An Overview of Machine Learning in Chatbots

Prissadang Suta[1]
[1]School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
Email: prissadang.sut@mail.kmutt.ac.th

Xi Lan[2], Biting Wu[2]
[2]Division of Engineering Science, University of Toronto, Ontario, Canada
Email: {xi.lan, mary.wu}@ mail.utoronto.ca

Pornchai Mongkolnam[1] and Jonathan H. Chan[1]
[1]School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
Email: {pornchai, jonathan}@sit.kmutt.ac.th

what why History / Assistants Challenges

***Abstract*— A chatbot is an intelligent system which can hold a conversation with a human using natural language in real time. Due to the rise of Internet usage, many businesses now use online platforms to handle customer inquiries, and many of them turn to chatbots for improving their customer service or for streamlining operations and increasing their productivity. However, there is still a gap between existing chatbots and the autonomous, conversational agents businesses hope to implement. As such, this paper will first provide an overview of chatbots and then focus on research trends regarding the development of human-like chatbots capable of closing this technological gap. We reviewed the literature published over the past decade, from 1998 to 2018, and presented an overview of chatbots using a mind-map. The research findings suggest that chatbots operate in three steps: understanding the natural language input; generating an automatic, relevant response; and, constructing realistic and fluent natural language responses. The current bottleneck in designing artificially intelligent chatbots lies in the industry's lack of natural language processing capabilities. Without the ability to properly understand the content and context of a user's input, the chatbot cannot generate a relevant response.**

***Index Terms*—chatbots, conversational agents, dialog system, human computer interaction**

I. Introduction

A chatbot, also known as a conversational agent, is a computer software capable of taking a natural language input and providing a conversational output in real time [1]. This human-chatbot interaction is typically carried out through a graphical user interface based on human-computer interaction (HCI) principles [2], [3].

The idea of an intelligent machine engaging in human interactions was first theorized by Alan Turing in 1950 [4], [5]. Shortly after, automated computer programs, referred to as "bots", were created to simulate human conversation. For example, ELIZA in 1966 matched user prompts to scripted responses, and Artificial Linguistic Internet Computer Entity (ALICE) in 1995 introduced natural language processing (NLP) to interpret user input [4]. Chatbots now exist in various messaging platforms, such as Facebook Messenger, Skype, and Kik, largely for customer service purposes [6].

Chatbots also evolved to interact via voice as well. Such chatbots are typically known as virtual assistants. In particular, the use of NLP led to the Big Four Voice Assistants: Apple's Siri (released as a standalone app in 2010, bundled into iOS in 2011, and added to the HomePod device in December 2017), Microsoft's Cortana (2013), Amazon's Alexa (released with its Echo products in 2014), and Google's Assistant (announced in 2016 [4], and has been an extension of Google's Voice Search product called Google Now since 2011) [7]. They are embedded in smartphones and smart home devices to control the Internet-of-Things (IoT) enabled devices. These assistants are now using voice recognition powered by AI to learn the words and phrases of the user's voice in order to interact with users in a personalized manner. For example, Audrey was the first documented speech recognition system in 1952, which recognized digits spoken by a single voice. Since then, Siri has improved to recognize users' voices and respond with personality. Although there are still improvements to be made, voice recognition technology is becoming increasingly used in business and commerce which can hear and understand what you are saying even in noisy environments [8]–[10]. In fact, these conversational interfaces were deemed one of the key breakthrough technologies of 2016 [4].

As evident, chatbots have become quite popular over the years. This is likely due to the rise of Internet users worldwide – there were 3.15 billion users in 2015, 3.39 billion in 2016, and 3.58 billion in 2017 [11]. There has also been a rise in e-commerce, as shown in Fig. 1, coupled with an increased demand for customer service on digital platforms [12]. According to Harvard Business Review, a mere five-minute delay could decrease a business's chances of selling to a customer. In fact, a ten-minute delay could reduce their chances by 400% [13]. However, a study done by Xu et al. examined one million conversations and found that the average response time was 6.5 hours [12]. To ameliorate this situation, some

improve

Manuscript received August 3, 2019; revised March 1, 2020.

doi: 10.18178/ijmerr.9.4.502-510

businesses began to employ chatbots to handle inquiries 24/7. Admittedly, there is still a gap between existing chatbots and chatbots intelligent enough to replace human representatives, but it is highly likely that chatbots will play a significant role in the digital future [4].

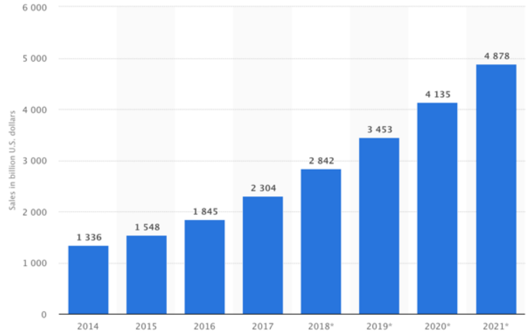


Figure 1. Retail e-commerce sales worldwide from 2014 to 2021 (in billion USD) [14].

Challenges

In addition to the chatbot's increasing popularity in business, there have been various publications providing an overview of existing chatbot platforms, architectures, and chatbot implementation methods [14]–[16]. One perspective [14] mentioned two challenges in the field of chatbot development: 1) Chatbots can only recognize specific sentence structures; 2) The responses generated by existing machine learning techniques are not always accurate or personalized. In this paper, we focus on summarizing existing machine learning techniques using the mind-mapping method.

This paper is organized as follows: methodology, mind-map presenting an overview of chatbots, detailed breakdown of the different aspects of chatbots, and conclusions.

