

# LIGHTBLUE: NURTURE YOUR PERSONAL CHATBOT

Xiang Zhang<sup>1</sup>, Yan Liu<sup>1</sup>, Gong Chen<sup>1</sup> and Sheng-hua Zhong<sup>2</sup>

<sup>1</sup>Department of Computing, The Hong Kong Polytechnic University,  
Hong Kong, China

<sup>2</sup>College of Computer Science and Software Engineering,  
Shenzhen University, Shenzhen, China

## ABSTRACT

*Chatbot has long been an important research topic in artificial intelligence and attracts lots of attention recently. Despite significant advancements in language ability, the interactions between users and chatbots are rather generic, short-term, and transnational. It has always been challenging to develop truly personal chatbots and even more difficult to establish long-term, affective connections. This paper first brings up “nurture” as a new interaction mode with chatbots. We introduce the nurture framework and accordingly design the learning algorithm and nurture functions. Then we present LightBlue – a platform that allows non-professionals to nurture personal chatbots from scratch. Experiments on both closed- and open-domain tasks validate the proposed framework and demonstrate a promising method for facilitating long-term interaction between users and chatbots.*

## KEYWORDS

*Personal Chatbot, Conversation Agent, Nurture, Human-chatbot Interaction, Long-term Relationship*

## 1. INTRODUCTION

Chatbot, a computer program which can communicate with humans in natural language, has been put into study since the birth of Artificial Intelligence (AI) [1]. It is also called as machine conversation system, virtual agent, dialogue system, and chatterbot [2]. A chatbot can be programmed to mimic humans, answer questions, serve as a personal assistant, and so on [3]. In recent years, the applications of chatbots boom in industry, emerging in many fields including education, social media, finance, catering, etc [4,5,6]. Siri as one of the most representative chatbot products, has attracted billions of users [6,7,8].

The work on chatbots can be divided into two groups: task-oriented chatbots and general chatbots. Task-oriented chatbots are designed to serve specific task, such as a hotel booking or a technical support service, and the conversation cannot go beyond the topic scope of the system. General chatbots are designed to pass the Turing test, or engage in social chit-chat with users [9], with no specific target or topic scope. They can be also used to engage users in open-domain human-computer conversations for entertainments or emotional companionship [10]. However, the current communication with chatbots are superficial and generic; users are more likely to regard them as tools or strangers [11].

Personal chatbots begin to receive the public attention in recent years. Rather than serving a broad audience, personal chatbots cater to a specific user, aiming to conduct a personal and in-depth conversation: with intimate knowledge of the user's preferences, needs, and habits, it provides necessary information and assistance [12]. This direction, which provides end-users with increased flexibility and control over their own personal chatbots, is yet to be explored. For example, companion chatbot, a kind of chatbot pursues long-term interaction and affectional connections with specific users, is still in its infancy [5,11].

The inherent difficulty in developing personal chatbots is balancing generality and specificity. On the one hand, it should be sufficiently flexible to be customized by all users; on the other hand, it should know the specific user to a certain degree to facilitate long-term and cumulative communication.

To provide a flexible conversation mechanism with both generality and specificity, the prerequisite is to be aware of the user's personal information. However, collecting user's information for each conversation not only requires a huge amount of work, but also is unavailable in many cases. Recently, many efforts have been put into building datasets providing users' information for dialogue samples. For example, PERSONA-CHAT dataset from Facebook AI research [13] contains over 10,000 dialogs on more than 1,000 different personas. However, the dialogs are collected from crowd-sourcing workers, thus can not fully represent real user-bot interaction. And the designed personas are hard to address the massive number of user profiles in real world.

User-specific conversation requires every sentence to be processed in a specific context, while the diversity of end-users makes the content of conversations unpredictable. The answers cannot be prepared in advance, as different users may interpret the same sentence differently. Furthermore, people would change their identities, preferences and status with time and the environment. Instead of preparing massive amounts of conversation samples, we emphasize the learning ability of a personal chatbot -- A personal chatbot should be able to continuously learn and grow from the interactions, and adapt to new changes after deployment.

Alan Turing has once raised the question in his proposal "Computing Machinery and Intelligence" – "*Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?*" [14]. The nature-versus-nurture debate concerning the development of human intelligence has lasted for a long time [15]. However, "nurture" hasn't received enough attention in today's data-driven machine learning methods. Inspired by Turing's "children and education" viewpoint, we propose to nurture a chatbot by users. That is, we first create a simple and general chatbot model. After deployment, it can be nurtured by end-users to become a personal chatbot. We define "nurture" as a new interaction mode with chatbots, which differentiates from traditional interactions in the following aspects:

- The learning algorithm grows gradually during the nurture process.
- Users can guide the algorithm to develop in a particular direction through the nurture process.

In this way, the learning algorithm is able to learn about the particular user and grow after deployment. By considering the end-users' initiative to engage in a cumulative and in-depth conversation, "nurture" contributes to the development of personal chatbots in the following aspects:

- In biological and social sciences, it is assumed that intelligence develops through learning culturally valued knowledge and skills in a human, social environment [16]. There are also powerful evidences showing the critical role of nurture in cognitive development [17].
- Nurture can be helpful to develop a close, interpersonal relationship, such as social bonds, between users and chatbots.
- Nurture plays a significant role in how humans learn to converse. In addition to passively observing others' conversations, we also actively adjust and correct our speech in response to feedback woven throughout our own conversations [18,19]. It has been proved that interaction, rather than exposure, is necessary for successful language acquisition [20,21].

In this paper, we define the nurture framework in detail and design the corresponding learning algorithm for nurture. Then we integrate the learning algorithm and nurture functions into a chatbot nurture platform called LightBlue, which enables nonprofessional users to nurture personal chatbots. There are three main contributions of this work:

- This paper is the first work of introducing the nurture framework as a new interaction mode between users and chatbots.
- Under the nurture framework, we further propose a corresponding learning algorithm for nurture and design a bunch of nurture functions.
- In this paper, we develop LightBlue – a chatbot nurture platform, integrating the learning algorithm and nurture functions.

The rest of the paper is organized as follows. We start by reviewing related work. Then we describe the nurture framework, the learning algorithm, nurture functions and the LightBlue chatbot nurture platform. Two experiments are conducted to verify the effectiveness of the nurture framework and we report the results in the following section. We visualize and analyze our experimental results and then discuss the features of the proposed framework. Finally, we conclude this paper.

## 2. RELATED WORK

In this section, we review the chatbot applications and technologies. There are several studies on learning after deployment and user-chatbot interactions that are related to our nurture framework design.

### 2.1. Chatbot Applications and Technologies

Following the first chatbot ELIZA [22] in 1960s, researchers have devoted to the study on chatbots. Many of the early works aimed to pass the Turing test [14]. For example, PARRY simulated paranoid behaviors and successfully fooled its judges [23]. In 1995, the Loebner Competition held its first unrestricted Turing test with no limits on the topics discussed. More general-purpose chatbots have been developed, such as MegaHAL [24], CONVERSE [25], ELIZABETH [26], A.L.I.C.E [27] and Mitsuku (Loebner prize winner in 2013, 2016, 2017, 2018) [28]. With the emergence of messaging platforms, not only have chatbots for socializing and small talk (e.g. A.L.I.C.E, Cleverbot, Simsim, Tay) gained popularity, but task-oriented chatbots have also attracted a lot of attention [29]. Bagousse estimates that chatbots in customer service will save businesses up to 11 billion dollars by 2025 [30]. Chatbot bot has never been so popular as recent days.

