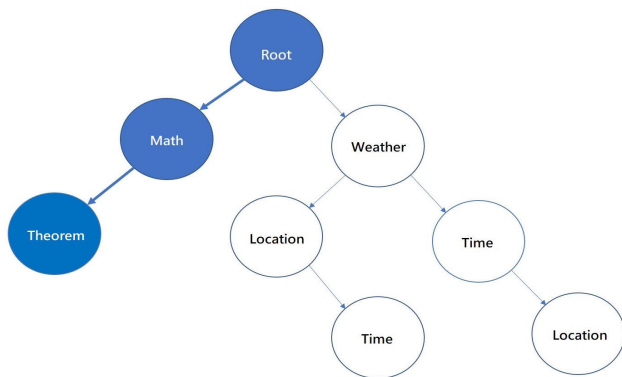


FIGURE 5. Topic-matching flow.

Fig 5 is the topic-matching flow in the dialogue module of our retrieval-based model. It shows the tree structure of our pre-defined topics. For example, if the user's input is about math, it will first be categorized as a Math topic. Then, more detailed math topics will be compared. If the input is about math theorems, it will be categorized as a theorem topic.

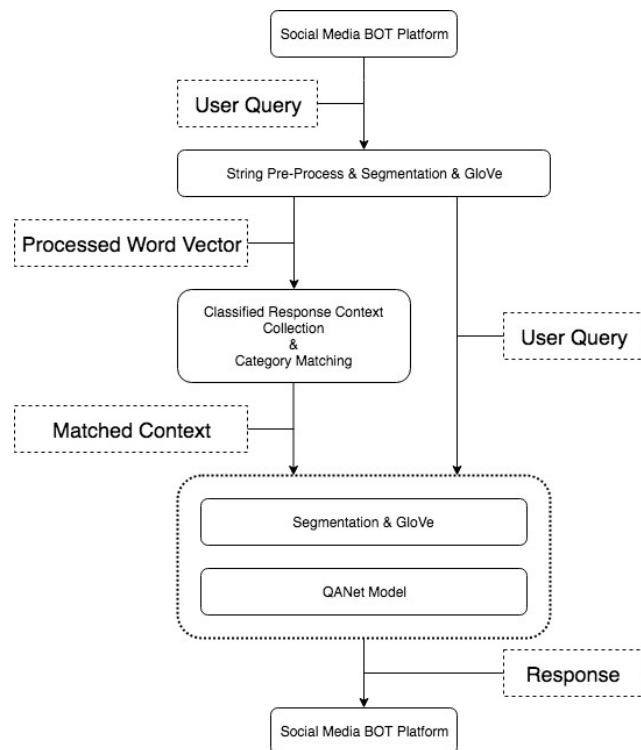


FIGURE 6. Workflow of the QANet model in our chatbot.

Fig 6 shows the QANet model and our article classification module in detail. First, a user's query is preprocessed and segmented. Words are changed into word vectors by GloVe used in our retrieval-based model. Then, word vectors are

compared with keywords from different topics of response contexts (article), e.g., suggestions for anime, songs, food, etc. After one best-matched keyword matches the user's query, the matched response context of that keyword and user query will be sent into QANet. QANet will select a sentence or words from the response context (article) as the response to the user.

C. DATA, PREPROCESSING AND TRAINING

In this section, the template of retrieval-based model, training data of the GloVe model and the QANet model will be described. Templates retrieval-based model contain pre-defined dialogue topics such as greetings, weather forecasts, study recommendations, course materials, and more.

The GloVe model used in the retrieval-based model and the QANet model is trained with data collected from the Chinese Wikipedia and textbooks published by the Ministry of Education. The data is comprised of 3 million articles from Wikipedia and 1,000 definitions from the Ministry of Education's related course materials. We then use openCC to convert Simplified Chinese into Traditional Chinese. After that, the data is processed by word segmentation and common words are removed. The training parameters are set as follows: size:300, alpha:0.025, window:5, and min_count:5.

Templates from the retrieval-based model contain professional, maintainable and E-Learning-related QA designed in collaboration with K-12 teachers from the E-Learning platform [37], [38], [39] on which our chatbot is based. We also provide direct web links to course videos for users alongside text responses. This helps improve learning efficiency without users having to search for course material on their own.

The QANet model is trained by the data collected and processed from the PTT forum and our conversations generated in-house. We also used translated SQuAD datasets [50] and translated conversation transcripts from the TV show, Friends. The translation is done by using Google Cloud translation API. SQuAD datasets provide a decent amount of data that fits the training format of QANet without extra processes. Since QANet's output response is not directly related to the content we used in training, we use as much Chinese data as we can in training to help the QANet model understand Chinese conversations. The total number of QA pairs currently used in training is around 60k and the number of articles is 20k. The training parameters are set as follows: batch size: 26; hidden layer: 94; training steps: 60000; article max length: 1000; Q&A max length: 100.

