Check for updates

Author Manuscript

PLEASE READ ALL 'COMMENT' TEXT BEFORE PREPARING YOUR ARTICLE. If you do not see the Comments, select View > Print Layout. Please delete them before submitting (Review > Delete All Comments in Document) so that peer reviewers see a clean copy of the manuscript.

Remember that you are writing for an interdisciplinary audience. Please be sure to discuss interdisciplinary themes, issues, debates, etc. where appropriate. Note that the WIREs are forums for review articles, rather than primary literature describing the results of original research.

If you have any questions, contact your editorial office.

WIREs Wiley Interdisciplinary Reviews

Article Title: A critical review of state-of-the-art chatbot designs and applications

Article Type:

○ OPINION ○ PRIMER ◉ OVERVIEW

○ ADVANCED REVIEW ○ FOCUS ARTICLE ○ SOFTWARE FOCUS

Authors:

[List each person's full name, ORCID iD, affiliation, email address, and any conflicts of interest.

Copy rows as necessary for additional authors. Please use an asterisk (*) to indicate the corresponding author.]

First author name Bei LUO, beiluo2-c@my.cityu.edu.hk, 0000-0001-8393-6506 Department of Information Systems City University of Hong Kong Tat Chee Avenue, Kowloon Hong Kong SAR.
Second author name Raymond Y.K. Lau*, raylau@cityu.edu.hk, 0000-0002-5751-4550

~~This is the author manuscript accepted for publication and has undergone full peer review but~~ has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/widm.1434

This article is protected by copyright. All rights reserved.

Department of Information Systems City University of Hong Kong Tat Chee Avenue,

Kowloon Hong Kong SAR

Third author name

Chunping Li, cli@tsinghua.edu.cn, 0000-0002-4521-0875

School of Software, Tsinghua University, Beijing, 100084, People's Republic of China

Fourth author name

Yain-Whar Si, fstasp@um.edu.mo, 0000-0001-8468-6182

Faculty of Science and Technology, University of Macau, China

Abstract

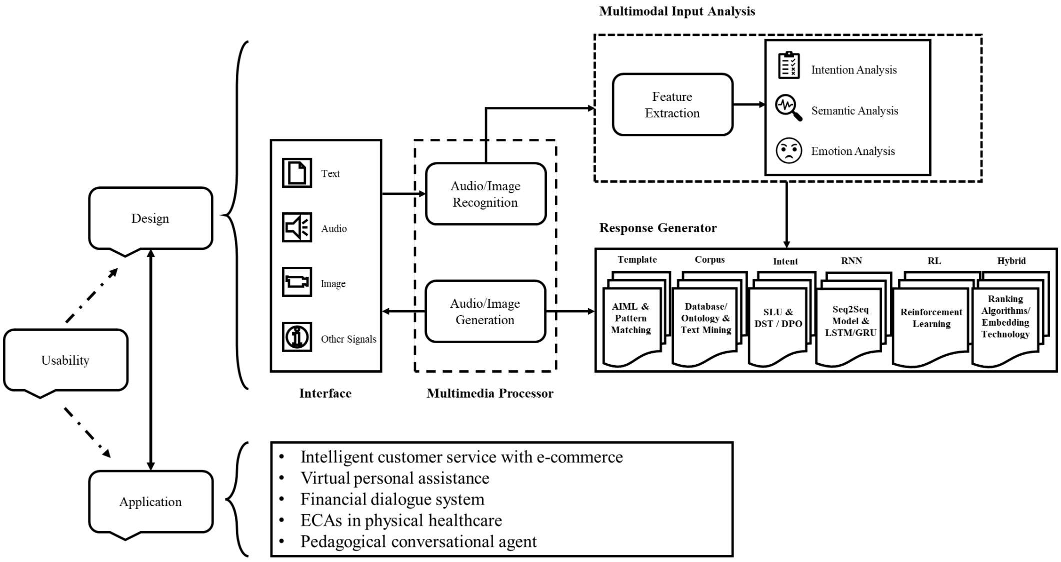
Chatbots are intelligent conversational agents that can interact with users through natural languages. As chatbots can perform a variety of tasks, many companies have committed numerous resources to develop and deploy chatbots to enhance various business processes. However, we lack an up-to-date critical review that thoroughly examines both state-of-the-art technologies and innovative applications of chatbots. In this review, we not only critically analyse the various computational approaches used to develop state-of-the-art chatbots, but also thoroughly review the usability and applications of chatbots for various business sectors. We also identify gaps in chatbot-related studies and propose new research directions to address the shortcomings of existing studies and applications. Our review advances both academic research and practical business applications of state-of-the-art chatbots. We provide guidance for practitioners to fully realize the business value of chatbots and assist in making sensible decisions related to the development and deployment of chatbots in various business contexts. Researchers interested in the design and development of chatbots can also gain useful insights from our critical review and identify fruitful research topics and future research directions based on the research gaps discussed herein.

Keywords: Chatbots; Conversational Agents; Deep Learning; Machine Learning; Chatbot Applications

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Graphical/Visual Abstract and Caption



Overview and discussion of the state-of-the-art chatbot technologies and applications.

AIML: artificial intelligence mark-up language; SLU: spoken language understanding; DST: dialogue state tracking; DPO: dialogue policy optimization; RNN: recurrent neural network; LSTM: long short-term memory; GRU: gated recurrent unit; RL: recurrent learning.

[1. INTRODUCTION]

Chatbots are conversational agents that can interact with users through natural languages and can also be described by the broader term of conversational user interfaces (Smestad, 2018). The term “chatbot” originated from “chatterbot”, a term initially proposed by Michael Mauldin in 1997 to describe robots with which humans could chat (Deryugina, 2010). The technology is also known by many other names such as dialogue system, conversational agent, conversational interface, virtual assistant, and personal assistant (Altinok, 2018).

As chatbots can perform many labour-intensive tasks at lower costs across a wide range of fields, such as customer service, healthcare, pedagogy, and personal assistance, many firms have

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

invested heavily in this burgeoning technology (Nagarhalli, Vaze, & Rana, 2020). To make reasonable investment decisions, managers need to thoroughly understand chatbot development and estimate the business value brought in by the technology. Thus, a review of the state-of-the-art chatbot technologies and applications can help track recent development.

Existing literature reviews on chatbots suffer from three notable problems. First, we lack a comprehensive chatbot classification scheme that naturally connects various chatbot technologies with different types of real-world applications. Current studies categorises chatbots into retrieval- and generation-based systems based on the mechanism of response generation (Song, Yan, Li, Zhao, & Zhang, 2016; Wu, Li, Wu, & Zhou, 2018; Wu, Wu, Xing, Li, & Zhou, 2017, 2018); the former retrieves predefined responses from static knowledge bases, and the latter generates new responses through adaptive learning and inference mechanisms. Other studies have tried to classify chatbots into task-oriented and non-task-oriented systems (Nuruzzaman & Hussain, 2018), but we believe that these classification methods are superficial and cannot capture the inherent characteristics of each type of chatbot in sufficient detail. Therefore, it is likely difficult for managers to use such superficial two-category classification schemes to determine which computational techniques are appropriate for specific business contexts.

Second, prior chatbot reviews (Abdul-Kader & Woods, 2015; Bradeško & Mladenić, 2012; Ramesh, Ravishankaran, Joshi, & Chandrasekaran, 2017) mainly discuss retrieval-based chatbots. Given the recent surge in the number of papers published on generation-based chatbots, an up-to-date review that assesses the state-of-the-art computational techniques used to design and develop generation-based chatbots is desirable.

Third, previous reviews have focused heavily on traditional computational techniques but less so on application contexts and usability analysis. Therefore, they efficiently highlight the technical characteristics of chatbots, but do not allow researchers and practitioners to fully realize the business values and potential problems associated with the deployment of chatbots for real-world applications. Therefore, there is a pressing need for an up-to-date review that covers both the state-of-the-art technologies and innovative applications of chatbots.

Here, we first provided an overview of the different computational approaches used to design and develop chatbots. We highlight the generic structure of chatbot systems based on the framework proposed by Abdul-Kader and Woods (2018), which comprises the ***interface,*** a ***multimedia***

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

processor, a ***multimodal input analysis, and*** a ***response generator.*** We then propose a new chatbot classification scheme that is relevant to the potential applications of chatbots and can facilitate future academic research. Based on the specific techniques utilized for response generation, there are six types of chatbots: template-, corpus-, intent-, recurrent neural network (RNN)-, reinforcement learning (RL)-based, and those with hybrid approaches. We also summarize the future directions of technology development in the fields of natural language processing (NLP), knowledge management, response generation, and multimedia interaction.

We next review the potential applications of chatbots. We identify the most popular fields of chatbot usage, such as intelligent customer service for e-commerce, virtual personal assistance, financial dialogue systems, embodied conversational agents (ECAs) in physical healthcare, virtual counselling services, and pedagogical conversational agents. We systematically review the existing usability studies of chatbots, which can provide insights for practitioners to fully realise the business value of chatbots. Following from our previous review of chatbot usability studies, although chatbots have some limitations in response propriety, they show promise in assisting humans in finishing specific tasks. Finally, based on a critical review of computational approaches, application fields, and usability, we identify gaps in current chatbot research and highlight future directions to guide researchers and practitioners in systematic chatbot design, development, and applications.

[2. A CRITICAL ANALYSIS OF VARIOUS COMPUTATIONAL APPROACHES USED TO DEVELOP CHATBOTS]

Abstracting the structure of a system is the first step toward understanding and further improving system designs. Usually, a structure is presented in the form of components and the relationships between them. Abdul-Kader and Woods (2015) divided the chatbot system into three constituents: a responder to control the interface, a classifier for input normalisation, and a graphmaster for knowledge storage and response retrieval. However, this structure no longer applies to the advanced chatbot designs for two reasons. First, the interaction mode between chatbots and users evolved from text only to multimedia, including voice, image, and video (Jonell, 2019; Lee, Chiang, Yeh, & Wen, 2020; Nuruzzaman & Hussain, 2018), making it necessary to add a new unit of multimedia processing to the existing structure. For example, the multimedia processing unit in a

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

voice-enabled chatbot transforms the audio signals into text and analyse either the tone or other sound characteristics to facilitate better response generation. Second, the graphmaster function does not apply to recent generation-based chatbots (Chen, Liu, Yin, & Tang, 2017; Sun, Chen, Pei, & Ren, 2018; Tran & Nguyen, 2018; Wen et al., 2015) as they do not need to search for answers in stored datasets. Instead, their models are pre-trained with dialogues and can directly generate outputs. Therefore, the graphmaster should be adapted to fit newly developed chatbots. Based on Abdul-Kader and Woods' efforts, we develop a modified structure for chatbot designs, as shown in Figure 1.

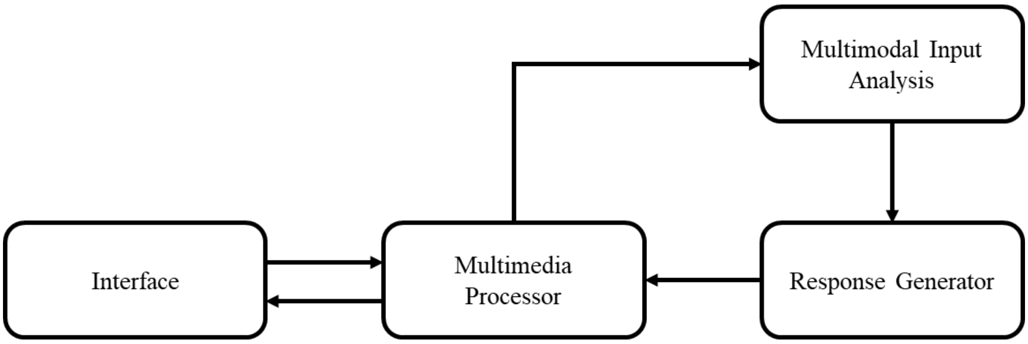


Figure 1. The generic architecture of chatbot systems.

A chatbot system comprises four parts: the interface, a multimedia processor, an multimodal input analysis, and a response generator. The interface manages multimedia interactions with users. It not only receives user inputs in various forms but also presents appropriate responses. As mentioned above, the multimedia processor transforms and extracts features from audio and video signals. The multimodal input analysis unit corresponds to a classifier and handle data pre-treatment. This module can be enhanced by advanced natural language understanding (NLU) techniques with functions such as semantic parsing, slot filling, intent identification and emotional analysis to support response generation. For retrieval-based chatbots, the response generator plays the same role as that of the graphmaster as it stores and retrieves responses, whereas for generation-based chatbots, it maps the normalised input transferred from the input pre-treatment unit to the output.

If the response generator is considered to be the core of a chatbot system, chatbots can be classified based on the specific techniques they utilise to create outputs into six categories: template-, corpus-, intent-, RNN-, RL-based, and those with hybrid approaches. In contrast to the two-category

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

classification method (i.e. retrieval- and generation-based chatbots, and task-oriented and non-task-oriented chatbots), our taxonomy has the advantage of demonstrating the characteristics of each type of chatbot based on the techniques used, allowing chatbot designers to blueprint the response generation model.

[2.1 Template-based chatbots]

Template-based chatbots select responses from predefined templates by comparing the relevance of the user query to the designed query patterns. These chatbots use two main techniques: Artificial intelligence mark-up language (AIML), a derivative of extensible mark-up language (XML), that organises and stores predefined pairs of templates and patterns, and a second technique that is used for a type of response searching called pattern matching, in which a set of word-by-word matching rules is used to determine best-matching results (Bradeško & Mladenić, 2012).

We use the chatbot Artificial Linguistic Internet Computer Entity (ALICE) to illustrate the specific techniques utilized by template-based chatbots. ALICE, developed by Wallace in 2003, is the oldest and most famous template-based chatbot. The AIML files of ALICE have many basic knowledge units called categories, each of which consists of a query pattern and a corresponding template. A basic category is shown in Figure 2. Based on the tags <pattern> and <template>, “10 Dollars” is stored as a pattern and its corresponding template is “Wow, that is cheap”. For flexible and high-quality match results, ALICE adopts wildcard symbols in patterns including _ and *. Pattern matching techniques run searches through all categories and determine whether the user query matches predefined patterns. The longest matched pattern is regarded as the most precise, and its corresponding template is selected as the output (Abushawar & Atwell, 2015).

```
<category>
<pattern>10 Dollars</pattern>
<template>Wow, that is cheap.
</template>
</category>
```

Figure 2. An example of category in artificial intelligence mark-up language (AIML).