During the last half century, chatbot technologies have varied from keywords matching, rule-based models, retrieve-based models, to generative models [31]. With the advancement of

computational power and the availability of big data, the current trend is to use neural networks to generate responses. One of the largely used structure in neural language generation is Sequence-to-Sequence (Seq2Seq) [32] model, adopted from machine translation. For example, [33] uses Sina Weibo dataset to train a neural responding machine based on Seq2Seq framework; Neural attention (or alignment) are then proposed to improve the Seq2Seq models by associating salient items in the source sequence with the generated item in the target sequence [34,35,36]; Zhou et al.[37] adds memory mechanism to address the emotion factor in large-scale conversation generation.

Deep learning has contributed a lot to develop intelligent chatbots in the sense of generating more human-like sentences and talking in a natural way [31]. However, these models are usually developed in the laboratory by professional engineers. No matter targeting a specific task or open-domain conversation, they're usually trained using large corpora of crowd-sourced or scraped conversations [38]. Thus, there are two common limitations: First, they tend to give generic answers – some “Jack of All Trades”-type answers will be applied to a different question, such as “*I don't know.*” [7,10]; Second, they usually lack consistent personality due to the large amount of mixed training data. And we wouldn't call these chatbots truly “understand” the dialogue but rather create an illusion of intelligence and empathetic understanding [39]. Since they lack the ability to further develop or learn about the user during interaction, it is hard to have long-term and deep conversation. Due to the same reason, users often lose interest after a short interaction [40].

## 2.2. Learning After Deployment

For personal chatbots, it's not feasible to collect a large amount chatting data between one particular user and the chatbot. Since users are free to communicate with the system in any way they wish, it is impossible to pre-plan the interaction logic and domain knowledge in advance [41]. A simple intuition is to allow chatbots to develop after deployment. For example, Huang et al. [42] proposes Evorus, a crowd-powered chatbot with can automate itself during deployment. Though Evorus learns to automatically select high-quality response over time, it learns from paid crowd workers rather than end-users. In another word, it can not learn or update without crowd oversight, which is not only labor-extensive, but also involve privacy issues.

A common way to learn from users is to require user's feedback during the interaction [13,43]. Typically, existing methods require paid annotators to provide scalar rewards or to use particular templates to so that the feedback can be used by the model [44,45,46]. For example, some chatbots learn actively during conversation in the question answering (QA) setting [47,48]. Some works also allow chatbots to learn directly from the natural dialogues, e.g., Hancock et al. [19] proposes the self-feeding chatbot capable of extracting new training examples from the conversations in which it participates. However, most of the work are targeting task-oriented chatbots or a specific scenario, but not designed for a general setting. And none of them allow users to develop chatbots from scratch, meaning that they still need large dataset to pre-train the model before deployment.

While learning from multiple paid workers or crowd-sourcing feedback may be effective for a specific task, the feedback can be chaotic and even conflicting when applied to a broader context. Tay, Microsoft's Twitter-based chatbot which learned from massive Twitter users, was forced to shut down for being incredibly racist [49].

## 2.3. User-Chatbot Interactions

Learning from a particular user can be very different from learning from crowd users in terms of data volume, dialogue content, interaction depth or consistency. Additional modes of interaction for efficient learning should be investigated. Current research on user-chatbot interactions focuses primarily on interface design in order to facilitate usage or improve the user experience. Jain et al. [50] provides a context view for chatbots to address a mismatch between the chatbot's state of understanding (also called context) and the user's perception of the chatbot's understanding. It also supports intuitive interactions with context values, allowing users to modify them simply and efficiently. Candello et al. [51] studies the influence of typefaces on the perception of humanness in chatbots. There's also an increasing interest in multi-modal interaction, such as combining visual information [52,53,54]. Some work also considers the interaction in special cases. Seering et al. [55] study chatbots supporting or engaging in group or multiparty interactions. And [56] explores the repair preferences for conversational breakdowns. Luger et al. [8] reveals a number of design challenges arising from the gap between user expectations and actual experience, such as how a chatbot can reveal its current state, or how to design system feedback and clearly communicate the goal of the system. However, the interaction between users and the underlying learning algorithm is still underexplored.

Though personal chatbots are not designed for specific tasks, the conversations will become specific and personal when deployed to the end-users. Thus, we introduce a new interactive mode – nurture, which enables non-professional users to train the learning algorithm using simple functions and to obtain personal chatbots.

## 3. SYSTEM DESIGN

In this section, we first define the nurture framework as a new interaction mode between users and chatbots. Following the nurture framework, we propose a novel computational model - pattern matching model for nurture, along with a set of simple yet effective nurture functions. Algorithms of those functions are explained formally. Finally, we put the proposed framework into practice by creating the LightBlue chatbot nurture platform, which can be directly used by non-professional end-users.

### 3.1. Nurture Framework

The basic idea of “nurture” is to allow end-users to guide the development of the intelligent algorithm and help it adapt to a new environment. Intuitively, it can be conducted in the same way that humans provide feedback, such as correcting incorrect responses, indicating a preference for certain actions, and awarding rewards. As a new interaction mode, “nurture” allows end-users to participate in the development of intelligent algorithms and enables joint development of engineers and non-professional users. In contrast to conventional user-chatbot interaction, which views chatbots as the interface to other functions, “nurture” goes beyond the interface and emphasizes interaction with the underlying learning algorithm. We define the nurture framework by three key components:

- Learning algorithm: an algorithm with a simple and general initial state, and can be nurtured in an open and dynamic environment through long-term interaction.
- Nurture functions: functions that can be used by nonprofessional users to guide the development of the learning algorithm.
- Feedback information: feedback from the chatbot that enables users to monitor the effect of nurture actions and the chatbot's state, and adjust their nurture strategy accordingly.

### 3.2. Pattern Matching Model for Nurture

Following the nurture framework, the learning algorithm should be able to grow in the open and dynamic environment. Specifically, in the application of personal chatbots, the algorithm should be able to adapt and grow through the interaction with unfamiliar users. Meanwhile, users can guide the development of the algorithm on a long-term and open-ended scale.

Here we adopt the classical idea of the pattern matching method used in chatbots, and develop it into a pattern matching model for nurture. The pattern matching technology can be traced back to the first chatbot ELIZA [22], which matches responses to pre-defined keywords. Based on this simple idea, we further develop heuristic pattern generation and update methods, allowing the chatbot to learn from new data incrementally. Its simplicity and transparency allow us to see how the new growth characteristics affect the original algorithm and to observe the role of nurture.

#### 3.2.1. Basic concepts and symbols

Figure 1 shows the important components in the pattern matching model. The following are detailed explanations of basic concepts in the model:

- *Message*: the user's input utterance.
- *Response*: the chatbot's output utterance to the user.
- *Basic unit*: word or punctuation.
- *Pattern*: a basic unit or a concatenation of basic units.
- *Goodness*: the degree to which a response is considered to be a good one.
- *Link*: the weighted edge between a pattern and a response.
- *Link activation*: the process by which the weight of a link is increased.
- *Number of activation(s)*: The number of link activation(s).
- *Dialog turn*: a message from the user and a following response from the chatbot.

