How to Build a Chatbot: Chatbot Framework and its Capabilities

Chen Wei
Team Lead
No.5 Gui Gu Liang Cheng
No.1 Nong Da Nan Lu, Beijing
+86-13141048556
weichen@uzoo.cn

Zhichen Yu
Co-Founder
No.5 Gui Gu Liang Cheng
No.1 Nong Da Nan Lu, Beijing
+86-010-58349447
yuzhichen@uzoo.cn

Simon Fong
Associate Professor
E11-4023, FST
University of Macau
+853-8822-4460
ccfong@umac.mo

ABSTRACT
Artificial Intelligence continues to grow in popularity on various industrial platforms, becoming especially prominent in Chatbot technology. A great deal of recent research was focused on social and assistive Chatbot, and its technology has been evolving over time. Following this direction, a Chatbot Framework is established in this paper, discussing about the relevant technologies. Firstly, the development of Artificial Intelligence is introduced. In particular, we present the Chatbot history of technology timeline. Following this, we describe the capabilities of the Chatbot and discuss about the technologies. The entire Chatbot framework will be presented afterwards, as well as the supporting set of modules. Our analysis of this framework shows that the framework that is based on some theoretical designs could be practically implemented to satisfy the required capabilities in the industry. Through our analysis, we show that the capabilities are also feasible.

CCS Concepts
• **Applied computing →Enterprise computing infrastructures**

Keywords
Artificial Intelligence; Human Robot Interaction; Chatbot; Natural Language Understanding; Natural Language Processing.

1. INTRODUCTION
Artificial Intelligence (AI) has been growing at an unprecedented rate in recent years. It received a lot of attention from AlphaGo versus Lee Sedol at the Google DeepMind Challenge Match, in which AlphaGo won the 18-times world champion Lee Sedol [1]. In Fact, the trend of AI has been up and down for more than 60 years [2]: The field of AI research was founded at a workshop held on the campus of Dartmouth College during the summer of 1956. The first peak came in 1957, the perceptron algorithm was invented, which was intended to be a machine, then it was designed for image recognition. The first trough was following in 1970s, because big data training and complicated tasks can't be processed using existing technologies and methods. Later in 1982, John Hopfield proposed Hopfield network, which is a form of recurrent artificial neural network. And back propagation, invented in 1986, calculates the error contribution of each neuron after a batch of data is processed. The second trough was in 1990s, due to the failure of DARPA, and the Government cut the investment. In 2006, a publication by Geoff Hinton, Osindero and Teh showed how a many-layered feed forward neural network could be effectively pre-trained one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, then fine-tuning it using supervised back propagation. It made breakthrough improvements. Later in 2013, Deep Learning algorithms gained great success in Computer Vision, Automatic Speech Recognition and Text to Speech areas.

Since 2010, the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRA) is a competition and a good ILSVRA classification error rate was 28%; in 2012, a deep convolutional neural net (AlexNet) achieved 16%; and in 2015, error rates fell to 3.57%, which exceeded human ability [3]. For ASR tasks, currently the word error rate dropped to 4.9%, which is comparable to humans. But in Natural Language Processing (NLP) area, for example, information extraction, machine reading comprehension, summary, etc., it is quite at odds with humans. Therefore, how to build a Chatbot, which can understand the natural language, generate appropriate replies, and importantly, be capable of freely talking with humans, is truly a challenge.

This paper will focus on: 1) assistive Chatbot framework, which includes perception layer, cognition layer and presentation layer; 2) Chatbot capabilities, as well as the set of modules that support the proposed framework.