Author Manuscript

Template-based chatbots are easy to create and configure, making them a popular choice for various small-scale applications.

[2.2 Corpus-based chatbots]

Though template-based chatbots are easy to adopt and deploy, they have apparent shortcomings when handling large-scale retrieval. AIML is essentially a format of text files and not a professional knowledge management technique. When the number of stored pattern-template pairs increases, the time required for word-by-word searching through all the files increases correspondingly. Consequently, template-based chatbots struggle to provide efficient and timely responses in contexts that require considerable knowledge.

This led to the development of corpus-based chatbots that substitute AIML with specialized knowledge management techniques. Correspondingly, the built-in pattern matching method in AIML is no longer used for corpus-based chatbots, which use new approaches for response retrieval.

A database is a common tool used to store and retrieve domain knowledge. Pudner et al. (2007) developed a method that could generate answers with specific knowledge and in natural languages. Specifically, the technique identifies relevant attributes and values from the user input, following which a subsystem in the chatbot automatically creates a structured query language (SQL) query statement based on the reasoning rules, and then sends it to the database for execution. Finally, the response sentences are constructed with the query results and displayed.

An ontology model or semantic web is a new, alternative method used to store domain knowledge. Compared to databases, the ontology models have the advantage of flexible knowledge management. Nazir et al. (2019) designed a semantic web to organise questions and answers collected in the fashion domain and provide answers to customer questions about brands and relevant products. Like SQL in databases, they used semantic query language (SPARQL) to manipulate the semantic web. The chatbot Kristina developed by Wanner et al. (2017) also builds a semantics-oriented knowledge base in ontology. Unlike traditional corpus-based chatbots that use predefined templates, Kristina exploits rule-based reasoning developed from semantic language analysis to integrate selected information into readable output sentences.

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A second type of corpus-based chatbots does not carefully curate knowledge(Cui et al., 2017; Prasomphan, 2019) but can still efficiently retrieve the correct responses because the dataset of query-response pairs is not stored as words but as vectors. Using the technique of word embedding, this type of chatbots can transform words into vectors that represent the semantic meanings of the words in a multidimensional space (Mikolov et al., 2013). For example, the relative positions of the words *King* and *Queen* in the vector space will be similar to those of M*ale* and *Female*. Then, by calculating the distance between the user input and query-response pairs, the query with the least distance is matched and its corresponding response is selected as the output.

Due to the strong capabilities of knowledge management, several applications that require formal domain knowledge use corpus-based chatbots. Corpus-based chatbots have several advantages over template-based chatbots. First, language tags and wildcard characters are not required for corpus-based chatbots, simplifying the chatbot creation process. Second, the techniques of knowledge management that corpus-based chatbots adopt (i.e., database, ontology, and semantic web) enable users to easily obtain specific and accurate information from large domains, which is difficult to achieve with template-based chatbots. Third, separate response retrieval makes response selection more flexible, increasing the potential to improve matching algorithms and enhance the response quality. Hence, corpus-based chatbots are well-suited to large-scale tasks that require specific knowledge.

[2.3 Intent-based chatbots]

Intent-based chatbots are widely used for task-oriented systems that feature multi-turn dialogues (Chen et al., 2017). Both the multimodal input analysis and response generator units of intent-based chatbots are augmented to satisfy task-oriented special requirements. Apart from data pre-training, the multimodal input analysis unit in these chatbots is responsible for semantic analysis. It usually uses advanced NLU techniques to parse user queries into pre-defined semantic slots (Chen et al., 2017) and utilizes machine learning approaches to classify the intents of input queries (Franco et al., 2020).

The knowledge bases of response generators in intent-based chatbots mainly comprise training phrases, intents, and responses (Rosruen & Samanchuen, 2019). In addition, the response

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

generators in intent-based chatbots incorporate a function called dialogue management (DM), which comprises dialogue state tracking (DST) and dialogue policy optimization (DPO) (Williams, Raux, & Henderson, 2016).

Task-oriented systems aim to support users in realizing certain tasks. During assistance, information should be exchanged between humans and chatbots in sequence (Mrkšic, Séaghdha, Wen, Thomson, & Young, 2017). For example, in an airline reservation system, the chatbot should first obtain basic reservation information such as the airports of departure and arrival and users' preferred departure times and prices, then send the filtered airline information to users, and finally confirm the order states. Therefore, conversations in task-oriented system are essentially flows of dialogues containing multi-turn utterances. To establish the sequence of dialogues, the techniques of DST, dialogue state identification, DPO, dialogue state processing, were developed rapidly for intent-based chatbots (Williams et al., 2016).

Rasa is a typical open-source intent-based chatbots (Bocklisch, Faulkner, Pawlowski, & Nichol, 2017), whose pipeline structure is shown in Figure 3. For generalization, we modified the system framework of Rasa with another similar one introduced by Williams et al. (2016) and integrated it with our proposed structure.

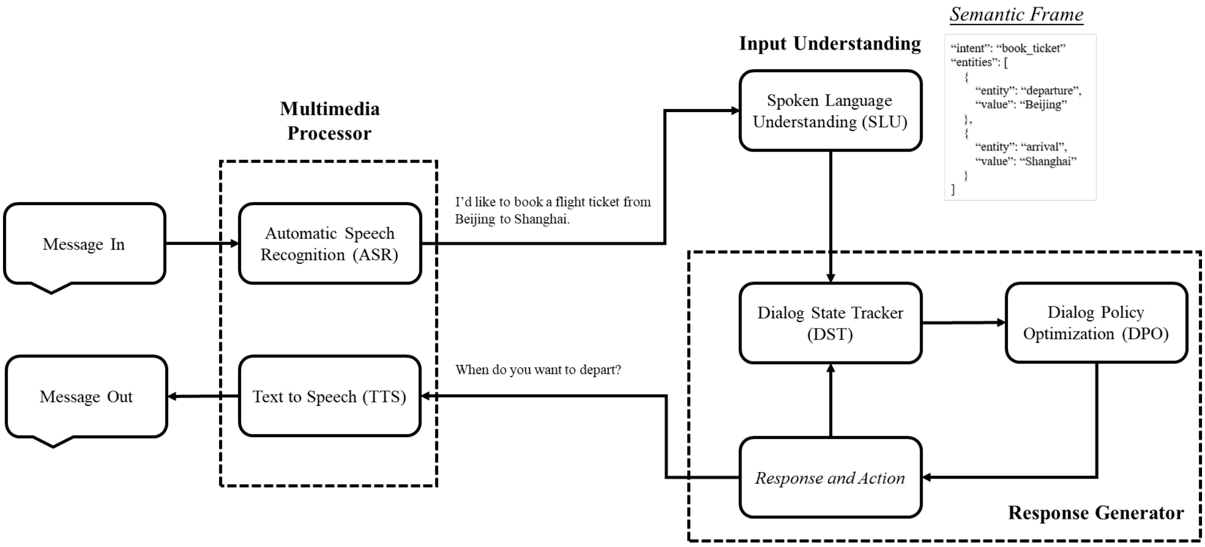


Figure 3. The pipeline mechanism of intent-based chatbots.

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

First, the automatic speech recognition (ASR) module transforms audio information into texts, realizing the function of a multimedia processer. Next, spoken language understanding (SLU) semantically parse user utterances to detect the intents of user queries, extract certain information from slots, and form a semantic frame. This module is paralleled with multimodal input analysis. Later, DST is used to estimate the state of current dialogues and DPO is used to determine the actions and responses; both these components belong to response generators. Finally, the appropriate messages are output.

Aside from researchers, technical companies make significant contributions to the development of intent-based chatbots. For instance, Dialogflow[1] (originally named Api.ai), developed by Google, and wit.AI[2], developed by Facebook, are used in academic research (Handoyo et al., 2018; Rosruen & Samanchuen, 2019).

[2.4 RNN-based chatbots]

The rapid development of deep learning, especially sequence-to-sequence (Seq2Seq) models (Cho et al., 2014; Sutskever, Vinyals, & Le, 2014), has allowed insights into the response generation of chatbots. Figure 4 illustrates the computational mechanism of the Seq2Seq model.

[1] https://dialogflow.cloud.google.com/

[2] https://wit.ai/

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

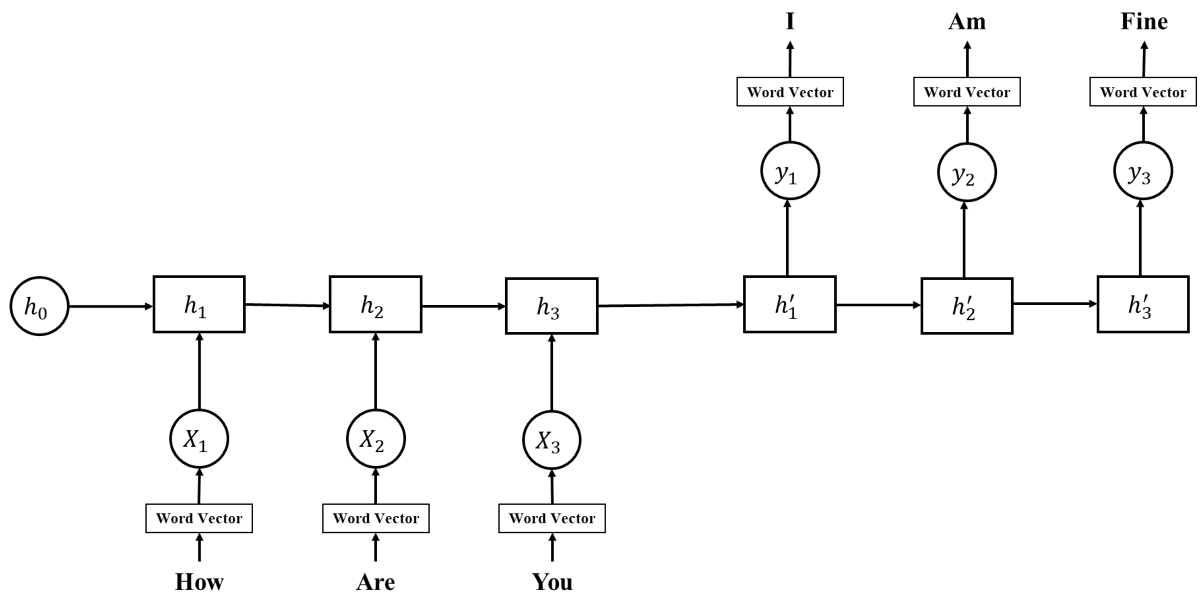


Figure 4. The computational mechanism of the sequence-to-sequence (Seq2Seq) model.

The Seq2Seq model, also called the encode-decode model, is a transformation of the RNN (Elman, 1990). Here, a sentence is split into a sequence of words, which serves as the input x_1 to x_3. Following a recurrent mechanism similar to the hidden layers in RNN, the output sequence y_1 to y_3 is generated as a complete sentence (Cho et al., 2014; Sutskever et al., 2014).

In practise, the recurrence of the hidden layer may result in calculation problems such as gradient explosion and vanishing, making accurate outputs difficult to achieve. Therefore, methods such as long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) and gated recurrent unit (GRU) (Chung, Gulcehre, Cho, & Bengio, 2014) are used to boost the computing power of the RNN as they can improve the output calculation of the hidden layer.

Chatbots based on a Seq2Seq model need not to match user queries with predefined query-response pairs. Instead, they are trained with datasets of dialogues and directly generate output. Thus, such chatbots are also called generation-based chatbots (Song et al., 2016; Wu, Li, et al., 2018; Wu et al., 2017; Wu, Wu, et al., 2018) and are markedly different from retrieval-based chatbots. The output of generation-based chatbots depends upon the model specification, training datasets, training process, and user input.

As many factors can influence the outcome of Seq2Seq models, the responses of neural-based chatbots are uncertain. Therefore, few RNN-based chatbots are used in task- or knowledge-

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

oriented scenarios. However, uncertainty can also be thought of as a representation of creativity. This implies that, RNN-based chatbots perform very well in chat-oriented applications, such as entertainment and mental health-related activities.

[2.5 RL-based chatbots]

As the name suggests, RL-based chatbots adopt RL as the core technique for response generation. Such chatbots are typically retrieval-based and make full use of the input context for optimal response searching (Silver et al., 2016). RL-based chatbots compare not only the similarities in current user input and pre-defined query-response pairs, but similarities in context as well.

The RL method is mainly based on the Markov decision process, typically depicted by

(1) a finite set of states $S = \{s_i\}$

(2) a finite set of actions$A = \{a_i\}$ that describes the change from one state to another.

(3) a policy $a = \pi(s)$ that specifies the probability of performing an action a when in state s;

(4) a transition model $T(s, a, s') = P_r(s'|s, a)$ that indicates the probability of the next state s' after taking action a from the current state s;

(5) a reward function $R(s, a, s')$ that specifies the instant reward after a certain state change. Given the above definitions, RL aims to find an optimal policy that maximises the cumulative reward:

$$Q_\pi(s, a) = \max_\pi E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots \vee s_t = s, a_t = a, \pi]$$

where Q_π is the maximum value of cumulative reward after time t, r_t is the reward of taking action *a* in state *s* at time *t*, and γ is the discounted factor that indicates the decreasing effect of states with time (Cuayáhuitl et al., 2019).

For RL-based chatbots, the state is a vector representation of a turn of talk. The representation could just be the sentence vectors derived using embedding techniques (Cuayáhuitl et al., 2019) or contains more information such as dialogue act, user sentiment, and generic user utterance (Serban et al., 2018). The context of communication forms a sequence of states. When the RL-based chatbot is trained by a large corpus of dialogues, it experiences numerous trials and errors, and can therefore select the optimal responses based on the history.

Compared with template- and corpus-based chatbots, the establishment of RL-based chatbots may be more complicated by its need for pre-defined dialogues to train policy. However,

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

unlike other chatbots that search responses from knowledge bases, RL-based chatbots find appropriate responses directly from well-trained policies (Serban et al., 2018), which can be considered a method of knowledge management.

[2.6 Chatbots with hybrid approaches]

Several researchers have tried to combine some of the diverse techniques of response generation to improve the performance of chatbots. We summarize these hybrid methods in Figure 5.

