Chatting with Bots:

A Quick Survey of Conversational Agents And Their Appropriate Applications

by

Maksim Kotov

A Thesis submitted to the Faculty

in partial fulfillment of the requirements for the

BACHELOR OF ARTS

Accepted

_____
Aaron Williams, Thesis Advisor

_____
Jackson Liscombe, Second Reader

_____
Amanda Landi, Third Reader

_____
Ian Bickford, Provost and Vice President

Bard College at Simon's Rock
Great Barrington, Massachusetts

May 2017

Acknowledgements

-I’d like to thank every single person who I talked to or interacted with during the course of the last year. Special mentions go out to:

My family and friends

My thesis committee: Aaron, Jackson and Amanda

Thesis lab f.a.v.m.o.m.s.

Table Of Contents

List of Figures

Abstract

Recent years have seen a wide-spread adoption of conversational interfaces. Conversational agents, commonly referred to as *chatbots,* are programs that can exhibit human-like conversational behavior. They are taking over messaging platforms, redefining marketing and replacing human workers in customer support centers. However, there exists healthy skepticism on whether chatbots can deliver on what has been promised. This thesis aims to cover the main points of mismatch between the actual technology and its public perception by providing historical context for understanding modern chatbots. We consult Turing's ground-laying work to figure out why people were interested in talking to computers in the first place, highlight the role of dialogue systems in the general field of AI research and follow conversational agents on their journey for recognition, funding and purpose. We apply our findings by analyzing potential applications for chatbots within the Simon’s Rock campus and community.

Outline

What follows is a detailed outline of what each chapter contains, as well as some of the themes that recur throughout the text.

Chapter 1 - Early Days aims to provide historical context for the discussion on potential applications of conversational agents. We take a closer look at the works of Alan Turing and John McCarthy as they set the foundation for research on artificial intelligence. We put the spotlight moves onto ELIZA, one of the first conversational agents to be developed, and analyze how it demonstrates the potential for chatbots.

Chapter 2 - Wider Adoption starts around the time when conversational agents are in the middle of their transition away from being used strictly for academics. A milestone event of that time period is the establishment of the annual Loebner Prize. We witness how the contest responds to criticism and evolves accordingly. The Loebner Prize becomes a platform that showcases the latest developments, as well as the most successful applications of the conversational technology. Afterwards, we follow a series of recurring disruptions to the way people interact with software and one another online: firstly, the rise of chat as conversational medium; secondly, the age of mobile apps; and, finally, messaging apps. The chapter concludes with an overview of how bots are integrated into particular platforms and services.

Chapter 3 - Building Examples uses all of the insights gained in the previous chapters to talk about good and bad real-life implementations of chatbots. Here we take a closer look at how certain processes and tasks within the college campus can be simplified by attaching a conversational interface to them. The proposed solution is a framework that supports a collection of modules, each responsible for its own well-defined task or area of expertise. We then take a closer look at two chatbot modules whose domains are evident from their titles:

Calendar Bot and *Lunch Bot.* After that, we consider possible extensions to the framework, e.g. other modules to be developed.

Chapter 1 - Early Days

"Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."
- Alan Turing, 1950

Chatbots have been overhyped since day one.

Chatbots have been overhyped even before they became *chat-based* robots, per se. Take ELIZA, a conversational computer program developed at MIT between 1964 and 1966 by Joseph Weizenbaum, one of the first of its kind. The goal of the program was to facilitate certain kinds of natural language conversation between man and computer possible (Weizenbaum, 1966). One was able to "converse" with ELIZA by typing one's own portion on a typewriter connected to a computer. The program would analyze user input and generate a response to be displayed back to the user on the typewriter. What follows is a transcript of a conversation between ELIZA and a young woman.

young woman: Men are all alike.

eliza: In what way?

young woman: They're always bugging us about something specific or other.

eliza: Can you think of a specific example?

young woman: Well, my boyfriend made me come here.

eliza: Your boyfriend made you come here?

young woman: He says I'm depressed much of the time.

eliza: I am sorry to hear you are depressed.

young woman: It's true. I'm unhappy.

eliza: Do you think coming here will help you not to be unhappy?

young woman: I need some help, that much seems certain.

eliza: What would it mean to you if you got some help?

young woman: Perhaps I could learn to get along with my mother.

eliza: Tell me more about your family.

young woman: My mother takes care of me.

eliza: Who else in your family takes care of you?

young woman: My father.

eliza: Your father.

young woman: You are like my father in some ways.

eliza: What resemblance do you see?

young woman: You are not very aggressive but I think you don't want me to notice that.

eliza: What makes you think I am not very aggressive?

young woman: You don't argue with me.

eliza: Why do you think I don't argue with you?

young woman: You are afraid of me.

eliza: Does it please you to believe I am afraid of you?

young woman: My father is afraid of everybody.

eliza: What else comes to mind when you think of your father?

young woman: Bullies.

eliza: Does that have anything to do with the fact that your boyfriend made you come here?

Creating a program that can engage in conversation is quite a broad concept. To narrow the scope of the problem, Weizenbaum puts “converse” in double quotes, possibly to

emphasize the artificial (read: *fake*) nature of the response. Later on, scholars such as Güzeldere and Franchi retrospectively defined what it meant for ELIZA to “converse”: “...when presented with sentences in English as their input, [ELIZA and other conversational agents of the earlier days] produce other grammatical sentences as their output, which sometimes manages to give the flavor of a conversation.” (Güzeldere and Franchi, 1995)

It wasn’t long until ELIZA became famous around the MIT campus. Unlike other programs that required the user to possess very specialized background knowledge, ELIZA could be appreciated on some level by anyone. After all, in order to engage with the program all one needed to know was how to hold a conversation.

Interestingly enough, even when told that the machine is a simple exercise in natural language conversation, most of the users tended to think that the machine was capable of engaging in natural conversations of any complexity and domain. Weizenbaum states:

> *“What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people”.* (1976)

What is the origin of these discrepancies between the creator’s intent and the public reception of his product? Perhaps it might have something to do with the framework for evaluating machine intelligence proposed by Alan Turing, known as *The Turing Test.*

Turing Test

Alan Turing (1912 - 1954) was a British computer scientist and mathematician. He made groundbreaking contributions to mathematics, cryptography, computer science and many other fields. While this document is mainly concerned with Turing’s ideas on computers and artificial

intelligence (AI), one should definitely consider exploring the rest on one's own time. For more information, the reader's attention is directed to a great book by Christof Teuscher, *Alan Turing: Life and Legacy of a Great Thinker*.

In 1950, Turing published his famous article, "Computing Machinery and Intelligence," in which he asks: "Can machines think?" In an effort to avoid getting stuck in philosophical arguments of what "machines" and "think" really mean, he proposes a new situation which he describes in terms of the *Imitation Game*.

In this imaginary (hypothetical) situation, two agents A (man) and B (woman) are conversing with C (the interrogator of any gender). C is isolated from A and B, and the communications are carried out via a typewriter. The goal of the game for C is to correctly identify which agent is which gender. For A, the object is to fool C into making the wrong identification, while B is trying to help the interrogator.

Turing then proceeds to ask the million dollar question: "What will happen if the machine –takes the part of A in this game?" He is particularly interested in how frequently the interrogator will correctly identify the participating agents in this new setup, as opposed to when the game is played between a man and a woman. He highlights that the new test "has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man" (Turing, 1950). Finally, an updated version of the original question is presented : "Are there imaginable digital computers which would do well in the imitation game?"

So why is this important? Retrospectively, the Turing Test became *the main* framework for "thinking" about machine intelligence. It also foreshadowed how integral the developments in *natural language processing* (NLP) will be for the whole field of AI research. However, some of the decisions that Turing made about the setup not only raised questions, but potentially set up the whole field of AI research on a failing trajectory.

Critique of the Test

It is safe to say that the debate on the legitimacy, importance and purpose of the Turing Test was paramount to the popularization of AI. However, it is hard to summarize the feedback that "Computing Machinery and Intelligence" received from the scientific (and non-scientific) community. The controversiality of the Turing Test will be explored further on.

Like any other debate of that size, it has produced an incredible amount of papers, articles and books that discuss its various aspects. Those texts are as filled with incredible insights and conclusions as they are with blatant misunderstandings and misreadings of Turing's ideas. Reviewing all of that literature is beyond the scope of this thesis, but for an inquisitive reader we recommend taking a look at "Turing Test, 50 Years Later" (Saygin et al., 2000), as well as the "Turing Test" entry in the The Stanford Encyclopedia of Philosophy (Oppy et al., 2016).

Meanwhile, we will focus primarily on aspects of Turing's paper that have influenced the opinions on what to expect from computers, conversational agents and NLP. One of these aspects has to do with the prediction that Alan Turing made:

> "*I believe that in about fifty years' time it will be possible, to programme computers... [to] play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning."* (Turing, 1950)

Most likely intended as a prediction, this part of Turing's paper is commonly mistaken as a "... criterion for having passed his Turing Test" (Harnad, 2006). The quote begins with Turing saying "I believe…", which clearly indicates its subjective nature. Furthermore, Turing states that "it will simplify the matter for the reader if I explain first my own beliefs in the matter." It is highly

unlikely that Turing intended that passage to be anything other than a benchmark, especially since at this point in the article he has already stated that the original question “Can machines think?” should be substituted with “Are there imaginable digital computers which would do well in the imitation game?” Perhaps, some confusion arises from how exact Turing’s prediction is, which may lead people to believe that this is the standard for the test. However, the “70 per cent” part is in there to emphasize that, in Turing’s opinion, computers will do better than “chance” (50/50 odds) at “fooling” people into thinking that they (people) are, indeed, conversing with another human.

