

Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University



Galymzhan Momonov

**Machine learning based chatbot for higher
education success**

THESIS

Presented in Partial Fulfillment for the
Degree of Master of Science in Computer Science
(degree code: 7M06102)

Department of Computer Sciences
Faculty of Engineering and Natural Sciences

Supervisor: **PhD Azamat Zhamanov**

Kaskelen, 2022

Suleyman Demirel University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty

Associate Professor, PhD Zhamanov A.



«_30_»_May_2022

Topic of the thesis:

Machine learning based chatbot for higher education success

Thesis submitted as part of the requirements for the award of the MSc in
“7M06102 - Computer Science”, SDU, 2020-2022

Head of Department

Associate Professor, PhD Cemil Turan

Academic Supervisor

Associate Professor, PhD Azamat Zhamanov

Master's student

Galymzhan Momonov

Kaskelen, 2022

Abstract

In this dissertation, a machine learning based chatbot is evaluated for its effectiveness in supporting undergraduate Compute Science students. Using Rasa open-source framework and behavioral science theories, a chatbot was created to support undergraduate students taking the Computer Networks 1 course at the School of Computer Science in a private selective university in Kazakhstan. A randomized experiment with 103 undergraduates majoring in Computer Science showed that the chatbot intervention had no effect on students' performance during the semester and in the final exam. Neither did the chatbot affect student attendance. At the same time, the chatbot increased course satisfaction by about 3 points. The results suggest that the chatbot intervention does improve students' satisfaction with the course while not affecting *academic outcomes*.

Аңдатпа

Бұл диссертацияда машиналық оқытуға негізделген чатбот оның компьютерлік ғылым бакалавриат студенттеріне қолдау көрсетудегі тиімділігі үшін бағаланады. Rasa ашық бастапқы негізін және мінсіз-құлық ғылымының теорияларын пайдалана отырып, Қазақстанның жеке таңдаулы университетінде Информатика мектебінде Компьютерлік желілер 1 курсының бакалавриат студенттеріне қолдау көрсету үшін чат-бот құрылды. Информатика мамандығы бойынша оқытын 103 магистрантпен рандомизацияланған эксперимент чатботтың араласуы студенттердің семестрдегі және қорытынды емтихандағы үлгеріміне әсер етпейтінін көрсетті. Чат-бот студенттердің сабаққа қатысуына да әсер еткен жоқ. Сонымен бірге чатбот курстан қанағаттануды шамамен 3 ұпайға арттырды. Нәтижелер чатбот араласуы академиялық нәтижелерге әсер етпей, студенттердің курсқа қанағаттануын жақсартатынын көрсетеді.

Аннотация

В этой диссертации чат-бот на основе машинного обучения оценивается с точки зрения его эффективности в поддержке студентов бакалавриата по информатике. Используя платформу с открытым исходным кодом Rasa и теории поведенческих наук, был создан чат-бот для поддержки студентов бакалавриата, изучающих курс «Компьютерные сети 1» в Школе компьютерных наук в частном селективном университете в Казахстане. Рандомизированный эксперимент со 103 студентами, специализирующимися в области компьютерных наук, показал, что вмешательство чат-бота не повлияло на успеваемость студентов в течение семестра и на выпускном экзамене. Чат-бот также не повлиял на посещаемость студентов. В то же время чат-бот увеличил удовлетворенность курсом примерно на 3 балла. Результаты показывают, что вмешательство чат-бота действительно повышает удовлетворенность студентов курсом, не влияя на академические результаты.

Acknowledgements

I would like to express my gratitude to my advisor Dr. Azamat Zhamanov for his guidance. I am also grateful to all the professors at the School of Computer Science at SDU for their teaching and support. I thank my fellow students for sharing this master in CS journey with me.

I also thank my family for being with me and supporting me during my master studies.

Contents

1	Introduction	8
2	Literature review	10
2.1	The history of chatbots	10
2.2	Artificially intelligent conversational chatbots: main types, programming issues and possible solutions	14
2.2.1	Main types of chatbots	15
2.2.2	AI chatbots: Programming issues and possible solutions	16
2.2.3	Preventing usability issues in chatbot programming . .	17
2.3	Review of research on using artificially intelligent chatbots in higher education	31
3	Chatbot design	35
3.1	Theoretical foundations for the chatbot	35
3.2	Technical details of the chatbot	36
3.3	Knowledge database for the chatbot	38
4	Data and methodology	43
4.1	Data	43
4.2	Randomized controlled trial	45
4.2.1	Randomization procedure and balance tests	46
4.2.2	Statistical estimation	46
5	Results	48
5.1	Student-chatbot interaction	48
5.2	Chatbot's effects on students' outcomes	48

6 Conclusion

52

References

55

1. Introduction

Recent advances in artificial intelligence (AI) technologies boosted the widespread adoption of AI conversational chatbots in many areas of human activity. AI chatbots have an enormous potential to be fruitfully utilized in higher education due to their capacity to flexibly support large groups of students at a low cost [1, 2, 3, 4].

The aim of this dissertation study is to create and evaluate the effectiveness of a machine learning based chatbot for use in higher education to support students taking Computer Science courses.

The significance of the study is in that machine learning based chatbots are increasingly used in education in developed countries, but in Kazakhstan and other developing countries such chatbots are essentially non-existent. This chatbot will contribute to the development of machine learning based chatbots and machine learning based educational technology in general in Kazakhstan and beyond.

Using Rasa open-source framework, a chatbot is created to support undergraduate students taking the Computer Networks 1 course at the School of Computer Science in a private selective university in Kazakhstan. The chatbot was created using behavioral science theories that suggest that reminders, information, and nudges influence decision-making and support individuals in making decisions that increase welfare.

To evaluate the chatbot's effect on students' outcomes, a randomized experiment is conducted with 103 undergraduates majoring in Computer Science. Administrative data from the university database and course instructors is used.

The rest of the dissertation is structured as follows: chapter 2 presents the literature review; chapter 3 describes the chatbot design covering the theoretical foundations of the chatbot, the technical aspects of how the chatbot was

created, and the knowledge base of the chatbot; chapter 4 provides an overview of the data and methodology; chapter 5 presents the results, and chapter 6 concludes and discusses study limitations, suggestions for chatbot developers and recommendations for future research.

2. Literature review

2.1 The history of chatbots

Chatbots, also called conversational agents, are computer programs that allow the computer to “talk” with the user. It is therefore a particular type of human-machine interface, capable of responding to requests made in writing or orally.

The history of the chatbot is therefore to be linked to that of the interactions between human beings and their creations. And if it is generally believed that the first conversational agents appeared in the 1960s, their origin can go back much earlier.

The beginnings of chatbots

When we talk about robotics in the broad sense, we cannot imagine finding traces of it before the industrial revolution. However, as early as the 18th century, a man began work that appeared to be the first steps towards chatbots. In 1780, a Frenchman Abbé Mical took part in a science competition organized by a Russian academy. His invention: two copper talking heads, capable of pronouncing four sentences. There was then no real interaction with the public, but the creations clumsily reproduced the human voice, giving the impression of conversing [5].

Westinghouse humanoid robots

Then, Europe experienced a great period of industrialization, bringing with it the beginnings of robotics. In this area, an American company has partic-

ularly distinguished itself: Westinghouse Electric and Manufacturing Corporation. His first creation in this field was called Televox, which appeared in 1927 [6]. This funny humanoid was capable of performing simple actions, depending on the signals received. He could accept a phone call, turn on a fan, etc.

Then, in 1937, Westinghouse launched a new robot: Elektro. Always with an appearance imitating a man, however, it presented new features. He could thus react to the voice of the user and walk, move his arms, or even smoke a cigarette. But above all, he could "speak", with a vocabulary of about 700 words. Elektro thus marked a decisive step towards modern chatbots [6].

The Turing test and the birth of artificial intelligence

However, the real turning point in the history of chatbots lies in the appearance of the famous "Turing test".

Alan Turing was an English mathematician, born in 1912. After brilliant studies, he devoted his energy to scientific research, particularly in the field of logic. It is to him that we owe the "Turing machine", which laid the foundations of the modern computer, as early as 1937 (the same year as the invention of Elektro). He is also famous for helping the British Army decrypt messages from the Enigma machine, used by Germany during World War II [7].

A few years later, he began to take an interest in artificial intelligence. He then proposes a new experience, which he calls "Imitation game", which will later take the name of "Turing test". This is an experiment to determine the ability of a machine to impersonate a human being. Its principle consists in connecting an individual with two interlocutors: a machine and a human being. The examiner is led to ask questions, the answers to which are provided by the other two actors. He must then determine which of the two is human. If it does not succeed, the machine is considered to have passed the Turing test, by reproducing human behavior. Subsequently, over the years, chatbots have been perfected, embedding more and more artificial intelligence elaborated. With the hope of successfully passing the Turing test. The most famous of them are called ELIZA, ALICE, or even Sophia. But only one would have passed the Turing test.

The history of chatbots has been punctuated by several more or less suc-

successful creations. And it would be impossible to list them all here. But here are some examples of conversational agents that, in their own way, have made an impression.

ELIZA, the first chatbot

The founding act of the birth of chatbots is commonly dated 1966. That year, Joseph Weizenbaum, a German-American professor at MIT, created ELIZA [8]. It is a computer program (and no longer a humanoid robot), now considered the first true conversational agent.

Its principle: imitate the behavior of a psychotherapist, using a dialog box. Patients are thus encouraged to confide in writing, and ELIZA responds to them. So much so that some individuals became attached to the machine, which could be irrefutable proof of a successful Turing test.

Nevertheless, ELIZA is then content to locate a key word in the patient's sentence, to then formulate a question. For example, if the user talks about sports, the robot can answer: "Since when do you like sports? A method that therefore had obvious limitations.

ALICE, reference for chatbots

However, ELIZA has inspired many scientists, even decades later. Including the American computer scientist Richard Wallace, creator of ALICE in 1995 [8].

The developer then takes up Joseph Weizenbaum's idea, perfecting the model. ALICE is thus able to identify keywords in user comments, but also relies on more sophisticated rules. In this way, the program is able to dialogue with a human being, without being content with only questions.

ALICE is still the benchmark for chatbots today. The conversational agent has also been awarded three times (in 2000, 2001 and 2004) with a Loebner prize, which highlights the programs that best meet the criteria of the Turing test. Nevertheless, ALICE never fully succeeded. Unlike the following chatbot.

Eugene Goostman, first chatbot to pass the Turing test

Eugene Goostman is a conversational agent, developed by three Russian computer scientists in 2001. Equipped with artificial intelligence, it is supposed to reproduce the behavior of a 13-year-old Ukrainian boy. In 2014, researchers at the University of Reading decided to confront it with the Turing test [9].