A typical combination introduces a ranking algorithm that selects the optimal response from candidate responses generated by several dialogue systems. Deep learning is commonly used to rank candidate responses based on the context. Wu et al. (2018) proposed a sequential matching framework based on an RNN. The framework considers the relationships between previous utterances and candidate responses and selects the optimal responses for the context. Mao et al. (2019) also designed hierarchical aggregation network of multi-representation based on bidirectional RNNs to match the information in context to alternative responses. The MILABOT developed by Serban et al. (2018) adopted deep RL to select the most appropriate responses after considering the context.

Technologies can also be combined by embedding one method into another. Song et al. (2016) proposed a generative dialogue system embedded in a corpus-based chatbot. Unlike normal RNN-based chatbots that regard the user input as the encode, the dialogue system of corpus-based chatbots is fed with their retrieval results. Wu et al. (2018) tried to extract topic information from context as a supplement for the matching mechanism. They first used topic modelling to obtain the topic words and then trained the RNN to weight the words for clustering. They incorporated topic matching when performing response retrieval. Some intent-driven chatbots do not pre-define responses and instead generate responses directly based on the abstract dialogue actions provided by DPO. This technique is called natural language generation (NLG) (Chen et al., 2017; Williams et al., 2016). Based on the report by Mikolov et al. (2010) on RNN language modelling, which is an RNN-based algorithm for speech recognition, Wen et al. (2015) introduced a novel neural-based approach for NLG. The approach used a convolutional neural network (CNN) and a backward RNN to re-rank the candidate responses obtained from an RNN whose architecture was similar to that of an

RNN language model. Moreover, Tran and Nguyen (2018) later extended the method by applying a gating mechanism.

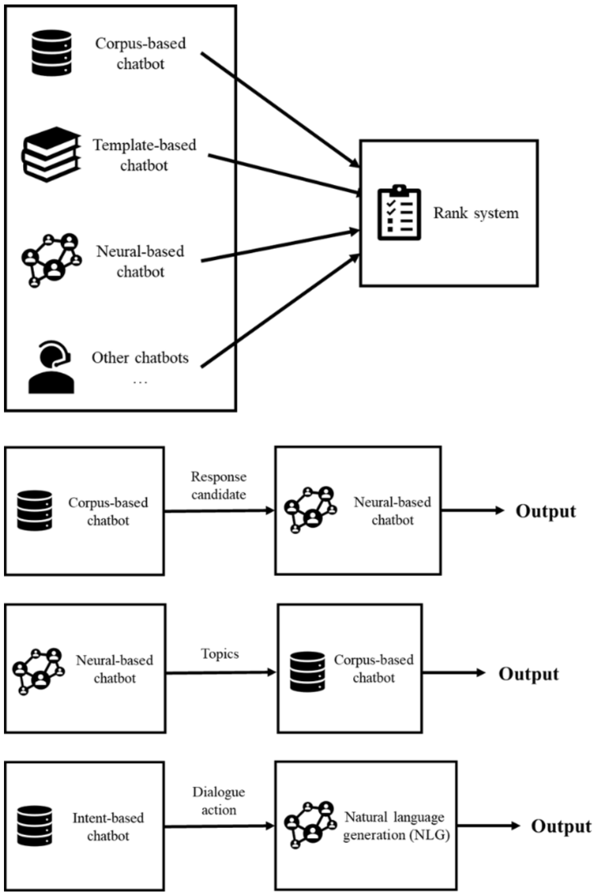


Figure 5. Typical approaches used by chatbots with hybrid techniques.

[2.7 Development of computational approaches for chatbots]

Except for the interface, the remaining three components of the chatbot system have undergone revolutionary developments in the past few decades. Multimodal input analysis has benefited from the flourishing of machine learning, especially NLP, and the ability of chatbots to understand user queries has significantly improved. Response generators have mainly developed along two directions: establishment and response generation. The growth of multimedia processors is even more remarkable as it started from scratch, with first-generation chatbots only capable of text communication.

[2.7.1 NLP]

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The capability of chatbots for query learning has kept pace with the quality of NLP technology. Word embedding, a potent distributed representation methods for words proposed by Mikolov et al. (2013), has been widely adopted in chatbot systems (Serban et al., 2018; Wu, Li, et al., 2018; Wu, Wu, et al., 2018; Zhao et al., 2019). On the basis of word embedding, Xue et al. (2018) introduced sentence embedding to quantify the sentence implication and character embedding to address the out-of-vocabulary problem. The recent emergence of bidirectional encoder representations from transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2019), a better substitute for the traditional word embedding technique of Word2Vec, led to the adoption of this novel technique by Bunk et al. (2020) for better user query learning.

Semantic analysis is especially crucial for intent-based chatbots as they need to infer a semantic frame from the user input. Yang and Liu (2015) established an RNN-based model for slot filling in SLU. Goo et al. (2018) proposed a slot-gated model for joint slot filling and intent prediction. Topic modelling also improve input recognition. Wu et al. (2018) extracted topic words by latent Dirichlet allocation (LDA) model, and then used the topic clues in a topic-aware attentive recurrent neural network to generate responses. Serban et al. (2018) also introduced a RL methods to select responses generated by multiple chatbot models based on identified topics.

[2.7.2 Knowledge establishment]

To increase the requirements for knowledge in chatbots, designers of template-based chatbots need to transform large discourse datasets into AIML files. To construct a better knowledge base, Marietto et al. (2014) introduced an intuitive design tool for data modelling called dialogue conceptual diagram (DCD). They also discussed the methodology of mapping the flow diagram with DCD to the rules of pattern matching and of the categories to AIML. However, the manual establishment of a knowledge base is still an immensely heavy workload for chatbot designers. Krassmann et al. (2019) proposed an automatic generation of AIML from text acquisition system to automatically extract knowledge from text corpora and convert it into AIML files. This design simplifies the manual work required to predefine chatbot knowledge base and allows normal users to customise their own chatbots by importing personal text.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Although neural-based chatbots are not required for knowledge management, they need large dialogue datasets to train the weights of the neural network model, by which the knowledge is represented. To simplify the preparatory work of acquiring knowledge, a new chatbot is expected to learn from well-trained chatbots. Arsovski et al. (2019) demonstrated a method that enables knowledge transfer between chatbots; this method utilises dialogues generated by existing chatbots to train the models of new chatbots. Mathur and Lopez (2019) solved the problem of large training datasets from another perspective. They improved the chatbot model using the GRU algorithm and bidirectional decoders, which can maintain good performance with little training data.

[2.7.3 Response generation]

To enhance the user experience and response quality of template-based chatbots, a high number of additional functions in the process of response generation have been proposed. Based on conversational analysis theory, Neves et al. (2006) developed an improved version of AIML, intentional AIML (iAIML), by incorporating tags for user intention treatment. They found that iAIML can realise consistent conversations between chatbots and users. Galvão et al. (2004) designed Persona-AIML, an extension of AIML that allows chatbots with different personalities. The Persona-AIML structure has five elements: traits, attitudes, mood, emotions, and physical states.

Several efforts have been made to precisely match knowledge for corpus-based chatbots. Wiak and Kosiorowski (2010) applied psycholinguistic rules to a chatbot system and stored the synonyms, hyponyms, and hypernyms in a database to exploit word relationships for better response searching. Lokman and Zain (2011) developed an algorithm called extension and prerequisite that constructs relationships between the sequences of user inputs and chatbot responses by replacing pronouns with consistent topics. The algorithm takes the context of a response into consideration and promises consistency of discourses.

Chatbots often fail to answer users' questions for three reasons: unknown concepts, out-of-domain tasks, and wrong fulfilment. Moore (2018) designed an end user development (EUD) method to tackle these issues. The core idea of EUD is to enable end users to "teach" chatbots the meaning of new concepts and the execution of new tasks. Abdul-Kader and Woods (2018) adopted multiple feature extraction methods to rank alternative responses. Their evaluation results suggest that the highest scored responses are the closest subjective matches.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

RNN-based chatbots have burgeoned following the rapid updating of deep learning. Bahdanau et al. (2015) first introduced the attention mechanism in Seq2Seq model, which allows models to distinguish key words by giving each word a weight based on its importance in a sentence. DeepProb, a broad information-directed chatbot designed by Yin et al. (2017), used the attention-based Seq2Seq model for user interactions and reported significant improvements in performance. A number of other researchers have made important contributions in overcoming the limitations of RNN-based chatbots. Shao et al. (2017) developed a practical approach, named the glimpse-model, to solve the problem of the standard Seq2Seq model generating long responses; the approach was sophisticated enough to track all of the information and fix the length of the decoder.

Several revolutionary enhancements in neural network structures have also increased the potential of RNN-based chatbots. For instance, Zhou et al. (2016) developed an attention-based bidirectional LSTM (bi-LSTM) network for relation classification that captured a sentence from two directions and reduced the impact of word sequence, as key words could appear at any position in a sentence. Xue et al. (2018) later adopted this idea in the design of their chatbot. Mnih et al. (2015) introduced the Deep Q-Network, a variant of deep q-learning, that harnessed the strengths of RL. Such advances enabled Guo (2015) to iteratively decode the output sequence, opening up a novel direction along which to improve the models of RNN-based chatbots.

User intention identification is a popular research topic in the field of neural-based chatbots that can enhance the response quality. Dialogue act (DA) classification is a useful method that can be used to this end. For example, the category “agree/accept” indicates positive attitudes, while the category “reject” indicates negative attitudes of users. As traditional DA classification only deals with human–human communication, Ahmadvand et al. (2019) developed a contextual dialogue act classifier (CDAC) to adapt to the specific characteristics of human–computer interactions. DeepProb, designed by Yin et al. (2017), recognises the user intention more directly as it actively asks users questions to learn their requirements. For intent-based chatbots, few-shot learning, a novel method that enables chatbots to learn the intents of user queries from a small sample, was also recently used for intent recognition (Peng, Li, et al., 2020; Peng, Zhu, et al., 2020).

[2.7.4 Multimedia interactions]

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

As people in the modern age become accustomed to visual interactions with digital devices, the provision of a graphical interface has become a promising avenue for future chatbots. Pirrone et al. (2008) extended AIML to graphical AIML (GAIML) with verbal and visual interaction modules. Building on GAIML, a conversational system called Graphbot was developed to examine the proposed framework. Graphbot enabled multi-modal communication between chatbots and users with a personalised interface, increasing not only conversation effectiveness but also user satisfaction.

Neural-based chatbots have unique advantages with respect to multimedia interaction, facilitating the success of image and sound processing in the field of deep learning. Lee et al. (2020) developed an emotion recognition algorithm based on voice-enabled chatbots, in which recognition is realised in two steps. First, the received user sound is transformed into a spectrogram by Mel-frequency cepstral coefficient (MFCC). Second, a CNN is adopted to analyse the spectrogram and classify it into five emotions. Jonell (2019) introduced a chatbot that can detect signals from users, such as skin temperature, respiratory rate, and pupil dilation. The captured physiological information is learned by the neural network to generate more adapted responses. Given that tones can also impact user experience of chatbots. Hu et al. (2018) designed a chatbot to recognise the tones of users and generate appropriate responses to different tones. Arsovski et al. (2020) designed the artificial intelligence snapchat visual conversation agent, a visual chatbot that can generate visual responses to visual stimuli, based on deep neural networks and latent semantic indexing and by comparing the vectors of captions generated from personalised image datasets to input images.

[2.8 Summary]

We summarise the taxonomy of chatbots described above in Table 1.

It is noteworthy that although RL-based chatbots do not use any specific knowledge management techniques, these chatbots are still categorized as retrieval-based chatbots because when communicating with users, it selects responses from a pre-defined knowledge base built using a well-trained policy.

Template-based chatbots are simple to create and configure but have significant disadvantages. Although automatic knowledge extraction can reduce the manual work required to create AIML files, organising a large knowledge base still comprises a complex and heavy workload for chatbot designers. Therefore, in most cases, template-based chatbots are used in situations in

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which the task is well-trained and less complicated. In the contrast, the good knowledge management for corpus-based chatbots make them suitable for applications that require large knowledge bases. The uncertainty characteristic of RNN-based chatbots is a double-edged sword: it enables chatbots to generate innovative responses that fit wide-ranging topics, but makes the response authenticity unreliable. Thus, RNN-based chatbots are ideal for open-domain tasks. Intent-based chatbots are suited to multi-turn response generation in task-oriented systems.

The relatively recent development of RL-based chatbots implies that the technology has yet matured. Ideally, RL-based chatbots should perform better than other chatbot types in terms of considering context. However, the measurements of states and rewards of RL in the service of chatbot systems remains insufficient and need further exploration.

Chatbots with hybrid approaches combine several sub-chatbot systems and therefore, show better performance. However, they require higher level of investment. Practitioners need to make business decisions only after carefully considering the cost and profit.

Table 1. A taxonomy of chatbots based on various computational approaches

	Template-based Chatbot	**Corpus-based Chatbot**	**Intent-based Chatbot**	**RNN-based Chatbot**	**RL-based Chatbot**	**Chatbot with hybrid approaches**
Techniques of Knowledge Management	AIML	Database and ontology	-	-	-	Depends on components
Techniques of Response Generation	Pattern matching	Text mining	SLU, DST, DPO, NLG	RNN and LSTM / GRU	RL	Ranking algorithm / embedded technology
Typical Chatbot	ALICE(Abushawar & Atwell, 2015)	KRISTINA(Wanner et al., 2017)	Dialogflow[3]	(Shao et al., 2017)	MILABOT(Serban et al., 2018)	(Song et al., 2016)
Characteristics	Simplicity	Well-performed knowledge management	Multi-turn responses	Wide-range communication topics; uncertain responses	Consideration of context	More complicated, but better performance
Suitable Application	Simple and task-oriented	Applications that require large knowledge bases	Multi-turn and task-oriented	Open domain and chat-oriented	Context-oriented	Applications that require high performance

[3] https://dialogflow.cloud.google.com/

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

						with large investment
Development Direction	Knowledge establishment and multimedia interaction	Knowledge management and response matching	User input understanding, dialogue state, and response generation	Response generation and multimedia interaction	Design for the appropriate application of RL chatbots	Ranking algorithm improvement / embedding technology

AIML: artificial intelligence mark-up language; SLU: spoken language understanding; DST: dialogue state tracking; DPO: dialogue policy optimization; NLG: natural language generation; RNN: recurrent neural network; LSTM: long short-term memory; GRU: gated recurrent unit; RL: recurrent learning; ALICE: Artificial Linguistic Internet Computer Entity.

