

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа ядерных технологий

Направление подготовки: 01.04.02 Прикладная математика и информатика

Отделение экспериментальной физики

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Development of a chatbot with natural language processing

УДК 004.522

Студент

Группа	ФИО	Подпись	Дата
ОВМ82	Исмоилов Нурулло Исроилович		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	к.ф-м.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Рыжакина Т.Г.	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Романова С.В.			

По английскому языку

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Маркова Н.А.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семёнов М.Е.	к.ф-м.н., доцент		

Томск 2020 г.

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа ядерных технологий

Направление подготовки 01.04.02 «Прикладная математика и информатика»

Отделение экспериментальной физики

УТВЕРЖДАЮ:

Руководитель ООП

_____ Семенов М. Е.

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Студенту:

Группа	ФИО
ОВМ82	Исмоилов Нурулло Исроилович

Тема работы:

Разработка чат-бота с обработкой естественного языка Development of a chatbot with natural language processing	
Утверждена приказом директора (дата, номер)	№86-25/с от 26.03.2020

Срок сдачи студентом выполненной работы:	31.05.2020
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.);</i></p>	<p>Объект исследования – чат-бот с поддержкой запросов на естественном (русском) языке. Кроссплатформенный чат-бот должен функционировать в режиме 24/7, с возможностью обучения на основании датасетов.</p> <p>The object of research is a chat bot with support for requests in the natural (Russian) language. Cross-platform chatbot should function 24/7, with the possibility of training based on datasets.</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none">1. Анализ существующих решений на рынке чат-ботов2. Рассмотреть различные типы чат-ботов3. Проанализировать инструменты, библиотеки и модели машинного обучения для разработки чат-бота с поддержкой запросов на естественном языке.4. Выполнить сбор, описание и анализ исходных данных и требований к функциональным возможностям чат бота.5. Разработать модуль обработки запросов и модуль ответов на естественном (русском) языке.6. Провести тестирование чат-бота, сделать выводы о работе. <ol style="list-style-type: none">1. Analysis of existing solutions in the chatbot market.2. Consider the different types of chat bots3. To analyze tools, libraries and machine learning models for developing a chat bot supporting queries in natural language.4. Collect, describe and analyze the source data and requirements for the functionality of the chat bot.5. To develop a request processing module and a

	<p>response module in a natural (Russian) language.</p> <p>6. Conduct a chatbot test, make conclusions about the work</p>
<p>Перечень графического материала <i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> 1. Строение биологического нейрона 2. Строение искусственного нейрона 3. Модель «Перцептрон» 4. Нейросеть 5. Система «End-to-end» 6. Рекуррентная нейросеть 7. Модель «Seq2seq» 8. Иллюстративные примеры результатов работы чат-бота 9. Мобильное приложение чат-бота 10. Виджет на сайт чат-бота <ol style="list-style-type: none"> 1. The structure of a biological neuron 2. The structure of artificial neuron 3. Perceptron model 4. Neural network 5. “End-to-end” system 6. Recurrent neural network 7. Seq2seq model 8. Illustrative example of a chatbot 9. Chatbot mobile application 10. Chatbot widget for a website

Консультанты по разделам выпускной квалификационной работы

(если необходимо, с указанием разделов)

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Рыжакина Татьяна Гавриловна
Социальная ответственность	Романова Светлана Владимировна

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	к.ф.-м.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0BM82	Исмоилов Нурулло Исроилович		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0BM82	Исмоилову Нурулло Исроиловичу

Школа	Инженерная школа ядерных технологий	Отделение (НОЦ)	Отделение экспериментальной физики
Уровень образования	Магистратура	Направление/специальность	01.04.02 Прикладная информатика и математика

Тема ВКР:

Разработка чат-бота с обработкой естественного языка	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	<p>Объект исследования: чат-бот с обработкой запросов на естественном языке.</p> <p>Рабочая зона: стол с компьютером</p> <p>Область применения: корпоративные сети средних и крупных предприятий.</p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<p>1. Правовые и организационные вопросы обеспечения безопасности:</p> <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<ul style="list-style-type: none"> – ТК РФ от 30.12.2001 N 197-ФЗ (ред. от 01. 04.2019) – ГОСТ 12.2.032 -78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
<p>2. Производственная безопасность:</p> <p>2.1. Анализ выявленных вредных и опасных факторов</p> <p>2.2. Обоснование мероприятий по снижению воздействия</p>	<ul style="list-style-type: none"> – Отклонение показателей микроклимата в помещении – Недостаток естественного освещения – Недостаточная освещенность рабочей зоны – Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека – Повышенная напряженность электромагнитного поля
<p>3. Экологическая безопасность:</p>	<ul style="list-style-type: none"> – Глобальное потепление