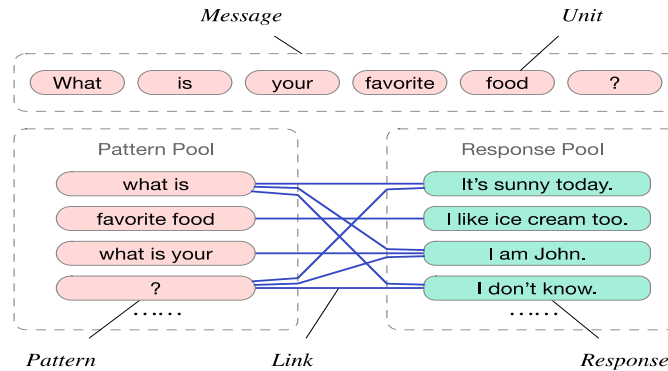


Figure 1. The architecture of pattern matching model for nurture.

Table 1 contains a list of frequently used notations in this paper.

Table 1. Summary of frequently-used notations.

Symbol	Definition
$P$	A set of patterns
$p_i$	The pattern $i$ in set $P$
$R$	A set of responses

$r_j$	The response $j$ in the set $R$
$l(p_i)$	Length of $p_i$ (i.e. the number of basic units it contains)
$g(r_j)$	The goodness of response $r_j$
$w(p_i, r_j)$	The weight of the link between $p_i$ and $r_j$
$s(p_i, r_j)$	The strength between $p_i$ and $r_j$
$t$	The dialog turn index

### 3.2.2. Chatting

Chatting is the most fundamental function of a chatbot model, which outputs in the form of natural language based on the dialog history. Figure 2 illustrates the workflow of the chatting function using the example input message “What is chatbot?”. The chatting function can be divided into two sessions: *response retrieval* and *update process*.

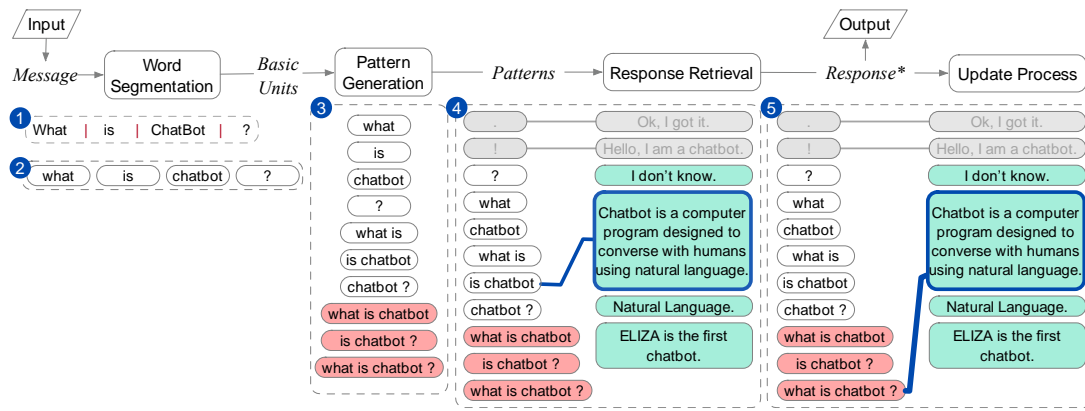


Figure 2. The workflow of the chatting function, illustrated with the input message "What is a chatbot?". The sentence is first divided into four basic units (words and punctuation). Following that, these basic units are used to generate related patterns. And links of those generated patterns are used to retrieve response\*, which is also the output of the chatbot. The links between generated patterns and response\* will be strengthened in the final stage.

In the *response retrieval* session, the user's input message is first decomposed into basic units. Then we define a heuristic function to generate patterns based on the basic units:

1. Each basic unit is a pattern.
2. If two patterns are adjacent to each other in the message, they are concatenated to form a new pattern.

Thus, for an input message, we can generate a set of patterns  $P = \{p_i\}(i = 1, \dots, n)$  according to the above rules. Then through the links between patterns and responses, we can get a set of  $m$  related responses  $R = \{r_j\}(j = 1, \dots, m)$  from the response pool. For the link between pattern  $p_i$  and response  $r_j$  at dialog turn  $t$ , we represent its weight by  $w_t(p_i, r_j)$ . Besides, we define the strength between pattern  $p_i$  and response  $r_j$  as follows:

$$s_t(p_i, r_j) = \frac{w_t(p_i, r_j)}{\sum_{k=1}^m w_t(p_i, r_k)} \times l(p_i) \quad (1)$$

where  $l(p_i)$  is the number of basic units in  $p_i$ . Then, we define the goodness of response  $r_j$  at dialog turn  $t$  as follows:

$$g_t(r_j) = \max_{i \in \{1, \dots, n\}} s_t(p_i, r_j) \quad (2)$$

Based on the goodness  $g_t(r_j)$ , the best response  $r^*$  is recognized as follows:

$$r^* = \operatorname{argmax}_{r_j \in \{r_1, \dots, r_m\}} g_t(r_j) \quad (3)$$

Then, the best response  $r^*$  is selected as the output response to the user.

After retrieval, here comes the *update process* session. When the response  $r^*$  is approved by the user, all patterns in set  $P$  and responses in set  $R$  will be linked. Then the link activation is performed for all links of  $r^*$  by updating the weight  $w_t(p_i, r^*)$  ( $i = 1, \dots, n$ ) as follows:

$$w_{t+1}(p_i, r^*) = w_t(p_i, r^*) + 1, \quad (4)$$

The weights of links between all patterns and responses are initialized as follows:

$$w_0(p_i, r_j) = 0 \quad (5)$$

where  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . By the *update process*, the algorithm continues to self-renewal during chatting which enables self-learning without supervision.

### 3.2.3. Nurturing Functions

Based on social scientists' research on learning in human communication [57], we further design three nurturing functions for the proposed nurturable pattern matching learning algorithm: *Change*, *Like*, and *Analysis*. This section discusses the details.

Via the *Change* function, users can override the chatbot's response  $r^*$  with a preferred response  $r'$ . If  $r'$  is a new response that is not in the current response set  $R$  the chatbot will add  $r'$  into  $R$ . The change function will increase the link weight associated with the preferred response  $r'$  iteratively until the goodness of  $r'$  exceeds that of  $r^*$ . The algorithm of the change function is provided in Algorithm 1.