[3. A REVIEW OF THE APPLICATION FIELDS OF STATE-OF-THE-ART CHATBOTS]

This section reviews chatbots that provide actual services in business markets and are used in diverse fields to support human work in various manners, such as customer service, personal assistance, finance, physical healthcare, mental healthcare, and education. Figure 6 presents some chatbots used for diverse applications.

In practise, a group of chatbots is referred to by the term ECA. Unlike conventional chatbots, ECAs usually have an embodiment or visual representation that enables multimodal communication methods such as speech, gesture, and facial expressions. In the following sections, we use ECA to denote a group of chatbots with higher anthropomorphism.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

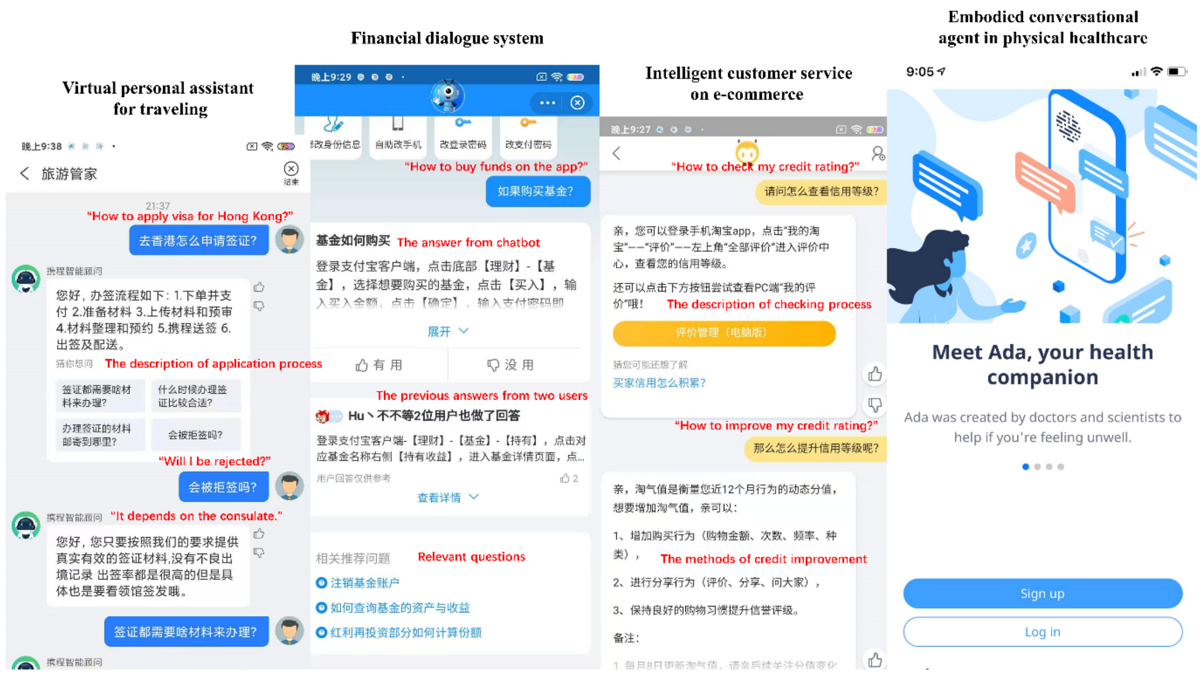


Figure 6. Chatbots used for diverse applications.

[3.1 Intelligent customer service for e-commerce]

The rapid development of e-commerce has markedly increased the workload of customer-support workers. This has necessitated the adoption of chatbots as a supplement to human services to reduce human costs and improve service quality. Cui et al. (2017) designed a virtual agent called SuperAgent for an e-commerce website to provide information about online products and services. SuperAgent generates answers from its knowledge bases extracted from detailed product information, frequently asked questions, and online reviews. It is a typical chatbot with hybrid approaches that uses a meta engine to rank and prioritize the responses generated from each sub-engine. Xue et al. (2018) improved the response generation model by incorporating user intentions. They designed a chatbot, Intuit Smart Agent (Isa) that adopted a neural-based intention classifier to provide higher-quality answers to users. Isa is a corpus-based chatbot that provides information by searching a large database of pre-defined answers and similar historical chats. Zhao et al. (2019) designed an intent-based virtual agent called MOLI that provides not only an employee intention classifier but also question type classification and semantic matching for better response generation.

Apart from technical improvement, new functions such as recommendation, negotiation (Jusoh, 2019), and emotional support (Kuramoto et al., 2018) have been added to chatbot systems.

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Furthermore, to answer customer questions accurately, chatbots now have access to larger databases. This implies that if chatbots are attacked by hackers, valuable information will be leaked (Bozic & Wotawa, 2018). To counter this risk, the Chatbot Security Control Procedure (CSCP) (Lai, Leu, & Lin, 2018) was proposed to monitor the behaviours of chatbots, avoid the leakage of confidential information, and ensure customer privacy.

Although chatbots have been widely adopted for customer service, there is still a notable gap between customer expectation and chatbot performance (Mimoun, Poncin, & Garnier, 2012). Kvale et al. (2020) summarized several key reasons for the failure of chatbots in customer service, such as the inability to interpret users' requests and provide engaging and convincing responses, and failure to handle complex queries or maintain long conversations in deference to the context. We also argue that despite ongoing research efforts in improving chatbot performance, human-chatbot communications are far from the levels of human-human interactions due to the lack of emotional exchange and the inability of chatbots to handle novel requests. Furthermore, we believe that other concerns regarding the deployment of chatbots in customer services, such as user privacy and data security, have not received enough attention from academic researchers.

[3.2 Virtual personal assistance]

Chatbots for personal assistance can be used to increase higher work efficiency. Kimani et al. (2019) designed a virtual agent called Amber, which is an RNN-based chatbot with a sensing framework (multimedia processor) that captures information from users' faces, speech, and app usage, an agent app (interface) to oversee information communication, and a dialogue framework (multimodal input analysis and response generator) to generate responses .It makes schedules for workers, prioritise tasks, sends reminders, and suppresses distractions caused by social media. Virtual personal assistants can be used to manage everyday life as well. Kadous and Sammut (2004) designed InCA, a rule-based virtual agent to provide basic assistance services such as appointment management, email reminders, and weather reports. Additionally, if the users want to travel or go out

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for a meal, the Messenger bot, an intent-based chatbot with the core of wit.ai[4] can order tickets (Handoyo et al., 2018).

However, we note that privacy concerns may prevent some individuals from using chatbots for virtual personal assistance (Ghosh & Eastin, 2020; Lau et al., 2018; Sweeney & Davis, 2020). Following the revelation that both Google and Amazon employed humans to listen to conversations between users and smart personal assistants in the early days of virtual assistance[5], individuals may refuse to use chatbots to avoid being monitored. Therefore, we strongly recommend that future designs of virtual personal assistance incorporate strong privacy controls.

[3.3 Financial dialogue system]

The rise of AI has led to the emergence of the financial technology industry and chatbots have become an important development in this filed. Many banks and other financial institutions have actively introduced chatbots to provide professional financial advice for various complex products and services (Beketov et al., 2018). Okuda and Shoda (2018) designed the Finplex Robot Agent Platform (FRAP), a corpus-based chatbot for Fujistu to support product sales and customer service. Doherty and Curran (2019) also proposed a corpus-based chatbot for banks that enabled users to obtain information from their personal accounts. Notably, the ability of chatbots to access user' accounts necessitates the urgent formulation of personal privacy protection and property safety policies. The CSCP mentioned above (Lai et al., 2018) is an important step in this direction.

To the best of our knowledge, there has not been much research on financial dialogue systems, resulting the lack of consensus on important subjects, such as mission-critical tasks, the establishment and organization of domain expertise, and technology standards. More research on financial dialogue systems is therefore required to promote chatbot development in this field.

[3.4 ECAs in physical healthcare]

[4] https://wit.ai/

[5] Amazon Workers Are Listening to What You Tell Alexa. Bloomberg, 2019-04-11.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the field of physical healthcare, a group of chatbots serves as a ***medical assistant for disease diagnosis***. The understanding of patient symptoms is paramount in the process of illness detection, and one of the main functions of these chatbots is to collect sufficient information from patients. For example, Ponathil et al. (2020) used a virtual agent to obtain information on family health history (FHx) from patients for the early diagnosis of genetic diseases. Compared to the traditional standard interface of FHx collection, which is usually a web-based questionnaire, people prefer chatbots because they provide better information quality, interface quality, and system usefulness. Chatbots are also qualified to safeguard the release of information. When a virtual agent developed by Van Heerden et al. (2017) was used to guide counselling sessions before HIV tests, the test takers said they felt comfortable talking with chatbots as the chatbots offered more privacy than humans. Both the above-described chatbots can be classified as intent-based as they possess three characteristics: 1) they generate immediate responses based on the current dialogue state; 2) they participate in multi-turn conversations; and 3) they finish a certain task.

In addition, some ECAs have the capacity for ***medical analysis***. Given the tremendous knowledge base required for medicine, it is difficult for any one person to know every disease and/or treatment. However, it is possible for machines to possess this knowledge. Madhu et al. (2017) designed an ECA that can predict diseases based on patient symptom descriptions and provide available treatments. The chatbot serves as a supplement of traditional medical diagnosis that assists patients in understanding their health problems and in checking the availability of prescriptions. The ECA proposed by Kasinathan et al. (2018), called Intelligent Healthcare Chatterbot (HECIA), can also offer diagnosis but is not as knowledgeable as the former agent. HECIA is designed to assist triage work in medical centres by providing a preliminary analysis of patient symptoms and routing patients to specific departments. Given the strong knowledge organization capability of corpus-based chatbots, all of the above-mentioned ECAs used for medical analysis belong to this category.

Chatbots can also be used as ***health managers*** to guide the prevention and control of diseases. For instance, Hauser-Ulrich et al. (2020) designed the corpus-based embodied conversational agent painSELfMAnagement (SELMA) to help patients cope with chronic physical and mental pain caused by diseases. Fadhil et al. (2019) introduced CoachAI, an intent-based chatbot that guides users on following healthy lifestyles and reducing the risk of chronic illness. Kadariya et al. (2019) developed the intent-based kBot, which can monitor medical adherence and other relevant

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

health signals in paediatric asthmatic patients and assist them in controlling the diseases. Many other health managers provide information on obesity (Huang, Yang, Chen, Wu, & Chen, 2019), Parkinson's disease (MacEclo et al., 2019), and cancer (Belfin, Shobana, Manilal, Mathew, & Babu, 2019).

Several prior studies have attempted to identify problems and propose future research directions to improve ECAs for physical healthcare. We have summarized a few here. First, the individual characteristics of target patients have been neglected (Kim, Park, & Robert, 2019). In particular, it is a huge challenge for elderly people to learn and accept chatbots (Montenegro, da Costa, & da Rosa Righi, 2019). Second, the relationships between chatbots and humans have not been investigated sufficiently (Kim et al., 2019). Third, the technology metrics of chatbots in healthcare vary greatly, and the lack of an unified objective standard may impede the advancement of this field (Abd-Alrazaq et al., 2020). Finally, the possible ethical problems of this field have not been given adequate attention. Luxton (2020) laid out several potential ethical risks that need to be addressed, including risk of bias, risk of harm, privacy, and inequitable access. Laranjo et al. (2018) also noted that the evaluation of patient safety was a rarity.

[3.5 Virtual counselling service]

In mental healthcare, some chatbots are used for the surveillance and prevention of mental diseases and extreme behaviours. For instance, Martínez-Miranda et al. (2019) designed a chatbot to collect the users' self-reported data for the analysis of suicide risk. Upon risk detection, the chatbot alerts relatives of the users. Although this chatbot selects utterances from multiple sources, including emotions from emotional models, information from past sessions, and cognitive-behaviour therapy (CBT) activities from pre-defined files, we consider it a corpus-based chatbot because all three kinds of responses are pre-defined and stored in XML files. Some chatbots are designed for diagnosis. Mujeeb et al., (2017) introduced a chatbot named Aquabot that efficiently diagnoses autism and achluophobia (fear of darkness) with an accuracy of 88%. When comparing with the advanced trivial diagnosis methods that require a lot of time and energy, automatic diagnosis techniques, with their higher accuracy and low resource requirement, offer more benefits to psychologists benefits from this automatic diagnosis technique for the high accuracy and saved resources. Other chatbots have been

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

developed for psychotherapy methods that target depression (Sharma, Puri, & Rawat, 2018) and other mental problems. Park et al. (2019) used one for motivational interviewing (MI), a psychotherapeutic approach used increasingly often in chatbot design, and found a positive effect of MI skills in stress management. These chatbots are also corpus-based but utilize diverse NLU techniques to improve their performance.

Provoost et al. (2017) surveyed the emerging ECAs used in clinical psychology. They studied 54 publications and found that chatbots in virtual counselling services were most applied for social skills training to alleviate the effects of autism. However, several existing chatbots were still in the pilot phase and not yet ready for practical application and evaluation. In the future, we foresee a requirement for more empirical studies to hasten the development of chatbots for virtual counselling service.

[3.6 Pedagogical conversational agent]

Some pedagogical conversational agents are designed for ***learning assistance***. As the main function of chatbots is to communicate with users, language learning has been one of the main avenues pursued. Haristiani et al. (2019) designed the virtual agent Gengbot based on a Japanese grammar dictionary with annotations in English. Gengbot helped students enrich their vocabulary. In addition, Marín (2015), Shawar (2017), and Ruan et al. (2019) designed similar chatbots to help students practise languages in a conversational manner.

Chatbots can also be used to learn many other subjects. For example, due to a shortage of skilful teachers to teach a programming language called Scratch in Thailand, Katchapakirin and Anutariya (2018) designed ScratchThAI to provide lessons. Pérez-Marín and Boza (2013) designed a virtual agent to instruct in exercises in secondary physics and chemistry; students are required to answer questions from the knowledge base of the chatbot. If they answer correctly, the chatbot assign questions of higher difficulty. and if they answer wrongly, the chatbot guides them to right solutions in a stepwise manner. Latham et al. (2010) designed a virtual agent for tutoring called Oscar, which can customise the tutoring content based on the learning style of each student.