The experiment takes place with 30 examiners, who must interact with the chatbot for 5 minutes. The organizers then consider that the test will be successful if the proportion of individuals fooled exceeds 30%. A result that will eventually reach 33%. Eugene Goostman would therefore be the first program to pass the Turing test [10].

A conclusion greeted with skepticism by the scientific community [11]. The reservations relate firstly to the time limit imposed for the conversation. But the media also question the supposed performance of artificial intelligence, after other inconclusive tests.

Sophia

Sophia is a different chatbot from the previous ones, while also being inspired by ELIZA. It is indeed a humanoid robot, like Televox or Elektro, but more realistic [12].

But the remarkable capabilities of this modern chatbot lie in its ability to interact. Using voice recognition and artificial intelligence technologies, Sophia can answer certain questions and hold a conversation on a few topics, such as the weather. Additionally, the robot can adopt a series of facial expressions, mimicking human behavior.

However, Sophia owes her notoriety above all to Saudi Arabia. In 2017, the country granted the chatbot Saudi nationality, a first for a robot [13]. A decision that obviously did not fail to arouse controversy. Several observers have notably asked whether Saudi nationality allowed Sophia to vote or marry.

Siri, Alexa, Google Assistant... The advent of voice assistants

Finally, it is impossible to conclude without mentioning the new form of chatbots: voice assistants. These programs, capable of understanding oral

commands and responding via voice synthesis, have experienced real growth in recent years. Thanks, in particular, to the giants of technology. It was Apple who first drew, in 2011, with Siri, present in particular in its iPhone. Then others followed suit: Amazon with Alexa, Google with Google Assistant, Microsoft with Cortana. Conversational agents who have become true virtual assistants. And which have allowed chatbots to enter the private sphere, beyond their professional applications.

2.2 Artificially intelligent conversational chatbots: main types, programming issues and possible solutions

Artificially intelligent (AI) chatbots are software which designed for communication and help users obtain information or advice via text or voice messages. Advances in the field of artificial intelligence have contributed to the wide adoption of chatbots. In particular, recent developments in machine learning, neural networks, and natural language processing (NLP) made chatbots smarter and more useful in many ways [14]. Thanks to these computer science advances, modern chatbots allow businesses to improve customer experiences at a relatively low cost. They serve users and reduce human employees' workload by effectively dealing with repetitive and frequently asked user questions or requests. Even when chatbots send user queries to human employees, human employee's time and energy are used more efficiently as they only deal with a few cases where their direct involvement is necessary. In other words, AI chatbots have become popular because they allow businesses to serve their clients well while also reducing the costs. AI chatbots are used in areas like education [15], medicine [16], banking [17], and many others.

Despite such progress in making chatbots useful and the wide adoption of chatbots, there are still programming issues associated with chatbot development. This section reviews recent literature in the field of chatbot programming and development and summarizes the key programming issues as well as ways to address them.

2.2.1 Main types of chatbots

Overall, chatbots have different degrees of complexity and can be classified by different criteria and architecture [18]. One way to group chatbots is into three main types based on their historical evolution as follows: 1. script-based chatbots, 2. database-based chatbots, and 3. natural language processing-based chatbots.

First, script-based chatbots (also referred to as button based chatbots) are designed to function in a predefined linear way. They have the simplest type of dialog system because developers create rigid tree-like structure where users choose between options. When interacting with such chatbots, users mostly don't type anything, they just press the provided buttons to get what they need or want. Such chatbots are often used in various kinds of online booking or delivery services. This type of chatbots is easy to create and use.

The second type of chatbots are database-based chatbots (also referred to as rule based chatbots). They rely on specific keywords from the user messages. Depending on what keyword is identified by the chatbot, it moves to the next stage of the conversation by using a response from the database connected to that specific keyword. The design and complexity of such chatbots might be very simple or more cumbersome, however the basic idea is that the chatbot retrieves the keyword from the user's message and with an "if-then" statement matches it with some response from the database or moves to another condition connected to keyword. Such chatbots are efficient when a simple form of a dialog system is needed, for example for answering frequently asked questions or other similar services.

The third type of chatbots are natural language processing based chatbots (also referred as AI based chatbots). These are more complex and more human like dialog systems which involve NLP, machine learning and deep learning models for different purposes. This type of chatbots generally consist of three parts which is the natural language understanding part, the dialog management part and the response generation part [14].

2.2.2 AI chatbots: Programming issues and possible solutions

Based on the review of the literature on chatbot development, the three major issues challenging programmers were identified.

1. Accounting for context. Human communication always takes place in some context. Even when using a chatbot for a specific purpose, for example, for getting information about student loan services, users interact with the chatbot in highly contextual ways. They have their unique backgrounds, circumstances, needs, language capabilities, and demands.

Solution: Natural language processing algorithms. One promising (though not magical) solution is to use natural language processing algorithms to account for context [19]. By extracting meaning from spoken or written messages, natural language processing algorithms can handle identity-specific encodings (pitch, tone, etc.), grammatical mistakes, self-corrections, and other aspects of speech that altogether create the context [20, 21].

2. Adapting to evolving conditions. Even after a chatbot has been built and piloted, there are things that change in the world or in the things users expect from the chatbot. New types of dialogs may need to be added into the dialog management part or new responses may have to be added in the response generation part.

Solution: Machine learning algorithms and human support. The potential solution here is the utilization of machine learning algorithms that feed on the responses handled by human employees. Specifically, if an unexpected question was received by the chatbot and the chatbot couldn't respond, it could send the question to the human employee's email. Later, when the human employee responds to the question, that email enters into the chatbot's response generation part and the dialog management part gets enhanced too. In sum, a knowledge database, a dialogue history corpus, and an external data source providing the chatbot with "intelligence" supported by machine learning algorithms address this programming issue [22]. Human-annotated corpora, discussion forums, and emails conversations are commonly used for these purposes [23]. Open-source conversational AI frameworks like Deep Pavlov hold much promise in tackling this issue in the future generations of chatbots.

2.2.3 Preventing usability issues in chatbot programming

Very often, conversations with chatbots fail for a variety of reasons. However, in most cases, the main reason is usability related programming and design issues. At this time, there is a considerable amount of literature describing various chatbots and we have built a considerable amount of expertise in developing chatbots in general. So this section summarizes this literature and aims to provide an overview of common usability issues that can be overcome by programmers relatively easily.

Here are the most common failures encountered when interacting with a chatbot and how to avoid them.

1. Deadlocks in Conversations

AI chatbots are still relatively rare and script-based chatbots tend to be more common due to their simplicity and general availability. So it is limitations of scripts used by chatbot developers that create deadlock conversations when the chatbot doesn't really have anything to say in response to users. Such an approach works relatively well for chatbots responsible for simple tasks, but sometimes companies borrow simple decision-making voices and schemas.

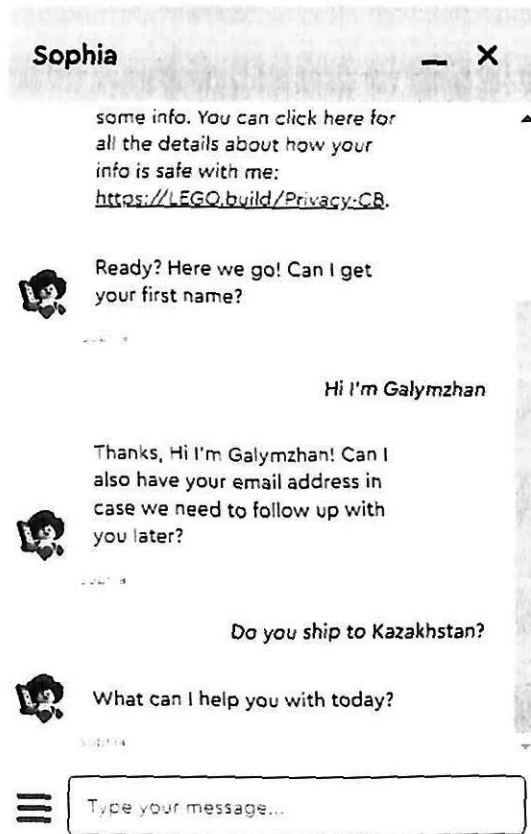


Figure 2.1: Screenshot of a chatbot not understanding a question (<https://www.lego.com/en-us/service>)

If programmers neglect to infuse the bot with some sort of personality or program it to deal with unplanned and unusual conversations, they risk boring the users.

Overly simplistic programming is another common pitfall that should be avoided. While tempting and easy for the programmer, such an approach is wrong in assuming that all or most of users will actually talk with the chatbot the way that was expected, especially if this way is simple and too straightforward. More pointed questions that are poorly handled are excusable, but if the chatbot fails to pick up variations of basic commands and phrases, it gets awkward.



Figure 2.2: Screenshot of a conversation with a robot that does not understand words (<https://shyn.kz/>)

One way to avoid such issues is to brainstorm several conversation flows. The best way to do so is to pilot a chatbot and then use the data generated through the pilot to develop the better knowledge base and conversation scenarios. Qualitative interviews or focus-group style research data is also a very useful approach as they generate a lot of real-life data [22] and help programmers to find out how target users actually talk, what language they use, etc.

This way the decision tree of a chatbot will be more nuanced and complex and responsive to a variety of possible conversations. The chatbot doesn't have to really recognize a large database of words, but by creatively deflecting irrelevant requests, one can prevent conversations from quickly reaching a dead end.

Today

Hi are you a chatbot?

7 minutes ago (not read)

...

The team will get back to you soon.
Please introduce yourself:

Hi, my name is...

...

Please introduce yourself

Hi, my name is...

Figure 2.3: Screenshot of conversation with a chatbot (<https://www.kimep.kz/en/>)

Here in Figure 2.3, the robot does not understand a basic question. There are programming solutions that can help developers overcome such issues. That is, the chatbot market is so large now that one can buy a package that will address most of the common conversation types. However, they tend to be in the English language, so if one needs a Kazakh- or Russian-speaking chatbot, the problem still needs to be addressed for each particular chatbot.

2. Information overload

What's more annoying than a friend briefly replying "OK. Cool" to a long message just sent to them? A chatbot who begins to tell us about his life in a message, when we haven't asked him anything.