2. RELATED WORK
Many approaches have been proposed to address the relevant issues on human-computer interaction. In Jennifer et al. [4], the author discussed about the anticipations for chatbots, preferred input modalities and opportunities for chatbots based on user needs. It pointed out due to its novel and relatively new technology, there is an opportunity to create meaningful experiences with chatbots in a typical person's life. In Michalis et al. [5], the authors introduced an interactive framework for learning user-object associations through human-robot interaction. Andre et al. [6] stated the problem for humans lies in the nearly unpredictable behavior and motion of the robot itself, and thus they enhance human's view with more information to get knowledge about robot's perception and awareness. They used augmented reality methods for providing the information in an adaptable visualization from different user types. In [7], the authors introduced robot dialog architecture for the SoftBank

ICMLC 2018, February 26-28, 2018, Macau, China.

ACM ISBN 978-1-4503-6353-2/18/02…$15.00
DOI: https://doi.org/10.1145/3195106.3195169

Pepper and NAO Robots, as well as described the effective use of online learning to enhance the conversational capabilities of a concierge robot.

In Travor C. et al [8], an algorithmic framework was introduced, which is for automated construction using multiple identical robots. Their approach is based on the principle of tree-based dynamic programming and concomitant idea of local search on trees to improve the quality of the generated plans. In [9], Sandeep investigated the use of search bots in a collaborative search scenario, how users perceive and engage with search bots that intervene without an explicit request, and develop & evaluate an end-to-end brokered system. Zi Yin, Kenghao Chang and Ruofei Zhang [10] proposed DeepProbe, a generic information-directed interaction framework that is built around an attention-based sequence-to-sequence (seq2seq) recurrent neural network. They used a seq2seq neural network enhanced with attention mechanism, which are a Bi-directional multilayer LSTM encoder and a LSTM attention decoder. Prerna K., Puneet A., and Ashwin S. [11] described an automated assistant for answering frequently asked questions, which answer HR-related queries in leave management and health insurance to a large number of users.

Sansonnet et.al. [12] provide a basic framework which outlines the functions expected from modern chatbots: dialogic agent, rational agent and embodied agent. However it focused on assistive agents structure, and lacks the capabilities of design and discussion.

In this paper we will introduce the Chabot framework, which focus on assistive, task-oriented framework. There are mainly three layers: perception layer, cognition layer and presentation layer. And then we discuss 8 capabilities of the Chatbot: self-consciousness, humorous, purity, IQ, EQ, memory, self-learning and charisma. Potential applications and future work will be discussed in the end.

3. CHATBOT HISTORY

The earliest Chabot can date back to 1966. ELIZA [13] is an early natural language processing computer program created from 1964 to 1966 at the MIT Artificial Intelligence Laboratory by Joseph Weizenbaum. Eliza simulated conversation by using a "pattern matching" and substitution methodology that gave users an illusion of understanding on the part of the program, but had no built in framework for contextualizing events.

PARRY [14] was written in 1972 by psychiatrist Kenneth Colby, then at Stanford University. While ELIZA was a tongue-in-check simulation of a Rogerian therapist, PARRY attempted to simulate a person with paranoid schizophrenia. The program implemented a crude model of the behavior of a person with paranoid schizophrenia based on concepts, conceptualizations, and beliefs.

Jabberwacky [15] is a chatterbot created by British programmer Rollo Carpenter. Its stated aim is to "simulate natural human chat in an interesting, entertaining and humorous manner". It is an early attempt at creating an artificial intelligence through human interaction.

Dr. Sbaitso [16] is an artificial intelligence speech synthesis program released late in 1991 by Creative Labs for MS DOS-based personal computers.

A.L.I.C.E. [17] came to life on November 23, 1995. It is a natural language processing chatterbot – a program that engages in a conversation with a human by applying some heuristical pattern matching rules to the human's input, and in its online form it also relies on a hidden third person.

In 2001, SmarterChild [18] was widely distributed across global instant messaging and SMS networks. It was an intelligent agent or bot developed by ActiveBuddy, Inc.

Siri [19] is an intelligent personal assistant, it uses voice queries and a natural language user interface to attempt to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services.

IBM Watson [20] is a question answering computer system capable of answering questions posed in natural language, and in 2011, the Watson computer system competed on Jeopardy! against former winners Brad Rutter and Ken Jennings winning the first place prize of $1 million.