Given that the use of chatbots in class can increase interactions and discussions among students, some chatbots are designed for ***class engagement***. To overcome the lack of participation

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in online courses, Caballé and Conesa (2019) tried to embed a chatbot into massive open online courses (MOOCs) to motivate students to engage in collaborative situations. Tegos et al. (2014) conducted a similar study, in which, based on the theory of academically productive talk, they designed a chatbot to motivate interactions in class. Chatbots can also be used for team formation. Xiao et al. (2019) designed a chatbot to infer the characteristics of students via deep communication. Based on the personal information learned from students, the chatbot assists in formulating groups and fostering teamwork.

It is clear that all these pedagogical conversational agents, whether for learning assistance or class engagement, are corpus-based. The specific requirements of a large knowledge base and precise knowledge outputs in the field of pedagogy imply that corpus-based chatbots are the most suitable technologies. However, these requirements also highlight a common problem of pedagogical chatbots: the lack of training corpora and domain expertise. Compared with chatbots used in customer service, there are few conversation records from classes or lessons. Therefore, establishing a chatbot that possesses domain expertise and can help students learn specific knowledge is difficult.

[4. USABILITY ANALYSIS]

Chatbots are an emerging research topic, and several papers have been published on their design and application, but studies examining their effects few. We summarise current research on the usability of chatbots in Table 2 and determine the focus of scholars in the current research phase of chatbots.

Table 2. Usability analysis of chatbots in various application fields

Application Field	Subordinate Field	Result	Examined Construct	Reference
Physical Healthcare	Health Management	People randomized to chatbot-intervened groups drink less alcohol and have more fruit intake, which indicates the usefulness of	Task Performance	Gardiner et al. (2017)

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

		chatbots in helping humans maintain a healthy lifestyle and manage stress.		
		Higher success rate of quitting smoking when chatbots intervened.	Task Performance	Perski et al. (2019)
		Chatbots are well-accepted among older and higher-educated patients	User Adoption	Philip et al. (2020)
	Medical Analysis	Chatbot can provide satisfactory answers about breast cancer and perform as well as the doctors	Task Performance	Bibault et al. (2019)
	Medical Assistant in Hospital	Most Internet users are willing to accept chatbots	User Adoption	Nadarzynski et al. (2019)
		Chatbots play an important role in finishing specific tasks such as appointment arrangement and providing medical information, but are weak in diagnosis and emotional communication.	Task Performance	Palanica et al. (2019)
Mental Healthcare		Empathetic chatbots can alleviate the negative impacts of social exclusion	Task Performance	de Gennaro et al. (2020)
		Chatbots can reduce anxiety by providing positive interventions	Task Performance	Greer et al. (2019)
		Chatbots are promising in stress management	Task Performance	Shamekhi et al. (2017)
		Chatbots based on cognitive behavioural therapy are effective in providing preventative mental health support	Task Performance	Suganuma et al. (2018)
		The same conclusion as for Suganuma et al. (2018)	Task Performance	Fitzpatrick et al. (2017)
		Companionship of chatbots can reduce loneliness	Task Performance	Ta et al. (2020)

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

		Chatbots in clinical psychology and psychotherapy is still in the experimental stage. Although chatbots have immense potential to feasibly, acceptably, and effectively guide psychological health, their usefulness in psychotherapy remains unclear.	Task Performance	Bendig et al. (2019)
Pedagogy	Student Participation in Class	Upon chatbot intervention condition, the discussions of students in a collaborative setting are more effective. Learning performance is also better when students can communicate with chatbots.	Task Performance	Tegos et al. (2015)
		The same conclusion as for Tegos et al. (2015)	Task Performance	Tegos et al. (2017)
		Chatbot intervention positively enhance learning outcomes. In particular, a directed intervention is better than an undirected one.	Task Performance	Tegos et al. (2016)
		The learning interests of students who collaborate with chatbots decrease more significantly than of those who collaborate with humans.	Task Performance	Fryer et al. (2017)
Customer Service		Credibility, competence, anthropomorphism, social presence, and informativeness significantly influence consumer's trust in chatbots	User Trust	Yen & Chiang (2020)
		Chatbot-related factors such as perceived expertise and responsiveness, environment-related factors such as brand perception, and user-related factors such as propensity to trust technology impact customer trust	User Trust	Nordheim et al. (2019)

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

	User trust in chatbots for customer service is affected by factors concerning specific chatbots, specifically the quality of interpretation of requests and advice, human-likeness, self-presentation, and professional appearance. Factors concerning the service context, specifically the brand of the chatbot host, the perceived security and privacy in the chatbot, and general risk perceptions regarding the topic of the request affect the users' trust as well.	User Trust	Følstad et al. (2018)
Personal Assistance	Only 41% of appropriate health-related responses and 39% of appropriate lifestyle-related responses are made by virtual assistants, indicating the limitations of chatbots in personal assistance.	Task Performance	Kocaballi et al. (2020)
	Users feel positive when interacting with chatbots. Among all of the emotions, interest is the most prominent.	User Attitude	Yang et al. (2019)

In Table 2, we summarize state-of-the-art research related to chatbot usability analysis, and focus on task performance, user adoptions of chatbots, user trust, and user attitudes. Based on the results mentioned above, it is known that although chatbots have some limitations in response propriety (Kocaballi et al., 2020) and collaboration with humans (Fryer et al., 2017), they hold promise in assisting humans in finishing specific tasks.

It is noteworthy that the examined constructs in the application field of customer service are significantly different from those in other fields. Given that the market for intelligent customer service is more mature, researchers are no longer satisfied by testing the direct task performance of chatbots. Instead, they now explore how chatbots affect human beings.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

[5. RESEARCH GAPS AND FUTURE RESEARCH DIRECTIONS]

Based on the above review of computational approaches, application fields, and usability analysis of chatbots, we identify gaps in current chatbot research and propose new directions to address these shortcomings.

[5.1 Knowledge establishment]

For intent-based chatbots, existing research on knowledge management is discrete and inconsistent. There is no unified schema for knowledge establishment, which is especially important for large-scale applications. For example, if a researcher wants to build an intent-based chatbot in a new field, they need to organize the potential intents from scratch. Thus, for intent-based chatbots it is necessary to share mutual design tools and methodologies, as done for the Entity-Relation Diagram in database design. Although template- and corpus-based chatbots utilize mature knowledge management technologies, the portability of the knowledge base is limited, representing a research gap in how to transplant the knowledge from an existing chatbot to a new one. We believe that researchers should develop a unified knowledge management schema to enable data transplantation, making the building of chatbots easier and removing the constraints of limited data.

[5.2 Response generation]

Chatbot designers have a long way to go in enhancing response generation. This is evidenced by the fact that, until now, no one has won the gold award for the Loebner Prize, the most famous Turing test competition in the field of chatbots. Apart from performance improvements in existing techniques (i.e. AIML, NLP, DST, DPO, NLU, Seq2Seq model, and RL), generative adversarial network (GAN), a fascinating deep learning model developed by Goodfellow et al. (2014), is a likely forerunner of a revolutionary chatbot technology. In a GAN, a generative model, i.e., the generator, captures data distribution, and a discriminative model, i.e., the discriminator, estimates the probability of a sample from training data. Both the generator and discriminator are trained simultaneously. This method is widely adopted in the field of computer vision for image synthesis, semantic image editing, and image super-resolution (Creswell et al., 2018). GAN can also potentially

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

be used in the dialogue system. Sun et al. (2018) proposed that using a discriminator to distinguish between human-generated and machine-generated dialogues as rewards in an RL task could enable chatbots to output human-like responses. However, it is still not clear how GAN can be effectively adapted to chatbot systems. We call for more relevant research on the subjective and believe that it can lead to a new branch of chatbot classification: seqGAN-based chatbots.

[5.3 Multimedia interaction]

The incentives of multi-modal inputs may derive from inclusivity and equality. In terms of disability, neurodiversity, and age, there are quite a group of people who are not capable of typing text. The cultural background and language proficiency may also impact the ability of English writing. Thus, it is important for chatbot designers to think about how to extend the diversity of input modes and support as broad user groups as possible.

A major problem in the utilisation of multimedia interactions is waste of information. Human expressions in the form of audition and vision contains more information than in the forms of words. Based on tones, facial expressions, and gestures, we learn the connotations behind the literal meanings. For example, when humans say “I am fine,” they sometimes mean they are really so but sometimes say so perfunctorily. This interpretation depends on the context and tone. Another example comes from the certain user groups mentioned above. People with low language proficiency may use inappropriate expressions, seemingly angry and offensive. And such a misinterpretation can be rectified via extra-linguistic markers such as tone, gaze, and smiling. However, current techniques in multimedia interaction are limited in processing such information. Future research should consider non-verbal information to better understand user queries and generate more appropriate responses.

[5.4 Interface]

In the present study, we learn from usability analysis that chatbot research mainly aims to achieve high task performance. However, human-computer interaction is deemed insufficient for the development of emerging IT artefacts. Many other factors determine the acceptance of chatbot systems, such as perceived ease of use, perceived usefulness (Davis, 1989), aesthetics (Cyr, Head, & Ivanov, 2006), perceived enjoyment (van der Heijden, 2004), and anthropomorphism (Schuetzler, Grimes, & Scott Giboney, 2020), and all of which are related to interface design. Nevertheless, few

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

chatbots have revealed their interactive user interfaces, and there is poor documentation of related research (Nuruzzaman & Hussain, 2018). The effects of these factors on the use of chatbots therefore remain unclear. We recommend more studies focusing on interface design and evaluation.

[5.5 Customization]

The user requirement of customization is important for service-oriented chatbots because modern users are accustomed to personalised services. The main technical challenge of customisation lies in enabling chatbots to deliver appropriate responses to specific users. Some existing approaches can be harnessed to realise this goal. For example, the GAIML designed by Pirrone et al. (2008) provides a personalised interface, and the Persona-AIML developed by Galvão et al. (2004) enables the use of chatbots with different personalities. However, there remain several unsolved problems, such as the identification of user preference, the optimal strategy for the most appropriate responses, the impact of personality on human-chatbot communications, and the design of specific personalities for users with corresponding preferences. The customization of chatbots is still in very early stages and requires much more research before it can be successfully implemented.

[5.6 Emotion-sensitive communication]

A significant distinction between humans and chatbots is that the latter lacks emotional resonance. Providing emotion-sensitive responses is therefore a potential research topic for chatbots. Although a few studies have used emotion detection (Kuramoto et al., 2018; Lee et al., 2020), they have not fully used the detected emotion to tailor the responses. We recommend that more research on the strategy and technology of emotion-sensitive response generation be conducted to achieve this goal.

[5.7 Convergence and divergence in application]

On the one hand, current research on chatbots is focused on healthcare, customer service, education, and personal assistance. Therefore, we argue that the focus should be broadened to more potential application fields. For example, chatbots can play a vital role in the development of novel concepts such as smart homes, intelligent vehicles, smart cities, and wisdom pensions.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

On the other hand, because the user requirements vary in different application fields, the development of chatbots in each field should be tailored to the requirements of the field. However, we find that only researchers working on chatbots in the online customer service field have been alerted to the problem of information security and user trust, whereas those in other fields have no consensus on the challenges of their respective fields. Therefore, we suggest that the process of chatbots design across various application fields be balanced between convergence and divergence.

[5.8 Usability Analysis]

We believe that a mature research discipline should balance technical development with usability analysis. The former helps design novel IT artefacts, while the latter helps provide feedback for improvements. However, research on tentative instances goes beyond usability analysis in the field of chatbot research. Based on our analysis, we believe that it is time to conduct exploratory research on the user's advanced requirements and thereby guide future design.

Figures and Tables

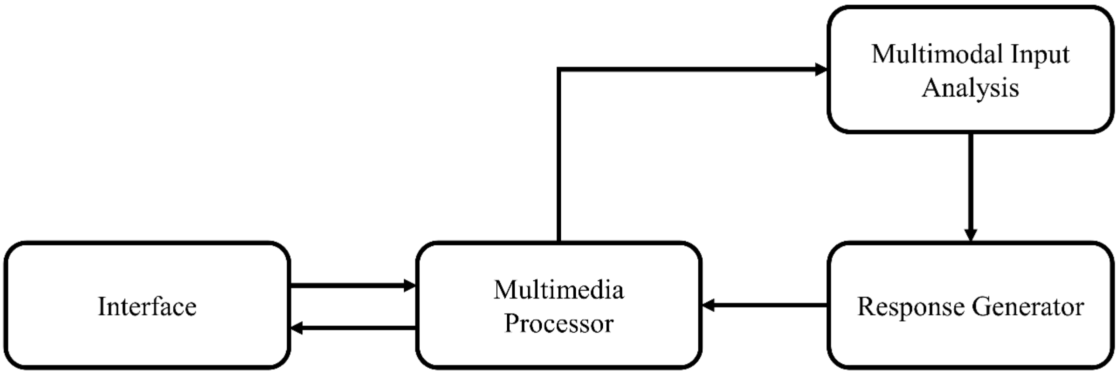


Figure 1. The generic architecture of chatbot systems.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

```
<category>

<pattern>10 Dollars</pattern>

<template>Wow, that is cheap.

</template>

</category>
```

Figure 2. An example category in artificial intelligence mark-up language (AIML).