II. Methodology

This paper will review works containing the listed keywords over the past two decades, published from 1998 to 2018, in online databases, such as Google Scholar, IEEE Xplore, ACM, and Web of Science. After searching for the keywords in the online databases, the two criteria used for selecting review papers were the publication date and the number of citations. It was important to review papers that were recent, especially since technology changes very rapidly over the years. Older papers were also reviewed to understand the history and the progression of chatbots over the years. After conducting the literature review, a mind-map was created to summarize the findings, as seen in Fig. 2. The mind-mapping method was used to visualize the relations between different concepts from various research papers. This method was chosen due to its capability of presenting complex relationships in a simple, visual form [17]. The branches stemming from the center represent the main characteristics, current architectures and systems, machine learning approaches, and applications of chatbots. Many of the published papers discussed architectures and systems, but few papers delved into the machine learning approaches in the context of chatbot development, which is the focus of this paper.

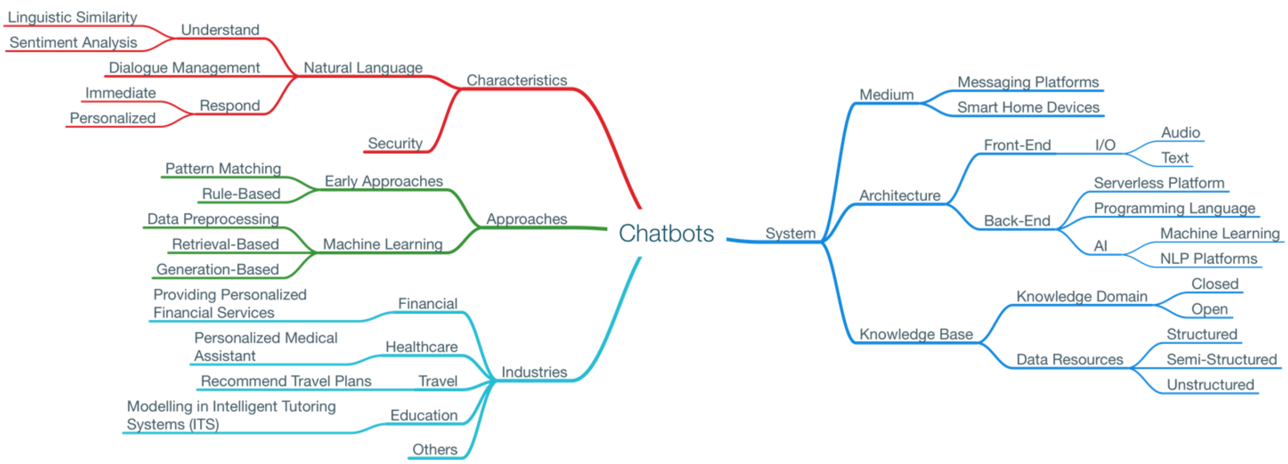


Figure 2. An overview of the properties of chatbots.

III. Chatbot Characteristics

Understanding the key characteristics of a chatbot is important for designing a chatbot. These characteristics were established through the study of people's expectations for chatbots [18], [19], and the method used in the study was comparing past human-human and human-chatbot conversations. As mentioned in a chatbot survey [16], the development of chatbots went from pattern matching and simple "Q&A" style to a more human-like way of carrying out and continuing conversations. This showed that advanced chatbots are expected to not only answer questions but also learn and improve themselves with each conversation, and eventually be able to respond appropriately in various contexts [12]. To create a smart chatbot capable of doing so, the main characteristics and capabilities needed are listed as follows:

A. *Communicate Using Natural Language*

Chatbots should understand and respond using human natural language. Advanced automated chatbots nowadays generally use Machine Learning (ML), coupled with Natural Language Processing (NLP) within the domain of Artificial Intelligence (AI) [20]. We found that Body. structure the anatomy of chatbots [21], used for seeking information, generally has three components: Natural Language Understanding (NLU) to categorize the user's intent, Dialogue Management (DM) to determine user's intent, and Natural Language Generation (NLG) to generate a response in natural language.

1) Understanding Natural Language Input.

NLP and ML are used to tackle the two most common problems in computational linguistics (a branch of linguistics in which computer science techniques are used for analyzing language). The first problem is sentiment analysis. It aims to first identify sentiments from the information input, which can be documents, sentences, or phrases, and then classify these sentiments into three polarity groups: positive, negative, or neutral. This polarity classification can be done using NLP approaches, such as n-grams, adjectives labeled, dependency relations and objective terms [22]. The other problem is linguistic similarity. It aims to represent linguistics using different lexicalization levels such as words, stem, named entities, and so on [23]. There are three main categories to determine useful distinctions in the study of language: pragmatic, the purpose of context; semantic, the meaning of context; and syntax, being grammatically correct.

2) Dialogue Management

After understanding the user input, the DM stage should choose an interaction strategy to determine a response based on the context of the conversation [24]. In other words, the dialogue management stage classifies the question type and determines the relevant category of answers the chatbot can use for responding. Many research papers presented ML approaches to accomplish this task. For example, Support Vector Machines (SVM) is one of the trending techniques for Q&A problems [25]–[27].

3) Responding with Natural Language

After determining the category of relevant answers, the chatbot must then: 1) construct the relevant, personalized response using natural language, and 2) respond with no time delay.