There are also examples of how Turing’s particular choice of words has stirred the conversation one way or another. For instance, describing the original question in terms of the “imitation *game*” definitely has more of an adverse effect than the author hoped. Particularly, the word “game” emphasizes competition and trickery - focusing on fooling the interrogator, rather than spawning artificial intelligence.

Harnad later writes that the way Turing presents the problem is similar to a party game, and that “… the unfortunate party-game metaphor again gave rise to needless cycles of misunderstanding in later writing and thinking about the Turing Test.” (Harnad, 2006)

A few other things need to be said in regards to Turing’s decision to substitute the “Can machines think?” question. As mentioned before, Turing decided to avoid defining the word “think” since he would have to frame the definition in terms of how the word is used commonly amongst people. So instead of surveying the population in order to define “think”, he chooses to construct a situation in which the *thinking process* of the machine is empirically observable. In other words, he chooses to focus on the performance aspect over the appearance. The interrogator C is placed in a separate room, which prevents C from seeing, hearing or touching participants A and B; C cannot request a physical demonstration either. However, is this valid universally? By solely focusing on what is displayed in the typewritten conversation logs, Turing unintentionally disregards any other performance capacity besides the verbal one (Harnad

2006). One could see how this stresses the need for a program to *converse* indistinguishably from a human, while not necessarily requiring the program to possess any structural data on the outside world - or an ability to obtain more data, for that matter.

Turing's Test is hypothetical in nature. Turing was hoping to avoid definitional haggling about imaginary counterexamples (Dennet, 2002). Rather than explicitly outlining measurable requirements for passing the test, he is inquiring whether it is possible to design a machine that could *potentially* pass the Turing Test. It's not a test to confirm scientific theories, but rather a philosophical quest. Despite the criticism of different restrictions that Turing imposes, the design of the test does not impose any strong restrictions on what methods should be used to power the machine's "thinking" process.

The setup of the Turing Test says a lot with respect to how computers were perceived in the middle of the twentieth century. It almost seems that Turing's sole goal was to plant it in people's minds that "thinking" and "intelligence" do not have to be defined in human terms. And from what we learned about ELIZA so far, Turing may have just succeeded. Not only did he, to some degree, normalize the idea of conversing with a computer, but he also introduced the idea that if an agent is capable of holding a conversation then the agent could be considered intelligent. Given the sheer credibility and influence that Turing's words had at that time, it is no wonder that people began attributing intelligence to ELIZA, seeing how the program was performing relatively well in dialogues.

ELIZA's inner workings

Weizenbaum is not only surprised by how overwhelmingly well his program is received, but also a bit startled by people's tendency to overestimate ELIZA's level of understanding of

the underlying dialogue. In January 1966 his article describing the program is published in the *Communications of the ACM,* where he attempts to explain how ELIZA's design allows her to participate in discussions conducted in natural language. Weizenbaum hopes that explaining ELIZA's tricks and design would help demystify his program.

ELIZA's architecture is two-tier, where the first tier is represented by a language analyzer and the second tier - by a script. A script is described as a set of transformation rules associated with certain keywords, usually centered around a certain theme. Therefore, each script enables the program to assume a specific conversational role. In support of his architecture, Weizenbaum highlights the importance of context to natural language understanding (Weinzenbaum, 1976). In this implementation, scripts are not hardcoded - it's simply data that can be imported into the program. This adds a great deal of versatility to the program, as the code could be reused to have the computer play virtually any role, as long as a *script* for this role can be written. Furthermore, the conversation can be conducted in languages other than English[1].

Weizenbaum outlines a few fundamental technical problems that ELIZA has to solve:

1) Finding and ranking keywords
2) Identifying minimal context in which the keywords appear
3) Choosing a correct response based on transformation rules associated with the discovered keywords, as described in the script, as well as carrying out the transformation
4) Responding intelligently when no applicable rule is found
5) Allowing a programmer to edit the contents of the script

The core mechanism of the program is in the transformation rules, which work in two steps. Firstly, the input gets decomposed according to certain criteria and is processed in order to see if it meets certain criteria and, therefore, qualifies for this particular transformation. Then,

[1] Scripts for ELIZA also exist in German and Welsh (Weizenbaum, 1966).

the decomposed input is reassembled in accordance with certain guidelines before it is presented to the user[2].

Based on his design and the mechanism of transformation, Weizenbaum chooses a suitable role for ELIZA to play – the role of a Rogerian psychotherapist who is conducting an initial interview with a patient. The first implementation is sometimes referred to as DOCTOR. This role is not a difficult one to play, since it allows one to assume "the pose of knowing almost nothing of the real world" (Weizenbaum, 1966). The way DOCTOR directs the conversation by rephrasing the patient's statements into prompts could lead one to assume that the DOCTOR had some purpose in doing so, while, in actuality, the program is simply following its algorithm. Nevertheless, the patient feels heard and understood. According to Weizenbaum, in order to justify to themselves how the program is able to do so, people begin attributing to DOCTOR all sorts of background knowledge and insights, as well as reasoning ability.

What Weizenbaum is referring to will later be described as *the ELIZA Effect* by Sherry Turkle, who defined as the tendency "to project our feelings onto objects and to treat things as though they were people" (Turkle, 2004). In this particular setting, it refers to when the users are treating program as if it had more intelligence than it actually does (Trappl *et al.,* 2002). The discovery of the ELIZA effect was paramount as it exposed exactly how powerful social engineering techniques can be in disguising a program to converse like a human.

The Dartmouth Workshop

As we saw previously, one shall be extremely cautious when choosing words to describe something potentially influential. Names stick, and they carry certain expectations of which people are, for the most part, unaware. However, the fact that humans are oblivious to these

[2] Weizenbaum claims that the idea behind transformation rules is deeply rooted in the nature of social interactions (McCorduck, p.293)

contextual details shall not undermine the important role names play in forming initial impressions, and, in particular, in deciding what people are going to expect.

Ironically, despite how much Turing worked on formalizing metrics of evaluation for artificially intelligent programs, he wasn't the one who coined the term *artificial intelligence*. The bragging rights for that one belong to John McCarthy, a pioneer of AI research. In the spring of 1956, McCarthy, in coalition with other researchers such as Dr. Marvin Minsky, decided to organize a two month long summer conference (known as *The Dartmouth Summer Research Project on Artificial Intelligence* or *The Dartmouth Workshop*), during which he proposed a "study of artificial intelligence." According to McCarthy, every aspect of every feature of intelligence can be described precisely enough that "a machine can be made to *simulate* [intelligence]" (McCarthy, 1955). There's a strong emphasis on simulation, ergo the resulting intelligence is *artificial.* Later on, it's described as "the artificial intelligence problem." The proposal provides a comprehensive review of what McCarthy *et al.* consider to be different aspects of artificial intelligence problem: automation, natural language processing, abstractions, creativity etc. Most of these concepts set the foundation for respective branches of AI research.

Even though no significant progress was made toward the stated goal of replicating intelligence within a machine (Roberts, 2016), some programs developed during the Dartmouth Workshop achieved narrow successes in performing discrete tasks. The research demonstrated enough potential to attract serious interest from investors, particularly from government agencies that saw applications in data analysis, drug research and many more (McCorduck, 2004). This newly secured funding allowed researchers to be comfortably occupied by their work, and also paid for much needed equipment. One example was a $2 million grant that sponsored the launch of the MAC project at MIT in July of 1963 (Crevier, 1993). Just a year later Weizenbaum began his work on ELIZA, powered by one of the resources of the MAC project - a "time-sharing" computer.

Chapter 2 - Wider Adoption

Computers that are able to converse in natural language were initially created as academic experiments. As a result, these programs were limited to interacting with their creators and a small group of users without being able to reach a greater audience. However, over time chatbots managed to gain traction amongst developers and users outside of academia. A large contributor to that was Dr. Hugh Loebner, who founded an annual installation of the Turing Test in 1991 which eventually transformed into an annual contest for the most human-like conversational program, also known as *The Loebner Prize (*LPC).

As a result, a new wave in the development of conversational agents began, marked by the following trends: a) involvement of amateur developers, b) an incredible multitude of industries that find applications for conversational agents, and c) a great variety of methods, algorithms, learning techniques and approaches when it comes to designing and building the program. In this chapter we will review various conversational agents (some of which participated in LPC), how they relate to trends in the development of chatbots, how consumers engage with these programs, and how that engagement changes with the introduction of mobile and messaging apps. We will specifically focus on analyzing the usefulness of particular applications of conversational agents and conversational interfaces. A big argument used by the opponents of the Loebner Prize is the lack of real-world application for the programs that participate in the contest (Shieber, 1992) – and, for that matter, conversational agents in general. However, as illustrated in this chapter, that is not necessarily true.

Setup of the Loebner Prize Contest (LPC)

In 1990 Hugh Loebner agreed, with The Cambridge Center for Behavioral Studies, to underwrite a competition in an attempt to stage a real-world test of a question posed by Turing, also known as the Turing Test (TT). This rather ambitious idea came into Loebner's head as he was considering how to potentially validate his own progress in developing a computer that could pass a Turing Test. Since there was no venue to do that, Loebner decided to establish a contest to do so.