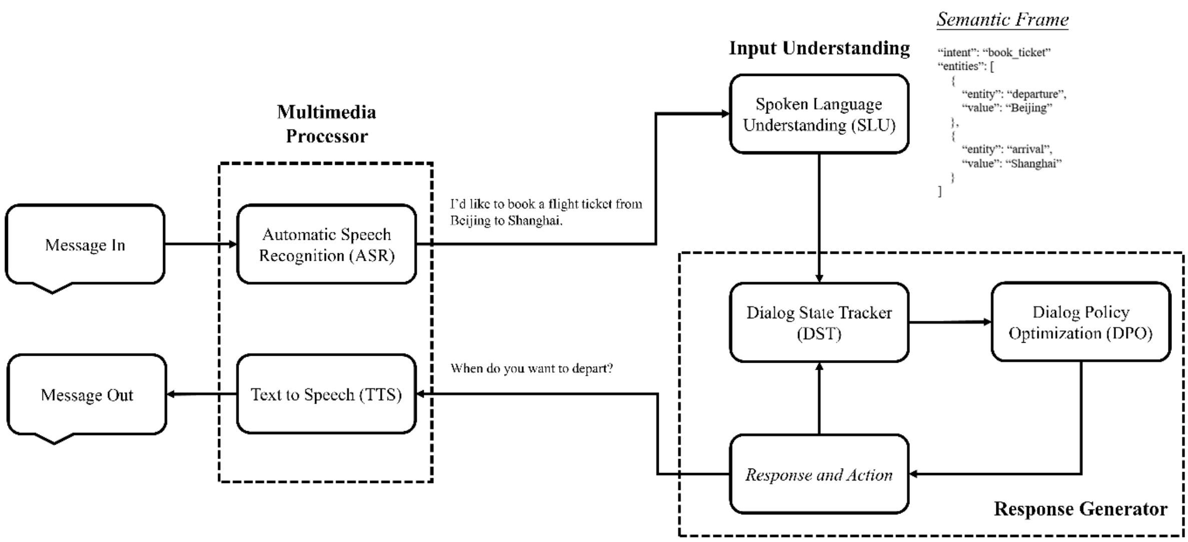


Figure 3. The pipeline mechanism of intent-based chatbots.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

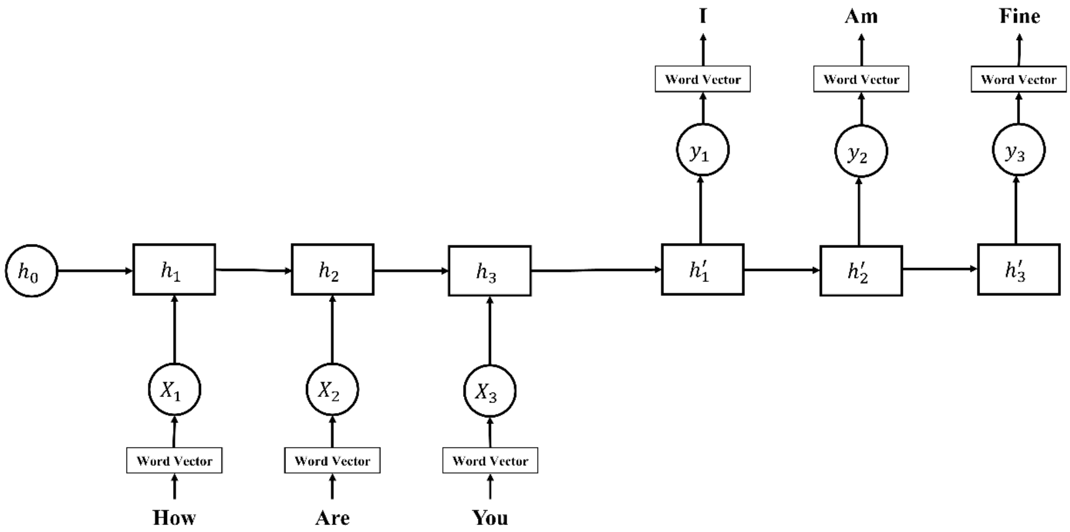


Figure 4. The computational mechanism of the sequence-to-sequence (Seq2Seq) model.

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

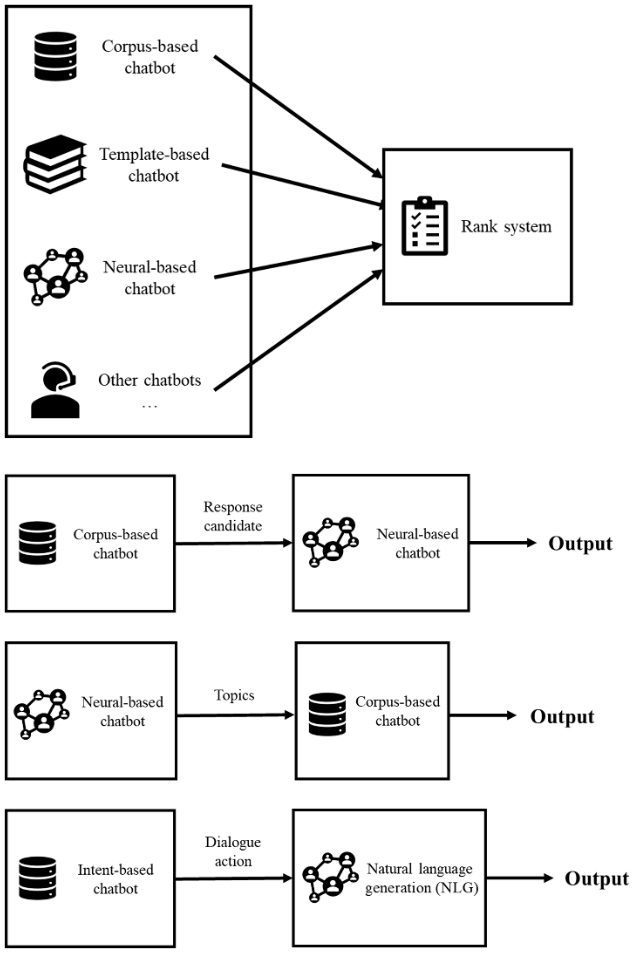


Figure 5. Typical approaches used by chatbots with hybrid techniques.

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

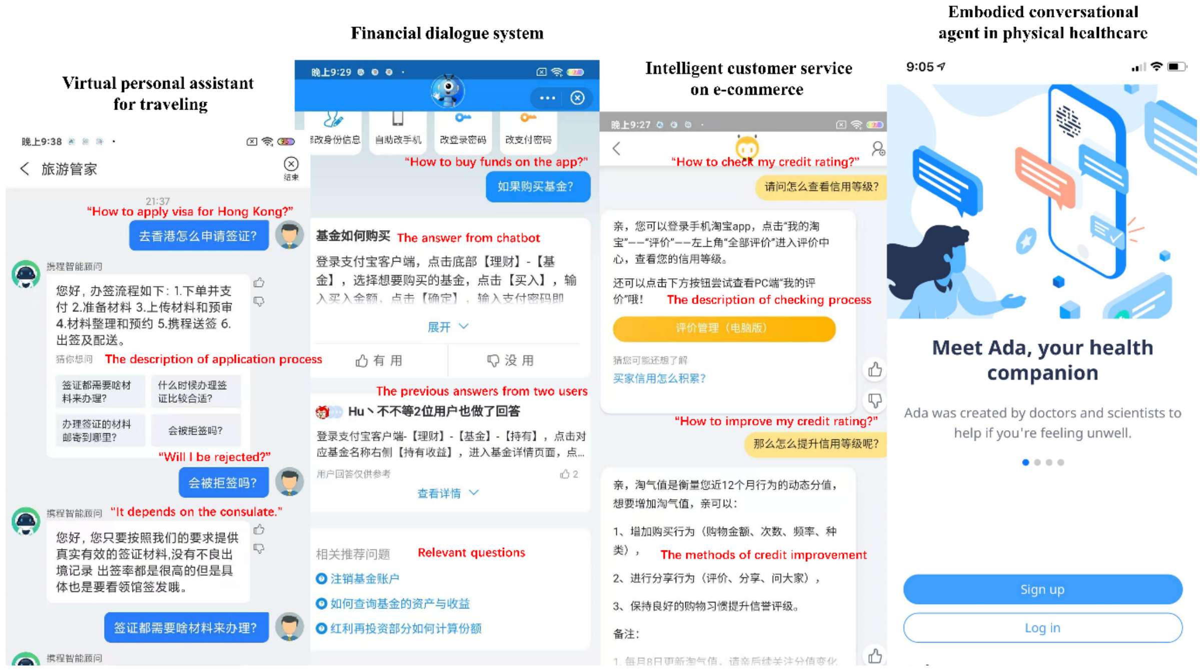


Figure 6. Chatbots used for diverse applications.

Sidebar title:

[Please include sidebars in the body of the text where appropriate]

Conclusion

We present a critical review of state-of-the-art chatbot technologies and applications. First, we extend the general architecture of chatbot systems based on the blueprint developed by Abdul-Kader and Woods (2015) and provide a new classification scheme of chatbots based on machine learning-based response-generating techniques. We propose six categories of chatbots, namely template-, corpus-, intent-, RNN-, and RL-based chatbots, and those with hybrid approaches. Moreover, we examine and analyse the literature on the contemporary applications of chatbots. Our review shows that current research mainly focuses on the applications of chatbots to healthcare, customer service, education, and personal assistance. Furthermore, we analyse existing usability studies of chatbots and find that although chatbots have weak appropriateness of response generation, they show

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

promise in improving the performance of various human tasks. Finally, we identify existing research gaps and propose future research directions to address the identified shortcomings of chatbots.

Funding Information

Our research work was partly supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project: CityU 11507219), the NSFC Basic Research Program (Project No. 71671155), and the CityU Shenzhen Research Institute.

Research Resources

Acknowledgments

Notes

References

Abd-Alrazaq, A., Safi, Z., Alajlani, M., Warren, J., Househ, M., & Denecke, K. (2020). Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. *Journal of Medical Internet Research*, *22*(6). https://doi.org/10.2196/18301

Abdul-Kader, S. A., & Woods, J. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications*, *6*(7), 72–80. https://doi.org/10.14569/ijacsa.2015.060712

Abdul-Kader, S. A., & Woods, J. (2018). Question answer system for online feedable new born Chatbot. Proceedings of the 2017, 2018 *Intelligent Systems Conference*, 863–869. https://doi.org/10.1109/IntelliSys.2017.8324231

Abushawar, B., & Atwell, E. (2015). ALICE Chatbot: Trials and Outputs. *Computación y Sistemas*, *19*(4), 1–8. https://doi.org/10.13053/CyS-19-4-2326.

Ahmadvand, A., Choi, J. I., & Agichtein, E. (2019). Contextual dialogue act classification for open-domain conversational agents. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1273–1276. https://doi.org/10.1145/3331184.3331375

Altinok, D. (2018). *An Ontology-Based Dialogue Management System for Banking and Finance Dialogue*

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Systems. Retrieved from http://arxiv.org/abs/1804.04838

Arsovski, S., Cheok, A. D., Govindarajoo, K., Salehuddin, N., & Vedadi, S. (2020). Artificial intelligence snapchat: Visual conversation agent. *Applied Intelligence*, *50*(7), 2040–2049. https://doi.org/10.1007/s10489-019-01621-2

Arsovski, S., Osipyan, H., Oladele, M. I., & Cheok, A. D. (2019). Automatic knowledge extraction of any Chatbot from conversation. *Expert Systems with Applications*, *137*, 343–348. https://doi.org/10.1016/j.eswa.2019.07.014

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *Proceedings of 3rd International Conference on Learning Representations*, 1–15.

Beketov, M., Lehmann, K., & Wittke, M. (2018). Robo Advisors: quantitative methods inside the robots. *Journal of Asset Management*, *19*(6), 363–370. https://doi.org/10.1057/s41260-018-0092-9

Belfin, R. V., Shobana, A. J., Manilal, M., Mathew, A. A., & Babu, B. (2019). A Graph Based Chatbot for Cancer Patients. *Proceedings of International Conference on Advanced Computing and Communication Systems*, 717–721. https://doi.org/10.1109/ICACCS.2019.8728499

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). Next Generation: Chatbots in Clinical Psychology and Psychotherapy to Foster Mental Health - A Scoping Review. *Verhaltenstherapie*, *29*(4), 266–280. https://doi.org/10.1159/000499492

Bibault, J. E., Chaix, B., Guillemassé, A., Cousin, S., Escande, A., Perrin, M., … Brouard, B. (2019). A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial. *Journal of Medical Internet Research*, *21*(11), 1–7. https://doi.org/10.2196/15787

Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). *Rasa: Open Source Language Understanding and Dialogue Management*. 1–9. Retrieved from http://arxiv.org/abs/1712.05181

Bozic, J., & Wotawa, F. (2018). Security testing for chatbots. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-99927-2_3

Bradeško, L., & Mladenić, D. (2012). A Survey of Chatbot Systems through a Loebner Prize Competition. *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies*, *2*, 34–37. Retrieved from https://www.researchgate.net/publication/235664166

Bunk, T., Varshneya, D., Vlasov, V., & Nichol, A. (2020). *DIET: Lightweight Language Understanding for Dialogue Systems*. Retrieved from http://arxiv.org/abs/2004.09936

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

Caballé, S., & Conesa, J. (2019). Conversational Agents in Support for Collaborative Learning in MOOCs: An Analytical Review. *Lecture Notes on Data Engineering and Communications Technologies*, *23*, 384–394. https://doi.org/10.1007/978-3-319-98557-2

Chen, H., Liu, X., Yin, D., & Tang, J. (2017). A Survey on Dialogue Systems: Recent Advances and New Frontiers. *Acm Sigkdd Explorations Newsletter*, *19*(2), 25–35. https://doi.org/10.1145/3166054.3166058

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing,* 1724–1734. https://doi.org/10.3115/v1/d14-1179

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 1–9. Retrieved from http://arxiv.org/abs/1412.3555

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial network: An overview. *IEEE Signal Processing Magazine*, *35*(1), 53–65. https://doi.org/10.19650/j.cnki.cjsi.J1804413

Cuayáhuitl, H., Lee, D., Ryu, S., Cho, Y., Choi, S., Indurthi, S., … Kim, J. (2019). Ensemble-based deep reinforcement learning for chatbots. *Neurocomputing*, *366*, 118–130. https://doi.org/10.1016/j.neucom.2019.08.007

Cui, L., Huang, S., Wei, F., Tan, C., Duan, C., & Zhou, M. (2017). Superagent: A customer service chatbot for E-commerce websites. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics,* 97–102. https://doi.org/10.18653/v1/P17-4017

Cyr, D., Head, M., & Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information and Management*, *43*(8), 950–963. https://doi.org/10.1016/j.im.2006.08.009

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, *13*(3), 319–339. https://doi.org/10.2307/249008

de Gennaro, M., Krumhuber, E. G., & Lucas, G. (2020). Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood. *Frontiers in Psychology*, *10*(January), 1–14. https://doi.org/10.3389/fpsyg.2019.03061

Deryugina, O. V. (2010). Chatterbots. *Scientific and Technical Information Processing*, *37*(2), 143–147. https://doi.org/10.3103/S0147688210020097

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*(Mlm), 4171–4186.