	<ul style="list-style-type: none"> – Загрязнение атмосферы и водного бассейна Земли вредными и ядовитыми веществами – Изменение ландшафта Земли
4. Безопасность в чрезвычайных ситуациях:	– Пожары

Дата выдачи задания для раздела по линейному графику	23.03.2020
---	-------------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ООД	Романова Светлана Владимировна			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0ВМ82	Исмоилов Нурулло Исроилович		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСООБЪЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
0BM82	Исмоилова Нурулло Исроиловича

Школа	ИЯШТ	Отделение	экспериментальной физики
Уровень образования	Магистратура	Направление/специальность	Прикладная математика и информатика

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Использование информации, представленной в российских и иностранных научных публикациях, аналитических материалах и изданиях, нормативно-правовых документах.
2. <i>Нормы и нормативы расходования ресурсов</i>	
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	Проведение предпроектного анализа. Определение целевого рынка и проведение его сегментирования. Выполнение SWOT-анализа проекта.
2. <i>Определение возможных альтернатив проведения научных исследований</i>	Определение целей и ожиданий, требований проекта. Определение заинтересованных сторон и их ожиданий.
3. <i>Планирование процесса управления НИ: структура и график проведения, бюджет, риски.</i>	Составление календарного плана проекта. Определение бюджета НИ
4. <i>Определение ресурсной, финансовой, экономической эффективности.</i>	Проведение оценки экономической эффективности исследования.

Перечень графического материала (с точным указанием обязательных чертежей):

1. <i>Оценка конкурентоспособности технических решений</i>
2. <i>Матрица SWOT</i>
3. <i>Оценка готовности проекта к коммерциализации</i>
4. <i>Календарный план-график и бюджет НИ</i>
5. <i>Оценка ресурсной, финансовой и экономической эффективности НИ</i>
6. <i>Сравнительная эффективность разработки</i>

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Рыжакина Татьяна Гавриловна	к.э.н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0BM82	Исмоилов Нурулло Исроилович		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код	Результаты обучения
P1	Самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять специальную лексику и профессиональную терминологию языка. Демонстрировать способность к самостоятельному обучению новым методам исследования, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения.
P5	Использовать на практике умения и навыки в организации исследовательских работ по использованию математических методов и моделирования для решения задач, способность организовывать и руководить работой команды.
P6	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.
P7	Разрабатывать стратегии и цели проектирования, критерии эффективности и ограничения применимости, новые методы, средства и технологии математического моделирования объектов и явлений в различных предметных областях
P8	Осуществлять авторское сопровождение процессов проектирования, внедрения, эксплуатации и модернизации моделирующих программных комплексов на всех этапах жизненного цикла

Реферат

Выпускная квалификационная работа: 92 листов, 17 рисунков, 31 таблицы, 38 использованных литературных источников.

Ключевые слова: чат-бот, обработка естественного языка, система «end-to-end», машинное обучение, нейронные сети.

Объект исследования: чат-бот.

Цель работы: разработать прототип чат-бота с обработкой естественного языка.

Работа включает: введение, пять разделов, список публикаций студента, а также список использованных литературных источников.

В разделе 1 «Обзор литературы» приведен краткий обзор литературы, необходимый для дальнейшей работы, сформулирована цель и задачи работы.

В разделе 2 «Теоретические аспекты» рассмотрены теоретические аспекты разработки чат-ботов, проведен анализ существующих аналогов и технологий для их разработки.

В разделе 3 «Программирование чат-бота» приведено описание процесса разработки, в том числе: подготовка данных для машинного обучения, разработка модели, разработка и тестирование прототипа чат-бота.

В разделе 4 «Социальная ответственность» оценены риски и опасности для здоровья и окружающей среды и рассмотрены способы их минимизации.

В разделе 5 «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» рассмотрена экономическая составляющая работы. Посчитаны затраты на реализацию и изучены конкуренты на рынке, а также посчитана рентабельность проекта.

Abstract

Final qualification work is stated on 92 pages, contains 17 figures, 31 tables and 38 references.

Keywords: chatbot, natural language processing, end-to-end system, machine learning, neural networks.

Object of research: chatbot.

Purpose: to develop a chatbot prototype with natural language processing.

The work includes: Introduction, five sections, a list of students' publications and references.

Section 1 "Literature review" provides a brief review of the literature, formulates goal and objectives of the research.

Section 2 "Theoretical aspects" discusses the theoretical aspects of developing chatbots and technologies for their development.

Section 3 "Programming a chatbot" describes the development process, including: preparing data for machine learning, developing a model, developing and testing a prototype chatbot.

Section 4 "Social responsibility" assesses the risks and hazards to health and the environment and considers ways to minimize them.

Section 5, "Financial Management, Resource Efficiency, and Resource Saving," discusses the economic component of the work. The costs of implementation were calculated and competitors in the market were studied, and the profitability of the project was also calculated.