Personalized: Writing Styles and Emotions. The responses generated by the chatbots should be grammatically correct and exhibit human-like behaviors and emotions. Chatbots should converse in certain manners based on the user's psychological state and behavior, and they should learn to adapt their writing style to the user's need and the context of the conversation. For example, Zhang et al. [28] proposed personalized response generators which use different responding styles. Their generators use lexical principles and a sequence to sequence framework with a recurrent neural network to detect the user's behaviour. In customer service, chatbots that can detect user's emotions and expected reactions are heavily needed since their users generally express their emotions in lieu of rationally stating their problems. Xu et al. [12] collected conversations between humans and chatbots over social media and reported that approximately 40% of those conversations were emotional. As a result, their chatbot, which used deep learning and information retrieval (IR) techniques, learned informal writing styles used in emotional conversations and can show as much empathy as human agents when chatting with the users. On a similar note, Chang et al. found that the word2vec technique could help the machine understand emotional texts from users [29].

Immediate: Realistic and Fast Response. Ideally, if the user's question requires searching through the internet, the chatbot should search and respond as quickly as most search engines.

B. Security

Security features should be applied to all the data in the chatbot databases for user security and privacy. This is particularly important in the field of personal-life care. For example, sensitive personal information like medical and health information should be encrypted to ensure confidentiality [30]. Moreover, security features in chatbots should only grant system or database access to verified intended users. For example, chatbots in the form of smart home devices should allow only the legal and registered residents in a house to command and control the devices [7], [31].

IV. Chatbot Systems

The system required to design, develop and implement a chatbot which can interpret the user's intent and provide proper responses to questions, is studied and broken down in the following section:

A. Communication Medium

This section details current chatbot communication platforms. Common media which allow users to input messages and communicate with a chatbot include messaging platforms, such as applications and cloud-based services (Slack bot, IBM Watson), and physical devices (Amazon Alexa, Google Home).

1) Messaging Platforms and their Features

The messaging platform is for communicating via online chat in real time. Many messaging platforms serve different purposes with diverse needs, but they all operate based on text, and their main purpose is dealing with customer service. R. Khan and A. Das [1] compared the available features across various messaging platforms (Facebook, Skype, Slack, Telegram, Microsoft Teams and Viber), as seen in Table I. These features include text message, carousel, cards, button, quick reply, webview, group chatbot, list, audio, video, GIF, image, and document or file. This is to help developers choose which platform to use for deploying their chatbots based on their specific needs.

TABLE I. MESSAGING PLATFORMS FEATURES COMPARISON [1].

Features / Platforms	Facebook	Skype	Slack	Tele-gram	MS Teams	Viber
Text message	✓	✓	✓	✓	✓	✓
Carousel	✓	✓	✓	Partial	✓	✓
Button	✓	✓	✓	✓	✓	✓
Quick reply	✓	✗	✗	✗	✗	✗
Web view	✓	✗	✗	✗	✗	✗
Group chatbot	✓	✗	✓	✓	✗	✗
List	✓	✓	✗	✗	✗	✗
Audio	✓	✓	✗	✓	✓	✓
Video	✓	✓	✗	✓	✗	✓
GIF	✓	✓	✓	✓	✓	✓
Image	✓	✓	✓	✓	✓	✓
Document/file	✓	✓	✗	✓	✓	✓

2) Smart Home Devices

In addition to commercial chatbots, which are used for customer service purposes, there are also chatbots that act as personal virtual assistants. These assistants help users accomplish daily tasks, such as booking a hotel, getting the latest news, checking the weather forecasts, or even buying products based on the user's personal preferences [30], [32]–[37].

In order for these virtual assistant chatbots to have more functionality, more communication channels were enabled, e.g. users can use voices to control many types of devices [31]. Current virtual assistants require a physical device in the environment to act as a host for the trained chatbots in the cloud. For example, Amazon Echo operates on Alexa, Google Home uses Google Assistant, Apple HomePod uses Siri, and Microsoft's virtual assistance with Cortana platform, and so on. A drawback with current virtual assistants is that they still have a closed knowledge domain, as they are only expected to carry out certain specified and preset tasks, such as turning on/off the lights, projectors, and air conditioning, but a conversation topic aside from that might not be recognized by the assistant. In the future, a more improved chatbot should be able to converse openly about a variety of topics [15].

B. System Architecture

The chatbot system is now leaning towards cloud-based, open-source, and serverless data operating systems. Some examples are OpenWhisk from IBM [38] and AWS Lambda from Amazon [32]. They provide an easy way to build a chatbot using a variety of programming languages, such as Javascript, Java, Python, etc. which enhances the functionality of backend system and thus delivers a personalized experience to each user. To build a closed-domain, serverless chatbot, Mengting et. al. [38] presented a generic architecture as shown in Fig. 3. It consists of four levels: the first level is Audio I/O which converts audio input to text and vice versa; the second level is Text I/O; the third level is many domain-specific chatbots, each of them has a specific task such as location-based weather reports, reminders, news, jokes, and others; the fourth level consists of third party services to deploy the chatbot. The chatbot can respond with text and/or audio depending on the request.

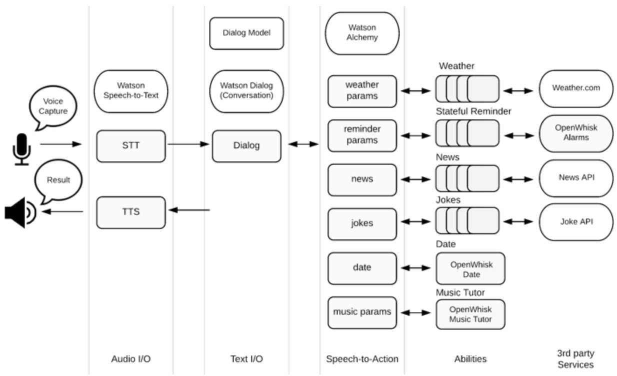


Figure 3. Chatbot architecture [38].