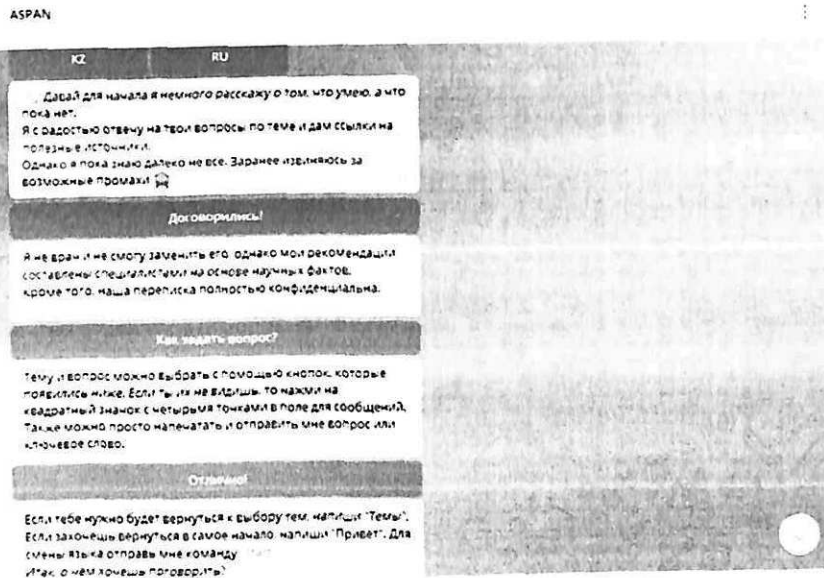


Figure 2.4: Screenshot of a large amount of chatbot welcome text (<https://uyatemes.kz/aspan>)

Instead of engaging in casual conversation with potential customers, chatbots like these overload users with tons of text. Such interaction is far from engaging. Another example, in Figure 5 below, uses too many emojis and makes the chatbot less than user-friendly.

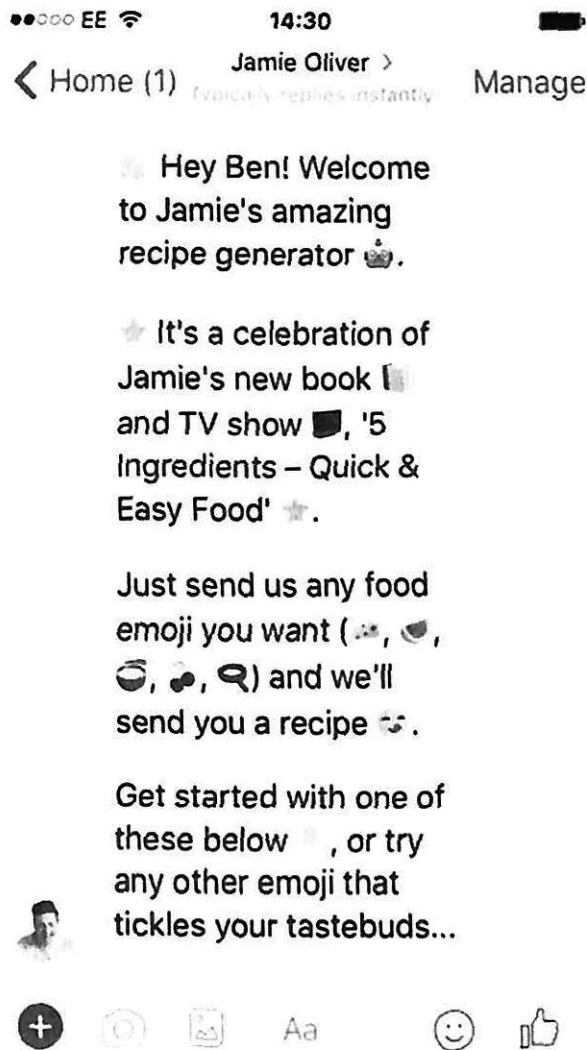


Figure 2.5: Screenshot of a large amount of chatbot welcome text (Screenshot of long text with emoji overload. Source: <https://d2vrvpw63099lz.cloudfront.net/chatbot-fails/jamie-oliver.jpg>).

Buttons and links are a useful tool in building the chatbot. However, using them too much may create a highly artificial and unfriendly chatbot experience.

Therefore, it is important to use buttons and links while integrating them into a natural-sounding language, typically used online. Smaller and shorter texts messages with only a few of essential links and buttons create a chatbot conversation experience that is the closest to real human conversations [24].

Appropriate uses of buttons are for selecting scheduling options or maybe other options essential for the service or product that the chatbot supports or provides. In Figure 6 below, the Seattle Balloon chatbot uses buttons in a user-friendly way by providing clear options for choosing the occasion for which the user is planning to use the balloon. Emojis used sparingly liven up the buttons

list. The geometry of buttons also makes it interesting to look at them. If the buttons were provided in an overly symmetric manner, the list of buttons would look boring and unengaging. This chatbot puts the user in control of his or her experience while swiftly guiding the conversation and making it efficient.



Figure 2.6: Screenshot of a chatbot showing buttons to help schedule meetings (www.seattleballooning.com).

3. Under- or over-developed personalities

Chatbots are of course robots created digitally, but they can and should sound more human, natural and real.

A tone that is too dry or too distant can break the entire communication strategy designed for a company. If websites are written in a lively and interesting language but the chatbot speaks in dry and boring language, such a combination only can put the company trying to communicate with users at a disadvantage.

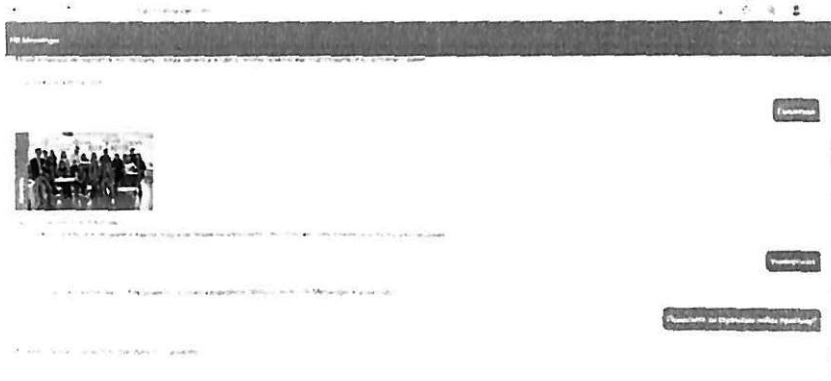


Figure 2.7: Screenshot of a conversation with a bot that is script-only (<https://chat.lrmessenger.com/>)

However, doing the opposite is not necessarily better. Chatbots that are too talkative and too comical can have the same negative effect on your customers.

Many chatbot perception surveys revealed that customers would like chatbots to crack jokes and express their emotions [25]. Nevertheless, it is important to keep in mind that if users say they like humour, it doesn't necessarily mean they will like any humour. A joke that will make a splash in Kazakhstan, may be misunderstood by those from India.

But a joke and a pun here and there probably won't hurt anyone, and sometimes the developers know how to use a lot of humor.

This excessive use of humor lengthens the response or resolution time for customer requests. Which isn't so funny anymore.

A solution for anonymous chatbots is to try to match their personality to the way your agents or target customers talk. Gather the phrases and words that agents and customers use most often and try to include them in your script. The more empathy and sympathy chatbots show, the better.

When writing the script, it is a good idea to create a profile for the chatbot. List their likes, dislikes, goals, and issues to guide you through the text design process.

This chatbot is an example of how a bot can mimic humans well by using enhanced personality features and humor.

4. No added value

In 2016, Business Insider reported that 80% of companies wanted a chatbot by 2020 [26]. To be more precise, this figure is based on 800 responses from business leaders from France, the Netherlands, South Africa and from the United Kingdom. It's not "every business", but it does bode well for years to come.

Yet with so many chatbots being used all over the Internet, many chatbots disappoint as they do not provide any added value but seem to be created just to make sure that website has a chatbot.

Figure 2.10 below shows an example of such a chatbot which responds similarly to basically any question posed. While not completely useless – as the chatbot does provide some information, or, rather, all information it has – the chatbot is far from doing anything value-added to what the website already does.

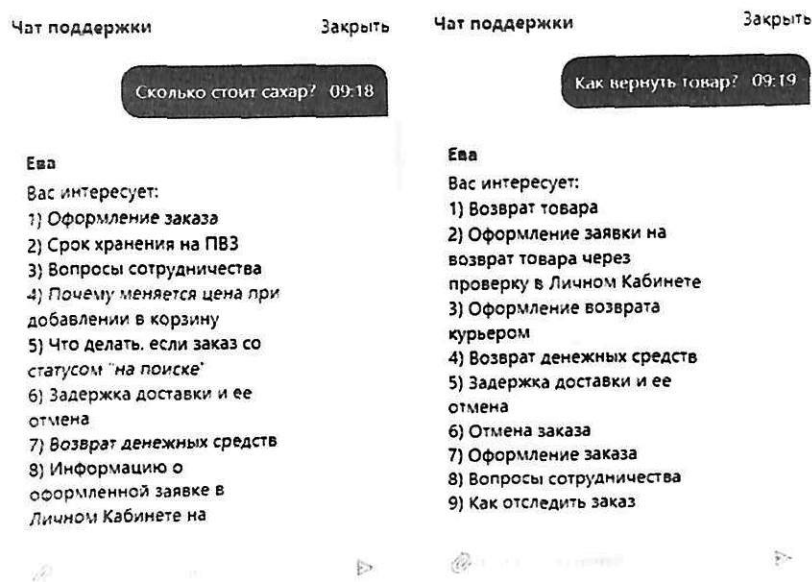


Figure 2.10: Screenshot of a florist robot repeating the same phrase (<https://kz.wildberries.ru/>)

A common flaw in these bots: the lack of strategy and objectives. Chatbots

are attractive because they are able to take on small, mundane tasks that once overwhelmed employees.

But this modern, high-tech representation of a digital assistant is what companies seem to concentrate too much. These chatbots may have a role to play, but their performance and added value are not enough.