In April 1993, Dr. Stuart Shieber authored *Lessons from a Restricted Turing Test,* in which he made the claim that the contest does not adhere to its own mission statement; it neither administers the TT correctly, nor advances the field of AI, and that the prize money could be allocated better. To people like Epstein, who say that the Loebner Prize at least generated publicity (Epstein, 1991), Dr. Shieber points to the difference between increased public understanding and publicity. While the former, from Shieber's point of view, is a "laudable goal", the latter is simply raising and dashing people's expectations of the technology created by superficial coverage.

Loebner writes back more than a year later. *In Response,* his article from June 1994, states that he isn't claiming "a patent on prizes", and that he is looking "forward with great anticipation to the Shieber Prize" - a statement from Loebner to his critic, in which he prompts the one who is critiquing to take action instead of simply criticizing. Loebner then goes on to elaborate on three main reasons behind establishing the LPC.

Firstly, Loebner wanted to establish a Turing Test by itself. And he succeeds with that, at least according to Weintraub, who refers to this event as "the first Turing Test in the world ... held at the Boston Computer Museum on a grand scale." (Weintraub, 1993) Shieber made his opposition known to this. He asserted that by limiting both the topic and the tenor

of the conversation, the contest attempts what should be known as a "restricted Turing Test." Passing restricted TT would not be an indication that a machine could pass unrestricted TT; that, again, defies the purpose of the contest (Shieber, 1993). Loebner, however, insists that the 1991 Loebner Prize contest is the first time the TT was ever tried. He also says that the word “lessons” in the title of Shieber’s article represents the positive nature of the contest for its contributions to advancing the field of AI, which leads us to his second point.

The second reason Loebner decided to establish the competition was because of his commitment to advancement of AI. He is a fervent utopian who believed that automation's ultimate goal should be total unemployment[3]. At the same time, he's realistic about the current state of affairs. Loebner argued that the contest could serve as a tool to measure the state of AI at any given moment in time. He promised that any inadequacy of earlier stages would incentivize outsiders to join in on the fun.

It turns out Loebner was right on that count. Mohan Embar, the creator of the 2012 Loebner Prize winner *Chip Vivant,* mentions in his personal accounts that he decided to apply to take part in the LPC ever since he knew about the contest, back in 2001. It took him 6 years to finally submit an application, and it might have taken him even longer if Mohan would not have found himself wondering, "Which is larger: an orange or the moon?" During his quest he stumbled upon "the most human-like chatterbots[4]," also known as the finalists of the 2007 Loebner Prize contest. Not one of those programs could adequately determine which of the two spherical objects is larger, and Embar interpreted that as a sign of weakness within the competition. He put his profession and clients in the background, and with his wife's support dove straight into developing a chatbot prototype that would eventually claim the main prize four competitions later, in 2012.

[3] The goal, then, is to design a society that can "equitably distribute the fruits of automation among its members."
[4] Used interchangeably with *chatbot.*

One must point out the perseverance with which Chip Vivant was created and released to the world. During his first try in 2008, Embar didn't make it past the initial screening to participate in the LPC. The program finished third twice --- in 2009 and 2011 --- and it was disqualified from the competition all together in 2010 due to a bug in the code. Nevertheless, all the necessary factors finally came together in 2012, and Embar took home the prize. He was particularly happy that he got to do it on his own terms, "without a fake backstory, fake typing errors and a preponderance of canned responses." (Embar, 2012)

As the third reason behind the establishment of LPC, Loebner briefly mentioned how his background in sociology allowed him to construct and conduct this social experiment in the form of a yearly competition to find the most human-like conversational agent. In opposition to Shieber, he insisted that LPC's recurring nature speaks to its credibility.

As to being called out on supporting development of tricky engineering hacks over systematic research in the area, Dr. Loebner makes a point that research would be boring if it revealed answers to the questions and problems directly intended. Therefore, it should be allowed to spawn in any direction, as there is a possibility that material will become useful down the road in the same discipline or possibly a different one, as an example of transferring mental models from one group of academia to another. He believes that his prize will "prove useful, nonetheless, perhaps in very unexpected ways." (Loebner, 1994)

Updates to the Loebner Prize format

As the tempers have settled down from that extremely polite academic argument, Loebner and the contest organizers were able to collect constructive feedback from participants and outside spectators. And they responded in the only right way -- by acting on it. Having witnessed how far the current state of the art systems are from passing unrestricted TT, Loebner decided that in the future LPC's artificial computer entries would be

rated based on which one of them is the closest to being human-like. However, he did not deny the benefits of unrestricted TT, and decided that from 1995 on that the tests would be unrestricted. Some other minor updates included resolving a couple of the security concerns and requiring all participants to be self-reporting.

Despite all that work, the LPC hasn't managed to make peace with opposition on all fronts. One of the most vocal opponents, Marvin Minsky (yes, same Minsky from the Dartmouth Workshop), promised $100 to the first person who gets Loebner to "revoke his stupid prize, save himself some money, and spare us the horror of this obnoxious and unproductive annual publicity campaign" (Minsky, 1995). What Minsky didn't know is that by announcing said target, he automatically agreed to co-sponsor the 1995 LPC (at least according to Loebner). Loebner figured that the only way to "revoke" the prize is by passing the Turing Test, and if a program passed the TT that year, the creators of that program would be the ones who "revoked" the prize. As an honorable man, Minsky would have to pay $100 to the person who writes a program that passes the TT. Consequently, that person would also be receiving the Loebner Prize. That way, Minsky would be contributing to the monetary prize of the contest, which makes him, by definition, a co-sponsor.

A big issue with Loebner Prize came up with respect to the publication of the winning entry. The following are the two possible ways to think about it: 1) employ a cloak of secrecy for the winner due to great commercial appeal for potential AI applications; 2) allow full transparency for the sake of the advancement in the field of AI.

Participating in Loebner Prize competition is quite enticing. A gold medal and $100,000 for the creators of the program that is able to pass an unrestricted Turing Test. However, when planning the setup of the first LPC, the organizers decided to settle for running a restricted version of the Turing Test, due to the state of language understanding technology at the time (Epstein, 1992). What that meant is that each entry was evaluated based on its ability to hold

conversation on a specific topic chosen by the participants themselves. Some example topics included “Whimsical Conversation”, “Liberal or Conservative?” and “Hockey”. Eventually, this domain restriction was lifted for the 1995 LPC (Loebner, 1994).

While the $100,000 jackpot is intended for a program that eventually passes the unrestricted version of the Turing Test, a bronze medal and a cash prize[5] awarded yearly to the most human-like bot, who also gets the title of the winner of LPC of that year.

Chatter Bot

The first few instances of the Loebner Prize were not too exciting: the same three programs finished first, second and third in 1991, 1992 and 1993 (Mauldin, 1994). The first place was taken by Joseph Weintraub’s program, which is a variation his *Personal Computer (PC) Therapist* conversational software; Kenneth Colby and his program that helps discuss bad marriage placed second, while the third place was claimed by Dr. Michael Mauldin’s program *Chatter Bot*[6].

Chatter Bot was designed to be a computer-controlled player bot inside a text-based multiplayer game TinyMUD, an implementation of a multi-user dungeon (MUD). MUDs provide worlds where players communicate with one another exclusively by typing. Therefore, an amazing opportunity to work on TT presents itself. Chatter Bot's intended tasks included being able to hold a conversation with other players, explore new worlds, and answer questions about navigation, as well as other players, rooms and objects. Chatter Bot and similar programs are more likely to succeed in a MUD world because of what Dr. Mauldin calls an *unsuspecting Turing Test*: players by default assume that the player they are chatting with is also a person. (Mauldin, 1994)

[5] In 2014, the cash prize was $4000.
[6] Eventually, the term *chatter bot* shrunk to chatbot.

Mauldin believes that the LPC does indeed further the field of AI. He notes, however, that a program that could pass unrestricted TT at some point in the future would have to leverage knowledge and tactics "based on long experience with typed communication with real people" (Mauldin, 1994). The two instances of typed conversational interactions he mentions are TinyMUDs and *Internet Relay Chat (IRC)* - a chat system developed by Jarkko Oikarinen in Finland in the late 1980s (Stenberg, 2011). IRC became popular as it allowed anybody with a connection to the Internet to participate in live discussions online. Interestingly enough, IRC allowed bots (software programs that performs automated tasks) to connect to the network as a client ("IRC Bot FAQ", 2003), which meant that these programs appeared to other users as just another user - similar to how *Chatter Bot* appeared to other MUD players. This initiated the process of introducing of bots into chat software, which marked an important milestone in the process of wider adoption of conversational agents.

Bots are moving online

Having realized that there is not enough data on how people interact with bots, developers decided to get their programs in front of as many people as possible. Luckily, the Internet was spreading in the mid 1990s. One particular advantage of hosting chatbots online is the program's ability to interact with many users at the same time, facilitated by the client/server model.

It wasn't too long until the method began delivering results. Albert One, the 1998 LPC winner, was arguably the first contestant to leverage the insights obtained from chatting with users online[7]. Some parts of the program were hosted online earlier in 1995 with the goal of exploring how people would interact with the bot. These efforts are summarized in a paper titled "FRED, Milton and Barry: the evolution of intelligent agents for the Web" (Caputo *et al.*, 1997),

[7] As discovered based on information obtained during preliminary research for the project.

coauthored by Robby Garner, Albert One's creator. Based on their findings, the authors classify people's reaction towards a conversational program into two main categories:

1) Attempting to self-administer a Turing Test in an effort to find out if the agent is actually a computer program;
2) Using the interaction as a medium to share their views or simply vent.