In our QANet's response context collection, we focused on generating relaxing, informative and entertaining responses to release stress and reduce feelings of isolation and detachment experienced by students. Knowing that our target users are K-12 students, we discuss the response contexts of QANet with K-12 teachers and make adjustments according to their feedback to ensure our chatbot responds with appropriate answers, since K-12 students could learn and imitate any responses from our chatbot. Other than that, the topics of response context suit the popular culture of K-12

students to make our chatbot a more attractive companion. As such, we defined topics of everyday conversations such as anime_suggestions, songs, food, anime_info_reviews, etc. Once all of the topics were defined, we then chose highly related keywords for each topic to be used to match word vectors of user input queries. After topics and keywords were defined, we started to generate common responses related to specific topics and consolidated those responses into a response context, which was an article consisting of conversations. Each response context is limited to a length of 1000 words to match parameters of the training set.

IV. EVALUATION

Our chatbot is trained using course materials from the chosen E-Learning platform. We mainly focused on Math concepts and Chinese History with some leisure and daily chat. Both Math and History account for 40 percent each of total data, and the rest are leisure and daily chat data. Two parts will be covered in this section. First, we describe the experiment design. Second, results of the experiment will be shown.

A. EXPERIMENT DESIGN

In this section, we will discuss how we evaluated our system in comparison to the teacher counselling services provided by the E-Learning platform on which our chatbot is based.

First, we showed the SQuAD Evaluation performance of the QANet model [50]. ExactMatch (EM) and F1 scores are evaluated by the evaluation script provided by [50]. This performance data is referenced as of July 12th, 2018, and with SQuAD version 1.1. The final F-1 score presented is the average F-1 score calculated from each precision and recall value from one specific QA test case. In every QA test case, there will be a specific precision and recall value. We used both the precision and recall value to calculate the F-1 score for that specific QA. The final F-1 score is (SUM of F-1 score of each QA test case / the total number of QA test cases) * 100 %.

Second, we compared the training time of the QANet model we implemented and a basic RNN-based seq2seq model with training datasets shown in section III, part C. Note that only QA in the training datasets are used in the training of RNN-based seq2seq model since seq2seq does not need extra article content. The parameters and training environment between the two models are the same.

Third, we conducted a response context design test where participants are asked to differentiate between responses generated by our chatbot and responses generated by a human. We generated 20 questions related to the topics in our response context collection. Each question has a default human-generated answer and the chatbot-generated answer which is generated by entering the same question into our chatbot. Please note that we only selected questions that our chatbot is capable of answering since our chatbot is still subject to a limited domain of responses. Also, this test is designed to gauge the accuracy and human-like nature of our chatbot-generated answers. The human-generated answers

TABLE 2. Translated example of a chatbot response context design test.

Question	Tell me how old are you?
Human-generated answer	I am 1 year and 2 months old.
Chatbot-generated answer	I am 1 year and 2 months old. My birthday is on Apr. 1st, 2017

were written by a college student who has never used our chatbot. Table 2 shows one of the questions and answers in this test. We recruited 30 participants with a college education to take this test. Each participant is shown two answers for each question. One answer is provided by the human and one answer is generated by our chatbot. Participants must choose the answer they think is written by a human. The purpose of this test is to check whether our chatbot can correctly classify a user's query to the right category and whether QANet can select the proper responses from the response context. It also tests whether our chatbot provides reasonable and human-like responses from our designed response contexts.

Finally, we used a questionnaire to evaluate and compare our chatbot with the E-Learning platform-provided teacher counselling services performed via messaging and telephone. The participants for this test are students using the E-Learning platform of [37], [38] and [39] which are mainly K-12 students. The number of students was 53. In the questionnaire, the scores for measuring how students feel range from 0 to 4. Zero means the student strongly disagrees or is very unsatisfied, and 4 means the student strongly agrees or is very satisfied. The questions are designed around three main categories: Feelings of isolation and detachment, course-related QA performance and user experience. In the first category, we collected student opinions on feelings of isolation and detachment based on three different scenarios: E-Learning platform only, E-Learning platform with teacher counselling services, and E-Learning platform with our chatbot service. This was a test to see whether our chatbot service could reduce feelings of isolation and detachment. The course-related QA performance category tested whether our chatbot provides thorough and easy-to-understand solutions or responses related to learning when compared to the teacher counselling service. We also wanted to measure how well our chatbot performed in terms of response speed and accuracy. User-experience questions are testing whether the chatbot provides an easy-to-use design and fun, human-like responses.

B. EXPERIMENT RESULTS

In this section, we show all experimental results and discuss the results.

First, the EM and F1 scores of SQuAD datasets evaluation are shown in Table 3. We also listed some top performers from the SQuAD leaderboard [50]. The table shows that the QANet model performs exceptionally well compared to other models, and is even on par with human performance. QANet currently holds the best performance model of

TABLE 3. EM and F1 scores of top models listed in squad leaderboard [50] and the performance result of the latest QANET mode. The performance of qanet model is referenced as of July 12th, 2018. squad version is 1.1.