Algorithm 1 Change function	Algorithm 2 Like function
<b>Input:</b> $r^*, \psi, f(\cdot), w_t(p_i, r')$ <b>Output:</b> $w'_t(p_i, r')$ 1: <b>if</b> $r' \notin R$ <b>then</b> 2: $R = R \cup \{r'\}$ 3: <b>end if</b> 4: compute $g_t(r^*)$ by Eq. (2) 5: compute $g_t(r')$ by Eq. (2) 6: <b>while</b> $g_t(r') \leq g_t(r^*)$ <b>do</b> 7: <b>for</b> $i = 1, \dots, n$ <b>do</b> 8: $w_t(p_i, r') = w_t(p_i, r') + 1$ 9: <b>end for</b> 10: compute $g_t(r')$ by Eq. (2) 11: <b>end while</b> 12: <b>for</b> $i = 1, \dots, n$ <b>do</b> 13: $w'_t(p_i, r') = w_t(p_i, r')$	<b>Input:</b> $r^*, \psi, f(\cdot), w_t(p_i, r^*)$ <b>Output:</b> $w'_t(p_i, r^*)$ 1: <b>for</b> $v = 1, \dots, \psi$ <b>do</b> 2: <b>for</b> $i = 1, \dots, n$ <b>do</b> 3: $w_t(p_i, r^*) = w_t(p_i, r^*) + f(v)$ 4: <b>end for</b> 5: <b>end for</b> 6: <b>for</b> $i = 1, \dots, n$ <b>do</b> 7: $w'_t(p_i, r') = w_t(p_i, r')$ 8: <b>end for</b> 9: <b>return</b> $w'_t(p_i, r')$
	Algorithm 3 Analysis function
	<b>Input:</b> response of ChatBot $r^*$ <b>Output:</b> pattern $p^*$ that results in $r^*$



14: <b>end for</b> 15: <b>return</b> $w'_t(p_i, r')$	1: <b>for</b> $i = 1, \dots, n$ <b>do</b> 2:     compute $s_t(p_i, r^*)$ by Eq. (1) 3: <b>end for</b> 4: $p^* = \underset{p_i \in \{p_1, \dots, p_n\}}{\operatorname{argmax}} s_t(p_i, r^*)$ 5: <b>return</b> $p^*$
---	---

Via the *like* function, user can encourage the good response of the chatbot. If a response is liked by the user, the link weight associated with that response is increased, and the chatbot is more likely to output that response in the future when a similar message is given. The like function can be used multiple times to enhance its effectiveness. The algorithm of the like function is provided in Algorithm 2. The weighted function  $f(\cdot)$  in Algorithm 2 can be determined by the user. For example, a simple case is linear weighting:  $f(v) = 1$  for  $v = 1, \dots, \psi$ , which means that every *like* is equally important.

*Analysis* is a function used to provide the explanation of the chatbot's response. The user can view the pattern that results in the output response  $r^*$  via the analysis function. The algorithm of *Analysis* is provided in Algorithm 3.

### 3.3. LightBlue – Chatbot Nurture Platform

In this section, we describe how we integrate the learning algorithm and nurture functions into the nurture platform – LightBlue. The details of implementation are explained, including its interface, interaction process and the process of user nurturing. With LightBlue end-users are able to nurture the chatbot while chatting with it.

#### 3.3.1. Interface

Figure 3 shows the interface of the LightBlue platform, which includes the chatting window, function buttons, chatbot image, and vocabulary. It can be mainly divided into two parts. On the left is the user-chatbot conversation interface, including the conversation history, the chatting box, nurture functions and common operation buttons. On the right shows the state of the learning algorithm, including the chatbot image and learned vocabulary. As users nurture the chatbot, we visualize the evolution of the learning model through the chatbot's image and vocabulary repertoire.

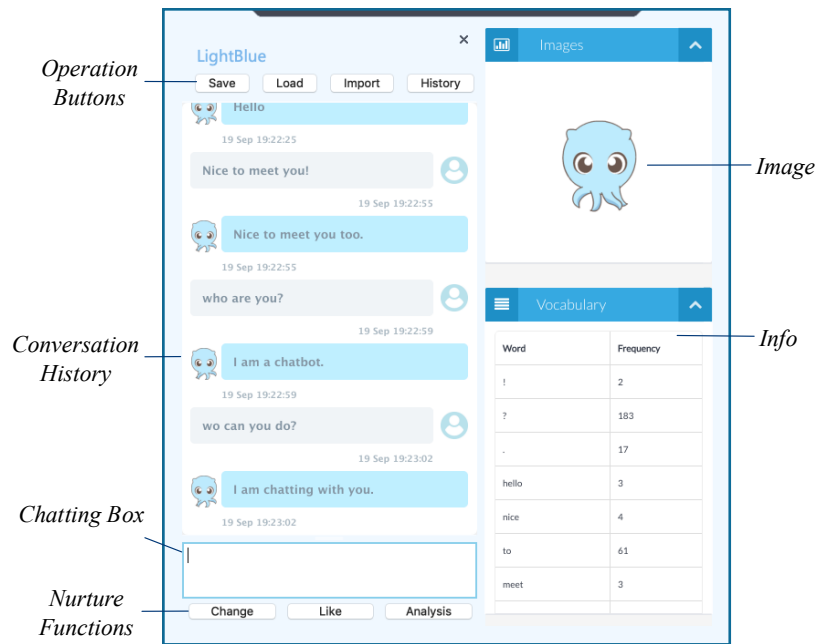


Figure 3. The interface of LightBlue. The left part contains the chatting window and functional buttons; and the right part contains an image of the chatbot and its vocabulary, which visualizes the progress of the nurture.

### 3.3.2. Interaction Process

This subsection describes how users interact with the underlying algorithm through the LightBlue platform. When the program is launched, a new initialized chatbot is automatically created. The *Save* and *Load* buttons allow you to save the current chatbot model and load an existing chatbot model. Users use the chatting box to send messages, and the chatbot responds using the chatting function described previously. The new dialogs will then be included in the conversation's history. As the conversation goes, the algorithm will learn and grow. And the right anthropomorphic chatbot image reflects the changing state. The right lower box contains more specific information.

### 3.3.3. Nurture Functions

We design three buttons to assist users in nurturing the chatbot based on the nurture function defined in the previous section. Users can click *like* button to express their preferences for the chatbots' responses. After receiving likes from users, the chatbot uses Algorithm 2 to increase the link weight of the preferred response. To change a response, the user can enter a preferred answer in the chatting box and click the *change* button. Then Algorithm 1 will be performed to revise the response. The *Analysis* button allows the user to determine why the chatbot provides a particular response. An example result of the analysis are showed in Figure 4, explaining the relevant information leading to the response.

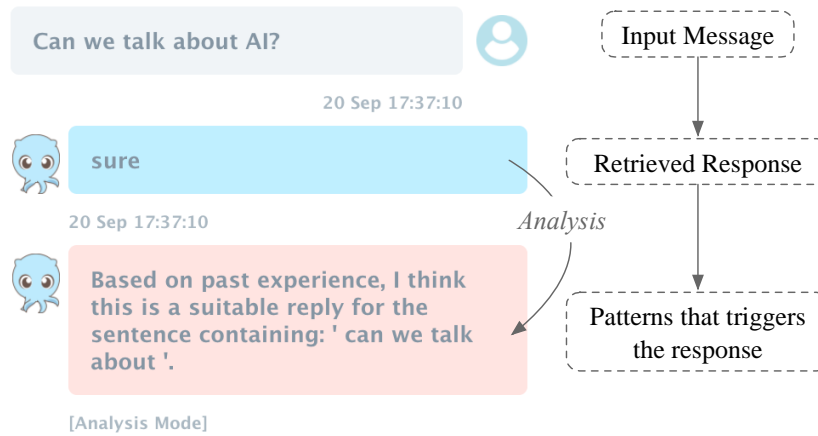


Figure 4. An example of the analysis function in LightBlue. The feedback from analysis shows the most contributing pattern “can we talk about” that leads to the retrieved response “sure”.