Doherty, D., & Curran, K. (2019). Chatbots for online banking services. *Web Intelligence*, *17*(4), 327–342.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E

Fadhil, A., Wang, Y., & Reiterer, H. (2019). Assistive Conversational Agent for Health Coaching: A Validation Study. *Methods of Information in Medicine*, *58*(1), 9–23. https://doi.org/10.1055/s-0039-1688757

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, *4*(2), e19. https://doi.org/10.2196/mental.7785

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service ? An Exploratory Interview Study. *Lecture Notes in Computer Science*, *11193*(December), 194–208. https://doi.org/10.1007/978-3-030-01437-7

Franco, M. F., Rodrigues, B., Scheid, E. J., Jacobs, A., Killer, C., Granville, L. Z., & Stiller, B. (2020). SecBot: A business-driven conversational agent for cybersecurity planning and management. *Proceedings of the 16th International Conference on Network and Service Management, CNSM 2020, 2nd International Workshop on Analytics for Service and Application Management, AnServApp 2020 and 1st International Workshop on the Future Evolution of Internet Protocols, IPFutu*. https://doi.org/10.23919/CNSM50824.2020.9269037

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Galvão, A. M., Barros, F. A., Neves, A. M. M., & Ramalho, G. L. (2004). Adding personality to chatterbots using the Persona-AIML architecture. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *3315*, 963–973. https://doi.org/10.1007/978-3-540-30498-2_96

Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., … Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling*, *100*(9),

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1720–1729. https://doi.org/10.1016/j.pec.2017.04.015

Ghosh, C., & Eastin, M. S. (2020). Understanding users' relationship with voice assistants and how it affects privacy concerns and information disclosure behavior. *International Conference on Human-Computer Interaction*, 381–392. https://doi.org/10.1007/978-3-030-50309-3_25

Goo, C. W., Gao, G., Hsu, Y. K., Huo, C. L., Chen, T. C., Hsu, K. W., & Chen, Y. N. (2018). Slot-gated modeling for joint slot filling and intent prediction. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *2*, 753–757. https://doi.org/10.18653/v1/n18-2118

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). *Generative adversarial networks*. https://doi.org/10.1145/3422622

Greer, S., Ramo, D., Chang, Y. J., Fu, M., Moskowitz, J., & Haritatos, J. (2019). Use of the chatbot "vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: Randomized controlled feasibility trial. *JMIR MHealth and UHealth*, *7*(10), 1–13. https://doi.org/10.2196/15018

Guo, H. (2015). *Generating Text with Deep Reinforcement Learning*. 1–10. Retrieved from http://arxiv.org/abs/1510.09202

Handoyo, E., Arfan, M., Soetrisno, Y. A. A., Somantri, M., Sofwan, A., & Sinuraya, E. W. (2018). Ticketing Chatbot Service using Serverless NLP Technology. *Proceedings of the 2018 5th International Conference on Information Technology, Computer and Electrical Engineering, ICITACEE 2018*, 325–330. https://doi.org/10.1109/ICITACEE.2018.8576921

Haristiani, N., Danuwijaya, A. A., Rifai, M. M., & Sarila, H. (2019). Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology*, *14*(6), 3158–3173.

Hauser-Ulrich, S., Künzli, H., Meier-Peterhans, D., & Kowatsch, T. (2020). A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): Pilot randomized controlled trial. *JMIR MHealth and UHealth*, *8*(4). https://doi.org/10.2196/15806

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/https://doi.org/10.1162/neco.1997.9.8.1735

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., … Akkiraju, R. (2018). Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. *CHI 2018*. https://doi.org/10.1097/00003727-199507000-

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

00013

Huang, C. Y., Yang, M. C., Chen, Y. J., Wu, M. L., & Chen, K. W. (2019). A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion. *IEEE International Conference on Industrial Engineering and Engineering Management*, *2019-Decem*, 1791–1795. https://doi.org/10.1109/IEEM.2018.8607399

Jonell, P. (2019). Using social and physiological signals for user adaptation in conversational agents: Doctoral consortium. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, *4*(c), 2420–2422.

Jusoh, S. (2019). Intelligent Conversational Agent for Online Sales. *Proceedings of the 10th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2018*, 2018–2021. https://doi.org/10.1109/ECAI.2018.8679045

Kadariya, D., Venkataramanan, R., Yip, H. Y., Kalra, M., Thirunarayanan, K., & Sheth, A. (2019). KBot: Knowledge-enabled personalized chatbot for asthma self-management. *Proceedings of the 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*, 138–143. https://doi.org/10.1109/SMARTCOMP.2019.00043

Kadous, M. W., & Sammut, C. (2004). InCA: A mobile conversational agent. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *3157*, 644–653. https://doi.org/10.1007/978-3-540-28633-2_68

Kasinathan, V., Xuan, F. S., Wahab, M. H. A., & Mustapha, A. (2018). Intelligent Healthcare Chatterbot (HECIA): Case study of medical center in Malaysia. *Proceedings of the 2017 IEEE Conference on Open Systems,* 32–37. https://doi.org/10.1109/ICOS.2017.8280270

Katchapakirin, K., & Anutariya, C. (2018). An Architectural Design of ScratchThAI A conversational agent for Computational Thinking Development using Scratch. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3291280.3291787

Kim, J., Park, S. Y., & Robert, L. (2019). Conversational Agents for Health and Wellbeing: Review and Future Agendas. *Workshop on Identifying Challenges and Opportunities in Human–AI Collaboration in Healthcare at the 22th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2019)*.

Kimani, E., Rowan, K., McDuff, D., Czerwinski, M., & Mark, G. (2019). A Conversational Agent in Support of Productivity and Wellbeing at Work. *Proceedings of the 8th International Conference on Affective*

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Computing and Intelligent Interaction, 332–338. https://doi.org/10.1109/ACII.2019.8925488

Kocaballi, A. B., Quiroz, J. C., Rezazadegan, D., Berkovsky, S., Magrabi, F., Coiera, E., & Laranjo, L. (2020). Responses of conversational agents to health and lifestyle prompts: Investigation of appropriateness and presentation structures. *Journal of Medical Internet Research*, *22*(2). https://doi.org/10.2196/15823

Krassmann, A. L., Flach, J. M., Da Silva Grando, A. R. C., Tarouco, L. M. R., & Bercht, M. (2019). A process for extracting knowledge base for chatbots from text corpora. *IEEE Global Engineering Education Conference, EDUCON*, *April-2019*, 322–329. https://doi.org/10.1109/EDUCON.2019.8725064

Kuramoto, I., Yoshikawa, Y., Baba, J., Kawabata, T., Ogawa, K., & Ishiguro, H. (2018). Conversational agents to suppress customer anger in text-based customer-support conversations. *Proceedings of the 6th International Conference on Human-Agent Interaction*, 114–121. https://doi.org/10.1145/3284432.3284457

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2020). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11970 LNCS*, 187–200. https://doi.org/10.1007/978-3-030-39540-7_13

Lai, S.-T., Leu, F.-Y., & Lin, J.-W. (2018). A Banking Chatbot Security Control Procedure for Protecting User Data Security and Privacy. *International Conference on Broadband and Wireless Computing, Communication and Applications*, 561–571. https://doi.org/10.1007/978-3-030-02613-4

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., … Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072

Latham, A. M., Crockett, K. A., McLean, D. A., Edmonds, B., & O'Shea, K. (2010). Oscar: An intelligent conversational agent tutor to estimate learning styles. *Proceedings of the 2010 IEEE World Congress on Computational Intelligence*. https://doi.org/10.1109/FUZZY.2010.5584064

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–31. https://doi.org/10.1145/3274371

Lee, M. C., Chiang, S. Y., Yeh, S. C., & Wen, T. F. (2020). Study on emotion recognition and companion Chatbot using deep neural network. *Multimedia Tools and Applications*, *79*(27–28), 19629–19657. https://doi.org/10.1007/s11042-020-08841-6

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

Lokman, A. S., & Zain, J. M. (2011). Enhancement algorithms for SQL-based chatbot. *Communications in Computer and Information Science*, *180 CCIS*(PART 2), 470–479. https://doi.org/10.1007/978-3-642-22191-0_41

Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98*(4), 285–287. https://doi.org/10.2471/BLT.19.237636

MacEclo, P., Pereira, C., Mota, P., Silva, D., Frade, A., & Madeira, R. N. (2019). Conversational agent in mhealth to empower people managing Parkinson's disease. *Procedia Computer Science*, *160*, 402–408. https://doi.org/10.1016/j.procs.2019.11.074

Madhu, D., Jain, C. J. N., Sebastain, E., Shaji, S., & Ajayakumar, A. (2017). A novel approach for medical assistance using trained chatbot. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017*, (Icicct), 243–246. https://doi.org/10.1109/ICICCT.2017.7975195

Mao, G., Su, J., Yu, S., & Luo, D. (2019). Multi-Turn Response Selection for Chatbots With Hierarchical Aggregation Network of Multi-Representation. *IEEE Access*, *7*, 111736–111745. https://doi.org/10.1109/access.2019.2934149

Marietto, M. D. G. B., De Aguiar, R. V., Botelho, W. T., & Dos Santos França, R. (2014). A new graphical representation of the chatterbots' knowledge base. *Lecture Notes in Computer Science*, *8610*, 513–524. https://doi.org/10.1007/978-3-319-09912-5_43

Martínez-Miranda, J., Martínez, A., Ramos, R., Aguilar, H., Jiménez, L., Arias, H., … Valencia, E. (2019). Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *Journal of Medical Systems*, *43*(8). https://doi.org/10.1007/s10916-019-1387-1

Mathur, S., & Lopez, D. (2019). A scaled-down neural conversational model for chatbots. *Concurrency Computation* , *31*(10), 1–10. https://doi.org/10.1002/cpe.4761

Mikolov, Tomáš, Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. *Eleventh Annual Conference of the International Speech Communication Association,* 1045-1048. https://doi.org/10.1080/00224065.1993.11979473

Mikolov, Tomas, Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 1–9.

Mimoun, M. S. Ben, Poncin, I., & Garnier, M. (2012). Case study-Embodied virtual agents: An analysis on

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reasons for failure. *Journal of Retailing and Consumer Services*, *19*(6), 605–612. https://doi.org/10.1016/j.jretconser.2012.07.006

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Montenegro, J. L. Z., da Costa, C. A., & da Rosa Righi, R. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, *129*, 56–67. https://doi.org/10.1016/j.eswa.2019.03.054

Moore, R. J. (2018). *Studies in Conversational UX Design*. https://doi.org/10.1007/978-3-319-95579-7

Mrkšic, N., Séaghdha, D., Wen, T. H., Thomson, B., & Young, S. (2017). Neural belief tracker: Data-driven dialogue state tracking. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 1777–1788. https://doi.org/10.18653/v1/P17-1163

Mujeeb, S., Hafeez, M., & Arshad, T. (2017). Aquabot: A Diagnostic Chatbot for Achluophobia and Autism. *International Journal of Advanced Computer Science and Applications*, *8*(9), 209–216. https://doi.org/10.14569/ijacsa.2017.080930

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, *5*, 1–12. https://doi.org/10.1177/2055207619871808

Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2020). A Review of Current Trends in the Development of Chatbot Systems. *Proceedings of the 6th International Conference on Advanced Computing and Communication Systems,* 706–710. https://doi.org/10.1109/ICACCS48705.2020.9074420

Nazir, A., Khan, M. Y., Ahmed, T., Jami, S. I., & Wasi, S. (2019). A Novel Approach for Ontology-Driven Information Retrieving Chatbot for Fashion Brands. *International Journal of Advanced Computer Science and Applications*, *10*(9), 546–552. https://doi.org/10.1017/CBO9781107415324.004

Neves, A. M. M., Barros, F. A., & Hodges, C. (2006). IAIML: A mechanism to treat intentionality in AIML chatterbots. *Proceedings of the International Conference on Tools with Artificial Intelligence,* 225–231. https://doi.org/10.1109/ICTAI.2006.64

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An Initial Model of Trust in Chatbots for Customer Service - Findings from a Questionnaire Study. *Interacting with Computers*, *31*(3), 317–335. https://doi.org/10.1093/iwc/iwz022

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Nuruzzaman, M., & Hussain, O. K. (2018). A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks. *Proceedings of the 2018 IEEE 15th International Conference on e-Business Engineering,* 54–61. https://doi.org/10.1109/ICEBE.2018.00019

Okuda, T., & Shoda, S. (2018). AI-based chatbot service for financial industry. *Fujitsu Scientific and Technical Journal*, *54*(2), 4–8.

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *Journal of Medical Internet Research*, *21*(4), 1–10. https://doi.org/10.2196/12887

Park, S., Choi, J., Lee, S., Oh, C., Kim, C., La, S., … Suh, B. (2019). Designing a chatbot for a brief motivational interview on stress management: Qualitative case study. *Journal of Medical Internet Research*, *21*(4). https://doi.org/10.2196/12231

Peng, B., Li, C., Li, J., Shayandeh, S., Liden, L., & Gao, J. (2020). *SOLOIST: Building Task Bots at Scale with Transfer Learning and Machine Teaching*. (May). Retrieved from http://arxiv.org/abs/2005.05298

Peng, B., Zhu, C., Li, C., Li, X., Li, J., Zeng, M., & Gao, J. (2020). *Few-shot Natural Language Generation for Task-Oriented Dialog*. 172–182. https://doi.org/10.18653/v1/2020.findings-emnlp.17

Perski, O., Crane, D., Beard, E., & Brown, J. (2019). Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digital Health*, *5*, 1–13. https://doi.org/10.1177/2055207619880676

Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A., & Micoulaud-Franchi, J. A. (2020). Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *Npj Digital Medicine*, *3*(1), 1–7. https://doi.org/10.1038/s41746-019-0213-y

Pirrone, R., Russo, G., Cannella, V., & Peri, D. (2008). GAIML: A new language for verbal and graphical interaction in chatbots. *Mobile Information Systems*, *4*(3), 195–209. https://doi.org/10.1155/2008/636924

Ponathil, A., Ozkan, F., Welch, B., Bertrand, J., & Chalil Madathil, K. (2020). Family health history collected by virtual conversational agents: An empirical study to investigate the efficacy of this approach. *Journal of Genetic Counseling*, (February), 1–12. https://doi.org/10.1002/jgc4.1239

Prasomphan, S. (2019). Improvement of chatbot in trading system for SMEs by using deep neural network. *Proceedings of the 4th International Conference on Cloud Computing and Big Data Analytics*, 517–522. https://doi.org/10.1109/ICCCBDA.2019.8725745

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

psychology: A scoping review. *Journal of Medical Internet Research*, *19*(5). https://doi.org/10.2196/jmir.6553

Pudner, K., Crockett, K., Ieee, M., & Bandar, Z. (2007). An Intelligent Conversational Agent Approach to Extracting Queries from Natural Language. *Proceedings of the World Congress on Engineering*, *1*, 305–310.