Model	EM Score	F1 Score
Human Performance [51]	82.304	91.221
QANet (ensemble) [31]	84.454	90.490
r-net (ensemble) [56]	84.003	90.417
MARS (ensemble) [57]	83.982	89.796
Reinforced Mnemonic Reader + A2D (ensemble model) [58]	82.849	88.764
SAN (ensemble model) [59]	79.608	86.496
FusionNet (ensemble) [60]	78.978	86.016
RaSoR + TR + LM (single model) [61]	77.538	84.163
BiDAF (ensemble) [62]	73.744	81.525
Document Reader (single model) [63]	70.733	79.535
SEDT+BiDAF (single model) [64]	68.478	77.971
FastQA [65]	68.436	77.070
BiDAF (single model) [62]	67.974	77.323
Match-LSTM with Ans-Ptr (ensemble) [66]	67.901	77.022

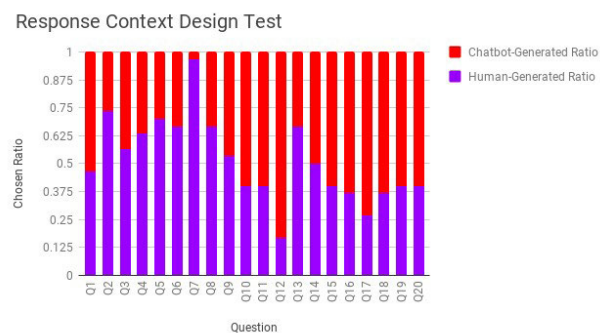
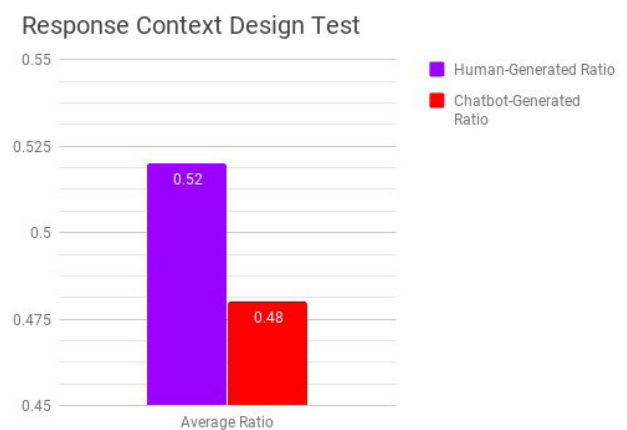
SQuAD1.1 datasets with 84.454 ExactMatch (EM) score and 90.490 F1 score. These scores prove that the QANet model performs exceptionally well. Our implemented QANet model returned a result of EM/F1 score of 70.8/80.1 at best with SQuAD v1.1 Datasets. Our result is not far from the best results of Google Brain and CMU's implementation.

TABLE 4. Training time of the QANET model and basic RNNBASED SEQ2SEQ MODEL.

Model	Total Training Time
QANet	6 hours
Basic RNN-Based Seq2seq Model	29 hours

Second, we compared the training time of the Chinese datasets to our trained QANet model and basic RNN-based seq2seq model. The learning rate was set to 0.001. The results in Table 4 show that QANet is nearly 5 times faster in training compared to the RNN-based seq2seq model. This advantage could help us train more data in the same time. Also, we could re-train and update our model more often to improve performance, since training time is not a major negative factor.

Third, we conducted a response context design test with participants differentiating between responses generated by our chatbot and responses generated by a human. Fig 7 shows the user-chosen result ratio of each question. Although some questions with human-generated answers were more natural and human-like, such as Q2-Q9, some answers generated by our chatbot were richer and full of personality, such as Q10-Q12 and Q15-20, and most participants preferred these over the human-generated answers. Fig 8 is the average chosen ratio between human-generated and chatbot-generated answers across all questions. The results show that although human-generated answers have an overall 52% average chosen ratio, our chatbot-generated answers are not far behind with an overall 48% average chosen ratio. This proves that our chatbot's responses are on par with how humans respond

**FIGURE 7.** Result of a chatbot response context design test. Red represents the ratio that a user chose the chatbot-generated answer. Blue represents the ratio that a user chose human-generated answer.**FIGURE 8.** Average ratio of a chatbot response context design test. Red represents the average ratio that a user chose the chatbot-generated answer. Blue represents the average ratio that a user chose the human-generated answer.

to a question. Most users will not be able to tell whether one is chatting with a bot or a human. This result also proves that our chatbot can produce human-like responses and conversation

experiences that could help reduce feelings of detachment in the user.

Having Feelings of Isolation and Detachment?

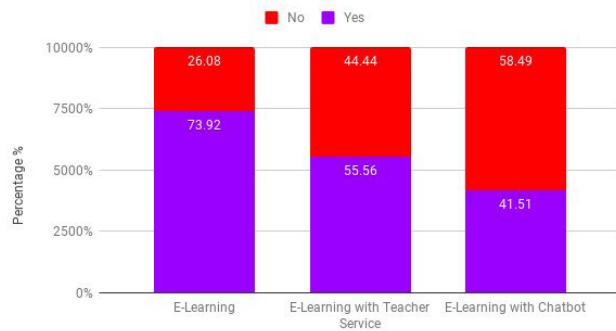


FIGURE 9. Ratio of how many participants have feelings of isolation and detachment in three different situations.