Even though this classification seems somewhat self-evident, the real value of the experiment lies in the fact that it validated the leap-of-faith assumption made by bot developers – people are interested in chatting with their programs online!

Now here is where things get interesting. In December 1993 Richard Wallace, who was a faculty at NYU at the time, began working on a very rather intriguing product – a robotic eye that was connected to the Internet and could be controlled by visitors online. The program was designed to respond to typed natural language commands like "tilt up" or "pan left". However, Wallace noticed that occasional malfunction in the program would trigger a rather emotional response from the users who'd type things like "You're stupid" (Thompson, 2002).

It was at that moment in time that Wallace had a realization. If users are trying to chat with the robot eye, why not have it talk back? Remembering the ELIZA bot, Wallace decided to build an updated version of the classic program, but with a twist: this one would have a grain of his personality embedded inside.

The first version of ALICE (that's what Wallace named it; stands for Artificial Linguistic Computer Entity) came online on November 23, 1995. Looking through the logs of conversations between users and ALICE, Wallace began noticing something peculiar: about 95% of user utterances began with one of 1800 words (Wallace, 2009). Apparently, the human input into ALICE was behaving in accordance with Zipf's Law, which states that in a given corpus of natural language utterances, the frequency of any word is inversely proportional to its

rank in the frequency table[8]. Furthermore, from ALICE's logs Wallace found that the inverse relationship between frequency and rank holds not only for words, but also on the level of entire queries (Wallace, 2009). What that meant for Wallace is that if he showed ALICE what to say back every time "she" could not come up with a response, he would eventually cover the majority of most common cases (Thompson, 2002). Wallace later writes: "If there was any significant innovation after ELIZA, it was this. There is a world of difference between writing 10,000 questions and answers for a bot, versus knowing in advance what the top 10,000 most likely questions will be. A.L.I.C.E. replies were developed directly in response to what people say." (Wallace, 2009)

Siri's disruption

An example of how a conversational interface is capable of disrupting entire industries is Siri. Debuting in 2011 as a part of Apple's iPhone 4S release, the program became available to anyone who purchased the phone, which is an unprecedented number of users for a conversational agent. The real achievement, however, was the fact that people began using Siri for help, with a particular goal in mind. Brooker claims it "works annoyingly well" (Brooker, 2011).

The origins of Siri can be traced back as far as 2003, when its creator Adam Cheyer was working on an intelligent agent at SRI international, a government-sponsored research lab in Menlo Park, CA. There he met Dag Kittlaus, a former Motorola executive, who eventually convinced Cheyer to port the technology from the intelligent assistant he'd been working on to a mobile application.

[8] A frequency table contains information on how many times a word occurs in the given corpus, sorted by descending frequency.

After only two weeks of being published in the App Store, Siri began generating significant enough traction that it attracted the attention of Steve Jobs himself. In April 2010, Apple acquired Siri for a reported $200 million (Shonfeld, 2010).

Siri successfully leveraged the advantage of the mobile phone interface. However, nothing groundbreaking in terms of conversational technology was introduced. Bedigian (2011) provides a great example that exposes a particular shortcoming of Siri. Consider the two dialogues below.

User: Siri, tell me a joke!

Siri: Two iPhones walk into a bar... I forget the rest.

****

User: Siri, what's the funniest joke you have ever heard?

Siri: I can't, I always forget the punchline.

While in the first dialogue Siri successfully captures the user's intent, her intuition is way off the second time. Clearly both of these canned responses are triggered by the presence of the same keyword - "joke". Meanwhile, in the second dialogue the program fails to recognize the context in which the keyword appears.

Despite what the public perception of it might have been, Siri was not meant to be another clever AI chatbot. Instead, it was supposed to disrupt mobile commerce. Original implementatie as a standalone app could bypass search engines, order flowers, buy movie tickets, etc. Cheyer went door-to-door in Silicon Valley to get permission to integrate Siri with services like Yelp, StubHub etc, racking up a total of 42 partner services. However, partnerships dissolved as Apple took over. Jobs' vision for the first iteration of the product was different than that of Kittlaus and Cheyer. Jobs wanted to make Siri the champion of

iOS, the ultimate tour guide. Arguably, that goal has been met, possibly at the expense of the original vision behind Siri. A 2012 survey concluded that the top 3 most used Siri functions are reminders, calendar and the alarm clock (“Global Siri Usage Statistics”, 2012).

Since then, Kittlaus and Cheyer have both left Apple to work on their new project – Viv. They hope Viv becomes what Siri was supposed to be – an open system.

Messaging platforms

“Chat apps will come to be thought of as the new browsers; bots will be the new websites. This is the beginning of a new Internet.”
– Ted Livingston, founder of Kik.

While Siri was continuing to underperform, messaging apps were gaining traction.

Despite unlimited selection options and constant arrival of new apps, comScore reports that most US smartphone owners download zero apps in a typical month ("The U.S. Mobile App Report", 2014). It also stated that the consumers were spending the majority of their time using a very limited selection of most important apps. In particular, people began spending unprecedented amounts of time on *messaging apps*. According to a Business Intelligence report circa 2015, messaging apps have surpassed social networks in terms of monthly active users (see Figure 1).

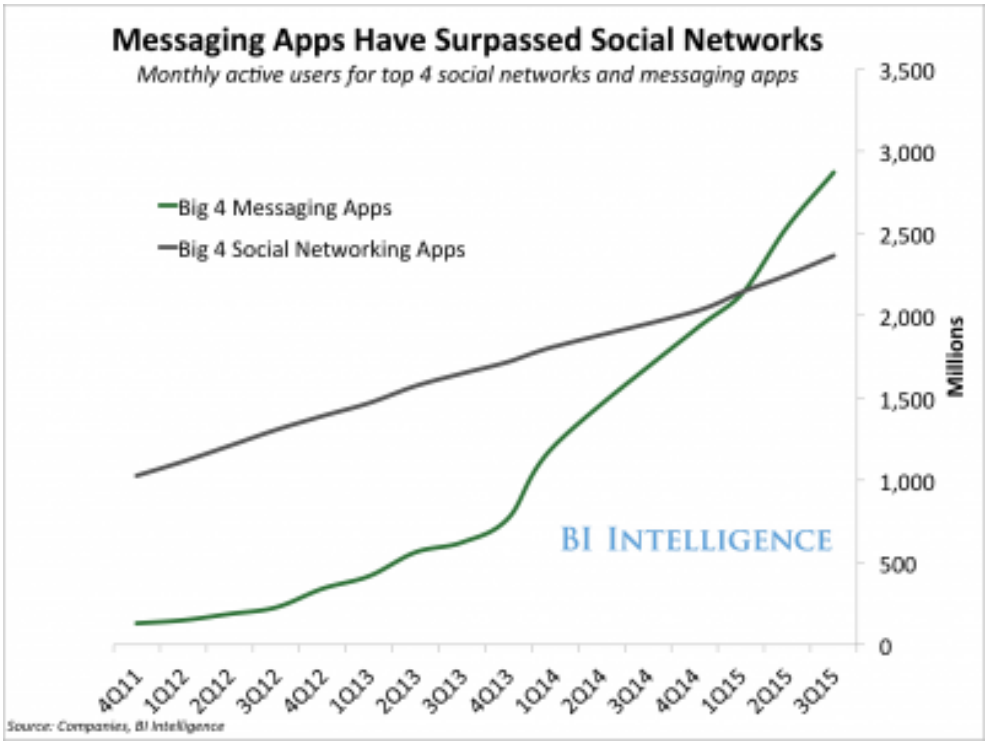


Figure 1. ***Monthly active users for top 4 social networks and messaging apps***

An opportunity presents itself for businesses to deploy chatbots to reach their customers where they spend most of their time. At this point in time, a few major messaging platforms such as Facebook Messenger, Kik, Telegram and many others support chatbot integrations. That means that these chatbots do not need to be downloaded and installed on user's phone as standalone applications. They live inside bigger apps, which facilitates smoother user experience since switching between bots does not require switching between apps.

On one hand, integrating with a particular messaging app allows a chatbot developer to access the platform's entire user base. The platform that leverages this advantage the most is Facebook, with its monthly active users on Messenger passing the 1.2 billion mark in 2017 (Constine, 2017). On the other hand, each platform comes with its own caveats and limitations. All of these factors come in play in the next chapter, where we discuss potential applications for chatbot technology within the Simon's Rock campus.