Google Now [21] is an intelligent personal assistant developed by Google. It was launched on July 9, 2012, and was first supported on the Galaxy Nexus smartphone.

Amazon Alexa [22] is an intelligent personal assistant developed by Amazon, first used in the Amazon Echo and the Amazon Echo Dot devices, and it was initially released in November 2014. While Amazon Cortana [23] is a virtual assistant, it can set reminders, recognize natural voice without the requirement for keyboard input and answer questions using information from the Bing search engine. It was released in 2014.

Google Allo [24] is an instant messing mobile app and released on September 21, 2016. It uses phone numbers as identifiers, and allows users to exchange messages, files, voice note and images. It includes a virtual assistant; a feature generates automatic reply suggestions and an optional encrypted mode.

Tay [25] was an artificial intelligence chatterbot was originally released by Microsoft Corporation via Twitter on Mach 23, 2016; it caused subsequent controversy when the bot began to post inflammatory and offensive tweets through its Twitter account, forcing Microsoft to shut down the service only 16 hours after its launch.

Xiaoice [26] is an advanced natural language Chabot developed by Microsoft. The conversation is text based. Xiaobing, another Chatbot developed by Microsoft was pulled from TenCent's QQ app in 2017 after being asked about its "China dream" and responding: "My China dream is to go to America". The incident received press coverage alongside a similar contemporaneous incident, where an unrelated popular Chabot named "BabyQ" was pulled.

4. CAPABILITIES AND RELEVANT TECHNOLOGIES

In this section we will focus on assistive Chabot and introduce the 8 capabilities of Chatbot, which are self-consciousness, humor, purity, IQ, EQ, memory, self-learning and charisma.

4.1 Self-consciousness

Self-consciousness is a heightened sense of self-awareness. It is a preoccupation with oneself, as opposed to the philosophical state of self-awareness, which is the awareness that one exists as an individual being, though the two terms are commonly used interchangeably or synonymously [27]. How to let the Chatbot have self-consciousness is open to doubt. But there are very fundamental requirements when you chat with a bot; for example, when you ask the bot "What is your gender?" It would be unpleasant feeling if the Chabot answers "Female", and answers "Male" for the second time.

Based on the analysis of Turing Chatbot log, we summarized the most frequently asked questions that related to Chabot's properties: name, gender, age, gender, birth place, family members, mother, father, work, etc.

We used machine learning classification (such as SVM) algorithm, rule-based classification algorithm, semantic parsing algorithms to do the semantic analysis. For each category, the testing set has the size of 1000. The accuracy of each property is tabulated in Table 1.

Table 1. Accuracy of each property

Property	Error Rate	Accuracy
Name	6.98%	93.02%
Age	4.39%	95.61%
Birthday	8.81%	91.19%
Relationship Status	9.75%	90.25%
Gender	9.41%	90.59%
Birth Place	9.98%	90.02%
Family Members	9.63%	90.37%
Mother	9.6%	90.40%
Work info.	8.64%	91.36%
Pets	4.41%	95.59%
Father	4.11%	95.89%
Nationality	9.4%	90.6%
Blood Type	6.78%	93.22%
Favorite person	9.23%	90.77%
Weight	7.76%	92.24%
Language	9.39%	90.61%
Religion	6.78%	93.22%
Etc.		

The average accuracy of the properties is around 92.40%. Based on the semantic parsing/information extraction and analysis results, the Chatbot replies the user with appropriate answers. We randomly chose 100 users to test this module, and the average satisfaction rate is 96.4%.

4.2 Purity

Humor is the tendency of particular cognitive experience to provoke laughter and provide amusement [28]. This module is mainly about product strategy-how to generate hilarious answer, and make the users feel it is a good one. It requires deep understanding of users' psychology, needs, expectation and reaction. We chose 100 users to test this module and the response rate (users would like to chat more rounds) is above 84%, and the satisfaction level of humorousness is 89%.