Finally, we collected questionnaires from all 53 participants. The first category we will discuss is feelings of isolation and detachment. Fig 9 shows how many of our participants had feelings of isolation and detachment in three different situations. The left bar in Fig 9 shows results from participants using only the E-Learning platform without other services. About 74% of participants experienced feelings of isolation and detachment while using the E-Learning platform. The middle bar in Fig 9 shows results from participants using the same E-Learning platform with teacher counselling services provided by the platform. The ratio of participants who experienced feelings of isolation and detachment dropped to about 55.5%, proving that interaction with a human teacher could reduce feelings of isolation and detachment (even though over 50% of participants still experience the issue). The right bar in Fig 9 shows results from participants using the same E-Learning platform with our designed chatbot. Replacing teacher counselling services with our chatbot could further reduce the ratio of participants who experience this issue to about 41.5%, proving that our chatbot could reduce feelings of isolation and detachment with greater effect than teacher counselling services. We think this is due to two factors: 1) Our chatbot responds almost instantly at any time compared to teacher services, which means users have more time to be with our chatbot. 2) Unlike teacher counselling services, our chatbot can chat casually with users. This is not possible for teacher counselling services due to its dedication to question answering and limited resources.

The second category we will discuss is the comparison between our chatbot’s E-Learning course-related QA performance to that of the teacher counselling service. Fig 10 shows the average user preference score of the chatbot and the teacher service. The top bar in Fig 10 shows how thorough and easy-to-understand the response of our chatbot is compared to the teacher counselling service. The average score (mean) is 2.61, meaning that our chatbot is comparable to the teacher service in this area (and actually scored

E-Learning Course-Related QA compared to Teacher Counseling Service

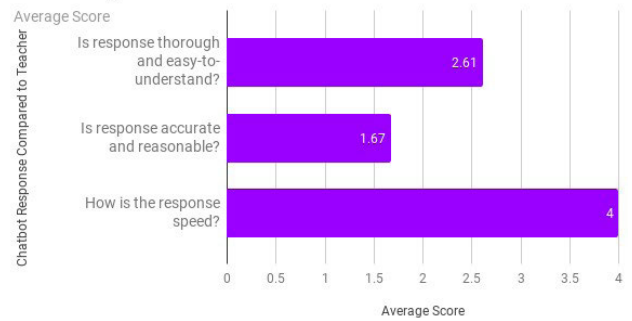


FIGURE 10. Average user preference scores in E-Learning courserelated question-answering performance between our chatbot and the teacher counselling service. A score of 0 means the chatbot performed much worse than the teacher counselling service. A score of 4 means the chatbot performed much better than the teacher counselling service.

E-Learning Course-Related QA compared to Teacher Counseling Service



FIGURE 11. User preference score distribution between our chatbot and teacher counselling services in E-Learning course-related question-answering performance. A score of 0 means the chatbot performed worse than the teacher counselling service. A score of 4 means the chatbot performed better than the teacher counselling service.

slightly higher). Also, Fig 11 confirms that most participants rated a score of 2 or 3 and the standard deviation is 0.927. We think the reason for this score is that our chatbot not only provides quality answers but also provides additional tutorial information, such as related course videos, at the user’s fingertips. The middle bar in Fig 10 compares the accuracy and reasonableness between our chatbot’s responses and the teacher counselling service. The average score (mean) is 1.67. Fig 11 shows that most participants rated a score of 1 or 2 and the standard deviation is 0.807. In this area, our chatbot falls short compared to teacher counselling services because our chatbot’s knowledge base is still not comparable to that of a teacher or counsellor. There are still many questions that our chatbot cannot answer correctly or properly, and we are continuously improving and extending our chatbot’s knowledge base. Lastly, the bottom bar in Fig 10 shows how fast our chatbot responds to users compared to the teacher counselling service. In this area, there is really no comparison between our chatbot and the teacher counselling service. All participants rated a score of 4 (see Fig 11) and the standard deviation is 0. This shows that a major advantage of our chatbot is that it can respond immediately at any time.

User Experience

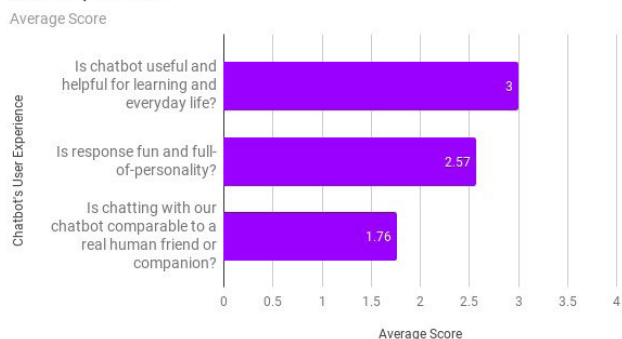


FIGURE 12. The average user experience scores of our chatbot. A score of 0 means the student disagreed greatly or was very unsatisfied. A score of 4 means the student agreed or was very satisfied.