## 4. RESULT

We conduct experiments in both closed-domain and open-domain settings to demonstrate the feasibility of the nurture framework for developing personal chatbots. In the closed-domain experiment, non-professional users are required to communicate with chatbots about a given topic and material. After verifying the effectiveness of the nurture framework, we further design an open-domain experiment to test if chatbots can adapt to different users and be applied to different task scenarios through nurture.

### 4.1. Initialization

As showed in Figure 5, every chatbot is initialized with three patterns (i.e. period, exclamation mark, and question mark), three responses (i.e. “Ok, I got it”, “Hello, I am a chatbot”, and “I don’t know”) and corresponding links between them. Initially, all links have a weight of 1. Thus, in the initial state, the vocabulary contains only three words: period (.), exclamation mark (!), and question mark (?).

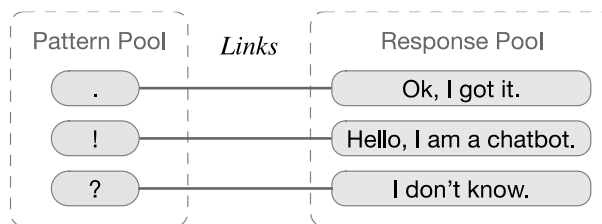


Figure 5. The initial state of Pattern Matching Model. It contains three patterns, three responses, and corresponding links between them.

### 4.2. Closed-domain Experiment

In this experiment, we selected “A brief history of artificial intelligence” as a specific topic, and provided all participants with a 2,000-word article and a 1,000-word vocabulary. Participants were asked to nurture their chatbots to use words in the vocabulary to answer questions related to the content of the article. After two weeks of training, 34 chatbots were collected. The

conversation history between users and chatbots is typically measured in thousands of turns, with a vocabulary ranging from hundreds to thousands.

#### 4.2.1. Response Quality

To determine whether the chatbot is capable of accurately answering questions, we invite three volunteers to read the article and then pose questions about the content. According to the content relevance, grammar correctness, and vocabulary, we finally screen 132 questions and randomly assign three chatbots to answer each question.

To grade the quality of the chatbots' responses, three categories are used: *correct*, *partly correct or relevant to the question*, and *not relevant to the question*. Three volunteers are assigned to categorize each response. To avoid ambiguity, we count only the responses that were assigned to the same category by all three volunteers.

Then we define the correct rate as the proportion of correct responses to all responses, and relative rate as the proportion relevant responses (including correct responses) to all the responses. The evaluation results of all 357 responses are shown in the Figure 6, where 83.5% responses are related to the questions and 66.4% responses correct with the best chatbot has the relative rate of 91.7% and correct rate 83.3%.



Figure 6. 66.4% of responses from 34 chatbots to 132 questions are correct, are 83.5% are related to the questions.

#### 4.2.2. Growth

To examine the growth ability of chatbots, we choose a chatbot with an average performance and also an average length of dialog history (about 3500 turns) from the 34 collected chatbots. Then we replicate the nurture process and test it with the same questions across time periods. With the increase of chatting dialogues, the chatbot's performance has improved (see Figure 7).

Due to the fact that users place a different emphasis on nurture, the growth rates of chatbots may vary at different stages. However, there is no obvious bottleneck of the growth, which demonstrates the robustness of our algorithm to different nurture methods and the characteristics of sustainable growth.

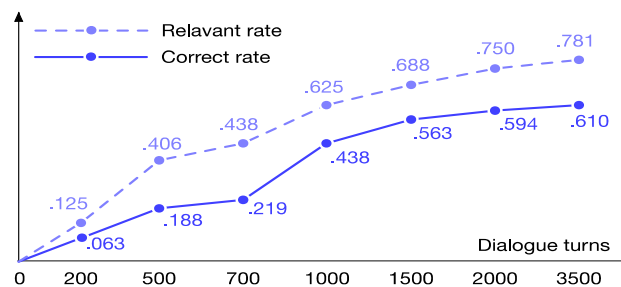


Figure 7. The response quality of the chatbot as nurture process goes.

### 4.2.3. Diversity

Even when the topic and content are the same, people's expressions vary widely due to the diversity of human language. We calculate the average rate of overlap between the conversation histories of each chatbot. Figure 8 shows the result, with an average overlap rate of 8.03%. And the overlap rate between conversation corpus and testing questions is even lower (less than 1% on average). Achieving high response quality, our proposed nurture framework demonstrates its prominent superiority in dealing with various conversational styles. Table 2 shows several examples where chatbots give diverse but reasonable responses to questions with objective answers in the material. These two questions are basic and simple questions which have clear answers in the article, and chatbots give out diverse responses. It's also an important observation that the same chatbot maintains a consistent conversational style. For example, Chatbot 14 tends to answer the question in a more casual manner, using examples or descriptions.

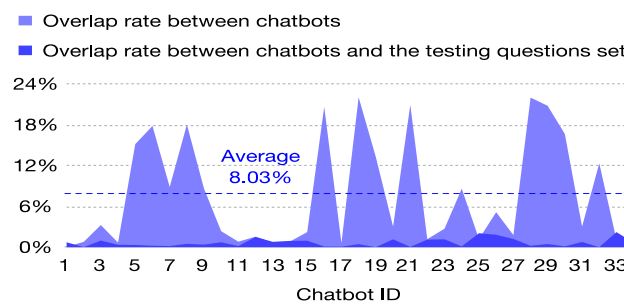


Figure 8. Overlap rate of chatbots' conversation history; and overlap rate between chatbots' conversation history and testing question set.

### 4.3. Open-Domain Experiment

In the open-domain experiment, we distribute LightBlue platform to 101 participants (mean age = 20.7, 36 females) for 21 days. The participants can nurture chatbots according to their preferences with unlimited topics, while basic greeting skills are required for all chatbots. To test the nurtured chatbots fairly and easily, we also provided a restricted vocabulary of 2024 words drawn from the common English vocabulary. After the experiment, we collected 98 chatbots, with conversation histories ranging from hundreds to tens of thousands turns.

Table 2. Diverse answers to the same questions related to the material. For the question with a clear answer in the article, human-nurtured chatbots are capable of providing diverse yet reasonable responses.

<b>Question</b>	<b>What is a virtual assistant?</b>
<i>Chatbot 10</i>	Virtual assistant is capable of accessing data and answering questions.
<i>Chatbot 9</i>	Serving as virtual assistants, the majority of these chatbots perform customer service or customer engagement functions, answering questions or offering suggestions for purchase.
<i>Chatbot 14</i>	Siri 2010, Google Now 2012, Alexa 2015, Cortana 2015, and Google Assistant 2016.
<b>Question</b>	<b>What is AI?</b>
<i>Chatbot 8</i>	It is the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

<i>Chatbot 21</i>	Artificial Intelligence: a.k.a machine intelligence, is intelligence demonstrated by the machine to display human behaviour.
<i>Chatbot 14</i>	If the user doesn't know they're speaking to a bot.