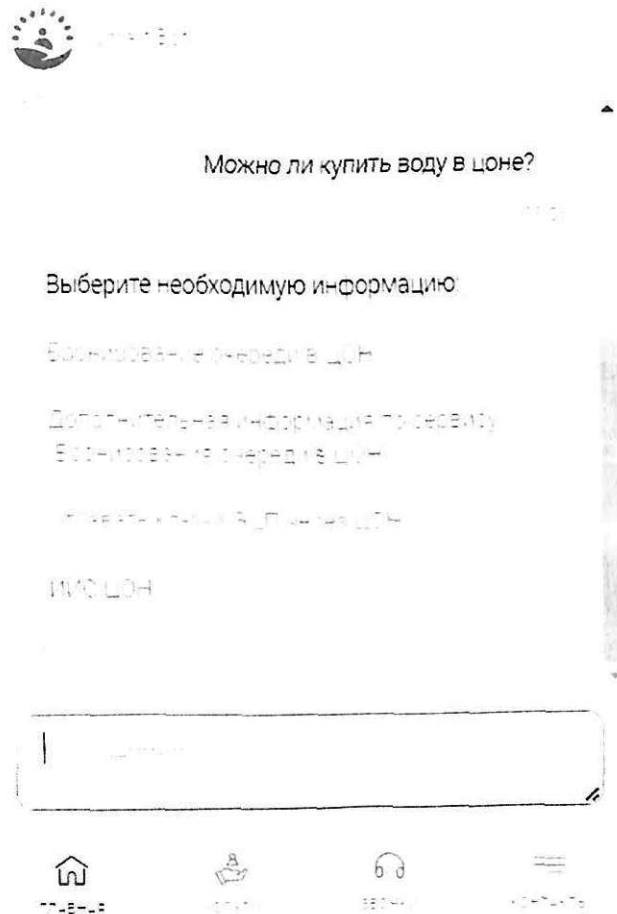


Figure 2.11: Screenshot of E-gov's virtual agent question about misunderstanding and suggesting irrelevant FAQ articles (www.egov.kz)

A solution here is to determine first if you really need a chatbot. If he is in a position to be able to tackle certain problems that your agents, your customers or your company are facing, clearly define his role and the missions he can realistically accomplish.

For example, if you need to set up an FAQ type chatbot, make sure that it is well versed in the questions customers frequently ask. PayPal has an entire page dedicated to email scams, but its chatbot doesn't recognize keywords.

If one assigns a single task to the chatbot, it is important to make sure it

completes it well. Regularly update the chatbot, so that it can identify and handle all new issues and requests. And make sure you're using a chatbot builder that's right for you. The OMQ bot and Landbot.io are very useful and simple to use for creating FAQ chatbots.

And above all, it is necessary to make sure the chatbot is connected to the live chat solution, so that any unresolved requests can be routed. This shows your customers that their queries are taken care of and they haven't just wasted their time.

5. Dysfunctions

This is the biggest failure of all. When scripts fail and programs go wrong, chatbots lose their way.

New chatbots are prone to errors in their early stages, much to the chagrin of customers. You risk ruining your brand image and frustrating customers if you deploy a bot that is not yet ready to take on the real world.

In high-stress situations, like bank account fraud or canceled flights, a malfunctioning chatbot is the last thing a customer wants to deal with.

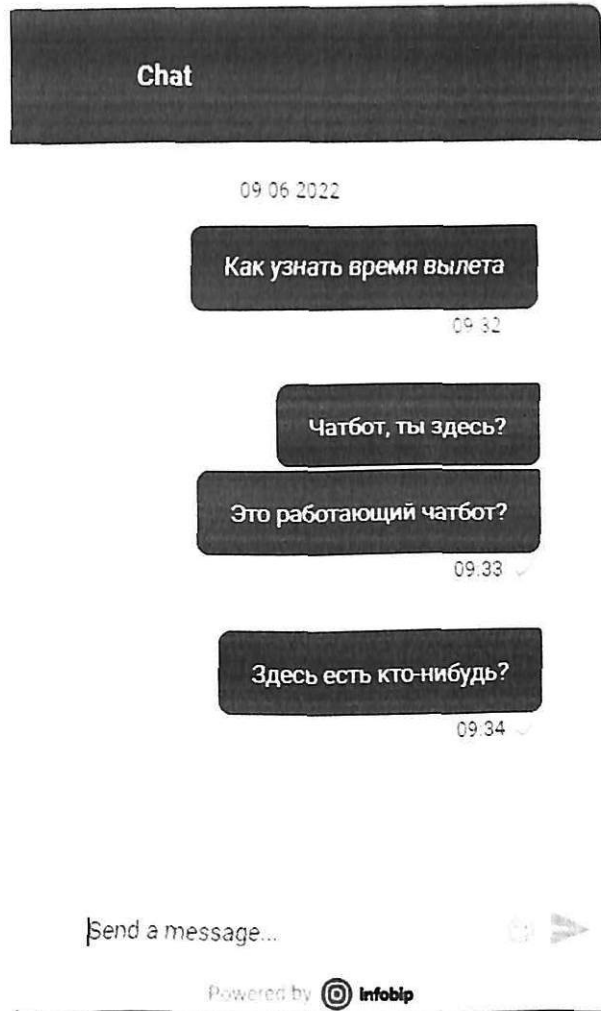


Figure 2.12: Screenshot of useless chat with an airline bot (<https://flyarystan.com/en/contact>)

The only solution here is to test, test and test the bot again and again before going live. It is recommended to ask people outside your development team or a few customers to test your bot and give you feedback.

In the event of a malfunction with the chatbot, lighten the air by getting the chatbot to acknowledge its flaws (“Sorry, I got lost. Let’s try again”) and offer options to move forward , For example:

Sometimes customers ask bots silly questions to test the scope of the conversation. When using a rules-based conversational structure for the chatbot, it is important to create a rule that helps the chatbot handle these kinds of diversions.

These can be keyword commands for the customer to use, forward-to-agent options, or button options.

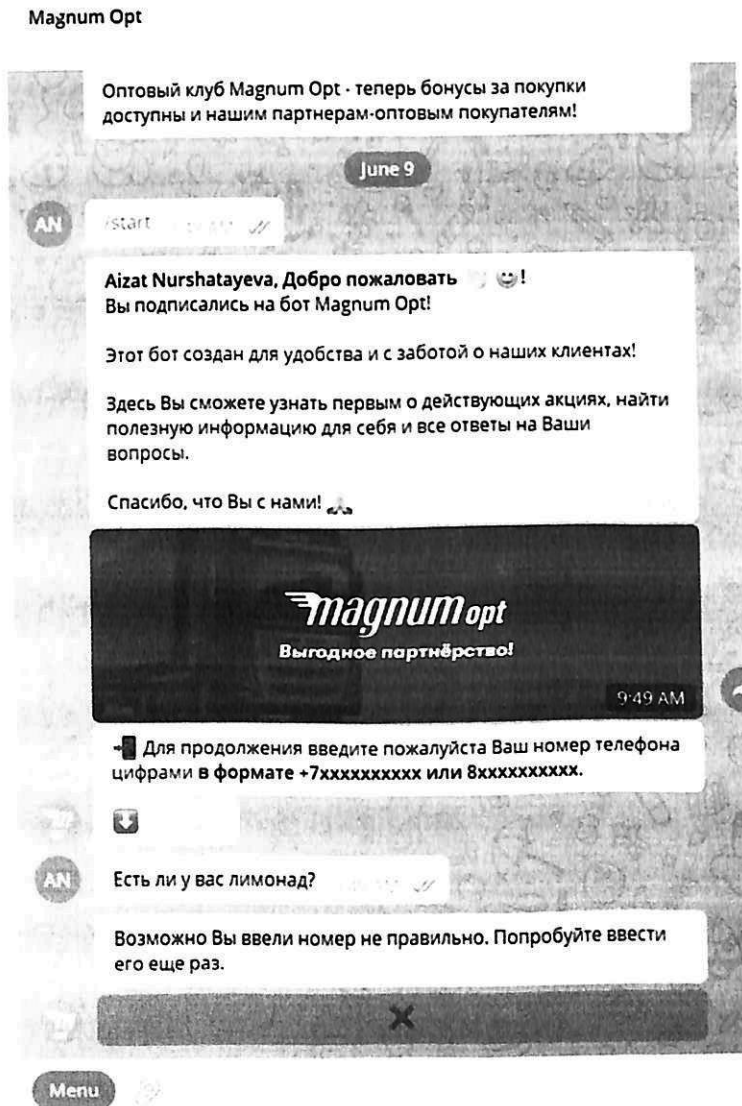


Figure 2.13: Screenshot of a conversation with a chatbot who knows how to stay on topic (Telegram Contact @magnumopt_bot)

If the chatbot is prepared to handle a few hitches here and there, it will be less likely to break down and compromise the smooth running of the exchange. If the chatbot consistently fails, it's best to try something else

In some cases, chatbots are not the solution to consider. Some of the biggest chatbot failures the internet has seen, like Tay from Microsoft, have come from bots that had a very narrow purpose or function.

2.3 Review of research on using artificially intelligent chatbots in higher education

Research on AI chatbots in higher education grew in the recent years yet is scattered across various disciplines and research methodologies which makes it difficult to systematically synthesize [24, 27]. Several reviews of the use of chatbots in education have been published recently. Winkler and Sollner [28] examine the applications of chatbots to improve learning processes and outcomes. Perez et al. [29] focus on chatbots' potential to improve learning and student services. Okonkwo and Ade-Ibijola [30] review chatbots' benefits and challenges in education. Wollny et al.[31] explore studies of chatbots focusing on applications, designs, evaluation, and educational effects.

While the reviews cited above synthesize an impressive amount of evidence, they tend to focus on secondary education and often adopt broad interpretations of chatbots. Arguably, higher education students have somewhat different needs than secondary education students, therefore, chatbot uses in higher education have their own unique features. Furthermore, most of the literature on chatbots in education describes various chatbots and their design features without any empirical research on chatbots' actual use. This review aims to systematically examine the burgeoning empirical literature on AI chatbots in higher education.

This section of the review is guided by the following research questions: 1) What are chatbots used for in higher education? 2) How do chatbots affect academic and non-academic outcomes of students in higher education?

The search was conducted in the following databases relevant for computer science, education research, and social sciences: Association for Computing Machinery (ACM) digital library, Institute of Electrical and Electronics Engineers (IEEE), Computer Society Digital Library (CSDL), Education Resources Information Center (ERIC), Scopus, and the National Bureau of Economic Research (NBER).

Since chatbots are relatively recent and the literature on chatbots in higher education is in its emerging stage, the main search term we used was "chatbot". To make sure the search is comprehensive, the search terms list was developed in consideration of possible studies that examine the use of chatbots to support a) students preparing to enter higher education, b) students already enrolled in

institutions of higher education, and c) students preparing for the labor market or for subsequent stages of their educational trajectories. The search terms therefore included “chatbot for students”, “chatbot for education”, “chatbot for K-12”, “chatbot for university”, “chatbot for learning”, “chatbot for graduate education”, and “chatbot for the job market”.

The papers included in the review had to meet the following criteria: 1) demonstrate empirical evidence involving human participants on the use of chatbots in or for higher education, 2) be published in peer-reviewed journals, 3) be published in 2010-2021, 4) be published in English. After applying these inclusion criteria, our final analytic sample included 6 papers.

Table 2.1 below presents an overview of the papers identified through the search and included in the final analytic sample.