Summarizing these papers in [32], [38], [39], we see that a chatbot development system can be broken down to front-end and back-end environments. These are to address the main tasks of chatbots proposed in the Characteristics section and the mind-map: understanding the user's natural language input, generating a proper response addressing the user's needs, and constructing the response with natural language. The front-end system is the chatbot's "mouth" and "ears/eyes", seeing what the user inputs are and responding. The back-end system is the chatbot's "brain," where algorithms and models are used to understand user intents and determine a suitable response.

Front-End System. The front-end consists of the user interface and the communication medium. As mentioned above, the input and output can consist of textual and audio data. Chatbots accepting audio inputs use a speech recognition engine that can convert speech-to-text (STT) and text-to-speech (TTS).

Back-End System. The back-end system is hosted on the cloud platform. It processes inputs and generates relevant responses. It is composed of a server, an AI engine, custom APIs, and a database through a REST API [32].

ML techniques which can help a chatbot better understand the context of conversation or construct suitable responses are in need. As of now, there are many NLP tools which can be used, such as Dialogflow from Google, formerly known as Api.ai (https://developers.google.com/actions/dialogflow/project-agent). More recently, to better model user behavior, many companies have developed AI platforms and released their APIs to the public. A few examples are: IBM Watson (https://www.ibm.com/watson/developer/), LUIS.ai, Microsoft's Language Understanding Intelligent Service (https://www.luis.ai/), Wit.ai (https://wit.ai/), and Amazon Lex (https://aws.amazon.com/lex/developers/).

C. Knowledge Base

The knowledge base is the information the chatbot refers to during the dialogue management stage when generating a response. The knowledge base is the "brain" of the chatbot engine, generally containing keywords/phrases and the responses associated with these keywords/phrases [40].

Data Resources. For chatbots to obtain knowledge, they need to extract data from different large-scale information sources and store this information to a knowledge base. These information sources could either be structured, semi-structured, or unstructured [40]. Most existing chatbots currently retrieve information from structured documents to build their knowledge base, since structured documents have labelled utterance-response (or Q-R) pairs the chatbot engine can store [41]. There are also new information retrieval approaches, such as a method called DocChat, proposed for extracting information from unstructured documents. Instead of the Q-R paris, the DocChat method selects the most relevant sentence from the document directly and thus improves the fluency in chatbots' responses. These retrieval approaches for unstructured data would aid the chatbot in developing a larger knowledge base.

Knowledge Domain. The knowledge base can either be open-domain or closed-domain (restricted domain). Chatbots with an open-domain knowledge base are generally conversational agents, capable of responding to a variety of user inputs. On the other hand, chatbots with a closed-domain knowledge base focus on specific domains, such as law, medicine, and programming. Closed-domain chatbots are generally goal-oriented -- the user is likely attempting to accomplish a task, such as asking a question, setting an alarm, or making a reservation. For a closed-domain question-answering chatbot, it is also important to consider the size of data available, the domain of interest itself, and the resources available for the domain when determining what techniques to use [42].

V. Early Approaches For Chatbot Development

This section discusses some of the earlier approaches used for chatbot development which do not use ML. This includes pattern matching and rule-based methods.

A. Pattern Matching

This approach is commonly used in question-answering bots. They generate predefined outputs and match them with a given input according to the characteristic variables of sentences. An example of a question answering bot is A.L.I.C.E. (Wallace, 2009), which stands for Artificial Linguistic Internet Computer Entity [43], [44]. The following sections introduce a common technique used for pattern matching.

Artificial Intelligence Markup Language (AIML) is a form of eXtensible Markup Language (XML), aimed to define rules for matching pattern and thus find the proper responses. An example is shown in Table II:

TABLE II. Example Of AIML Code [43].

```
<category>
<pattern>What the user says</pattern>
<template>What the bot responds</template>
</category>
```

The <pattern> tag is used to match in user's input. The <template> tag is used to respond to the pattern.

B. Rule-Based

The rule-based or template-based approach is used to map sentences with the pattern associated with the collected input database. Some chatbots applying these techniques are: ELIZA (created by Weizenbaum in 1966) and PARRY (created by Colby in 1975) chatbot. For ELIZA, the textual input searches for the keywords, which are then assigned a rank. The input is transformed into a "keystack," where keywords with the highest rank are at the top. The keyword of the highest rank is used to determine the category of responses most related to the keyword [43]. This helps the bot determine a relevant response. PARRY has a similar structure to ELIZA but uses a better controlling structure and has the ability to understand language, since it has a mental model that can simulate the bot's emotions [45].

VI. Machine Learning Approaches For Chatbot Development

This section discusses existing modern NLP techniques that can be used when developing a chatbot. Recently, NLP techniques have been combined with ML, as ML improves the chatbots' performance of finding patterns from large amounts of data. The following sections discuss the current stages of NLP and ML techniques applied in the field of chatbot model development.

A. Data Preprocessing

An arbitrary user input in human language needs to be processed to be "clean data," which is the data a chatbot machine can understand. There are many preprocessing methods currently used in chatbots, such as stopwords removal, removing capitalization, and labelling. There are three features to consider when preprocessing the data:

Lexical. The lexical feature is also called the "word form" feature because it focuses on each word rather than the sentence structure or grammar [46]. It uses the bag of n-grams method to group words together [47]. Three preprocessing methods using this lexical approach are word level n-grams, stemming, and lemmazation [44]. The word level n-grams method groups together n consecutive words and looks for word n-grams which might indicate the category of the question (for example, if the unigram "city" is used, the question is likely asking for a location). Stemming reduces words to their grammatical roots by removing suffixes [47]. This, however, fails when words change endings in plural form (for example, leaf and leaves would have different stems). Lemmazation is a more accurate approach which can identify the correct roots by referring to a lexical database of English [48].