Chapter 3 - Building Examples

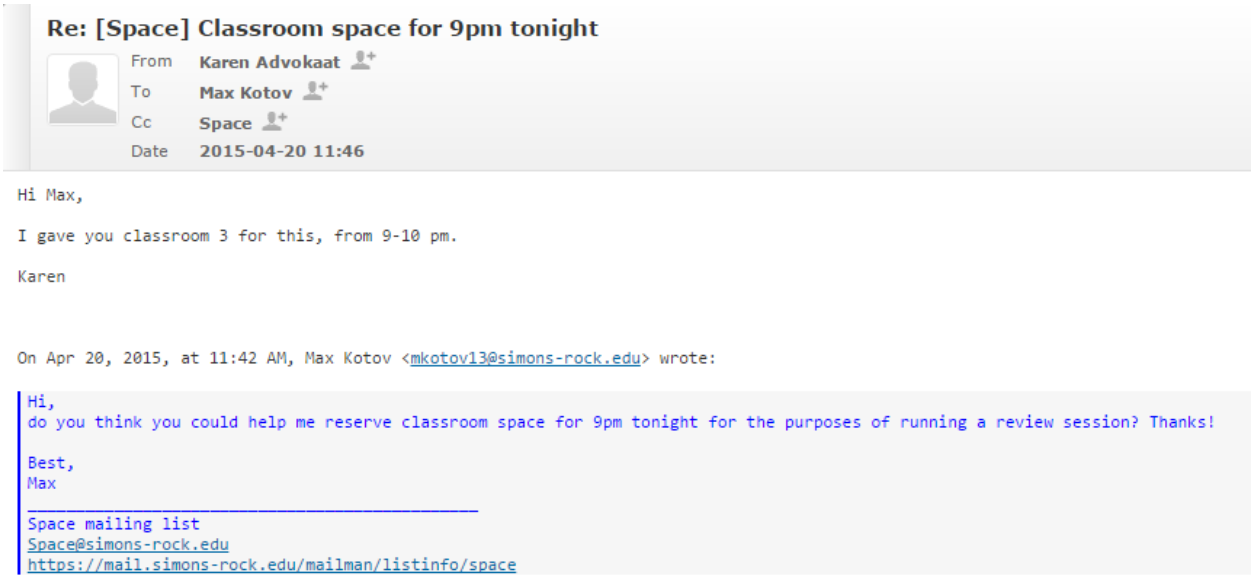
In the previous sections we have set up a solid historical context to discuss chatbots, their implementations, the niche they are attempting to fill, as well as their future prospects. That means that now we can move on to the more fun part - actually building one. (There is hardly a better way to learn about a technology than using it).

As a general theme for my bots, I decided to focus on how this technology can be leveraged to improve/optimize certain aspects of the college environment here at Bard College at Simon's Rock. It should be noted that there has been significant strides made in that direction, mainly the installation of the Rocket Chat (open source chat-based enterprise communication software) in the summer of 2016. According to the chat server's admin, Dennis Chen (2015), "chat" has served over 150,000 messages during its first year. Rocket Chat platform supports bot integration, so there is a great chance that the bots will eventually be available to chat on there. However, as of April 2017 Rocket Chat (referred to around campus simply as “chat”) has yet to be adopted by college students and faculty outside of the Computer Science department. Hosting the bots exclusively on Rocket Chat will strip them of their competitive advantage over web browsing and traditional mobile applications gained because of how easily they can be accessed via messaging platforms.

In the remainder of this chapter we consider chatbots for reserving rooms, answering simple questions about the academic calendar and surveying what is served for lunch at the dining hall. This thesis has made an emphasis on finding the appropriate applications for the chatbot technology. In order to further the contrast between a good and a bad implementation, we will review examples of both.

Space Bot

The process of booking a space for an event (lecture, club meeting etc.) involves several steps and a small group of people. In the current arrangement, if you are interested in hosting an event, you have to email your request to space@simons-rock.edu (referred to later simply as *Space)*. This request arrives to a shared inbox monitored by several people. Depending on their availability, one person from the group responds with either a confirmation or with a better suggestion in terms of the location. Provided below is an example of how one could request to book a room.



Re: [Space] Classroom space for 9pm tonight

From Karen Advokaat
To Max Kotov
Cc Space
Date 2015-04-20 11:46

Hi Max,

I gave you classroom 3 for this, from 9-10 pm.

Karen

On Apr 20, 2015, at 11:42 AM, Max Kotov <mkotov13@simons-rock.edu> wrote:

> Hi,
> do you think you could help me reserve classroom space for 9pm tonight for the purposes of running a review session? Thanks!
>
> Best,
> Max
>
> _____
> Space mailing list
> Space@simons-rock.edu
> https://mail.simons-rock.edu/mailman/listinfo/space

Figure 2. ***Booking a room for a review session using Space@simons-rock.edu***

The idea of improving the current Space workflow using chatbots was proposed during the mid-year thesis committee meeting. However, its implementation was never carried out. What follows is an explanation of how we came to realize that attaching a conversational interface to this system would not noticeably improve the user experience.

There seemed to be a sense of general consensus around campus -- including the people who monitor Space themselves -- that Space could use an upgrade. This system is prone to occasional errors due to coordination of multiple people; it also lacks proper data storage and manipulation functionality. However, as we began collecting opinions with regards to what could be improved about the current system, the necessity of adding a conversational interface continued getting bleaker and bleaker.

A potential chatbot that could fully automate this workflow would require two things:

1. Information on classroom availability;
2. Well-specified criteria for which requests should be approved

While the first already exists and could potentially be tapped into, facilitating automatic approval is rather difficult. And if the bot does not have an ability to automatically pre-approve requests, the user is going to be left hanging after going through a rather extensive line of questioning. That kind of negative user experience is the reason why users get so discouraged by poorly designed implementations of chatbots.

Rather, the request would have to be collected from user input and transferred to a human capable of making the call. While the almost seamless conversational UX[9] powered by these hybrid human-bot solutions might be a widely accepted practice in current implementations of automated customer support systems for enterprise, it simply does not make sense in the case of Space@SRC. The goal of this project is to make the lives of campus staff easier, not to create additional superfluous tasks.

An important lesson to learn is that one should avoid attaching a conversational interface just for the sake of doing so.

[9] UX stands for user experience.

Calendar Bot

The dates in the academic calendar come up in conversation on the Simon's Rock campus quite often (making travel plans, planning homework, etc.). At the moment in time that I'm writing this, retrieving the dates information is rather arduous. Students or faculty must login into a private website with their Simon's Rock credentials, then they navigate to the "Academics" portion of the website, then they search for the calendar and download or open a pdf file, then they search within the pdf for the specific date. It can take several minutes to complete this task, and laptop and desktop users often end up with multiple copies of the pdf downloaded on their computer[10].

This seems like a good application for bots. A conversational interface will allow people to describe what they're looking for without having to know the precise terminology.

Building with Motion AI

For my first bot I was looking for a platform that would ideally help me build a bot with little to no coding involved. After doing a little bit of research and reaching out to the chatbot-building community online, I came across Motion AI.

One feature that instantly sets Motion AI apart from other platforms is its visual approach to designing conversational systems. Bots in Motion AI consist of interconnected blocks called Modules. Each bot response correlates to a module. There a few different types of modules, and they differ from one another based on how the response is generated and what type of data is supposed to be extracted from the user's following response. Complex conversations are modeled through the use of Connections (see Figure 2).

[10] It was brought to my attention later that an updated version of the academic calendar is located on the Simon's Rock website. However, the process of getting there is not any easier.

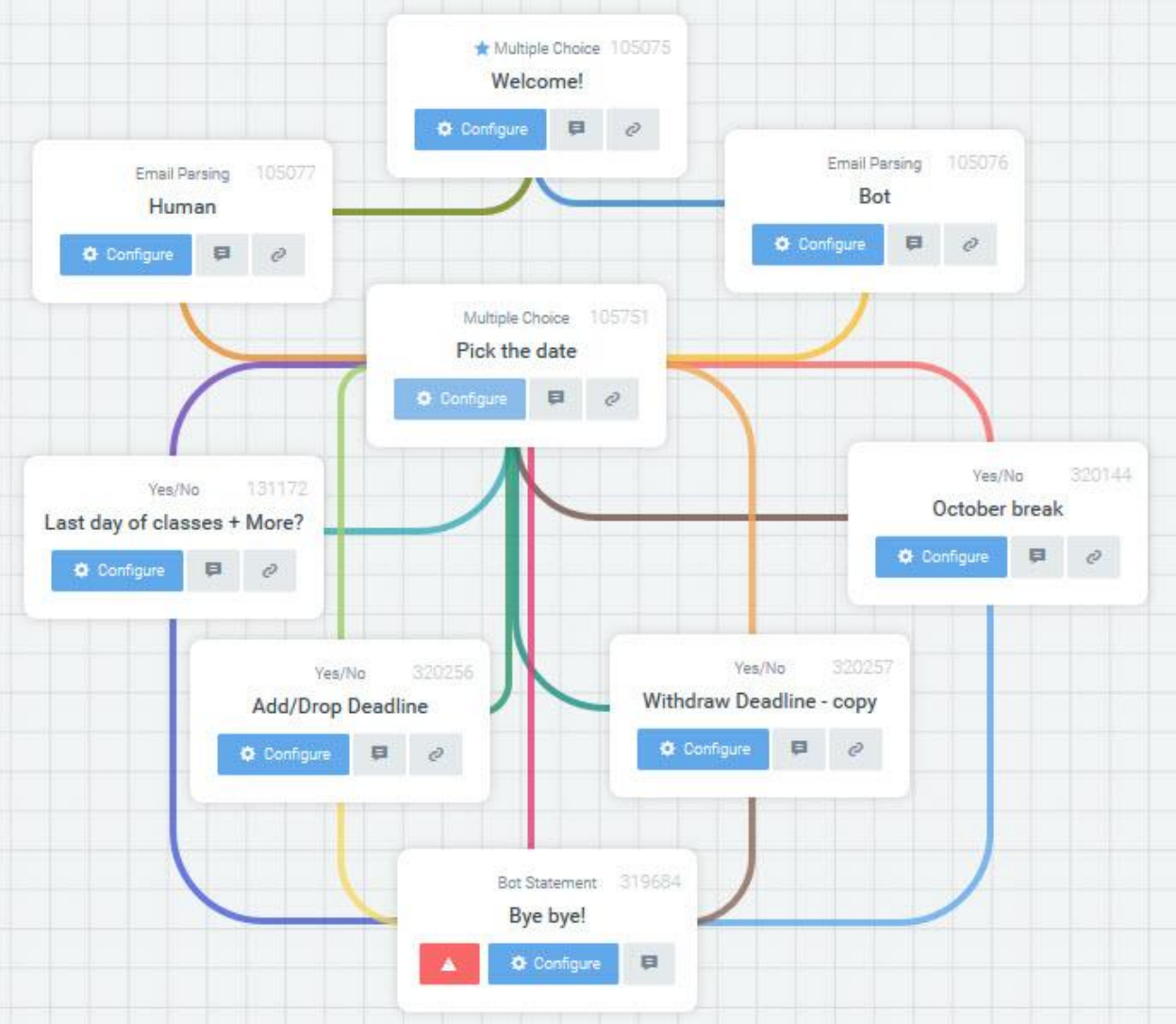


Figure 3. ***This conversational flow chart represents how users interact with the bot***

The resulting diagram of the proposed conversational flow is displayed in Figure 2.