Ayedoun et al. [32] developed and evaluated a prototype of a chatbot encouraging willingness to communicate among university students studying English as a foreign language in Japan. Their chatbot prototype simulated a conversation at a restaurant. A total of 5 undergraduate and graduate students at a Japanese university interacted with the chatbot prototype. Before and after participants’ interaction with the chatbot, the researchers measured their willingness to communicate, interactions with the chatbot, nervousness, degree of immersion, and fluency of conversation. The findings showed that confidence and desire for communication increased and nervousness decreased after interacting with the chatbot. Overall, the authors conclude that the chatbot increased participants’ willingness to communicate.

Study	Research design	Country	Sample size
Ayedoun et al. [32]	Pre-/post-analysis	Japan	5
Ciechanowski et al. [33]	RCT	Poland	31
Fryer et al. [34]	Pre-/post-analysis	Japan	122
Lin & Chang [35]	Mixed methods	Canada	28
Page & Gehlbach [36]	RCT	USA	7,489
Nurshatayeva et al. [37]	RCT	USA	4,442

Table 2.1: Studies included in the review

Ciechanowski et al. [33] explored the psychophysiological reactions to chatbots and willingness to collaborate with a chatbot. Psychophysiological re-

actions were measured using electromyography, respirometer, electrocardiography, and electrodermal activity. Willingness to collaborate with chatbots was examined using the theory of planned behavior survey, the social presence survey, and the anthropomorphism scale. Two chatbots were designed for the Kozminski Academy to assist new students with enrollment related issues. One chatbot was text-based, the other chatbot had an avatar reading out the responses in addition to presenting them as text on screen. Participants (n=31) were randomly assigned to either of these chatbots. The overall conclusion of this study is that participants enjoyed their interactions with both chatbots. However, they were more positive about the text-based chatbot.

Fryer et al. [34] examine the effects of using chatbots on task and course interest of students studying English as a foreign language at a private university in Japan. A total of 122 students from various majors were randomly assigned either to interacting with a chatbot or a human partner for one week. After the first week, the treatments were switched, that is, the chatbot group started chatting with a human and vice versa. The findings showed that the chatbot decreased students' task interest. Structural equation modeling showed that task interest with a human partner only contributed to increasing course interest. Notably, students' task interest was high at the start of interactions with the chatbot, but decreased sharply after the first task suggesting that chatbot's novelty effect faded quickly and did not contribute to sustained interest in learning.

Lin Chang [35] present a chatbot aimed at improving university students' writing skills. Their mixed methods study with Canadian undergraduate students showed that the chatbot improved students' essay outline performance and was helpful in facilitating the communication between instructors and students.

Page Gehlbach [36] conducted a randomized controlled trial with 7,489 students entering Georgia State University to estimate the effect of an AI chatbot on enrollment and academic outcomes in the first semester of university studies. The chatbot aimed to support students in transitioning from high school to college and enrolling in Georgia State University. The experiment results showed that the chatbot increased success with pre-enrollment tasks and raised enrollment by 3.3 percentage points among students who expressed early

commitment to study at Georgia State University.

Nurshatayeva et al. [37] replicated the study by [36] at East Carolina University. A total of 4,442 students entering East Carolina University were randomly assigned to either receiving the chatbot support or to business-as-usual supports provided by the university. The results suggest that the chatbot had no effect on overall enrollment rates but increased enrollment among first-generation students by 3 percentage points.

The present review showed that research papers evaluating chatbots in higher education focus on chatbots teaching English as a foreign language, developing writing skills, and supporting students in their transition from high school to university.

Overall, based on the reviewed studies, one can conclude that chatbots positively affect academic and non-academic outcomes of students in higher education. Chatbots assist students in improving their English language skills and other skills like writing. Chatbots also have the capacity to support students in the completion of various administrative tasks and nudge them to enroll on time in a university of their choice.

Notably, the reviewed papers tend to use rigorous research designs to evaluate the effects of the chatbots on the outcomes of interest. Randomized controlled trials were common perhaps reflecting the considerable research methods capacity of the researchers studying chatbots in education. Sample sizes in the two U.S.-based RCT studies are quite large reflecting the scale-up potential of chatbots in contrast to other types of educational technologies. Overall, the treatment effects of chatbots yield themselves particularly well to evaluations involving advanced quantitative research designs. The area of chatbots for higher education is ripe for more research.

3. Chatbot design

3.1 Theoretical foundations for the chatbot

The literature review above showed that the most advanced chatbots for supporting students in higher education are created using behavioral science [37, 38, 39]. Reminders, information provision, and support with completion of specific steps are the core nudges used in the literature [40]. Specifically, the chatbots using nudges or behaviorally informed features that help users make better decisions for their welfare are the most promising to improve students' outcomes at the tertiary education level [41].

Behaviorally informed interventions, including chatbots, are not only the cutting-edge solution offered by behavioral research, these interventions have also been evaluated rigorously and they work. Randomized controlled trials using behavioral interventions showed that such interventions are effective in improving many outcomes of interest, such as health-related and savings decisions [41, 42]. Furthermore, in higher education specifically, the growing evidence base shows the promise of nudges and other behaviorally informed information interventions in improving students' academic and non-academic outcomes [39, 43]. Given the theoretical rigor and promising practical effects, in this dissertation research, the chatbot was created using these behavioral insights.

Notably, behaviorally informed chatbots have been studied in a series of cutting-edge social science research papers and are only emerging in the field of computer science and computer science education. So, by creating a behaviorally informed chatbot, this dissertation research benefits from the interdisciplinary perspective.

3.2 Technical details of the chatbot

To develop the chatbot, I used the RASA open-source framework for automating text and voice-based conversations [44, 45]. Rasa has modular architecture and this allows Rasa to be easily integrated with other systems.

RASA architecture contains two main parts which are NLU or natural language understanding and Core or dialog management part [46, 47]. These two parts are designed to be separate for testing and evaluation purposes. Rasa NLU takes inputs from users while Rasa Core makes decisions based on the input received; these characteristics make Rasa Core comparable to an ear and Rasa Core to the brain of the chatbot technology [45].

Moreover, both parts NLU and Core are in a continuous development state [48, 49]. For instance, Rasa Core can be utilized as a dialogue manager with other services and not necessarily together with Rasa NLU. The code of both NLU and Core is created in Python, however, both Core and NLU can expose HTTP APIs and therefore can be integrated in projects built with different programming languages [44].

These two main parts contain several smaller components and incoming text messages from the user go through the so-called sequence of components which is called a pipeline. Each component can be customized and needs to process input information from the previous component and produce some output to the next component [44].

Figure 3.1 below demonstrates Rasa's architecture in more detail. The "Interpreter" block represents the NLU component in Rasa. The incoming messages from users go through the NLU. The NLU then produces structured output containing both the original message received and information on its intent and entities if any.

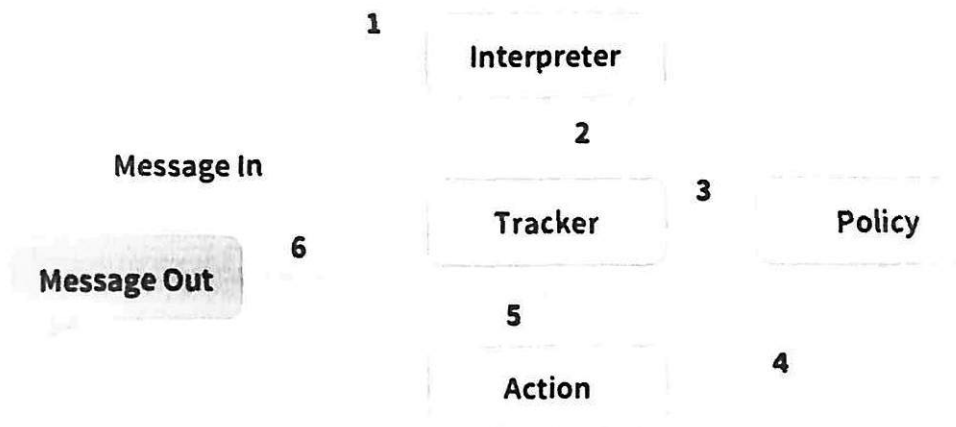


Figure 3.1: Rasa architecture [50]

Next, the “tracker” keeps up the conversation and receives the output from the interpreter. Next, this output goes to the “policy” when Rasa Core makes a decision about which action to perform next. The “tracker” keeps the log of the selected action. Finally, “action” denotes that the chatbot provides an appropriate answer to the user.

The class `rasa.core.policies.policy` makes the decision regarding which action to take in each conversation. Rasa Core’s main training policies include, but are not limited to, MemoizationPolicy, MappingPolicy, KerasPolicy, TED-Policy, EmbeddingPolicy, Form Policy, FallbackPolicy and TwoStageFallbackPolicy [45].

In order to get optimal results and performance developers tend to combine these policies in their chatbots using `config.yml`. setting up the Max History and Data Augmentation parameters is also important as they affect the performance of the model. Rasa Core makes policy decisions based on priority and selects the highest priority policy. The calculation of priority is carried out by estimating a confidence score for each priority, that is, policies having the highest confidence score have the highest priority. Supervised reinforced learning is typically used to enhance the efficacy of the chatbots.

In order to customize the chatbot, I integrated into the pipeline a deep learning model by arranging the inputs and outputs and classifying upcoming intents with the Sklearn Intent Classifier. This is a common approach for chatbot customization [51, 52, 53, 54, 55].

Conversation Driven Development in Rasa

Conversation-Driven Development (CDD) is the process of listening to the users and using those insights to improve the chatbot. It is the overarching best practice approach for chatbot development.

Developing great chatbots is challenging because users will always say something unanticipated. The principle behind CDD is that in every conversation users are telling us—in their own words—exactly what they want. By practicing CDD at every stage of bot development, one can orient chatbots towards real user language and behavior.

CDD includes the following actions [50]:

- Share the chatbot with users as soon as possible
- Review conversations on a regular basis
- Annotate messages and use them as NLU training data
- Test that the chatbot always behaves as you expect
- Track when the chatbot fails and measure its performance over time
- Fix how the chatbot handles unsuccessful conversations

CDD is not a linear process; you'll circle back to the same actions over and over as you develop and improve your bot.

Rasa offers Rasa X, a purpose-built tool for CDD.

In this dissertation, I attempted using the conversation driven development. However, due to the small number of participants and their limited interaction with the chatbot, I had insufficient data for complete implementation of the CDD.

3.3 Knowledge database for the chatbot

The knowledge base of the chatbot was created using two primary documents: 1) the university's academic calendar, and 2) the course syllabus. From the university's official academic calendar, the main dates and events were extracted and programmed into proactive chatbot messages, e.g., as reminders about approaching deadlines. From the course syllabus, course-specific information was developed into proactive messages, e.g., as course announcements,

reminders and tips for completing course assignments.