#### 4.3.1. Diversity

We test 98 chatbots using six commonly used English greetings. The results in Table 3 shows that, on average, 85% of responses are appropriate, and the average overlap rate of responses is about 14%. This reflects the diversity of training data and training methods from different users, which results in the diversity of chatbots.

Table 3. Correct rate and overlap rate of responses to common greetings.

Greetings	Correct rate	Overlap rate
<i>Hello!</i>	0.88	0.32
<i>Good morning!</i>	0.88	0.18
<i>Nice to see you.</i>	0.79	0.07
<i>How are you?</i>	0.93	0.06
<i>How's everything?</i>	0.86	0.04
<i>Have a nice day!</i>	0.77	0.21
<b>Average</b>	<b>0.85</b>	<b>0.14</b>

#### 4.3.2. Consistency

Nowadays, one of the most common challenges or complaints about chatbots is their lack of consistency. That is, responses to different questions may be inconsistent. We select 8 common topics which may result in inconsistency, including food, education, weather, mood, hobbies, color, living information and friends. We ask more than ten questions about each topic, and the same questions may be repeated to the chatbot if they yield different responses. If all of the responses to a specific topic are consistent, the chatbot is considered consistent in that topic. Given that not all chatbots are capable of discussing the eight topics, we only examine those reasonable and related responses.. It is worth noting that for *LightBlue*, we test all the nurtured chatbots, and the consistent answers are those consistent within one chatbot. We regard *LightBlue* as a consist chatbot on one topic if all answers are consist (which can be different for different chatbots). For the chatbot from the Internet, we only test the model once.

Table 4 summarizes the results of the consistency evaluation, demonstrating that human-nurtured chatbots (*LightBlue*) always provide consistent responses, whereas the average consistency rate for general Internet chatbots is low (below 50%). As a task-oriented chatbot, *Siri* also maintains a consistent personality. However its topic scope is somehow limited compared to others since it is not designed for general chatting. And in many cases, especially unrelated to pre-defined tasks, it tends to respond with “*I'm sorry, I'm afraid I don't have an answer to that.*”.

Table 4. Consistent rates of responses to contextual topics.

Chatbot	Topics	Consistent Answers	Consistent Rate
<i>LightBlue</i>	8	8	<b>1</b>
<i>Mitsiku [58]</i>	8	3	0.38
<i>Cleverbot [59]</i>	8	3	0.38

<i>Eviebot [60]</i>	8	4	0.50
<i>SimSimi [61]</i>	8	1	0.13
<i>Siri [62]</i>	5	5	<b>1</b>

### 4.3.3. Adaptability to different task scenarios

In addition to basic greetings, users can train chatbots to perform a variety of tasks based on their preferences. By their self-feedback after the experiment, the application scenarios designed for personal chatbots include domain knowledge (e.g. math), entertainment (e.g. playing games, telling jokes), personal assistant (e.g. notes), adviser (e.g. career advice), customer service (e.g. booking tickets), chit-chat, and emotional companions (see Table 5). The majority of users expressed satisfaction with their chatbots' performance in the designed scenario. This again demonstrates the robustness of the nurture framework in adapting to a variety of tasks, corpora, and users.

Table 5. Customized tasks in open-domain experiment.

<b>Customized tasks</b>	<b>Number of chatbots</b>	<b>Average history length (turns)</b>
<i>Chit-chat</i>	50	1020
<i>Personal assistant</i>	4	2613
<i>Emotional companion</i>	11	4978
<i>Domain knowledge</i>	15	5038
<i>Entertainment</i>	4	1198
<i>Adviser</i>	5	1969
<i>Customer service</i>	6	3104

### 4.3.3. Social Bond

While it is difficult to quantify the social bond directly, it is reflected in the depth, quantity, and sentiment of dialogues. To better explore the connections between users and chatbots, we examine only chatbots that have been nurtured for more than two weeks.

Throughout the experiment, users are asked to record a new session whenever they change the subject or end the conversation. During two or three weeks of nurturing, our chatbots averaged 21 conversation-turns per session (CPS), which is larger than the CPS of most chatbots conversations between users and nearly equal to the CPS of the 6th generation of XiaoIce [63]. For comparison, the CPS of the 1st generation of XiaoIce is 5. Because each session's conversational content is focused on a single subject, the large CPS indicates that the conversation has progressed to a certain level of depth.

We further examine how the amount of conversation changes throughout the nurture process. To eliminate the influence of start and end periods, we take a week in the middle as the observation window. Figure 9 shows the daily conversation turns between users and chatbots during 19th Nov to 25<sup>th</sup> Nov. With the process of nurture, the conversation between users and chatbots increases gradually. It also indicates that the users' interests are gradually increasing and their bonds are getting tighter.

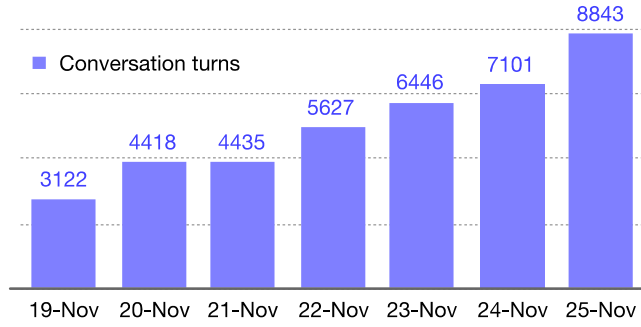


Figure 9. The number of daily conversation turns between users and chatbots from 19th Nov to 25th Nov.

To analyse the sentiment of the conversation, we examine the frequency with which positive and negative words are used. As a reference, the Sentiment Lexicon [64] is used, which contains a list of approximately 6800 positive and negative opinion or sentiment words in English. For simplicity, we refer to the conversation history of all chatbots nurtured on the LightBlue platform as the LightBlue Corpus. As shown in Table 6, we compare the positive and negative word frequencies in the LightBlue Corpus to those in other commonly used datasets for chatbot training. The result shows that datasets containing human-chatbot dialogues (LightBlue and ConvAI) have a lower Negative-Positive ratio, implying a higher level of positivity. And LightBlue has the highest Positive Frequency and the lowest Negative-Positive ratio. It is noteworthy that for both positive frequency and negative frequency, LightBlue Corpus is higher than other datasets. The high frequency of positive and negative words indicates that users are more emotionally engaged when interacting in a nurture-based context.

Table 6. The frequency of positive words, the frequency of negative words, and the Negative-Positive ratio in LightBlue and public datasets.

Dataset	Positive Frequency	Negative Frequency	N/P Ratio
<i>LightBlue</i>	0.244	0.039	0.159
<i>Corpus ConvAI [65]</i>	0.051	0.012	0.233
<i>Twitter [66]</i>	0.038	0.028	0.734
<i>Cornell movie dialogues [67]</i>	0.019	0.017	0.896
<i>Open subtitles [68]</i>	0.028	0.020	0.724

## 5. DISCUSSION

This chapter summarizes the features of LightBlue and discusses its limitations.

### 5.1. Features of LightBlue Chatbot Nurture Platform

Through LightBlue chatbot nurture platform, non-professional users can easily nurture and customize a chatbot. Despite of the simple design of the learning algorithm and small amount of dialogues, human-nurtured chatbots achieve satisfying performance in both closed-domain and open-domain experiments. Below we further summarize the features of LightBlue and how the nurture framework benefits developing personal chatbots.