User Experience

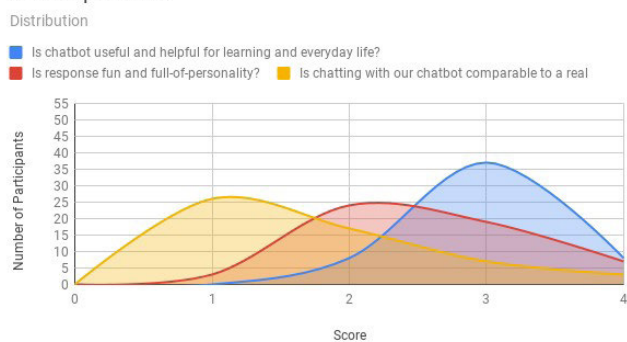


FIGURE 13. Distribution of user experience scores of our chatbot. A score of 0 means the student strongly disagreed or was very unsatisfied. A score of 4 means the student strongly agreed or was very satisfied.

The third category we will discuss is user experience. Fig 12 shows the average user experience score of our chatbot. The top bar in Fig 12 is how useful and helpful our chatbot is for learning and everyday life. The score (mean) for this area is 3 and the standard deviation is 0.555. Fig 13 shows that all participants rated a score of 2 and above. This result demonstrates that our chatbot is very useful and helpful in both learning and everyday life. The middle bar in Fig 12 is whether our chatbot responds with fun and full-of-personality answers. The average score (mean) is 2.57. The standard deviation is 0.797, and Fig 13 shows most participants rated our chatbot with a score of 2 or greater. This shows that our chatbot is enjoyable to chat with, and its responses are fun and entertaining. The bottom bar in Fig 12 shows whether or not users find that chatting with our chatbot is comparable to a real human friend or companion. The average score (mean) is 1.76. The standard deviation is 0.897, and Fig 13 shows that most participants disagree. There might be several reasons for this result. First, our chatbot only performs average in providing correct and reasonable responses, which is due to its still-growing knowledge base (as shown in Fig 12). Second, some of our chatbot’s responses are too complete and precisely answered. For example, when someone asks what kind of beverage would you like? Most people simply reply with the drink they like. Our chatbot, however, replies with something like, “The beverages I like to drink are...” Due to

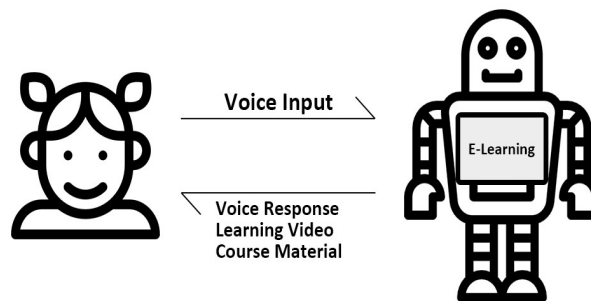


FIGURE 14. Our vision for K-12 E-Learning.

the limitations of our current trained QANet model, responses must be this precise and standardized or the model’s accuracy will drop significantly. For most people, it is very unusual to hear someone reply so precisely. This is one reason why our chatbot still needs improvement in order to become more like a real human companion. Third, our chatbot currently still uses bot platforms from social media platforms with text input, and most functions must be operated through a user’s smartphone or tablet, which is quite a human-like interaction. In our vision, users can use voice-only to talk and interact with our chatbot which is a more natural human interaction. We are working toward this vision and continually improving our current chatbot design.

V. CONCLUSION AND FUTURE WORK

E-Learning is a wonderful way to learn. It utilizes many convenient platforms, such as smartphones and tablets, so students can learn and study anytime, anywhere. However, learning on E-Learning platforms lacks interaction between students or teachers which often causes students to experience feelings of isolation and detachment. Evaluation results show that our proposed hybrid chatbot is able to reduce these negative feelings. Responses from our chatbot are relatively close to what a real human would say, and chatting with our chatbot is a fun and entertaining experience. Compared with the teacher counselling services provided by the E-Learning platform on which our chatbot is based, our chatbot does have advantages in chatting with students. These advantages include interesting conversations and instant responses at any time of day or night. However, our chatbot still falls short of real human teachers in solving learning problems due to the fact that its current datasets and knowledge base are still in infancy, a huge constraint for any kind of chatbot. It is important that a chatbot designed for educational use has rich, accurate responses for course-related questions. We must invest more effort into adding more course-related and educational materials to our chatbot. Both the response contexts and the hybrid model still have room for improvement. Our experiments and questionnaires need tweaks as well, since the questions are not fully optimized for the comprehension ability of K-12 students. The questions also need to better consider the psychology of user feelings. The number of participants is still small, and we are working on increasing K-12 participants who have adequate experience using the

E-Learning platform on which our chatbot is based. However, we do have a general picture of how our chatbot performs as an E-Learning assistant. Even though most users feel that our chatbot is not like a real human companion, we think our chatbot does perform well as a dedicated E-Learning companion and assistant.