The chatbot used in this study was designed to support students both academically and socialization-wise. Table 2 presents the overview of the chatbot's pre-scheduled text messages. In March 2022, the chatbot was introduced to the treatment group students and the chatbot-student interaction began. Throughout March and April of 2022, the chatbot sent students relevant information and reminders related to their core course assignments, such as quizzes and lab tasks. During March and April 2022, the chatbot also sent attendance booster messages reminding students of the upcoming class sessions. Students also received professional tips from the chatbot, these included messages with links to useful websites, articles about how the skills developed in the course they were taking could help them professionally in the future, and other similar information.

Starting from March 2022 through May 2022, the chatbot sent motivational and social messages in order to support students through the exam session which takes place in May. Motivational messages included messages aimed to inspire students to study in the course and aim for a productive career in Computer Science. Social messages mostly consisted of congratulatory messages on national holidays, such as Nauryz (March 22), Unity Day (May 1), Defender of the Fatherland Day (May 7). Social messages also included messages related to the university-level and faculty-level events and activities.

Type of Messages	Description	When sent
1. Introduction	Chatbot introduction	March 2022
2. Course assignments	Messages about course assignments, reminders of deadlines	March-April 2022
3. Attendance boosters	Nudges and reminders aimed at improving student attendance	March-April 2022
4. Professional tips	Messages with information for professional development	March-April 2022
5. Motivational	Inspirational messages to support students' interest in the course and profession more broadly	March-May 2022
6. Social	Miscellaneous messages, e.g. congratulations on the Nau-ryz	March-May 2022

Table 3.1: Overview of the chatbot's pre-scheduled messages

Figure 3.2 below shows screenshots of how the proactive messages sent by the chatbot looked for the users.

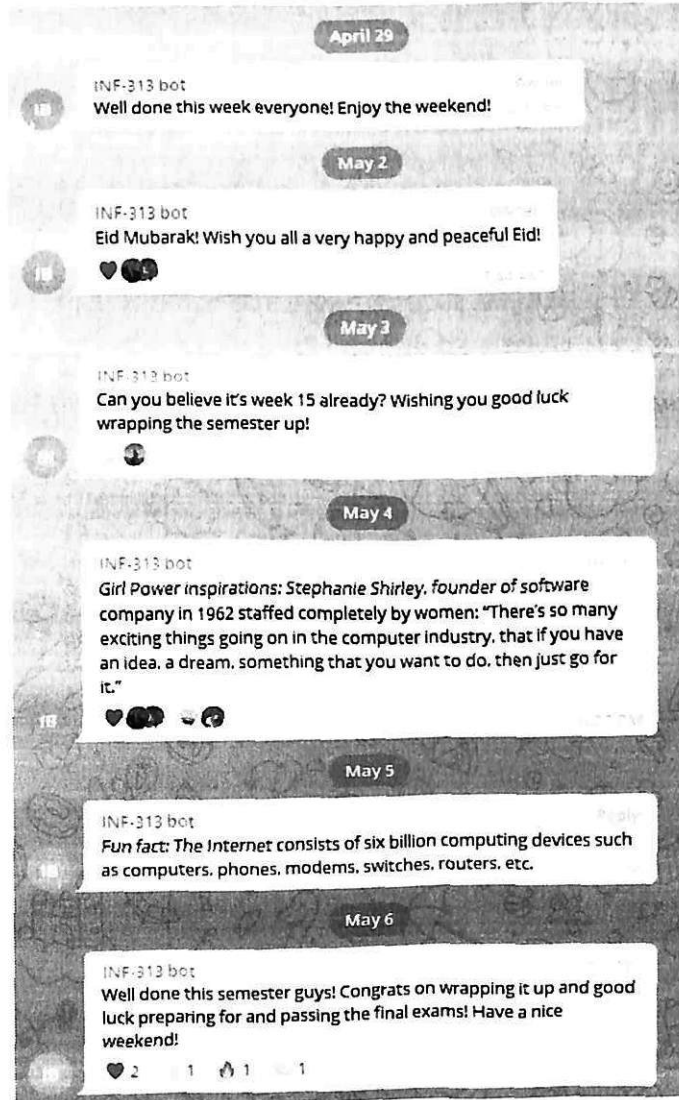


Figure 3.2: Chatbot in action: Proactive messages to users

Figure 3.3 below shows the interaction log of the users with the chatbot.

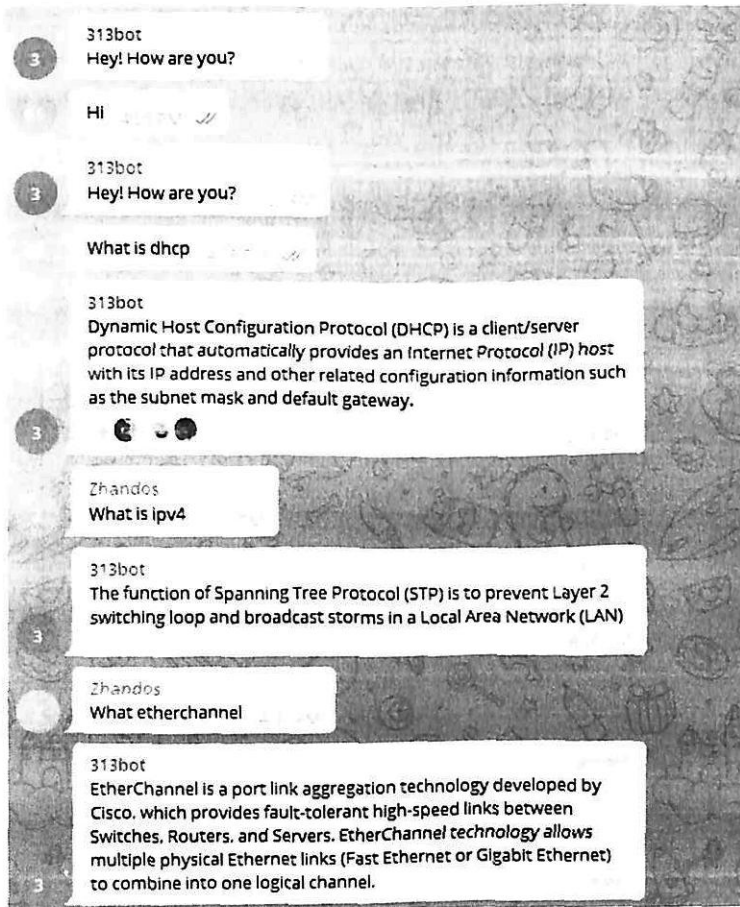


Figure 3.3: Chatbot in action: Interactions with users

4. Data and methodology

4.1 Data

The baseline student data was obtained from the university’s database and includes basic demographics and high school outcomes. The list and description of variables obtained from the university’s database is presented in Table 4.1 below.

Variable name	Description	Coding
1 male	Gender	1= male, 0= female
2 age	Age	Numeric
3 langschool	Language of instruction in high school	KZ=Kazakh, EN= English, RU=Russian
4 altynbelgi	Indicator for whether the student received the Altyn Belgi medal of distinction in high school	1=received Altyn Belgi, 0=didn't receive Altyn Belgi
5 hsgpa	High school GPA	Numeric
6 year	Year of study at the university	Numeric
7 englishlevel	Level of English proficiency	1=A1, 2=A2, 3=B1, 4=B2, 5=C1, 6=C2
8 gpa	GPA at university	Numeric

Table 4.1: Baseline variables

The chatbot was intended to provide nudging and support to 3rd-year undergraduate students taking the INF313 Computer Networks 1 course at the School of Computer Science in a private selective university in Kazakhstan. All students major in Information Systems and their language of instruction is English. All students are citizens of the Republic of Kazakhstan and 99% of them are of Kazakh ethnicity. In terms of financial aid, all students in the

dataset are funded by the government scholarships.

Table 4.2 below displays some more basic descriptive statistics for the pooled sample. Most students are male (75%). On average, students are about 19 years old though there are students aged as little as 17 years and as much as 24 years.

The majority of students (86%) graduated from secondary schools with the Kazakh language of instruction. Less than 3% of the students in the sample studied in secondary schools where English was the language of instruction. And about 12% of students studied in schools where instruction was delivered in Russian.

Academic-wise, there are about 20% of students who received the Altyn Belgi (golden medal) distinction for their secondary studies. The average high school GPA is quite high at about 4.7 out of 5 suggesting that overall, the incoming students at this university did relatively well in high school.

The students mostly are in their 2nd year of studies at the university, have B1-B2 English proficiency levels. Their average GPA earned at the university so far is about 2.6, although the dataset includes students with GPAs as low as 0.22 and as high as 3.8 (out of 4).

	Obs.	Mean	SD	Min	Max
Male	104	.75	.44	0	1
Age	104	18.9	0.95	17	24
Kazakh language high school	104	85.58	0.35	0	1
English language high school	104	2.88	0.17	0	1
Russian language high school	104	11.54	0.32	0	1
Received Altyn Belgi	104	0.19	0.40	0	1
High school GPA	75	4.68	0.40	3.1	5
Year at university	104	2.125	0.33	2	3
English level	90	3.52	0.99	1	6
GPA at university	104	2.55	0.67	0.22	3.8

Table 4.2: Descriptive statistics for the pooled sample

4.2 Randomized controlled trial

A randomized controlled trial research design was used to evaluate the impacts of the chatbot use on students' outcomes. The randomized controlled trial approach is based on the counterfactual conceptual framework, whereby the participants randomly assigned to the treatment and control groups comprise valid counterfactual and differ only in the treatment received [56, 57, 58]. Statistical comparison of the outcomes of the treatment and control groups allows us to estimate the causal effect of the treatment in such research designs [57].

Variable	Treatment group Mean (SD)	Control group Mean (SD)	Difference (SE)
Male	0.73 (0.45)	0.77 (0.43)	-0.04 (0.09)
Age	18.94 (0.96)	18.88 (0.94)	0.06 (0.19)
Kazakh language high school	0.85 (0.36)	0.87 (0.34)	-0.02 (0.07)
English language high school	0.04 (0.19)	0.02 (0.14)	0.02 (0.03)
Russian language high school	0.12 (0.32)	0.12 (0.32)	0 (0.06)
Received Altyn Belgi	0.23 (0.43)	0.15 (0.36)	0.08 (0.08)
High school GPA	4.75 (0.40)	4.63 (0.40)	0.12 (0.09)
Year at university	2.15 (0.36)	2.10 (0.30)	0.06 (0.07)
English level	3.47 (1.00)	3.57 (0.99)	-0.11 (0.21)
GPA at university	2.54 (0.67)	2.55 (0.67)	-0.01 (0.13)

Table 4.3: Balance tests

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ According to t-tests and proportion tests, there are no statistically significant differences between the treatment and control groups.