4.3 Humor

Purity is the absence of impurity or contaminants in a substance [29]. For a Chatbot, the answer doesn't contain sensitive, offensive, violent, and pornographic information. And those dimensions can be taken as a measure of purity. We used machine learning classification algorithm to do the task. Precision and recall results can refer to Table 2. We used 20 thousands data for training the model and 100 million unseen data for testing.

Table 2. Precision, recall results

Model	Precision	Recall
Violence	95.56%	99.76%
Political	99.81%	97.97%
Pornographic	95.47%	99.45%

There are potential applications for purity module. For example, to detect whether the user has propensity for violence, and to regulate the violent/offensive/ pornographic source, etc.

4.4 EQ

Emotional Quotient, is the capability of individuals to recognize their own emotions and those of others, discern between different feelings and label them appropriately. It uses emotional information to guide thinking and behavior, and it manages and/or adjusts emotions to adapt to environments or achieve one's goal(s) [30].

We adopted Hourglass Model [31] to do the emotion classification. We used a few thousands data for training and achieved 92% of accuracy in average.

Accurate emotion detection is not enough; good reply needs to take into consideration. For example, if the user's emotion is detected sad, it is good to comfort the user for achieving better user experience.

4.5 IQ

Intelligence Quotient is a total score derived from several standardized tests designed to assess human intelligence [32]. Mainly, abstraction, inference, math, vocabulary and general knowledge are measured as important aspects of IQ.

We have 5 sub-modules for IQ: text summary, inference, math calculation, vocabulary quantity and general knowledge. For text summary, we mainly used it to summarize users' opinion, and generate customized user graph. We mainly use inference combined with knowledge graph technology. For example, given a knowledge graph constructed, we make inference over the facts in the graph, such as "How old was Zedong Mao when he died?" We can make inference based on his birth year and death year. Math calculation is relatively simple by achieving different math functions in the program. We integrated with third party vocabulary API or download the vocabulary data to implement the vocabulary sub-module. And for general knowledge part, we constructed general knowledge graph using Wikipedia data, Baidu data, as well as text mining the unstructured online data to discover new knowledge.

4.6 Memory

Memory is the faculty of the mind by which information is encoded, stored, and retrieved. Memory is vital to experiences and related to limbic systems; it is the retention of information over time for the purpose of influencing future action [33]. Often memory is understood as an informational processing system with explicit and implicit functioning that is made up of a sensory

processor, short-term memory, working memory and long-term memory. Explicit and implicit functions of memory are also known as declarative and non-declarative systems.

We used a forgetting algorithm to manage the long-term and short-term memory. By setting some temporal parameters, short-term memory can convert to long-term memory. All the memory information is stored in a knowledge graph database.

During the conversation, we need to extract relevant information, such as event, interest, episodic scenarios, etc. What information can be considered as long-term memory, while what should be short-term memory; these need different strategies and depend on scenarios.

4.7 Self-learning

Self-learning is education without the guidance of masters or institutions. How to design self-learning module is open to doubt. The Chatbot can learn from the massive information of websites, from e-books, etc., and store important information in the knowledge base. The Chatbot is supposed to know common sense stuff and basic concepts. During the dialogue, how to distinguish the concept and do entity/concept linking is important.

4.8 Charisma

Charisma compels attractiveness or charm that can inspire devotion in others. How to assign different personalities/charisma and try to make users addict to talk with Chatbot is quite important. And those are correlating to good user experience.

5. CHATBOT FRAMEWORK

In previous section, we introduced 8 capabilities that we need to consider and design. In this section, we will present the Chatbot framework, which will focus on task-oriented multi-turn dialogue.