In our future work, we want to improve our bot's ability to understand and solve more learning-related problems. Our designed topics in the dialogue module, the hybrid model's performance, the response contexts, and the evaluation design still have room for improvement. The concept of combining models for different purposes in one chatbot can be used in other domains, and the combination or number of models can be altered to suit different applications. Our future vision for a K-12 E-Learning assistant is that of a robot like Pepper [53] that could interact with K-12 students via voice as shown in Fig 14.⁴ All K-12 E-Learning material and courses can be acquired by voice interaction with a robot. In addition to responding to the close-ended question, it can further answer the open-ended question. This creates an interactive learning experience similar to a traditional classroom with the advantage of E-Learning. Also, the interaction experience is more human-like since it is based on voice conversation.

REFERENCES

- [1] *The Summit Learning Platform*. Accessed: Nov. 5, 2017. [Online]. Available: <https://www.summitlearning.org/program/online-platform#>
- [2] B. Klimova and P. Maresova, "Cloud computing and e-learning and their benefits for the institutions of higher learning," in *Proc. IEEE Conf. e-Learn., e-Manage. e-Services (IC3e)*, Langkawi, Malaysia, Oct. 2016, pp. 75–78.
- [3] H. Garcia-Gonzalez, J. E. L. Gayo, and M. Paule-Ruiz, "Enhancing e-learning content by using semantic Web technologies," *IEEE Trans. Learn. Technol.*, vol. 10, no. 4, pp. 544–550, Oct. 2017.
- [4] P. J. Munoz-Merino, C. D. Kloos, and M. Munoz-Organero, "Enhancement of student learning through the use of a hinting computer e-Learning system and comparison with human teachers," *IEEE Trans. Educ.*, vol. 54, no. 1, pp. 164–167, Feb. 2011.
- [5] *Docebo-eLearning-Trends-for-2018 Whitepaper*. Accessed: Dec. 25, 2017. [Online]. Available: <https://www.docebo.com/resource/whitepaper-elearning-trends-2018/>
- [6] S. K. Dn. *E-Learning Trends for 2018*. Accessed: Dec. 26, 2017. [Online]. Available: <https://elearningindustry.com/elearning-trends-for-2018-9>
- [7] R. Austin, M. Sharma, P. Moore, and D. Newell, "Situating computing and virtual learning environments: E-Learning and the benefits to the students learning," in *Proc. 7th Int. Conf. Complex, Intell., Softw. Intensive Syst.*, Taichung, Taiwan, Jul. 2013, pp. 523–528.
- [8] M. Nadrljanski, D. Vukić, and D. Nadrljanski, "Multi-agent systems in e-learning," in *Proc. MIPRO*, Opatija, Croatia, 2018.
- [9] A. Bernik, G. Bubas, and D. Radošević, "Measurement of the effects of e-learning courses gamification on motivation and satisfaction of students," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, Opatija, Croatia, May 2018, pp. 0990–0995.
- [10] M. Montebello, "E-learning paradigms: A model to address known issues," in *Proc. Comput. Conf.*, Jul. 2017, pp. 1180–1189.
- [11] H. Al-Samraie, H. Selim, T. Teo, and F. Zaqout, "Isolation and distinctiveness in the design of e-learning systems influence user preferences," *Interact. Learn. Environ.*, vol. 25, no. 4, pp. 452–466, May 2017.
- [12] B. Mourad, A. Tarik, A. Karim, and E. Pascal, "System interactive cyber presence for e learning to break down learner isolation," *Int. J. Comput. Appl.*, vol. 111, no. 16, pp. 35–40, Feb. 2015.
- [13] AlphaGo. *Deepmind*. Accessed: Jan. 2, 2018. [Online]. Available: <https://deepmind.com/research/alphago/>
- [14] B. du Boulay, "Artificial intelligence as an effective classroom assistant," *IEEE Intell. Syst.*, vol. 31, no. 6, pp. 76–81, Nov. 2016.
- [15] C. Ramos, J. C. Augusto, and D. Shapiro, "Ambient intelligence—The next step for artificial intelligence," *IEEE Intell. Syst.*, vol. 23, no. 2, pp. 15–18, Mar. 2008.
- [16] *Google Assistant*. Accessed: Jan. 2, 2018. [Online]. Available: <https://assistant.google.com/>