4.2.1 Randomization procedure and balance tests

For a randomized controlled trial, random assignment to treatment and control is essential. We used Python 3 for randomizing the students taking the INF313 Computer Networks 1 course.

The balance tests presented in Table 5 show that there are no statistically significant differences in any of the covariates across the treatment status. This suggests that randomization worked and the main assumptions of the randomized controlled trial are satisfied. That is, any differences in outcomes are attributable to the treatment assignment.

4.2.2 Statistical estimation

To analyze the results of the randomized controlled trial, multiple linear regression analysis will be used [59]. Multiple linear regression allows estimation of the treatment effects as the coefficients on the treatment variable. The regression also allows adding multiple additional covariates effectively controlling for variation in the outcome variable associated with each of those covariates. In addition, adding covariates increases the precision of the coefficients on the treatment variable thus making it a preferable option over simple tests like t-tests.

The following equation was estimated using the multiple linear regression:

$$Y_i = \alpha + \beta_1 TREATMENT_i + X_\gamma + \epsilon_i$$

where Y_i stands for the outcome variable for each student i ; $TREATMENT_i$

variable indicating whether the student i is assigned to the treatment or control group, X_i a vector of student characteristics including gender, age, language of instruction in secondary school, an indicator for whether the student received the Altyn Belgi distinction award, high school GPA, English proficiency level, and GPA; and E_i represents the error term.

The effects of the chatbot intervention on four outcome variables are explored. The first outcome variable is pre-final grades. These grades range from 0 to 60 and represent the grades earned by the students during the semester. In the data, these grades are calculated by more than one instructor: the lecturer and the tutorial instructor. The grades are the sum of grades for various assignments such as lab assignments and midterms.

The second outcome variable is the final exam grade. This is calculated out of 40 points and is estimated by the main course lecturers. The exam included a written test with a coding assignment.

The third outcome variable is the number of absences. Since the lectures in the course were online, only attendance in tutorial classes was taken into account. The absences are simply counted and represent a discrete variable.

The final outcome is a measure of course satisfaction taken from the course satisfaction survey by the course instructor. The variable contains students' responses to a simple question "How satisfied are you with the course?" The answers ranged from 0 representing "not at all satisfied" to 10 representing "completely satisfied."

5. Results

5.1 Student-chatbot interaction

The chatbot interacted with the students from March through May 2022. The chatbot delivered the messages via Telegram taking advantage of the fact that Telegram is the preferred medium for student-instructor communication at this university so students are used to getting course-related information via Telegram specifically.

All the proactive messages were sent as scheduled. However, there were technical glitches with the chatbot's response component that prevented full student-chatbot interaction in March. Starting in mid-April, the technical issue was resolved and the chatbot became responsive to students' messages.

Therefore, it should be noted that the treatment effects should be interpreted with caution as the chatbot intervention was somewhat limited. The reasonable interpretation pertains to a beta-version of the chatbot rather than to a complete chatbot intervention with full functionality.

5.2 Chatbot's effects on students' outcomes

Tables 5.1 and 5.2 present the main results.

In table 5.1, equation 1 is estimated without the vector of student covariates. That is, the constant represents the average values of the outcome variables for the control group while the coefficient on the "treatment" variable represents the average values for the treatment group.

	(1) Prefinal grades	(2) Exam	(3) Absences	(4) Satisfaction
Treatment	1.262 (3.259)	0.634 (1.693)	-0.210 (1.111)	3.103*** (0.405)
Constant	34.52*** (2.315)	12.71*** (1.203)	4.902*** (0.789)	5.194*** (0.288)
Observations	103	103	103	103
R-squared	0.001	0.001	0.000	0.367

Table 5.1: Chatbot effects (the model without covariates)
Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The estimates in Table 5.1 overall show that the chatbot had no effects on the prefinal grades, exam scores, and absences. Column 1 shows that while the treatment students received about 1.3 points higher pre-final grades, this difference is not statistically significant. Similarly, the treatment students received about 0.6 points higher exam scores, but yet again, this difference is not statistically significant.

Finally, the treatment students have, on average, 0.2 fewer absences, but the difference across groups is not statistically significant.

At the same time, the chatbot increased student satisfaction in the treatment group. The treatment group students reported approximately 3 points higher satisfaction with the course and the difference is statistically different ($p < 0.001$).

Next, equation 1 is estimated including the student-level covariates. These estimates are presented in Table 5.2 below. There are a lot of missing values in some covariates, therefore, only 70 observations are in the data used for estimating the complete equation 2. Nevertheless, the results are similar to those in Table 5.1.

	(1) Prefinal grades	(2) Exam	(3) Absences	(4) Satisfaction
Treatment	0.657 (3.360)	-0.554 (1.840)	-1.158 (1.072)	3.141*** (0.522)
Male	3.143 (4.036)	5.396** (2.210)	0.354 (1.287)	-0.0757 (0.627)
Age	0.0656 (2.239)	0.207 (1.226)	-0.295 (0.714)	0.462 (0.348)
Kazakh language high school	-0.395 (8.403)	-1.693 (4.602)	-3.111 (2.680)	1.077 (1.305)
Russian language high school	7.148 (10.08)	0.757 (5.522)	-1.846 (3.216)	2.659* (1.566)
Received Al-tyn Belgi	0.362 (4.724)	1.819 (2.587)	2.542* (1.507)	0.458 (0.734)
High school GPA	-3.238 (5.258)	-0.769 (2.879)	2.051 (1.677)	-0.0523 (0.817)
English level	0.421 (1.984)	0.928 (1.087)	-0.818 (0.633)	-0.00447 (0.308)
GPA at university	19.17*** (3.072)	9.383*** (1.682)	-5.872*** (0.980)	-0.0113 (0.477)
Constant	-5.202 (55.83)	-17.45 (30.57)	20.85 (17.81)	-4.443 (8.672)
Observations	70	70	70	70
R-squared	0.445	0.422	0.440	0.446

Table 5.2: Chatbot effects (the model with covariates)

Standard errors in parentheses. *** $p < 0.01$. ** $p < 0.05$. * $p < 0.1$

Specifically, the chatbot had no effects on pre-final grades, on exam scores, and tutorial attendance. Still, the chatbot increased course satisfaction by about 3 points ($p = 0.001$).

Equation 1 was also estimated for various subgroups: for low-achieving, for high-achieving, for males, for females. In all cases, the estimates were similar, suggesting that the chatbot had no subgroup effects. These estimates support

the interpretation that the chatbot increased course satisfaction both in the full sample and within the subgroups investigated.

6. Conclusion

In this dissertation, a chatbot to support students majoring in Computer Science was developed using the Rasa open-source framework. The chatbot was then evaluated using a randomized controlled trial with 103 undergraduate students taking the INF313 Computer Networks 1 course at the School of Computer Science in a private selective university in Kazakhstan.

The randomized controlled trial evaluation showed that the chatbot intervention had no effect on students' performance during the semester and in the final exam. Neither did the chatbot affect student attendance. At the same time, the chatbot increased course satisfaction by about 3 points.

The results suggest that the chatbot intervention does improve students' satisfaction with the course while not affecting academic outcomes. This appears to be in line with the recent studies of chatbots. In contrast with the existing literature on chatbot-supported nudging[36, 37, 60], we find no subgroup effects of the chatbot intervention.

The estimates suggest reserved optimism about chatbot support of students. While being a cheap and scalable support tool, it is important to understand that chatbots may have limited effects on academic outcomes.

Yet, increased course satisfaction may be an important outcome that higher education institutions may wish to pursue. So chatbots may be recommended for use in universities aiming to enhance student satisfaction. It is also possible that after extended use, satisfaction with the course may increase students' effort and therefore performance. It is also possible that course satisfaction may eventually increase student retention rates and their interest in the profession.

Limitations

This study has several limitations. First, the chatbot intervention was a demo version with the interactive component becoming functional only in the second half of the experiment due to technical glitches. With improvements and with greater knowledge base developed based on the student-chatbot interaction data (and not just based on the proactive messages), it is possible that the chatbot may indeed potentially enhance students' academic outcomes. If not grades, attendance could probably be boosted via timely and individualized reminders. Longer duration of the intervention, for example, during the whole semester, could also have different effects on students. In fact, many chatbot interventions in education, for example [34, 36], tend to be longer-term, over the period of several months.

Second, the study uses data from one university only. It is possible that the university students have sufficient support both within the course taught and at the level of their department and school. Therefore, the findings may be different in another context where, for example, student support is less readily available.

Third, the chatbot was developed in English as the research site was a university where the language of instruction is English. However, it is possible that students are not used to receiving such interventions in English. The chatbot that speaks English may have been perceived as somewhat distant and not relatable. Given the recent evidence suggesting that the linguistic characteristics of chatbots matter [25], the choice of the chatbot language is an important factor in determining its success.

Finally, only short-term academic outcomes and one course satisfaction outcome is studied. The chatbot could have potentially have effects on a wider range of outcomes.

Suggestions for chatbot developers

The following suggestions for chatbot developers are based on the experience of creating the chatbot for this study.

The first suggestion will be to work closely with the course instructors and

with other stakeholders in building the knowledge base of the chatbot. It is essential that administrative information, course information, and various contextual rules are taken into account when creating the chatbot content.

The second suggestion is to carefully select the language of the chatbot. While selecting English makes a lot of things easier given that most chatbot tools are geared for English, selecting the language which is native to students may make the chatbot much more relevant and useful for students. In the context of this dissertation study, Kazakh language should likely be selected as most students speak Kazakh. Yet this leaves the question of how to best serve Russian speaking students who may be less fluent in Kazakh.

Suggestions for future research

In future research, the following should be addressed. First, the chatbot theory of action should be clearly formulated. Why the chatbot is expected to affect this or that outcome should be made clearer. The chatbot cannot be expected to be affecting any academic outcome or any non-academic outcome. Perhaps the research should move towards what outcomes chatbots can reasonably affect.

Next, more replications in various contexts should be conducted in order to understand how context may influence chatbot effects. At this point, what we know is whether chatbots affected or didn't affect specific outcomes in a specific university. Whether or not a chatbot will be effective in a specific new context remains an empirical question.

Further, given the positive effects of the chatbot on course satisfaction in this study, it appears reasonable to explore the potential of chatbots in increasing non-academic measures such as motivation, interest in the discipline or profession, self-efficacy, etc.