Syntactic. Methods using syntactic principles include part-of-speech (POS) tagging and chunking. POS tagging [46] labels each word in a sentence with its part of speech (noun, pronoun, verb, etc.). Chunking is then used to partition the sentence into non-overlapping, non-recursive segments. Each partition has a chunk tag, which is its class label. The question classifier model then uses the POS tags, the surrounding context, and the class label to identify the question type [49].

Semantic. Whereas syntax focuses on sentence structure, semantics focuses on the meaning of words. One method using semantic principles is Named Entity Recognition (NER) [50]. Most NER implementations use a coarse-grained hierarchical classifier consisting of a layered semantic hierarchy of answer types [51]. One paper designed an open-domain question answering chatbot using a two-layered hierarchy containing 6 coarse classes and 50 fine classes to answer 500 questions in the TREC competition. They classified user questions into different question types (with 98.8% accuracy), generated expected answer types, extracted keywords, and reformulated questions into semantically equivalent questions [51].

Vector Representations. Vector representation maps high dimensional word features to low dimensional feature vectors. Based on certain rules and relationships, words are represented by vector coordinates. For example, related words are closer together. The common techniques for vector representation are:

- **word2vec** [29]: Each word is represented by a vector in a specified vector space containing continuous bag-of-word (CBOW) and skip-gram (SG) architectures.
- **doc2vec** [52]: The vector representation for paragraphs and documents is found by taking the weighted average of all the words in the document.
- **Global Vectors (GloVe)** [6]: The global corpus statistics for the unsupervised learning of word representations which outperform other models on word analogy, word similarity, and named entity recognition (NER). The source code from Stanford can be found at http://nlp.stanford.edu/projects/glove/

B. *Retrieval-Based*

A group of researchers used the information retrieval technique to tackle one of the difficult problems for chatbots: the short text conversation. By collecting short conversations on social media and using them to train different models, such as the translation model, latent space model (linear model), deep learning model (non-linear model), and topic-word model, they reported that the retrieval-based model can perform more “intelligently” than some of the older approaches. This model collects data from many social media sources, such as Q&A forums, and find the differences between user inputs and questions online with the cosine similarities method [53].

Another research group used unstructured documents and examined their features on different levels, such as word level, phrase level, sentence level, document level, relation level, type level, and topic level. This allowed them to respond to utterances in addition to question-response (Q-R) pairs in which the response R is a short text and only depends on the last user utterance Q. Their method selects a sentence from given documents directly, by ranking all possible sentences based on features designed at different levels of granularity. They compared their chatbot’s performance with a chitchat engine from China called XiaoIce, and found that their chatbot generated more formal and informative responses, whereas XiaoIce generated more colloquial responses. They also found that their chatbot generated either an equally relevant or more relevant response than XiaoIce in 109 out of 156 conversations, which is promising [41].

C. *Generation-Based*

The generation-based methods use an encoder-decoder framework [41]. Some researchers proposed deep learning technique using the sequence to sequence (seq2seq) model to predict the next sentence in a conversation given previous sentences using two recurrent neural network (RNN), one being an encoder and the other being the decoder [54]. The encoder output provides necessary information to the decoder to generate a sequence element by element. The decoder takes a sequence as the input and generates a sequence output [43].

Long Short-Term Memory (LSTM) networks are an extension of RNN which maximize the probability of generating a response given the previous conversations [55]. A. Xu et al. [2] proposed a technique using LSTM networks to generate responses. Their process starts by converting the user’s input into its vector representations with the word2vec method, and then using LTSM to learn the mapping from sequence to sequence. It consists of two LSTM neural networks. The first one is an encoder which maps variable-length inputs to a fixed-length vector. The second LSTM neural network is a decoder which then maps this vector to a variable-length output. This is similar to the two RNN networks in the seq2seq model.

VII. Chatbot in Industries

In various industries, chatbots are becoming a ubiquitous component of customer service. The usages of chatbots in different fields are summarized in Table III below. They are used in Customer Relationship Management (CRM) which helps companies stay connected to both current and potential customers for increased customer retention [56]. Both commercial and non-profit companies can improve their profitability if they understand their users’ needs better.

TABLE III. The Role Of Chatbots In Various Industries.

Industries	Description
Healthcare	Personalized medical assistant relies on AI algorithms to hold daily conversations, provide health-related information, and recommend activities and restaurants to the elderly [33]. As

Industries	Description
	purposed by this paper [34], an LSTM model can be used to extract semantic information from the elderly's inputs. The chatbot's responses were generated by Euclidean distance for matching patterns. These chatbots often use frameworks which have four layers. Data layer: record the data processing progress and store the labeled data collected from multiple sensory components. Information layer: mapping on lifelong ontology, Knowledge layer: personalized behavioral predictions, and Service layer: the results of health service recommendation for cloud computing environments [30].
Travel	These chatbots can recommend travel plans based on personal preferences from travel history that was gathered from previous flight, hotel, and car rental bookings. It then generates a recommendation using collaborative filtering with rating scores deployed on Alexa Skills market [32].
Education	Chatbots can be used to teach students basic computer science concepts [35]. One paper proposed Intelligent Tutoring Systems which are computer environments which adapt to the needs of the individual learner [36]. In particular, Open Learner Modelling allows the system and student to jointly negotiate the learner model. This allows both the student to reflect on their learning and the learner model to improve its accuracy.
Financial	Since the financial industry is increasingly deregulated, many financial transactions are now digitized. This leaves financial businesses large amounts of financial and personal data to leverage to deliver a variety of new services online [37]. For example, chatbots can be used to help financial advisors and strategists with decision making based on previous financial transactions or trends.