Ramesh, K., Ravishankaran, S., Joshi, A., & Chandrasekaran, K. (2017). A Survey of Design Techniques for Conversational Agents. *INFORMATION, COMMUNICATION AND COMPUTING TECHNOLOGY*, *750*, 336–350. https://doi.org/10.1007/978-981-10-6544-6_31

Rosruen, N., & Samanchuen, T. (2019). Chatbot Utilization for Medical Consultant System. *Proceedings of the 3rd Technology Innovation Management and Engineering Science International Conference*, (Im), 1–5. https://doi.org/10.1109/TIMES-iCON.2018.8621678

Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019). Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *Proceedings of the 6th 2019 ACM Conference on Learning at Scale, L@S 2019*, 1–4. https://doi.org/10.1145/3330430.3333643

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875–900. https://doi.org/10.1080/07421222.2020.1790204

Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., … Bengio, Y. (2018). *A Deep Reinforcement Learning Chatbot*. 1–40. Retrieved from http://arxiv.org/abs/1801.06700

Shamekhi, A., Bickmore, T., Lestoquoy, A., & Gardiner, and P. (2017). Augmenting Group Medical Visits with Conversational Agents for Stress Management Behavior Change. *Lecture Notes in Computer Science*, *10171*, 55–67. https://doi.org/10.1007/978-3-319-55134-0

Shao, L., Gouws, S., Britz, D., Goldie, A., Strope, B., & Kurzweil, R. (2017). Generating high-quality and informative conversation responses with sequence-to-sequence models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2210–2219. https://doi.org/10.18653/v1/d17-1235

Sharma, B., Puri, H., & Rawat, D. (2018). Digital Psychiatry - Curbing Depression using Therapy Chatbot and Depression Analysis. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, (Icicct), 627–631. https://doi.org/10.1109/ICICCT.2018.8472986

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Smestad, T. L. (2018). Personality Matters! Improving The UserExperience of Chatbot Interfaces - Personality provides a stable pattern to guide the design and behaviour of conversational agents. *MS Thesis*, (June).

Song, Y., Yan, R., Li, X., Zhao, D., & Zhang, M. (2016). *Two are Better than One: An Ensemble of Retrieval- and Generation-Based Dialog Systems*. *1*, 1–11. Retrieved from http://arxiv.org/abs/1610.07149

Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An Embodied Conversational Agent for Unguided Internet-Based Cognitive Behavior Therapy in Preventative Mental Health: Feasibility and Acceptability Pilot Trial. *JMIR Mental Health*, *5*(3), e10454. https://doi.org/10.2196/10454

Sun, X., Chen, X., Pei, Z., & Ren, F. (2018). Emotional Human Machine Conversation Generation Based on SeqGAN. *Proceedings of the 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018*. https://doi.org/10.1109/ACIIAsia.2018.8470388

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, *4*(January), 3104–3112.

Sweeney, M. E., & Davis, E. (2020). Alexa, are you listening?: An exploration of smart voice assistant use and privacy in libraries. *Information Technology and Libraries*, *39*(4), 1–21. https://doi.org/10.6017/ITAL.V39I4.12363

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., … Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, *22*(3), 1–10. https://doi.org/10.2196/16235

Tegos, S., Demetriadis, S., & Karakostas, A. (2014). Leveraging conversational agents and concept maps to scaffold students' productive talk. *Proceedings of the 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014*, 176–183. https://doi.org/10.1109/INCoS.2014.66

Tegos, S., Demetriadis, S., & Karakostas, A. (2015). Promoting academically productive talk with conversational agent interventions in collaborative learning settings. *Computers and Education*, *87*, 309–325. https://doi.org/10.1016/j.compedu.2015.07.014

Tegos, S., Demetriadis, S., Papadopoulos, P. M., & Weinberger, A. (2016). Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions. *International*

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Journal of Computer-Supported Collaborative Learning, *11*(4), 417–440. https://doi.org/10.1007/s11412-016-9246-2

Tegos, S., Demetriadis, S., Tegos, S., & Demetriadis, S. (2017). International Forum of Educational Technology & Society Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Journal of Educational Technology & Society*, *20*(1), 99–111.

Tran, V. K., & Nguyen, L. M. (2018). Semantic Refinement GRU-Based Neural Language Generation for Spoken Dialogue Systems. In *Communications in Computer and Information Science* (Vol. 781). https://doi.org/10.1007/978-981-10-8438-6_6

van der Heijden, H. (2004). User Acceptance of Hedonic Information Systems. *MIS Quarterly*, *28*(4), 695–704. https://doi.org/10.2307/25148660.

Van Heerden, A., Ntinga, X., & Vilakazi, K. (2017). The potential of conversational agents to provide a rapid HIV counseling and testing services. *Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science, FADS 2017*, *2018-Janua*, 80–84. https://doi.org/10.1109/FADS.2017.8253198

Wallace, R. (2003). The Elements of AIML Style. In *Alice AI Foundation*. Retrieved from https://files.ifi.uzh.ch/cl/hess/classes/seminare/chatbots/style.pdf

Wanner, L., André, E., Blat, J., Dasiopoulou, S., Farrùs, M., Fraga, T., … Vrochidis, S. (2017). Kristina: A knowledge-based virtual conversation agent. *International Conference on Practical Applications of Agents and Multi-Agent Systems*, 284–295. https://doi.org/10.1007/978-3-319-59930-4_23

Wen, T. H., Gašić, M., Kim, D., Mrkšić, N., Su, P. H., Vandyke, D., & Young, S. (2015). Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference*, 275–284. https://doi.org/10.18653/v1/w15-4639

Wiak, S., & Kosiorowski, P. (2010). The use of psycholinguistics rules in case of creating an intelligent chatterbot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6114 LNAI*(PART 2), 689–697. https://doi.org/10.1007/978-3-642-13232-2_85

Williams, J. D., Raux, A., & Henderson, M. (2016). The Dialog State Tracking Challenge Series: A Review. *Dialogue & Discourse*, *7*(3), 4–33. https://doi.org/10.5087/dad.2016.301

Wu, Y., Li, Z., Wu, W., & Zhou, M. (2018). Response selection with topic clues for retrieval-based chatbots.

Neurocomputing, 316, 251–261. https://doi.org/10.1016/j.neucom.2018.07.073

Wu, Y., Wu, W., Xing, C., Li, Z., & Zhou, M. (2017). Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 496–505. https://doi.org/10.18653/v1/P17-1046

Wu, Y., Wu, W., Xing, C., Li, Z., & Zhou, M. (2018). A sequential matching framework for multi turn response selection in retrieval based chatbots. *Computational Linguistics*, *45*(1), 163–197. https://doi.org/10.1162/COLI

Xiao, Z., Zhou, M. X., & Fu, W. T. (2019). Who should be my teammates: Using a conversational agent to understand individuals and help teaming. *Proceedings of the 24th International Conference on Intelligent User Interfaces, Proceedings (IUI' 19)*, *Part F1476*, 437–447. https://doi.org/10.1145/3301275.3302264

Xue, Z., Ko, T. Y., Yuchen, N., Wu, M. K. D., & Hsieh, C. C. (2018). Isa: Intuit smart agent, a neural-based agent-assist chatbot. *IEEE International Conference on Data Mining Workshops, ICDMW*, 1423–1428. https://doi.org/10.1109/ICDMW.2018.00202

Yang, Xi, Aurisicchio, M., & Baxter, W. (2019). Understanding affective experiences with conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 1–12. https://doi.org/10.1145/3290605.3300772

Yang, Xiaohao, & Liu, J. (2015). Using word confusion networks for slot filling in spoken language understanding. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2015-Janua*(3), 1353–1357.

Yen, C., & Chiang, M. C. (2020). Trust me, if you can: a study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. *Behaviour and Information Technology*, *0*(0), 1–18. https://doi.org/10.1080/0144929X.2020.1743362

Yin, Z., Chang, K. H., & Zhang, R. (2017). DeepProbe: Information directed sequence understanding and chatbot design via recurrent neural networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 2131–2139. https://doi.org/10.1145/3097983.3098148

Zhao, G., Zhao, J., Li, Y., Alt, C., Schwarzenberg, R., Hennig, L., … Xu, F. (2019). MOLI: Smart conversation agent for mobile customer service. *Information (Switzerland)*, Vol. 10. https://doi.org/10.3390/info10020063

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

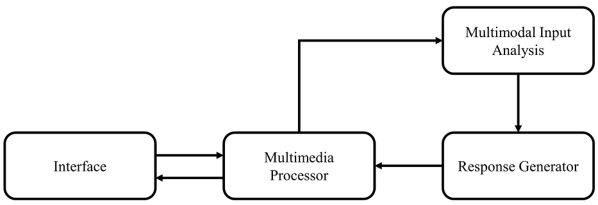
Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers*, 207–212. https://doi.org/10.18653/v1/p16-2034

Further Reading

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript



WIDM_1434_Fig1.tif

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

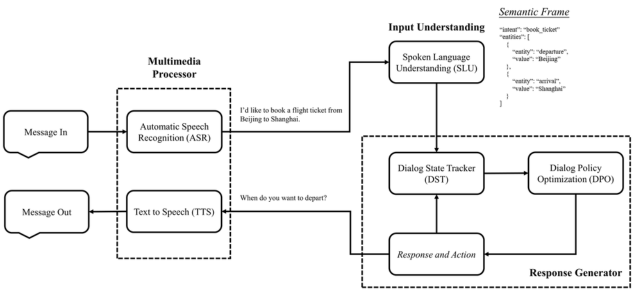
Author Manuscript

```
<category>

<pattern>10 Dollars</pattern>

<template>Wow, that is cheap.

</template>

</category>
```

WIDM_1434_Fig2.tif

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

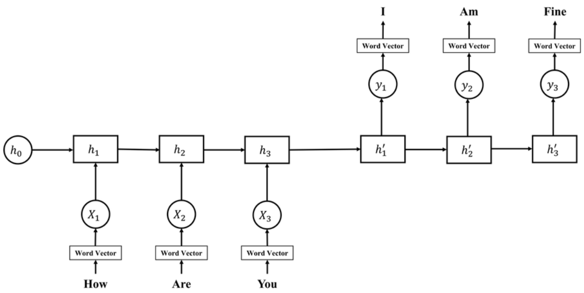
Author Manuscript



WIDM_1434_Fig3.tif

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

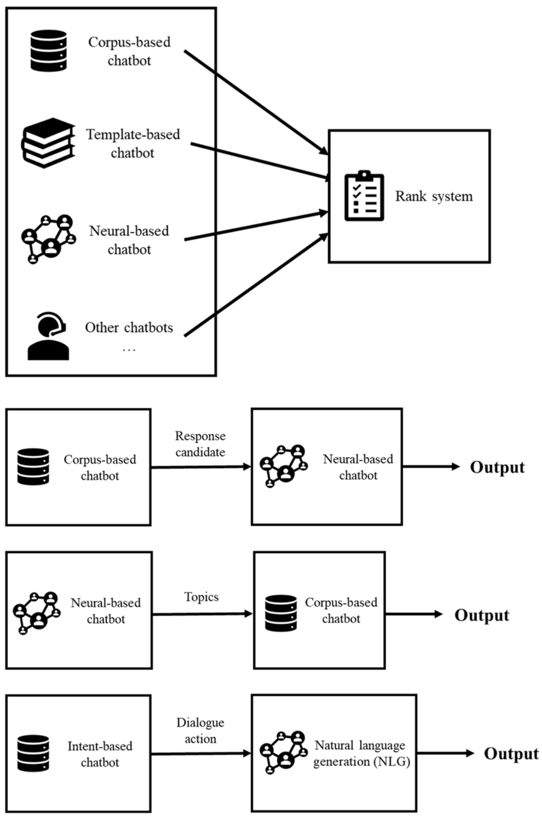
Author Manuscript



WIDM_1434_Fig4.tif

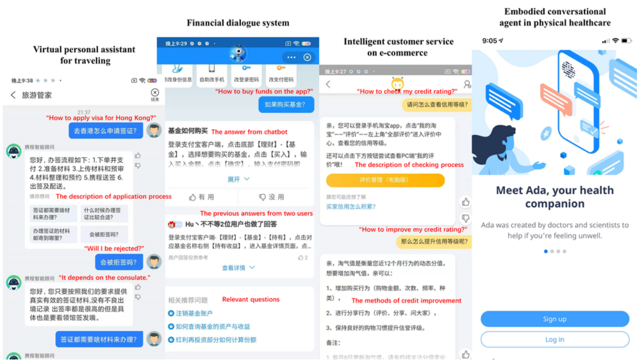
19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript



WIDM_1434_Fig5.tif

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

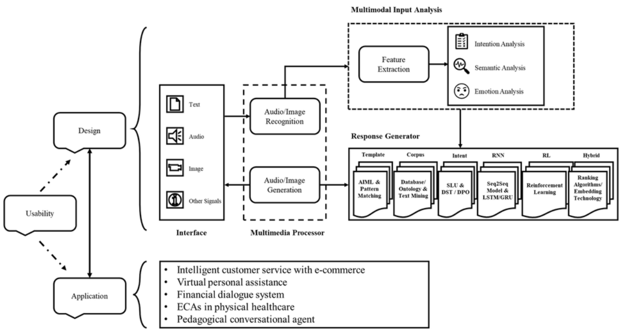


WIDM_1434_Fig6.tif

Author Manuscript

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Author Manuscript



WIDM_1434_graphical abstract.tif

19424795, 2022, 1, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1434, Wiley Online Library on [30/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License