### 1. Adaptability and Growth

Users can nurture chatbots to perform a variety of tasks via the LightBlue platform, with application scenarios ranging from entertainment to personal assistants. The learning algorithm also demonstrates growth characteristics during the nurturing process. And there's no obvious bottleneck of the growth trend as the conversation corpus grows. The capacity for sustained growth will also facilitate long-term interaction between users and chatbots.

### 2. Robustness

Through the two experiments (closed-domain and open-domain), we find that LightBlue is robust to a variety of tasks, nurture corpora, nurture methods, and end-users. That is, it does not require professionals or particular process to nurture a chatbot.

### 3. Diversity and Consistency

The main challenges faced by nowadays general chatbots are uniform answer and inconsistent personality. Our chatbots, when nurtured by different end users, can be naturally personalized and maintain diversity, overcoming the uniform responses caused by big data. Additionally, because the conversation corpus originates from a single user, it naturally has a distinct style and is less prone to inconsistencies.

### 4. Social functions

According to statistics, the public will lose interest in a general chatbot after a short interaction. However, the participants in our two experiments continue to communicate with their chatbots for 2-3 weeks, with conversations ranging from hundreds to tens of thousands of turns. Not only does their conversation frequency increase over time, but they also maintain a high CPS (around 21), which is significantly higher than the average user-chatbot conversation on the Internet. And the interaction involves more emotion than common dialogues. This suggests a new social role for chatbots in our lives, lays the groundwork for long-term user-chatbot relationships, and even facilitates emotion functions such as affective bonding.

## 5.2. Limitations

In this work, we only implement one learning algorithm – the pattern matching model for nurture. As a retrieval algorithm, It has inherent limitations such as the inability to generate new questions and answers. At the same time, the hyperparameters and initialization of the model may influence the experimental results. Further investigation of experimental environments is necessary.

Because the experiment participants are all college students who are reasonably familiar with computer operations and share similar cultures and habits, the experiment results may be influenced by group characteristics. Additional subjects with diverse backgrounds, such as middle school students, workers, and the elderly, will be needed to evaluate the robustness of the LightBlue platform.

## 6. CONCLUSION

This work introduces “nurture” as a new mode of interaction between humans and chatbots, and defines the nurture framework for developing personal chatbots. We further propose a novel learning algorithm and design nurture functions accordingly. Then we implement the LightBlue

chatbot nurture platform and distribute it to non-professional users. Experiments in both closed-domain and open-domain are designed to verify the effectiveness of the nurture framework. Through the LightBlue platform, non-professional users can develop diverse and personalized chatbots. They're able to perform diverse tasks with small amount of training data while remaining robust to end-users, application scenarios, and nurture methods. Nurture, as an interaction mode, demonstrates a new way for chatbots to perform social functions and promotes long-term connections between users and chatbots.

## ACKNOWLEDGEMENTS

This work is supported by Innovation and Technology Fund from the Innovation and Technology Commission of Hong Kong, project number ITS/110/19.

## REFERENCES

- [1] S. A. Abdul-Kader and J. Woods, "Survey on chatbot design techniques in speech conversation systems," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 7, 2015.
- [2] B. A. Shawar and E. Atwell, "Chatbots: are they really useful?," in *Ldv forum*, vol. 22, pp. 29–49, 2007.
- [3] K. Ramesh, S. Ravishankaran, A. Joshi, and K. Chandrasekaran, "A survey of design techniques for conversational agents," in *International Conference on Information, Communication and Computing Technology*, pp. 336–350, Springer, 2017.
- [4] R. Dale, "The return of the chatbots," *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817, 2016.
- [5] S. Chandel, Y. Yuying, G. Yujie, A. Razaque, and G. Yang, "Chatbot: efficient and utility-based platform," in *Science and Information Conference*, pp. 109–122, Springer, 2018.
- [6] R. M. Thomas, S. Punna, M. C. K. K. Reddy, and B. RamanaMurthy, "Survey on artificially intelligent chatbot," *Journal of Applied Science and Computations*, vol. 6, no. 1, pp. 85–94, 2019.
- [7] M. Nuruzzaman and O. K. Hussain, "A survey on chatbot implementation in customer service industry through deep neural networks," in *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, pp. 54–61, IEEE, 2018.
- [8] E. Luger and A. Sellen, "Like having a really bad pa: the gulf between user expectation and experience of conversational agents," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5286–5297, ACM, 2016.
- [9] B. Chantarotwong, "The learning chatbot," *Final year project.[Online]: <http://courses.ischool.berkeley.edu/i256/f06/projects/bonniejc.pdf>*, 2006.
- [10] R. Yan, "' chitty-chitty-chat bot": Deep learning for conversational ai.," in *IJCAI*, vol. 18, pp. 5520–5526, 2018.
- [11] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, *et al.*, "What makes a good conversation?: Challenges in designing truly conversational agents," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 475, ACM, 2019.
- [12] F. Daniel, M. Matera, V. Zaccaria, and A. Dell'Orto, "Toward truly personal chatbots: on the development of custom conversational assistants," in *Proceedings of the 1st International Workshop on Software Engineering for Cognitive Services*, pp. 31–36, ACM, 2018.
- [13] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2204–2213, 2018.
- [14] C. Machinery, "Computing machinery and intelligence-am turing," *Mind*, vol. 59, no. 236, p. 433, 1950.
- [15] R. J. Sternberg and E. Grigorenko, *Intelligence, heredity and environment*. Cambridge University Press, 1997.
- [16] S. Scarr, "Biological and cultural diversity: The legacy of darwin for development," *Child Development*, vol. 64, no. 5, pp. 1333–1353, 1993.