Most conversations are held on text-based platforms like email and online chat. An important variant on these conversational machines is the ability to think. It is why industries are moving towards a modern chatbot which uses AI technology to interact with a human more intelligently. In past years, most chatbots in the industries could only perform simple tasks because they are programmed to respond to a predefined list of questions. In order to become self-learning chatbots, which is what they may do in the future, they need to be trained using data from their past conversations and update its knowledge base autonomously to deliver personalized responses [57], [58].

VIII. Conclusions

In this paper, we used a mind-mapping approach to present an overview of chatbots, after reviewing papers published from 1998 to 2018. This can help researchers develop a better understanding of the current implementation techniques and usages of chatbots. This is important because chatbots are becoming increasingly popular, especially for customer service in the industry and as an intelligent virtual assistant for personal use.

This paper outlines many machine learning techniques which could improve the performance of chatbots because they allow chatbots to learn and adapt through experience. Having the ability to improve itself with every interaction will likely improve the chatbot's capability of understanding the content and context of the user's input, which would help the chatbot generate a more accurate, relevant response.

However, existing chatbots have a few limitations. The main challenge for a chatbot right now is understanding the context in a conversation and generating a relevant response. Hence, future intelligent chatbots should: 1) implement improved natural language processing techniques to accurately recognize the content of the user input; 2) learn to understand the context of conversations and respond accordingly with emotions or personalized content. The ultimate goal of chatbots is to replicate human-human interaction, which requires improved machine learning and natural language processing techniques.

Turing Test

The current trend in chatbot development suggests that chatbots will continue to be improved with advanced technologies driven by ML- and NLP-based AI. As previously mentioned, the Turing test, conducted by a human conversational interrogator, is the most popular test for determining if a chatbot has achieved human-level intelligence [5], [59], The human judges ask questions to determine if one of participants is not human, so if a chatbot can pass the test, it demonstrates human-level communication capabilities. The goal of chatbots is to one day pass the Turing test and achieve human conversational capabilities, which we believe will happen.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

PS and JHC designed the study; PS analyzed data and wrote the paper; XL and BW analyzed the data and co-wrote the paper; PM and JHC commented on the manuscript at all stages; all authors discussed the results. All authors had approved the final version.

Acknowledgment

This research is supported by the Petchra Pra Jom Klao Doctoral scholarship, King Mongkut's University of Technology Thonburi (KMUTT), Thailand. Our thanks to all members of Data Science and Engineering Laboratory (D-Lab), School of Information Technology (SIT) at KMUTT for their useful comments and feedback.

References

[1] R. Khan and A. Das, "Introduction to chatbots," in *Build Better Chatbots*, Berkeley, CA: Apress, 2018, pp. 1–11.

[2] A. Følstad and P. B. Brandtzæg, "Chatbots and the new world of HCI," *Interactions. ACM.org*, vol. 24, no. 4, pp. 38–42, Jun. 2017.

[3] A. Schlesinger, K. P. O'Hara, and A. S. Taylor, "Let's talk about race: Identity, chatbots, and AI," in *Proc. the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 2018, pp. 1–14.

[4] R. Dale, "The return of the chatbots," *Natural Language Engineering*, vol. 22, no. 5, pp. 811–817, Sep. 2016.

[5] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 49, pp. 433–460, 1950.

[6] P. B. Brandtzaeg and A. Følstad, "Why people use chatbots," in *International Conference on Internet Science (INSCI 2017)*, 2017,

vol. 10673 LNCS, pp. 377–392.

[7] M. B. Hoy, "Alexa, Siri, Cortana, and More: An introduction to voice assistants," *Medical Reference Services Quarterly*, vol. 37, no. 1, pp. 81–88, Jan. 2018.

[8] M. Pinola, "History of voice recognition: from Audrey to Siri," *ITBusiness.ca*, 2011. [Online]. Available: https://www.itbusiness.ca/news/history-of-voice-recognition-from-audrey-to-siri/15008. [Accessed: 12-Aug-2018].

[9] M. Saba, "A brief history of voice recognition technology," *Call Analytics, Call Intelligence, Call Recording*. [Online]. Available: https://www.callrail.com/blog/history-voice-recognition/. [Accessed: 13-Aug-2018].

[10] H. Sim, "Voice assistants: This is what the future of technology looks like," *Forbes*. [Online]. Available: https://www.forbes.com/sites/herbertrsim/2017/11/01/voice-assistants-this-is-what-the-future-of-technology-looks-like/#389fc513523a. [Accessed: 13-Aug-2018].

[11] Statista, "Number of internet users worldwide 2005-2017 | Statista," *The Statistics Portal*, 2017. [Online]. Available: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/. [Accessed: 16-Jul-2018].

[12] A. Xu, Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju, "A new chatbot for customer service on social media," in *Proc. of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 2017, pp. 3506–3510.

[13] A. Galert, "Chatbot report 2018: Global trends and analysis," 2017. [Online]. Available: https://chatbotsmagazine.com/chatbot-report-2018-global-trends-and-analysis-4d8bbe4d924b. [Accessed: 17-Jul-2018].

[14] Statista, "Retail e-commerce sales worldwide from 2014 to 2021 (in billion U.S. dollars)," *Statista*, 2017. [Online]. Available: https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/. [Accessed: 19-Jul-2018].