A few bugs were detected as soon as the first outside user interacted with the bot. It also became obvious that there needs to be a way to access the bot quickly. Motion AI provided the address for where the bot can be found, which saved me from having to embed the code into a webpage or integrating the bot into any one messaging platform. However, the address was quite long, so I had to use a URL shortener. The Calendar Bot now lives at http://tinyurl.com/calendarbot.

Below is an example of a conversation between Calendar Bot and a user.

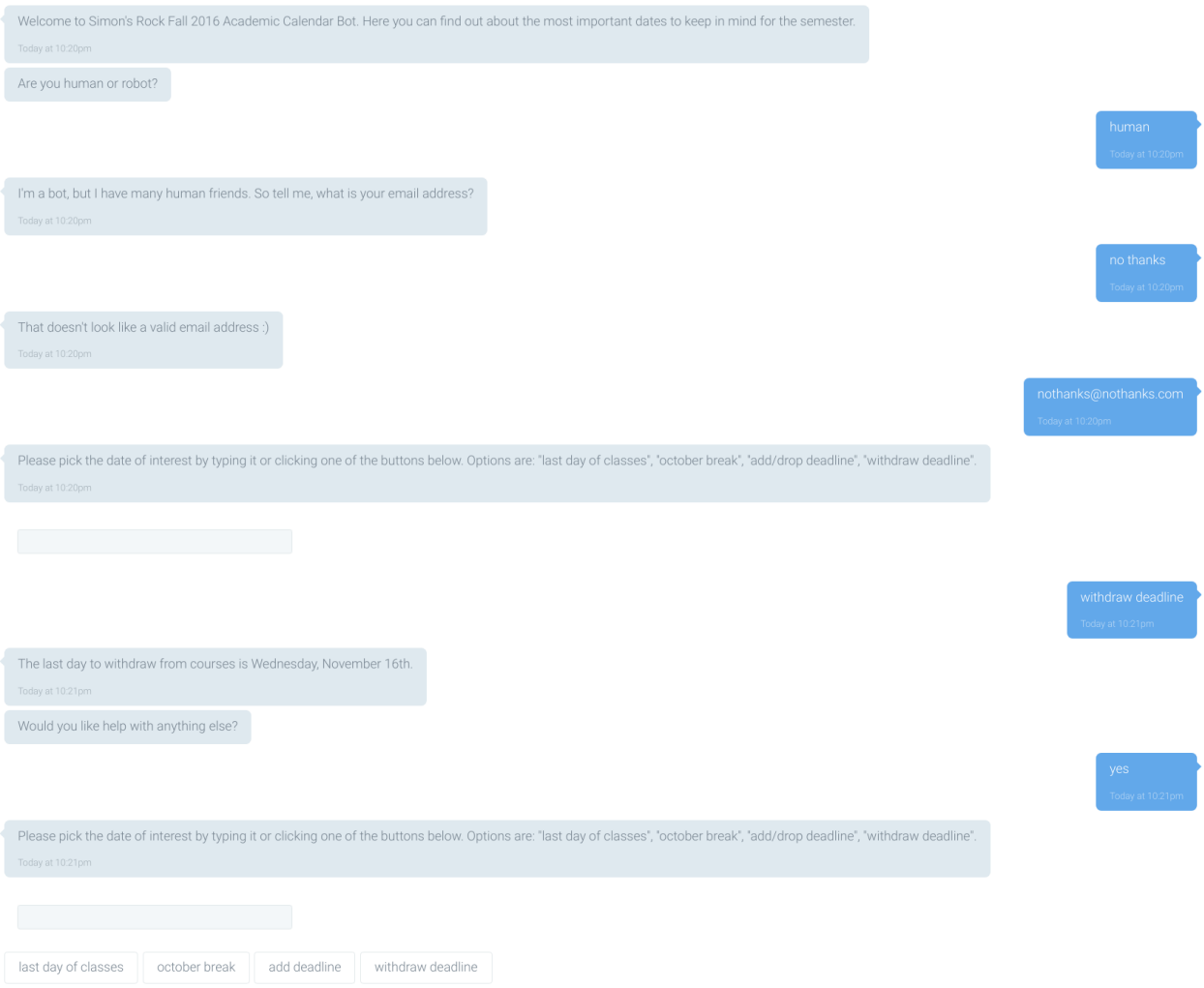


Figure 4. ***A sample dialogue between a user and Calendar Bot***

Lunch Bot

Another important aspect of campus life is the dining hall. While trying to organize my day and deciding when I should eat, I would often find myself texting friends to find out what's at the dining hall. So it seemed almost natural to implement a chatbot to provide this information.

There are two things that this bot needs to be able to do. Firstly, it needs to obtain the information about the food. Secondly, it needs to be able to provide that information to the user via a conversational interface.

Processing language with Wit.AI

Due to the limited amount of premade conversational blocks, and no possibility for automation, the Lunch Bot was not implemented with Motion AI. Instead, we used Wit.AI, a platform for natural language understanding. Its two main features are *Intent Parser*, which translates user input into actions, and *Stories*, which helps developers outline the proposed flow of conversation between the bot and the user.

One of the advantages of the service highlighted by the company is a large database of previously trained entities. There's a form on Wit.AI's front page that demonstrates how the system will handle certain input. For example, a phrase "Open the door please!" gives the result seen in Figure 5.

Figure 5. ***Wit.AI's interpretation of "Open the door please!"***

What about food? Can Wit successfully determine that the user is trying to figure out what's at the dining hall? Turns out, not so well, as seen in Figure 6.

Figure 6. ***Wit.AI's response to "What's for lunch?"***

Since the system did not have a predefined way of handling our proposed input, we had to manually tune it. Suppose a user types: "*What's for lunch?"* That natural sentence should be understood as intent "food" and meal "lunch". Once an intent is detected, the action `getFood` should be executed. If all goes well, this action is expected to create a `food` key in the context. At that stage, the `getFood` action hasn't been implemented yet. We're simply telling Wit our expectations for the action. The actual information is collected by the means of web scraping.

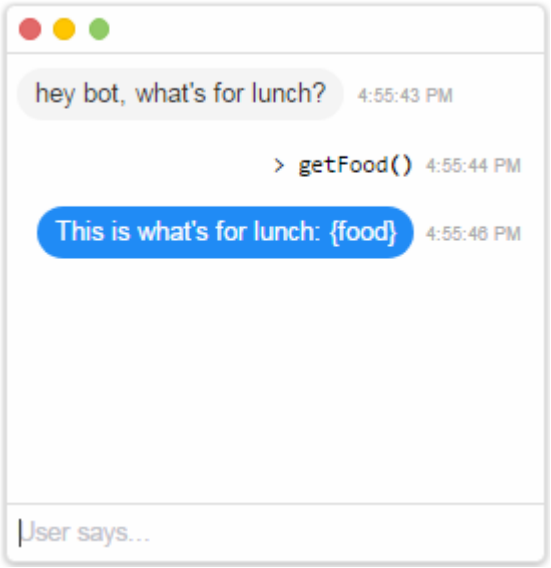


Figure 7. ***Designing the dialogue via Wit.AI***

Scraping data using ParseHub

Currently, one can find what's for lunch on the Chartwells website at https://new.dineoncampus.com/simonsrock/whats-on-the-menu. We need to use a web scraper to collect the necessary information since Chartwells does not have an API.

ParseHub provides a very intuitive graphical interface for web scraping, which allows us to focus on the conversation part of the bot and to avoid having to analyze the source code of the page. It also provides an extensive API, which comes in handy for the setup that we have.

Currently Lunch bot is still under construction.

Conclusion

In this thesis we have seen that Chatbots and their public perception have come a long way in their relatively short history. It was not that long ago that Alan Turing's groundbreaking thinking even introduced the idea that one day people would be able to communicate with a machine. Some specifics of the Turing Test he proposed are still being debated up to this day. 15 years later, Joseph Weizenbaum's ELIZA revealed people's tendency to attempt to relate to whatever it is they are "talking" to. Weizenbaum's experiment also showed how one could possibly leverage such an unusual reaction from people in order to make a conversational program sound more human. These social engineering techniques later on became widely applied by chatbot makers.

We also traced how conversational agents slowly moved away from being purely academic and progressively made their way into many aspects of people's daily lives. There are a few people whose contribution to this shift cannot be overestimated. One of them is Hugh Loebner. While in the academic community might disregard the legitimacy of the annual contest he established, we saw ourselves how the Loebner Prize helped popularize the idea of conversational agents. An important lesson that to learn is that a problem as big as natural language understanding cannot be solved behind closed doors.