- [17] *Google I/O*. Accessed: Jan. 2, 2018. [Online]. Available: <https://events.google.com/io/>
- [18] Google. *Google AI Blog: Google Duplex*. Accessed: Jun. 2, 2018. [Online]. Available: <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>
- [19] H.-Y. Shum, X.-D. He, and D. Li, "From Eliza to XiaoIce: Challenges and opportunities with social chatbots," *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 1, pp. 10–26, Jan. 2018.
- [20] Apple. *Siri*. Accessed: Jan. 6, 2018. [Online]. Available: <https://www.apple.com/ios/siri/>
- [21] Amazon. *Alexa*. Accessed: Jan. 6, 2018. [Online]. Available: <https://developer.amazon.com/alexa>
- [22] Microsoft. *XiaoIce*. Accessed: Jan. 12, 2018. [Online]. Available: <http://www.msxiaoice.com/>
- [23] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "EEG-based mobile robot control through an adaptive brain-robot interface," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 44, no. 9, pp. 1278–1285, Sep. 2014.
- [24] *Forbes Asia Bot*. Accessed: Jan. 13, 2018. [Online]. Available: <https://www.messenger.com/t/forbesasia>
- [25] *TaxiGo Bot*. Accessed: Jan. 15, 2018. [Online]. Available: <https://www.messenger.com/t/taxigo.tw>
- [26] *Got It Study*. Accessed: Jan. 20, 2018. [Online]. Available: <http://expert.gotitapp.co/>
- [27] *Deakin Genie*. Accessed: Jan. 20, 2018. [Online]. Available: <https://www.youtube.com/watch?v=zsRPuU53E74>
- [28] *Jill Watson*. Accessed: Jan. 20, 2018. [Online]. Available: <https://pe.gatech.edu/blog/meet-jill-watson-georgia-techs-first-ai-teaching-assistant>
- [29] *IBM Watson*. Accessed: Jan. 20, 2018. [Online]. Available: <https://www.ibm.com/watson/>
- [30] A. C. Graesser, X. Hu, V. Rus, and Z. Cai. *AutoTutor*. Accessed: May 15, 2018. [Online]. Available: <http://ace.autotutor.org/IISAutotutor/index.html>
- [31] A. W. Yu, D. Dohan, M. Luong, R. Zhao, K. Chen, M. Norouzi, and Q. V. Le, "QANet: Combining local convolution with global self-attention for reading comprehension," in *Proc. ICLR*, Vancouver, BC, Canada, 2018, pp. 1–16.
- [32] B. F. Klimova and P. Poulova, "Tutor as an important e-learning support," *Procedia Comput. Sci.*, vol. 3, pp. 1485–1489, Jan. 2011.
- [33] Facebook. *Messenger*. Accessed: Jan. 30, 2018. [Online]. Available: <https://messenger.fb.com/>
- [34] *Telegram*. Accessed: Jan. 30, 2018. [Online]. Available: <https://telegram.org/blog/bot-revolution>
- [35] *Microsoft Bot Framework*. Accessed: Jan. 30, 2018. [Online]. Available: <https://dev.botframework.com/>
- [36] *Chatfuel*. Accessed: Jan. 30, 2018. [Online]. Available: <https://dev.botframework.com/>
- [37] E. H. Wu, S. Chuang, C.-Y. Shih, H.-C. Hsueh, S.-S. Huang, and H.-P. Huang, "A flexible and lightweight user-demand DRM system for multimedia contents over multiple portable device platforms," *Softw. Pract. Exper.*, vol. 47, no. 10, pp. 1417–1441, Oct. 2017.
- [38] H.-P. Huang, "TW K-12 academy: A finger-driven interact learning system with smart-pen assistance on iOS devices," M.S. thesis, Dept. Comput. Sci. Inform. Eng., Nat. Central Univ., Zhongli, Taiwan, 2016.
- [39] W.-T. Chen, "Toward smart self-driven learning environment: Integrating DRM system in K12 academy," M.S. thesis, Dept. Comput. Sci. Inform. Eng., Nat. Cent. Univ., Zhongli, Taiwan, 2017.
- [40] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Designing and implementing a human-robot team for social interactions," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 4, pp. 843–859, Jul. 2013.
- [41] R. P. Schumaker and H. Chen, "Interaction analysis of the ALICE chatterbot: A two-study investigation of dialog and domain questioning," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 1, pp. 40–51, Jan. 2010.
- [42] G. An, M. Shafiee, and D. Shamsi, "Improving retrieval modeling using cross convolution networks and multi frequency word embedding," 2018, *arXiv:1802.05373*. [Online]. Available: <http://arxiv.org/abs/1802.05373>