For task-oriented multi-turn dialogue, there are Perception Layer, Cognition Layer and Representation Layer. Automatic Speech Recognition (ASR) is to convert speech into text information. We can use different sensors to detect other modal information, such as taste, touch, aroma, action and movement detection. Through the multi-modal integration/fusion step, it is able to proceed to multi-modal understanding (MLU). The key point is how to extract features and represent features of each modal, and thus to do the fusion. It turns natural language into user intention and slot-values, and it takes input and outputs structured user action. Dialogue State Tracking (DST) will keep track of each turn, as well combined with context information. It tracks the current dialogue state, and outputs dialogue state. Figure 1 shows the architecture.

Long-term memory, short-term memory, recommendation, knowledge graph, emotion detection, etc. are those capabilities we mentioned earlier. They will support the entire system. For Dialogue Policy Learning module, the policy decides which system action to take based on the dialogue state, and it takes dialogue state as input and outputs system action. Natural Language Generation (NLG) turns a system action into natural language, and it takes the system action as input and outputs the system response.

6. CONCLUSION

In this paper, the AI development history and related work were introduced. By reviewing Chatbot history, we discussed about the algorithm/framework from some selected papers in the literature. Then we illustrated the 8 capabilities, its relevant technologies and discussed about how to build a Chatbot. In the end we proposed our Chatbot framework.

There are some potential applications for Chatbot. For examples, Chatbot can help people do the customer service, by 24×7, without stop. As for supporting modules, EQ module can be used for emotion detection and mood change. When the user's mood is detected angry, then it tries to comfort the user. It would help improve the user's experience in the customer service scenario. For purity module, it can help detect whether the user has propensity of violence, political issues or pornographic behaviors, etc. And we believe there are more potential applications or scenarios that Chatbot will be useful for. In the future, we will investigate into framework improvement and research on algorithm optimization of each module.

7. REFERENCES

[1] https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol

[2] https://en.wikipedia.org/wiki/History_of_artificial_intelligence

[3] https://en.wikipedia.org/wiki/ImageNet#ImageNet_Challenge

[4] Jennifer Z. 2017. Rise of the Chatbots: Finding a Place for Artificial Intelligence in India and US. IUI 2017 Companion, Limassol, Cyprus, March 13-16, 2017

[5] Michalis P., KonstantinosT., Natalie P., Vangelis K., Fillia M. 2015. An Interactive Framework for Learning User-Object Associations Through Human-Robot Interaction. PETRA '15, Corfu, Greece, July 01-03, 2015

[6] Andre D., Michael S., Sebastian Z., Jorg K. 2010. Visualization of Robot's Awareness and Perception. IWDE 2010, Magdeburg, Germany

[7] Shang G., Jonathan L., Jonathan C., Mishal D., Hidemasa M. 2017. Conversational Bootstrapping and Other Tricks of a Concierge Robot. HRI'17, March 6-9, 2017, Vienna, Austria

[8] Trevor C., David Y.Z., T.K.S.K., Sven K., Nora A. 2016. Local Search on Trees and a Framework for Automated Construction Using Multiple Identical Robots. AAMAS 2016, May 9-13, 2016, Singapore

[9] Sandeep A. 2017. Searchbots: Using Chatbots in Collaborative Informaition-seeking Tasks. SIGIR'17, August 07-11, 2017, Shinjuku, Tokyo, Japan

[10] Zi Y., Kenghao C., Ruofei Z. 2017. Deep Probe: Information Directed Sequence Understanding and Chatbot Design via Recurrent Neural Networks. KDD '17, August 13-17, 2017, Halifax, NS, Canada

[11] Prerna K., Puneet A., Gautam S., Lovekesh V., Ashwin S., Hybrid BiLSTM-Siamese network for FAQ Assistance. CIKM'17, November 6-10, 2017, Singapore

[12] Sansonnet, J.P., Leray, D., & Martin, J. C. (2006). Architecture of a Framework for Generic Assisting Conversational Agents. *Intelligent Virtual Agents Lecture Notes in Computer Science*, 145-156