We also considered the development of bots to serve the Simon's Rock community. In particular, we examined three specific applications: room reservations, academic calendar queries, and meal reports. We concluded that the first problem would be better solved with other technologies, and we provided initial steps to solving the latter problems using Motion AI and Wit.AI, respectively. While these small-domain service bots can solve individual niche problems, the ultimate goal is to begin developing a highly cognitive agent with an ability to understand and manage data from all kinds of sources and areas. At this moment in time, the technology behind chatbots is still new, and it will likely experience major changes. But one thing is for sure

-- properly implemented conversational interfaces facilitate amazing user experiences. And user is king.

Bibliography

"College Chatbot Service AdmitHub Raises $2.95M in Seed Funding to Guide Students Through School (EdSurge News)." *EdSurge*. N.p., 11 Jan. 2017. Web. 12 May 2017.

"Advanced Natural Language Processing Tools for Bot Makers – LUIS, Wit.ai, Api.ai and Others (UPDATED)." *Stanfy | San Francisco Mobile App Developers*. N.p., n.d. Web. 29 Apr. 2017.

"Alan Turing at 100." *Harvard Gazette*. N.p., 13 Sept. 2012. Web. 26 Apr. 2017.

Alton, Larry. "How Consumer-Focused AI Startups Are Breaking Down Language." *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Barnett, Emma. "Google's Eric Schmidt: Apple's Siri Could Pose 'Threat'." 7 Nov. 2011. *www.telegraph.co.uk*. Web. 8 May 2017.

Bedigian, Louis. "Is Siri Seriously Awesome or Seriously Flawed?" *Benzinga*. N.p., 14 Oct. 2011. Web. 14 May 2017.

Billings, Lee. "Rise of Roboethics." *Seed* 16 July 2007. Web. 24 Apr. 2017.

Bora, Anita. "What a Chatterbot!" *Reddif Guide To The Net*. N.p., 10 June 2003. Web. 28 Apr. 2017.

"Bot-making Service Motion.ai Now Supports Node.js." *VentureBeat*. N.p., 1 Dec. 2016. Web. 12 May 2017.

"Brillig Understanding - About Us." N.p., n.d. Web. 10 May 2017.

Brooker, Charlie. "Charlie Brooker: The Dark Side of Our Gadget Addiction." *The Guardian* 1 Dec. 2011. *The Guardian*. Web. 8 May 2017.

Business, Author: Steven Levy Steven Levy. "Siri's Inventors Are Building a Radical New AI That Does Anything You Ask." *WIRED*. N.p., n.d. Web. 8 May 2017.

"C.L.A.U.D.I.O. Personality Test Bot (A.L.I.C.E. AI Foundation)." N.p., n.d. Web. 12 May 2017.

Caputo, Luigi, Robby Garner, and Paco Xander Nathan. "FRED, Milton and Barry: The Evolution of Intelligent Agents for the Web." *Advances in Intelligent Systems*. Ed. F. C. Morabito. Amsterdam, The Netherlands: IOS Press, 1997. 400–407. *ACM Digital Library*. Web. 12 May 2017.

"Chatbot Rose Wins 2015's Loebner Artificial Intelligence Prize." 19 Sept. 2015. *www.bbc.com*. Web. 25 Apr. 2017.

"Chip Vivant Wins the 2012 Loebner Prize Competition! | Chip Vivant." N.p., n.d. Web. 19 Apr. 2017.

"Clippy's Back: The Future of Microsoft Is Chatbots." *Bloomberg.com*. *www.bloomberg.com*. Web. 27 Apr. 2017.

Colby, K. M., J. B. Watt, and J. P. Gilbert. "A Computer Method of Psychotherapy: Preliminary

Communication." *The Journal of Nervous and Mental Disease* 142.2 (1966): 148–152. Print.

comScore, Inc. "The U.S. Mobile App Report." 2014. Whitepaper.

Constine, Josh. "Facebook Messenger Hits 1.2 Billion Monthly Users, up from 1B in July." *TechCrunch*. N.p., n.d. Web. 17 May 2017.

Contributor. "Siri Is Only The Beginning." *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Crevier, Daniel. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York, NY: Basic Books, 1993. Print.

"David Levy Wins Loebner 2009." N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

Dennett, Daniel C. "Can Machines Think?" *Alan Turing: Life and Legacy of a Great Thinker*. Ed. Christof Teuscher. Springer Berlin Heidelberg, 2004. 295–316. *link.springer.com*. Web. 25 Apr. 2017.

"Divabot 2015." N.p., n.d. Web. 10 May 2017.

"Dr. Thomas Whalen's Biography." N.p., n.d. Web. 12 May 2017.

Dvorsky, George. "How Much Longer Before Our First AI Catastrophe?" *io9*. N.p., 6 Apr. 2017. Web. 6 Apr. 2017.

Dwoskin, Elizabeth. "Siri's Creators Say They've Made Something Better That Will Take Care of Everything for You." *Washington Post* 4 May 2016. Web. 8 May 2017.

Effron, Lauren. "Siri Lost in Translation With Accents." *ABC News*. N.p., 28 Oct. 2011. Web. 8 May 2017.

Epstein, Robert. "The Quest for the Thinking Computer." *AI Magazine* 13.2 (1992): 81. Print.

Gartner. *Gartner Says Organizations That Integrate Communities Into Customer Support Can Realize Cost Reductions of Up to 50 Percent*. N.p., 2017. Web. 19 Apr. 2017.

Gay, Alicia. "Apple's Latest iPhone Has No Problem Pointing Users to Viagra, But Comes Up Blank on Birth Control, Abortion." *American Civil Liberties Union*. N.p., 30 Nov. 2011. Web. 8 May 2017.

Güzeldere, Güven, and Stefano Franchi. "Dialogues with Colorful Personalities of Early Ai." *SEHR* 4.2 (1995): n. pag. Web. 5 Apr. 2017.

Harnad, Stevan. "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence." *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Ed. Robert Epstein and Grace Peters. Kluwer, 2006. *cogprints.org*. Web. 25 Apr. 2017.

Haveliwala, Taher. "Search Facilities for Internet Relay Chat." *JCDL Poster Session, 2002*. Portland, Oregon: N.p., 2002. *ilpubs.stanford.edu:8090*. Web. 19 Apr. 2017.

"Home Page of the Loebner Prize." N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

"IRC Bot FAQ." *IRCHelp.org*. N.p., June 2003. Web. 12 May 2017.

"John Battelle On The Future Of Search." *Search Engine Land*. N.p., 19 Mar. 2010. Web. 8 May 2017.

"Joseph Weintraub Information." N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

Kahn, Jennifer. "It's Alive!" *WIRED* Mar. 2002. Web. 26 Apr. 2017.

Kantrowitz, Alex. "My Life With Facebook's M." *BuzzFeed*. N.p., n.d. Web. 29 Apr. 2017.

Kelly, Meghan. "Siri Ads 'false and Misleading,' According to Class Action Lawsuit." *VentureBeat*. N.p., 14 Mar. 2012. Web. 8 May 2017.

Kittlaus, Dag. "A Cambrian Explosion In AI Is Coming." *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Knecht, Tobias. "A Brief History Of Bots And How They've Shaped The Internet Today." N.p., 21 Sept. 2016. Web. 12 May 2017.

Livingston, Aiden. "Bots Will Replace People before They Replace Apps." *VentureBeat*. N.p., 16 Jan. 2017. Web. 11 May 2017.

Loebner, Hugh. "In Response." *Communications of the ACM* (1994): n. pag. Web. 6 Apr. 2017.

Lun, Erwin van. "Chatbot Do-Much-More, Intelligent Toys | Virtual Assistant Do-Much-More | Virtual Agent Do-Much-More | Chat Bot Do-Much-More | Conversational Agent Do-Much-More | (5029)." *Chatbots.org*. N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

Lunden, Ingrid. "Confirmed: Nuance Has Bought Virtual Assistant Specialist VirtuOz To Ramp Up Its Enterprise Services." *TechCrunch*. N.p., n.d. Web. 14 May 2017.

MacNaught, Stacey. "Global Siri Usage Statistics." *Blogsession*. N.p., 14 Nov. 2012. Web. 14 May 2017.

Maher, Jimmy. "Eliza, Part 3 The Digital Antiquarian." N.p., n.d. Web. 13 May 2017.

"Making Conversational Interfaces Easier to Build." *Google Developers Blog*. N.p., n.d. Web. 8 May 2017.

Matney, Lucas. "Kik Already Has over 6,000 Bots Reaching 300 Million Registered Users." *TechCrunch*. N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

Mauldin, Michael. "ChatterBots, TinyMuds, and the Turing Test: Entering the Loebner Prize Competition." *PROCEEDINGS OF THE TWELFTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*. N.p., 2017. Web. 19 Apr. 2017.

McCorduck, Pamela. *Machines Who Think: a Personal Inquiry into the History and Prospects of Artificial Intelligence*. 25th anniversary update. Natick, Mass: A.K. Peters, 2004. Print.

"Meet Telltale Core Engineer Bruce Wilcox." *Telltale Community*. N.p., 12 Sept. 2011. Web. 25 Apr. 2017.

"Meet XiaoIce, Cortana's Little Sister." N.p., 5 Sept. 2014. Web. 14 May 2017.

"Microsoft Word - Ontology-overview-semantic-desktop-v3-ac.doc - Ontology-overview-semantic-desktop.pdf." Web. 6 Apr. 2017.