[15] K. Nimavat and T. Champaneria, "Chatbots: An overview types, architecture, tools and future possibilities," *International Journal of Scientific Research and Development*, vol. 5, no. 7, pp. 1019–1024, Oct. 2017.

[16] A. Deshpande, A. Shahane, D. Gadre, M. Deshpande, and P. M. Joshi, "A survey of various chatbot implementation techniques," *International Journal of Computer Engineering and Applications*, vol. XI, 2017.

[17] O. S. Synekop, "Effective writing of students of technical specialties," *Advanced Eduucation*, no. 4, pp. 51–55, 2015.

[18] M. C. Jenkins, R. Churchill, S. Cox, and D. Smith, "Analysis of user interaction with service oriented chatbot systems," in *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 76–83.

[19] V. Ravi and S. Kamaruddin, "Big data analytics enabled smart financial services: Opportunities and challenges," in *International Conference on Big Data Analytics (BDA 2017)*, 2017, vol. 10721 LNCS, pp. 15–39.

[20] F. Halper, "Advanced analytics: Moving toward AI, machine learning, and natural language processing," *TDWI Best Practices Report*, 2017. [Online]. Available: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/tdwi-advanced-analytics-ai-ml-nlp-109090.pdf. [Accessed: 07-May-2018].

[21] S. Quarteroni, "Natural language processing for industry: ELCA's experience," *Informatik-Spektrum*, vol. 41, no. 2, pp. 105–112, Apr. 2018.

[22] V. Ng, S. Dasgupta, and S. M. N. Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in *Proc. COLING-ACL '06 Proceedings of the COLING/ACL on Main conference poster sessions*, 2006, pp. 611–618.

[23] P. Molino, L. M. Aiello, and P. Lops, "Social question answering: Textual, user, and network features for best answer prediction," *ACM Transactions on Information Systems*, vol. 35, no. 1, pp. 1–40, Sep. 2016.

[24] J. Cahn, "CHATBOT: Architecture, design, and development," University of Pennsylvania, 2017.

[25] S. J. Yen, Y. C. Wu, J. C. Yang, Y. S. Lee, C. J. Lee, and J. J. Liu, "A support vector machine-based context-ranking model for question answering," *Information Sciences*, vol. 224, pp. 77–87, Mar. 2013.

[26] D. Tomás and J. L. Vicedo, "Minimally supervised question classification on fine-grained taxonomies," *Knowledge and Information Systems*, vol. 36, no. 2, pp. 303–334, Aug. 2013.

[27] T. C. Zhou, M. R. Lyu, and I. King, "A classification-based approach to question routing in community question answering," in *Proc. of the 21st International Conference Companion on World Wide Web - WWW '12 Companion*, 2012, pp. 783–790.

[28] W. Zhang, T. Liu, Y. Wang, and Q. Zhu, "Neural personalized response generation as domain adaptation," Jan. 2017.

[29] C. Y. Chang, S. J. Lee, and C. C. Lai, "Weighted word2vec based on the distance of words," in *Proc. of 2017 International Conference on Machine Learning and Cybernetics, ICMLC 2017*, 2017, vol. 2, pp. 563–568.

[30] K. Chung and R. C. Park, "Chatbot-based heathcare service with a knowledge base for cloud computing," *Cluster Computing*, pp. 1–13, 16-Mar-2018.

[31] C. J. Baby, F. A. Khan, and J. N. Swathi, "Home automation using IoT and a chatbot using natural language processing," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, 2017, pp. 1–6.

[32] A. Argal, S. Gupta, A. Modi, P. Pandey, S. Shim, and C. Choo, "Intelligent travel chatbot for predictive recommendation in echo platform," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 176–183.

[33] D. Madhu, C. J. N. Jain, E. Sebastain, S. Shaji, and A. Ajayakumar, "A novel approach for medical assistance using trained chatbot," in *Proc. of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017*, 2017, pp. 243–246.

[34] M. H. Su, C. H. Wu, K.-Y. Huang, Q. B. Hong, and H. M. Wang, "A chatbot using LSTM-based multi-layer embedding for elderly care," in *2017 International Conference on Orange Technologies (ICOT)*, 2017, pp. 70–74.

[35] L. Benotti, M. C. Martínez, and F. Schapachnik, "Engaging high school students using chatbots," in *Proc. the 2014 conference on Innovation & Technology in Computer Science Education - ITiCSE '14*, 2014, pp. 63–68.

[36] A. Kerly, P. Hall, and S. Bull, "Bringing chatbots into education: Towards natural language negotiation of open learner models," *Knowledge-Based Systems*, vol. 20, no. 2, pp. 177–185, Mar. 2007.

[37] F. Corea, "How AI Is Transforming Financial Services," in *Applied Artificial Intelligence: Where AI Can Be Used In Business*, Springer, Cham, 2018, pp. 11–17.

[38] M. Yan, P. Castro, P. Cheng, and V. Ishakian, "Building a Chatbot with Serverless Computing," in *Proc. the 1st International Workshop on Mashups of Things and APIs - MOTA '16*, 2016, pp. 1–4.

[39] A. M. Rahman, A. Al Mamun, and A. Islam, "Programming challenges of chatbot: Current and future prospective," in *5th IEEE Region 10 Humanitarian Technology Conference 2017, R10-HTC 2017*, 2017, vol. 2018–Janua, pp. 75–78.

[40] K. B. Reshmi S, "Empowering chatbots with business intelligence by big data integration," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, pp. 627–631, 2018.

[41] Z. Yan *et al.*, "DocChat: An information retrieval approach for chatbot engines using unstructured documents," in *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 516–525.