- [17] R. Winston and R. Chicot, "The importance of early bonding on the long-term mental health and resilience of children," *London journal of primary care*, vol. 8, no. 1, pp. 12–14, 2016.
- [18] M. A. Bassiri, "Interactional feedback and the impact of attitude and motivation on noticing l2 form," *English Language and Literature Studies*, vol. 1, no. 2, p. 61, 2011.
- [19] B. Hancock, A. Bordes, P.-E. Mazare, and J. Weston, "Learning from dialogue after deployment: Feed yourself, chatbot!," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3667–3684, 2019.
- [20] J. Berko-Gleason, "Insights from child language acquisition for second language loss," *The loss of language skills*, vol. 23, no. 1, pp. 12–23, 1982.
- [21] M. G. Werts, M. Wolery, A. Holcombe, and D. L. Gast, "Instructive feedback: Review of parameters and effects," *Journal of Behavioral Education*, vol. 5, no. 1, pp. 55–75, 1995.
- [22] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
- [23] K. M. Colby, *Artificial paranoia: a computer simulation of paranoid process*. Pergamon Press, 1975.
- [24] J. L. Hutchens, "How to pass the turing test by cheating," *School of Electrical, Electronic and Computer Engineering research report TR97-05*. Perth: University of Western Australia, 1996.
- [25] B. Batacharia, D. Levy, R. Catizone, A. Krotov, and Y. Wilks, "Converse: a conversational companion," in *Machine conversations*, pp. 205–215, Springer, 1999.
- [26] B. A. Shawar and E. Atwell, *A comparison between Alice and Elizabeth chatbot systems*. University of Leeds, School of Computing research report 2002.19, 2002.
- [27] B. AbuShawar and E. Atwell, "Alice chatbot: Trials and outputs," *Computación y Sistemas*, vol. 19, no. 4, pp. 625–632, 2015.
- [28] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino, and Y. Matsuo, "Towards an open-domain conversational system fully based on natural language processing," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 928–939, 2014.
- [29] P. B. Brandtzaeg and A. Følstad, "Why people use chatbots," in *International Conference on Internet Science*, pp. 377–392, Springer, 2017.
- [30] A. L. Bagousse, "Customer Service Chatbots Are Redefining Customer Service," 2021. <https://blog.smart-tribune.com/>
- [31] R. Yan, Y. Song, X. Zhou, and H. Wu, "Shall i be your chat companion?: Towards an online human-computer conversation system," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 649–658, ACM, 2016.
- [32] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, pp. 3104–3112, 2014.
- [33] L. Shang, Z. Lu, and H. Li, "Neural responding machine for short-text conversation," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1577–1586, 2015.
- [34] Z. Yin, K.-h. Chang, and R. Zhang, "Deepprobe: Information directed sequence understanding and chatbot design via recurrent neural networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2131–2139, ACM, 2017.
- [35] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015.
- [36] A. Parikh, O. Täckström, D. Das, and J. Uszkoreit, "A decomposable attention model for natural language inference," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2249–2255, 2016.
- [37] H. Zhou, M. Huang, T. Zhang, X. Zhu, and B. Liu, "Emotional chatting machine: Emotional conversation generation with internal and external memory," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [38] I. V. Serban, R. Lowe, P. Henderson, L. Charlin, and J. Pineau, "A survey of available corpora for building data-driven dialogue systems: The journal version," *Dialogue & Discourse*, vol. 9, no. 1, pp. 1–49, 2018.
- [39] M. Dibitonto, K. Leszczynska, F. Tazzi, and C. M. Medaglia, "Chatbot in a campus environment: Design of lisa, a virtual assistant to help students in their university life," in *International Conference on Human-Computer Interaction*, pp. 103–116, Springer, 2018.

- [40] J. Zhang, N. M. Thalmann, and J. Zheng, “Combining memory and emotion with dialog on social companion: A review,” in *Proceedings of the 29th international conference on computer animation and social agents*, pp. 1–9, ACM, 2016.
- [41] R. Yan, Y. Song, and H. Wu, “Learning to respond with deep neural networks for retrieval-based human-computer conversation system,” in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 55–64, 2016.
- [42] T.-H. K. Huang, J. C. Chang, and J. P. Bigham, “Evorus: A crowd-powered conversational assistant built to automate itself over time,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 295, ACM, 2018.
- [43] J. Li, A. H. Miller, S. Chopra, M. Ranzato, and J. Weston, “Learning through dialogue interactions by asking questions,” *arXiv preprint arXiv:1612.04936*, 2016.
- [44] V. Rieser and O. Lemon, *Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation*. Springer Science & Business Media, 2011.
- [45] H. Zhang, H. Yu, and W. Xu, “Listen, interact and talk: Learning to speak via interaction,” *arXiv preprint arXiv:1705.09906*, 2017.
- [46] B. Liu, G. Tür, D. Hakkani-Tür, P. Shah, and L. P. Heck, “Dialogue learning with human teaching and feedback in end-to-end trainable task-oriented dialogue systems,” in *NAACL-HLT*, 2018.
- [47] J. E. Weston, “Dialog-based language learning,” in *Advances in Neural Information Processing Systems*, pp. 829–837, 2016.
- [48] J. Li, A. H. Miller, S. Chopra, M. Ranzato, and J. Weston, “Dialogue learning with human-in-the-loop,” *arXiv preprint arXiv:1611.09823*, 2016.
- [49] R. Price, “Microsoft is deleting its AI chatbot's incredibly racist tweets,” *Business Insider*, 2016.
- [50] M. Jain, R. Kota, P. Kumar, and S. N. Patel, “Convey: Exploring the use of a context view for chatbots,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 468, ACM, 2018.
- [51] H. Candello, C. Pinhanez, and F. Figueiredo, “Typefaces and the perception of humanness in natural language chatbots,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3476–3487, ACM, 2017.
- [52] B. Huber, D. McDuff, C. Brockett, M. Galley, and B. Dolan, “Emotional dialogue generation using image-grounded language models,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 277, ACM, 2018.
- [53] Y. Matsuyama, A. Bhardwaj, R. Zhao, O. Romeo, S. Akoju, and J. Cassell, “Socially-aware animated intelligent personal assistant agent,” in *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue*, pp. 224–227, 2016.
- [54] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, *et al.*, “Simsensei kiosk: A virtual human interviewer for healthcare decision support,” in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 1061–1068, International Foundation for Autonomous Agents and Multiagent Systems, 2014.
- [55] J. Seering, M. Luria, G. Kaufman, and J. Hammer, “Beyond dyadic interactions: Considering chatbots as community members,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 450, ACM, 2019.
- [56] Z. Ashktorab, M. Jain, Q. V. Liao, and J. D. Weisz, “Resilient chatbots: Repair strategy preferences for conversational breakdowns,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 254, ACM, 2019.
- [57] E. A. Schegloff, G. Jefferson, and H. Sacks, “The preference for self-correction in the organization of repair in conversation,” *Language*, vol. 53, no. 2, pp. 361–382, 1977.
- [58] mitsuku, “mitsuku.” Web Program, Sep 2019. Retrieved Sep 21, 2019 from <https://www.pandorabots.com/mitsuku/>.
- [59] Cleverbot, “Cleverbot.” Web Program, Sep 2019. Retrieved Sep 21, 2019 from <https://www.cleverbot.com/>.
- [60] Eviebot, “Eviebot.” Web Program, Sep 2019. Retrieved Sep 21, 2019 from <https://www.eviebot.com/en/>.
- [61] simsimi, “simsimi.” Web Program, Sep 2019. Retrieved Sep 21, 2019 from <https://www.simsimi.com/Home>.

- [62] J. R. Bellegarda, “Spoken language understanding for natural interaction: The siri experience,” in *Natural Interaction with Robots, Knowbots and Smartphones*, pp. 3–14, Springer, 2014.
- [63] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, “The design and implementation of xiaoice, an empathetic social chatbot,” *Computational Linguistics*, vol. 46, no. 1, pp. 53–93, 2020.
- [64] B. Liu and M. Hu, “Opinion mining, sentiment analysis, and opinion spam detection,” *Dosegljivo: https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html#lexicon.[Dostopano 15. 2. 2016]*, 2004.
- [65] V. Logacheva, M. Burtsev, V. Malykh, V. Polulyakh, and A. Seliverstov, “Convai dataset of topic-oriented human-to-chatbot dialogues,” in *The NIPS’17 Competition: Building Intelligent Systems*, pp. 47–57, Springer, 2018.
- [66] C. Tan, L. Lee, and B. Pang, “The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter,” in *Proceedings of ACL*, 2014.
- [67] C. Danescu-Niculescu-Mizil and L. Lee, “Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs.,” in *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*, 2011.
- [68] P. Lison and J. Tiedemann, “Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles,” 2016.