Finally, if a language other than English will be chosen as the interaction language for the chatbot, programming issues related to using languages other than English should be addressed. Utilization of conversational AI frameworks like Deep Pavlov is recommended to make a functional chatbot speaking a language other than English. [57, 58, 59]

References

- [1] G. Momonov. “Artificially intelligent conversational chatbots: Main types, programming issues and possible solutions”. In: *Proceedings of IYSW* 10.10 (2021), pp. 350–358.
- [2] A.-A. Georgescu. “Chatbots for education—trends, benefits and challenges,” in: *Conference proceedings of eLearning and Software for Education «(eLSE) 2.14* (2018), pp. 195–200.
- [3] S. Yang and C. Evans. “Opportunities and challenges in using AI chatbots in higher education”. In: *Proceedings of the 2019 3rd International Conference on Education and E-Learning* (2019), pp. 79–83.
- [4] G.-H. Hwang J. Wang and C.-Y. Chang. “Directions of the 100 most cited chatbot-related human behavior research: A review of academic publications”. In: *Computers and Education: Artificial Intelligence* (2021), p. 100023.
- [5] G. Recchia. “Investigating AI Narratives with Computational Methods,” in: *A History of Imaginative Thinking about Intelligent Machines* (2020), p. 382.
- [6] N. Sharkey and A. Sharkey. “Electro-mechanical robots before the computer”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 223.1 (2009), pp. 235–241.
- [7] R. M. French. “The Turing Test: the first 50 years,” in: *Trends in cognitive sciences* 4.3 (2000), pp. 115–122.
- [8] M. Goyal V. Sharma and D. Malik. In: *An intelligent behaviour shown by chatbot system, International Journal of New Technology and Research* 3.4 (2017), p. 263312.

- [9] N. Sharkey and A. Sharkey. "Electro-mechanical robots before the computer". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 223.1 (2009), pp. 235–241.
- [10] K. Warwick and H. Shah. "Can machines think? A report on Turing test experiments at the Royal Society," in: *Journal of experimental Theoretical artificial Intelligence* 28.6 (2016), pp. 989–1007.
- [11] J. Copeland. "Why Eugene Goostman did not pass the Turing Test," in: (2014).
- [12] M. T. ZEMČÍK. "A brief history of chatbots," in: *DEStech Transactions on Computer Science and Engineering* 10 (2019).
- [13] A. Atabekov and O. Yastrebov. "Legal status of artificial intelligence across countries: Legislation on the move," in: *European Research Studies* 21.4 (2018), pp. 773–782.
- [14] N. Ali. "Chatbot: A conversational agent employed with named entity recognition model using artificial neural network," in: *arXiv preprint* (2020), p. 04248.
- [15] M. Cortiñas R. Chocarro and G. Marcos-Matás. "Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics," in: *Educational Studies* (2021), pp. 1–19.
- [16] L. Laranjo A. S. Miner and A. B. Kocaballi. "Chatbots in the fight against the COVID-19 pandemic," in: *NPJ digital medicine*, 3.1 (2020), pp. 1–4.
- [17] J. Trivedi. "Examining the customer experience of using banking chatbots and its impact on brand love: The moderating role of perceived risk," in: *Journal of internet Commerce*, 18.1 (2019), pp. 91–111.
- [18] J. Cahn. "CHATBOT: Architecture, design, development," in: *University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science*, (2017).
- [19] Z. Callejas M. McTear and D. Griol. "The Conversational Interface: Talking to Smart Devices". In: *Springer International Publishing* (2016).

- [20] W. Minker R. E. Gruhn and S. Nakamura. "Statistical pronunciation modeling for non-native speech processing." In: *Springer Science Business Media* (2011).
- [21] G. Tur and R. De Mori. "Spoken language understanding: Systems for extracting semantic information from speech." In: *John Wiley Sons*, (2011).
- [22] S. A. Abdul-Kader and J. C. Woods. "Survey on chatbot design techniques in speech conversation systems," in: *International Journal of Advanced Computer Science and Applications*, 6.7 (2015).
- [23] C. I. Nass and S. Brave. "Wired for speech: How voice activates and advances the human-computer relationship." In: *MIT press Cambridge*, (2005).
- [24] T. Araujo. "Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions," in: *Computers in Human Behavior*, 85 (2018), pp. 183 189.
- [25] A. P. Chaves et al. "It's how you say it: Identifying appropriate register for chatbot language design," in: *Proceedings of the 7th International Conference on Human-Agent Interaction*, (2019), pp. 102–109.
- [26] Business Insider. "80% of businesses want chatbots by 2020." In: <https://www.businessinsider.com/80-percent-of-businesses-want-chatbots-by-2020-2016-12?r=DEIR=T> (accessed. (2016-2020)).
- [27] I. Roll and R. Wylie. "Evolution and revolution in artificial intelligence in education," in: *International Journal of Artificial Intelligence in Education* 26.2 (2016), pp. 582 599.
- [28] R. Winkler and M. Söllner. "Unleashing the potential of chatbots in education: A state-of-the-art analysis," in: (2018).
- [29] J. Q. Pérez, T. Daradoumis, and J. M. M. Puig. "Rediscovering the use of chatbots in education: A systematic literature review," in: *Computer Applications in Engineering Education*, 28.6 (2020), pp. 1549 1565.
- [30] C. W. Okonkwo and A. Ade-Ibijola. "Chatbots applications in education: A systematic review," in: *Computers and Education: Artificial Intelligence*. (2021), p. 100033.

- [31] S. Wollny et al. "Are we there yet?," in: *A systematic literature review on chatbots in education, Frontiers in artificial intelligence*, 4 (2021).
- [32] Y. Hayashi E. Ayedoun and K. Seta. "A conversational agent to encourage willingness to communicate in the context of English as a foreign language," in: *Procedia Computer Science* 60 (2015), pp. 1433-1442.
- [33] L. Cicchanowski et al. "In the shades of the uncanny valley: An experimental study of human-chatbot interaction," in: *Future Generation Computer Systems*. 92 (2019), pp. 539-548.
- [34] L. K. Fryer et al. "Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners," in: *Computers in Human Behavior* 75 (2017), pp. 461-468.
- [35] M. P.-C. Lin and D. Chang. "Enhancing Post-secondary Writers' Writing Skills with a Chatbot," in: *Journal of Educational Technology Society*, 23.1 (2020), pp. 78-92.
- [36] L. C. Page and H. Gchlback. "How an artificially intelligent virtual assistant helps students navigate the road to college," in: *AERA Open* 3.4 (2017), pp. 1-12.
- [37] A. Nurshatayeva et al. "Are artificially intelligent conversational chatbots uniformly effective in reducing summer melt? Evidence from a randomized controlled trial," in: *Research in Higher Education* 62.3 (2021), pp. 392-402.
- [38] R. H. Thaler and C. R. Sunstein. "Nudge: Improving decisions about health, wealth, and happiness," in: *new Haven CT: Yale University Press* (2008).
- [39] R. H. Thaler. "From cashews to nudges: The evolution of behavioral economics". In: *American Economic Review* 108.6 (2018), pp. 1265-87.
- [40] P. Oreopoulos et al. "Improving non-academic student outcomes using online and text-message coaching". In: *Journal of Economic Behavior Organization* 171 (2020), pp. 342-360.
- [41] A. M. Lavecchia, H. Liu, and P. Oreopoulos. "Behavioral economics of education: Progress and possibilities," in: *Handbook of the Economics of Education*. 5 (2016), pp. 1-74.

- [42] R. Chetty. “Behavioral economics and public policy: A pragmatic perspective”. In: *American Economic Review* 105.5 (2015), pp. 1–33.
- [43] B. L. Castleman, S. Schwartz, and S. Baum. “Decision making for student success: behavioral insights to improve college access and persistence”. In: *New York: Routledge* (2015).
- [44] T. Bocklisch, N. Pawlowski J. Faulkner, and A. Nichol. “Rasa: Open source language understanding and dialogue management”. In: *arXiv preprint* 05181 (2017).
- [45] R. K. Sharma and M. Joshi. “An analytical study and review of open source chatbot framework, rasa”. In: *International Journal of Engineering Research* 9.6 (2020).
- [46] A. Singh, K. Ramasubramanian, and S. Shivam. “Introduction to microsoft bot rasa and google dialogflow”. In: *Building an Enterprise Chatbot: Springer* (2019), pp. 281–302.
- [47] G. Shekhar, R. D’Souza, and K. Fernandes. “AI Driven Contextual Virtual Teaching Assistant Using RASA”. In: *Proceedings of the 21st Annual Conference on Information Technology Education* (2020), pp. 346–346.
- [48] T. N. T. Mai and S. Maxim. “Enhancing Rasa NLU model for Vietnamese chatbot”. In: *International Journal of Open Information Technologies* 9.1 (2021), pp. 31–36.
- [49] A. Jiao. “An intelligent chatbot system based on entity extraction using RASA NLU and neural network”. In: *Journal of Physics: Conference Series* 1487.1 (2020), p. 012014.
- [50] RASA Technologies. “Rasa Architecture”. In: <https://rasa.com> (accessed 2022) ().
- [51] I. Drori et al. “Automatic machine learning by pipeline synthesis using model based reinforcement learning and a grammar”. In: *arXiv preprint* 10345 (2019).
- [52] S. Larson et al. “An evaluation dataset for intent classification and out-of-scope prediction”. In: *arXiv preprint* 02027 (2019).

- [53] C. Segura et al. “Chatbol, a chatbot for the Spanish La Liga”. In: *9th International Workshop on Spoken Dialogue System Technology* (2019), pp. 319–330.
- [54] J.-W. Chang. “Enabling progressive system integration for AIoT and speech-based HCI through semantic-aware computing,” in: *The Journal of Supercomputing* 02027 (2021), pp. 1–37.
- [55] X. Liu et al. “Benchmarking natural language understanding services for building conversational agents,” in: *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction* (2021), pp. 165–183.
- [56] P. E. Cairns and A. L. Cox. “Research methods for human-computer interaction”. In: *Cambridge University Press* (2008).
- [57] G. Imbens and D. B. Rubin. “Causal inference for statistics, social, and behavioral sciences”. In: *New York, NY: Cambridge University Press* (2015).
- [58] T. D. Cook W. R. Shadish and D. T. Campbell. “Experimental and quasi-experimental designs for generalized causal inference.” In: *Boston, MA: Houghton Mifflin Company* (2002).
- [59] G. James, D. Witten, and T. Hastie. “An introduction to statistical learning”. In: *Springer* (2013).
- [60] L. C. Page, J. Lee, and H. Gehlbach. “The conditions under which college students are responsive to nudging”. In: *Annenberg Institute* (2020).