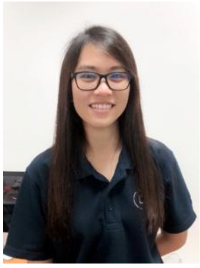
[42] D. Mollá and J. L. Vicedo, "Question answering in restricted domains: an overview," *Computational Linguistics*, vol. 33, no. 1, pp. 41–61, Mar. 2007.

[43] K. Ramesh, S. Ravishankaran, A. Joshi, and K. Chandrasekaran, "A survey of design techniques for conversational agents," in *Information, Communication and Computing Technology (ICICCT)*, 2017, pp. 336–350.

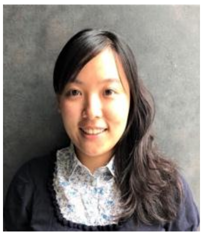
[44] S. Reshmi and K. Balakrishnan, "Implementation of an inquisitive chatbot for database supported knowledge bases," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 41, no. 10, pp. 1173–1178, 2016.

[45] H. Shum, X. He, and D. Li, "From Eliza to XiaoIce: Challenges and opportunities with social chatbots," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 10–26, Jan. 2018.

[46] J. Le, Z. Niu, and C. Zhang, "Question classification based on fine-grained PoS annotation of nouns and interrogative pronouns," in *Pacific Rim International Conference on Artificial Intelligence (PRICAI 2014): Trends in Artificial Intelligence*, 2014, pp. 680–693.
[47] M. Mishra, V. K. Mishra, and S. H.R., "Question classification using semantic, syntactic and lexical features," *International journal of Web & Semantic Technology*, vol. 4, no. 3, pp. 39–47, 2013.
[48] C. D. Manning, P. Raghaven, and H. Schuetze, "Stemming and Lemmatization," in *Introduction to Information Retrieval*, 2009, pp. 22–34.
[49] L. Zhu, L. S. Chao, D. F. Wong, and X. D. Zeng, "A noun-phrase chunking model based on SBCB ensemble learning algorithm," in *Proc. - International Conference on Machine Learning and Cybernetics*, 2012, vol. 1, pp. 11–16.
[50] D. Molla, M. Zaanen, and D. Smith, "Named entity recognition for question answering," in *Proc. the Australasian Language Technology Workshop 2006*, 2006, pp. 51–58.
[51] X. Li and R. Dan, "Learning question classifiers," in *COLING '02 Proc.s of the 19th international conference on Computational linguistics*, 2002, vol. 1, pp. 1–7.
[52] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. of the 31st International Conference on Machine Learning*, vol. 32, no. 2, pp. 1188–1196, May 2014.
[53] Z. Ji, Z. Lu, and H. Li, "An information retrieval approach to short text conversation," pp. 1–21, Aug. 2014.
[54] O. Vinyals and Q. Le, "A neural conversational model," in *Proc. of the 31 st International Conference on Machine Learning*, 2015, vol. 37, pp. 233–239.
[55] J. Li, W. Monroe, A. Ritter, D. Jurafsky, M. Galley, and J. Gao, "Deep reinforcement learning for dialogue generation," in *Proc. of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1192–1202.
[56] A. M. Seeger and A. Heinzl, "Human versus machine: Contingency factors of anthropomorphism as a trust-inducing design strategy for conversational agents," in *Lecture Notes in Information Systems and Organisation*, vol. 25, Springer, Cham, 2017, pp. 129–139.
[57] F. Sweis, "Building and training self-learning chatbots: Developers, you can drive the chatbot revolution," *ComputerWorld*, 2017. [Online]. Available: https://www.computerworld.com.au/article/631249/building-training-self-learning-chatbots-developers-can-drive-chatbot-revolution/. [Accessed: 14-Aug-2018].
[58] L. Vishnoi, "How the development of AI has advanced the technology available for chatbots," *Forbes Technology Council*, 2018. [Online]. Available: https://www.forbes.com/sites/forbestechcouncil/2018/05/23/how-the-development-of-ai-has-advanced-the-technology-available-for-chatbots/#2038c11fc213. [Accessed: 14-Aug-2018].
[59] K. Warwick and H. Shah, "Passing the turing test does not mean the end of humanity," *Cognitive Computation*, vol. 8, no. 3, pp. 409–419, 2016.



Prissadang Suta, received her B.S. in Computer Engineering from Mae Fah Luang University, Chiang Rai, Thailand, in 2010 and received her M.S. in Software Engineering, School of Information Technology (SIT), King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand, in 2015. She is currently a Ph.D. student in Computer Science at SIT, KMUTT. Her research interested is in intelligent chatbot.



Xi Lan is currently in her fourth year of undergraduate study in Engineering Science, University of Toronto, Toronto, Canada. Her specialization in her program is robotics engineering. Her interests are in computer vision, circuit design and network security.

Biting Wu is currently in her fourth year of undergraduate study in Engineering Science, University of Toronto, Toronto, Canada. In her program, her specialization is in mathematics, statistics, and finance. She has worked with and is interested in various machine learning applications, such as image classification, analytics, and modelling.

Pornchai Mongkolnam holds a Ph.D. in computer science from Arizona State University and currently works at School of Information Technology (SIT) at KMUTT in Thailand. He is the Head of Data Science and Engineering Laboratory (D-Lab) at SIT, KMUTT.

Jonathan H. Chan is an Associate Professor at the School of Information Technology, KMUTT, Thailand. Jonathan received his B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto, Canada. He is a senior member of IEEE, ACM, APNNS and INNS. His research interests include intelligent systems, biomedical informatics, machine learning, and data science.