"MiQ507Epst5p.indd - FROM_RUSSIA_WITH_LOVE-Epstein-Sci_Am_Mind-Oct-Nov2007.pdf." Web. 6 Apr. 2017.

Oppy, Graham, and David Dowe. "The Turing Test." *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta. Spring 2016. Metaphysics Research Lab, Stanford University, 2016. *Stanford Encyclopedia of Philosophy*. Web. 6 Apr. 2017.

Pandorabots. "Tracking Your Bot's Conversations." *pandorabots-blog*. N.p., 14 Nov. 2014. Web. 13 May 2017.

Panzarino, Matthew, and Fitz Tepper. "Dag Kittlaus, Co-founder of Siri, Will Give the World's First Demo of a Next Gen AI Called Viv at Disrupt NY." *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Perez, Sarah. "Viv, Built By Siri's Creators, Scores $12.5 Million For An AI Technology That Can Teach Itself." *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Pescovitz, David. "Sons and Daughters of HAL Go on Line." *The New York Times* 18 Mar. 1999. *NYTimes.com*. Web. 12 May 2017.

Phillips, Eve M. (Eve Marie). "If It Works, It's Not AI : a Commercial Look at Artificial Intelligence Startups." Thesis. Massachusetts Institute of Technology, 1999. *dspace.mit.edu*. Web. 26 Apr. 2017.

Quora. "Why Chatbots Struggle So Much With Retention." *Forbes*. N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

"Releasing Bot Engine V1." *Wit.AI Blog*. N.p., 28 June 2016. Web. 28 Apr. 2017.

"Review of ``Artificial Intelligence: A General Survey''." N.p., n.d. Web. 26 Apr. 2017.

Richard Wallace. *Be Your Own Botmaster: The Step by Step Guide to Creating, Hosting and Selling Your Own A.I. Chat Bot on Pandorabots.* S.l: ALICE A.I. foundations, Inc., 2003. Print.

Roberts, Jacob. "Thinking Machines: The Search for Artificial Intelligence." *Chemical Heritage Foundation Magazine - Distillations* 14 July 2016. Web. 19 Apr. 2017.

Robino, Giorgio. "How to Build Your First Chatbot Using ChatScript." *freeCodeCamp*. N.p., 2 Aug. 2016. Web. 10 May 2017.

Rodriguez, Ashley. "Microsoft's AI Millennial Chatbot Became a Racist Jerk after Less Than a Day on Twitter." *Quartz*. N.p., n.d. Web. 10 May 2017.

Rosenberg, Scott. "How Kik Predicted The Rise of Chat Bots." *Backchannel*. N.p., 17 June 2016. Web. 12 May 2017.

Russel, Jon. "Kik Introduces Promoted Chats To Let Brands On Its Messaging App And Finally Make Money | TechCrunch." N.p., 19 Apr. 2017. Web. 19 Apr. 2017.

Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 2nd ed. Upper Saddle River, N.J: Prentice Hall/Pearson Education, 2003. Print. Prentice Hall Series in Artificial Intelligence.

Saenz, Aaron. “Cleverbot Chat Engine Is Learning From The Internet To Talk Like A Human.” *Singularity Hub*. N.p., 13 Jan. 2010. Web. 26 Apr. 2017.

---. “We Live in a Jungle of Artificial Intelligence That Will Spawn Sentience.” *Singularity Hub*. N.p., 10 Aug. 2010. Web. 6 Apr. 2017.

“Samsung to Acquire Viv, the Next Generation Artificial Intelligence Platform.” N.p., n.d. Web. 8 May 2017.

“Sassy Woman or Machine? Tech Giants Divided over Digital Assistants.” *Reuters* 24 Nov. 2015. *Reuters*. Web. 19 Apr. 2017.

Saygin, Ayse Pinar, Ilyas Cicekli, and Varol Akman. “Turing Test: 50 Years Later.” *Minds and Machines* 10.4 (2000): 463–518. *link.springer.com*. Web. 10 May 2017.

Schofield, Jack. “Let’s Talk About Sex ... with Robots.” 16 Sept. 2009. *The Guardian*. Web. 19 Apr. 2017.

Schonfeld, Erick. “Silicon Valley Buzz: Apple Paid More Than $200 Million For Siri To Get Into Mobile Search.” *TechCrunch*. N.p., n.d. Web. 8 May 2017.

Shieber, Stuart. “Lessons from a Restricted Turing Test.” *Communications of the Association for Computing Machinery* 37.6 (1994): 70–78. Print.

Shim, Richard. “Handspring to Charge for Support Calls.” *CNET*. N.p., 15 Feb. 2002. Web. 26 Apr. 2017.

“Snapshot.” Web. 29 Apr. 2017.

Solo. *The Time Scale of Artificial Intelligence: Reflections on Social Effects*. N.p., 2017. Web. 6 Apr. 2017.

Solon, Olivia. “Alibaba Founder Jack Ma: AI Will Cause People ‘more Pain Than Happiness’.” *The Guardian* 24 Apr. 2017. *The Guardian*. Web. 25 Apr. 2017.

Soper, Taylor. “Why People in China Love Microsoft’s Xiaoice Virtual Companion, and What It Says About Artificial Intelligence - GeekWire.” N.p., 25 Nov. 2015. Web. 19 Apr. 2017.

Stenberg, Daniel. “History of IRC (Internet Relay Chat).” N.p., 29 Mar. 2011. Web. 12 May 2017.

Swan, Rachel. “Ask Me Anything: Having a Forced Conversation with an Artificial Intelligence.” *SF Weekly*. N.p., n.d. Web. 10 May 2017.

Teuscher, Christof, ed. *Alan Turing: Life and Legacy of a Great Thinker*. Berlin ; New York: Springer, 2004. Print.

Thompson, Clive. “Approximating Life.” *The New York Times* 7 July 2002. *NYTimes.com*. Web. 12

May 2017.

Titlow, John Paul, and John Paul Titlow. “Tired of Waiting For Siri API, Developers Take Matters Into Their Own Hands.” *Fast Company*. N.p., 16 Sept. 2014. Web. 8 May 2017.

Trappl, Robert, Paolo Petta, and Sabine Payr, eds. *Emotions in Humans and Artifacts*. Cambridge, Mass: MIT Press, 2002. Print.

Tripathy, Abinash. “How Chatbots Will Make Voice the Primary User Interface.” *VentureBeat*. N.p., 5 Apr. 2017. Web. 12 May 2017.

Turing, A. M. “Computing Machinery and Intelligence.” *Mind* 59.236 (1950): 433–460. Print.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit*. 20th anniversary ed., 1st MIT Press ed. Cambridge, Mass: MIT Press, 2005. Print.

“Ultra Hal Assistant.” *PCWorld*. N.p., 17 June 2009. Web. 26 Apr. 2017.

Wallace, Richard. “Zipf’s Law.” *A.L.I.C.E. AI Foundation*. N.p., 1996. Web. 13 May 2017.

Wallace, Richard S. “The Anatomy of A.L.I.C.E.” *Parsing the Turing Test*. Ed. Robert Epstein, Gary Roberts, and Grace Beber. Springer Netherlands, 2009. 181–210. *link.springer.com*. Web. 13 May 2017.

Ward, Mark. “Music Site Chatbot Wins AI Loebner Contest.” *BBC News* 16 Sept. 2013. *www.bbc.com*. Web. 10 May 2017.

Weinzenbaum, Joseph. “ELIZA--A Computer Program For the Study of Natural Language Communication Between Man and Machine.” *Communications of the ACM* 9.1 (1966): n. pag. Web. 24 Apr. 2017.

Weizenbaum, Joseph. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: Freeman, 1976. Print.

Wilcox, Bruce. “Beyond Façade: Pattern Matching for Natural Language Applications.” Online Magazine on Video Games. *Gamasutra*. N.p., 18 Apr. 2017. Web. 18 Apr. 2017.

“Wit.ai Raises $3M Seed Round Led by Andreessen Horowitz.” N.p., n.d. Web. 28 Apr. 2017.

Appendix A - List of all winners of the Loebner Prize

Year	Winner	Program
1991	Joseph Weintraub	PC Therapist
1992	Joseph Weintraub	PC Therapist
1993	Joseph Weintraub	PC Therapist
1994	Thomas Whalen	TIPS
1995	Joseph Weintraub	PC Therapist
1996	Jason Hutchens	HeX
1997	David Levy	Converse
1998	Robby Garner	Albert One
1999	Robby Garner	Albert One
2000	Richard Wallace	Artificial Linguistic Internet Computer Entity (A.L.I.C.E.)
2001	Richard Wallace	Artificial Linguistic Internet Computer Entity (A.L.I.C.E.)
2002	Kevin Copple	Ella
2003	Juergen Pirner	Jabberwock
2004	Richard Wallace	Artificial Linguistic Internet Computer Entity (A.L.I.C.E.)

2005	Rollo Carpenter	George (Jabberwacky)
2006	Rollo Carpenter	Joan (Jabberwacky)
2007	Robert Medeksza	Ultra Hal
2008	Fred Roberts	Elbot

Table 1. The list of winners of Loebner Prize from 1991 up to 2016.

Year	**Winner**	**Program**
2010	Bruce Wilcox	Suzette
2011	Bruce Wilcox	Rosette
2012	Mohan Embar	Chip Vivant
2013	Steve Worswick	Mitsuku
2014	Bruce Wilcox	Rose
2015	Bruce Wilcox	Rose
2016	Steve Worswick	Mitsuku

Table 1 (Continued). The list of winners of Loebner Prize from 1991 up to 2016.