⁴Icon on Fig 14 are made by Freepik, Nikita Golubev, Icon Works from <http://www.flaticon.com/>

- [43] Z. Yu, A. Papangelis, and A. Rudnicky, "TickTock: A non-goal-oriented multimodal dialog system with engagement awareness," in *Proc. AAAI Spr. Symp.*, Palo Alto, CA, USA, 2015, pp. 108–111.
- [44] *Pizza Hut Messenger Bot*. Accessed: Jan. 13, 2018. [Online]. Available: <https://www.messenger.com/t/PizzaHutPak>
- [45] *CNN News Messenger Bot*. Accessed: Jan. 13, 2018. [Online]. Available: <https://www.facebook.com/messages/t/cnn>
- [46] A. Kannan, P. Young, V. Ramavajjala, K. Kurach, S. Ravi, T. Kaufmann, A. Tomkins, B. Miklos, G. Corrado, L. Lukacs, and M. Ganea, "Smart reply: Automated response suggestion for email," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 955–964.
- [47] M. D. G. Bruno Marietto, R. V. Aguiar, G. D. O. Barbosa, W. T. Botelho, E. Pimentel, R. D. S. Franca, and V. L. da Silva, "Artificial intelligence markup language: A brief tutorial," *Int. J. Comput. Sci. Eng. Surv.*, vol. 4, no. 3, pp. 1–20, Jun. 2013.
- [48] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. 27th Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 3104–3112.
- [49] T. Hu, A. Xu, Z. Liu, Q. You, Y. Guo, V. Sinha, J. Luo, and R. Akkiraju, "Touch your heart: A tone-aware chatbot for customer care on social media," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, Montreal, QC, Canada, Apr. 2018, pp. 1–12.
- [50] *SQuAD*. Accessed: Jul. 12, 2018. [Online]. Available: <https://rajpurkar.github.io/SQuAD-explorer/>
- [51] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ questions for machine comprehension of text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, TX, USA, 2016, pp. 1–10.
- [52] A. Tammewar, M. Pamecha, C. Jain, A. Nagvenkar, and K. Modi, "Production ready chatbots: Generate if not retrieve," in *Proc. 32nd AAAI*, New Orleans, LA, USA, 2018, pp. 739–745.
- [53] J. C. Vaz and P. Oh, "Transformer-DARwIn: A miniature humanoid for potential robot companion," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2018, pp. 1–5.
- [54] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543.
- [55] Fxsjy. *GitHub: Fxsjy/Jieba*. Accessed: Jan. 25, 2018. [Online]. Available: <https://github.com/fxsjy/jieba>
- [56] M. R. A. Natural Language Computing Group. (May 2017). *R-NET: Machine Reading Comprehension With Self-Matching Networks*. [Online]. Available: <https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/r-net.pdf>
- [57] *Y R NLP, MARS (Ensemble): Multi-Attention Readers*, YuanFuDao Research NLP Group, Beijing, China, Jun. 2018.
- [58] M. Hu, Y. Peng, Z. Huang, X. Qiu, F. Wei, and M. Zhou, "Reinforced mnemonic reader for machine reading comprehension," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Stockholm, Sweden, Jul. 2018, pp. 4099–4106.
- [59] X. Liu, Y. Shen, K. Duh, and J. Gao, "Stochastic answer networks for machine reading comprehension," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, Melbourne, VIC, Australia, 2018, pp. 1694–1704.
- [60] H.-Y. Huang, C. Zhu, Y. Shen, and W. Chen, "FusionNet: Fusing via fully-aware attention with application to machine comprehension," in *Proc. 6th ICLR*, Vancouver, BC, Canada, 2018, pp. 1–20.
- [61] S. Salant and J. Berant, "Contextualized word representations for reading comprehension," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, New Orleans, LA, USA, 2018, pp. 554–559.
- [62] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, "Bidirectional attention flow for machine comprehension," in *Proc. ICLR*, Toulon, France, 2017, pp. 1–13.
- [63] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading Wikipedia to answer open-domain questions," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, Vancouver, BC, Canada, 2017, pp. 1870–1879.
- [64] R. Liu, J. Hu, W. Wei, Z. Yang, and E. Nyberg, "Structural embedding of syntactic trees for machine comprehension," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Copenhagen, Denmark, 2017, pp. 815–824.
- [65] D. Weissenborn, G. Wiese, and L. Seiffe, "Making neural QA as simple as possible but not simpler," in *Proc. 21st Conf. Comput. Natural Lang. Learn. (CoNLL)*, Vancouver, BC, Canada, 2017, pp. 271–280.
- [66] S. Wang and J. Jiang, "Machine comprehension using match-LSTM and answer pointer," in *Proc. ICLR*, Toulon, France, 2017, pp. 1–11.



ERIC HSIAO-KUANG WU (Member, IEEE) received the B.S. degree in computer science and information engineering from National Taiwan University, in 1989, and the master's and Ph.D. degrees in computer science from the University of California, Los Angeles (UCLA), in 1993 and 1997, respectively. He is currently a Professor of computer science and information engineering with National Central University, Taiwan. His primary research interests include wireless networks, mobile computing, and broadband networks. He is also a member of the Institute of Information and Computing Machinery (ICM)



CHUN-HAN LIN received the M.S. degree in computer science and information engineering from National Chiao-Tung University, Taiwan. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Engineering, National Central University. His research interests include wireless networks, mobile computing, streaming technology, and location-based services.



YU-YEN OU is currently an Associate Professor with the Department of Computer Science and Engineering, Yuan-Ze University, Taiwan. His research interests include bioinformatics, machine learning, and data mining.



CHEN-CHUNG LIU is currently a Chair Professor with the Department of Computer Science and Information Engineering, National Central University, Taiwan. His research interests include investigation of collaborative learning processes in online and classroom settings.



WEI-KAI WANG was born in Taiwan, in 1993. He received the B.S. and M.S. degree in computer science and information engineering from National Central University, Taiwan, in 2016 and 2018, respectively. His research interests include natural language processing, neural networks, and machine learning.



CHI-YUN CHAO was born in Taiwan, in 1993. He received the Bachelor of Science degree and the master's degree in computer science and information engineering from the National Central University of Taiwan, in 2016 and 2018, respectively. His research interests include machine learning and natural language processing.

...