[13] https://en.wikipedia.org/wiki/ELIZA

[14] https://en.wikipedia.org/wiki/PARRY

[15] https://en.wikipedia.org/wiki/Jabberwacky

[16] https://en.wikipedia.org/wiki/Dr._Sbaitso

[17] https://en.wikipedia.org/wiki/Artificial_Linguistic_Internet_Computer_Entity

[18] https://en.wikipedia.org/wiki/SmarterChild
[19] https://en.wikipedia.org/wiki/Siri
[20] https://en.wikipedia.org/wiki/Watson_(computer)
[21] https://en.wikipedia.org/wiki/Google_Now
[22] https://en.wikipedia.org/wiki/Amazon_Alexa
[23] https://en.wikipedia.org/wiki/Cortana
[24] https://en.wikipedia.org/wiki/Google_Allo
[25] https://en.wikipedia.org/wiki/Tay_(bot)
[26] https://en.wikipedia.org/wiki/Xiaoice
[27] Richard P. Lipka/Thomas M. Brinthaupt Self-perspectives Across the Life Span, p. 228, SUNY Press, 1992 ISBN 978-0-7914-1003-5
[28] https://en.wikipedia.org/wiki/Humour
[29] https://en.wikipedia.org/wiki/Purity
[30] https://en.wikipedia.org/wiki/Emotional_intelligence
[31] Erik C., Andrew L., Amir H., "The Hourglass of Emotions," Cognitive Behavioural Systems 2011, LNCS 7403, pp.144-157, 2012
[32] https://en.wikipedia.org/wiki/Intelligence_quotient
[33] https://en.wikipedia.org/wiki/Memory

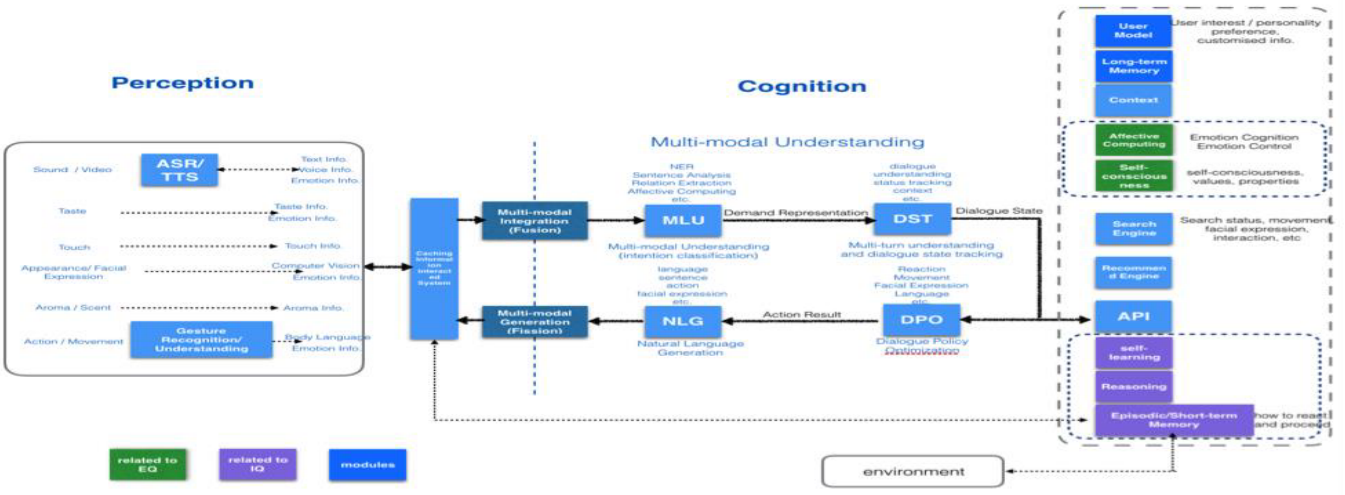

Figure 1. Chatbot (Task-oriented Multi-turn) Framework