Definitions

Application programming interface (API) – is a computer interface which defines interactions between multiple software intermediaries. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc.

Artificial neuron – it is a mathematical function conceived as a model of biological neurons. Artificial neurons are elementary units in an artificial neural network.

Biological neuron – is an electrically excitable cell that communicates with other cells via specialized connections called synapses. It is the main component of nervous tissue in human and animal.

Chatbot – a program that simulates a person's speech behavior when communicating with one or more interlocutors.

End-to-End System – a full package of machine learning system that processes data, learns from it and deliver user-friendly results.

Graphics processing unit (GPU) – is a specialized electronic circuit designed by rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device.

Long-Short Term Memory (LSTM) – an artificial recurrent neural network (RNN) architecture used in the field of deep learning.

Natural language – a complex system of rules in accordance with which speech activity occurs, i.e. generation and understanding of texts.

Natural language processing – one of the areas of artificial intelligence and mathematical linguistics, which is studying the problems of computer analysis and synthesis of natural languages.

Neural network –a network that simulates human or animal brain with lots of densely interconnected artificial neurons, so that a computer program can learn things, recognize patterns and make decisions in a humanlike way.

Neuroplasticity – The process of changing humans' or animals' brain by gaining experience throughout lifetime.

Recurrent neural network (RNN) - is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. These kinds of networks are used when next data depends on previous data.

Term Frequency – Inverse Term Frequency (TF-IDF) value is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.

Tokenization – the process of demarcating and possibly classifying sections of a string of input characters.

TABLE OF CONTENTS

INTRODUCTION	15
1 LITERATURE REVIEW	18
1.1 Introduction to a research area	18
1.2 Corporate tasks analysis and development tools.....	20
1.3 Dataset Preprocessing	23
1.4 Purpose and tasks research.....	25
2 THEORETICAL ASPECTS	26
2.1 Neural network	26
2.2 About chatbots	26
2.3 Relevance of work	28
2.4 Analyze the existing tools for developing chatbot.....	28
2.5 Development tools and libraries.....	29
2.6 Dataset.....	32
2.7 A training model	32
2.8 Natural language processing	33
3 PROGRAMMING A CHATBOT	35
3.1 Text analysis steps	35
3.2 Building a model	37
3.3 Building and testing a chatbot.....	38
4 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	44
4.1 Правовые и организационные вопросы обеспечения безопасности.....	45
4.1.1 Эргономические требования к рабочему месту	45
4.1.2 Особенности законодательного регулирования проектных решений.....	45
4.2 Производственная безопасность	46
4.2.1 Отклонение показателей микроклимата в помещении	47
4.2.2 Недостаток естественного освещения	49
4.2.3 Недостаточная освещенность рабочей зоны	50

4.2.4	Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека	51
4.2.5	Повышенный уровень электромагнитных излучений	52
4.3	Экологическая безопасность	54
4.4	Безопасность в чрезвычайных ситуациях	55
4.5	Выводы по разделу «Социальная безопасность»	57
5	ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	58
5.1	Предпроектный анализ	58
5.1.1	Потенциальные потребители результатов исследования	58
5.1.2	Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения	59
5.1.3	SWOT-анализ	61
5.1.4	Оценка готовности проекта к коммерциализации	63
5.1.5	Методы коммерциализации результатов научно-технического исследования.....	65
5.2	Инициация проекта.....	66
5.3	Планирование управления научно-техническим проектом	68
5.3.1	Иерархическая структура проекта	68
5.3.2	План проекта	68
	CONCLUSIONS	85
	Список публикаций студента	88
	REFERENCES	89

INTRODUCTION

Human or animal brains can change during lifetime by gaining experience. This statement is called neuroplasticity. Because of it we can learn different things. Sometimes we can learn very amazing things. Some examples of them are given below.

Usually, information from eye retina goes to lateral geniculate nucleus and then to primary visual cortex. Medial geniculate nucleus is responsible for hearing and forwards information from ear to auditory cortex. Indian neuroscientist Mriganka Sur provided an experiment in MIT. During this experiment he changed the direction of neurons in some ferrets' brain, so that information from retina goes to medial geniculate nucleus, which is normally processes audio information. Despite these ferrets were able to process visual information [1].

Another experiment provided by Paul Bach-y-Rita is known as “sensory substitution”. During this experiment blind people are set down in a special chair whose back was covered by vibrating panels. This chair transmitted information from video camera into vibrations. Despite the fact that camera was only 400 pixels people were able to distinguish between people. Later Paul Bach-y-Rita found that information can be transmitted better with human saliva than tissue. Then the camera was installed in helmet and the vibrating plate in mouth of people. In the matter of hours people was able to distinguish objects [2]. Now this device is known as Brain port [3].

These experiments show that brain is not strictly programmed, but it is very adaptive mechanism which can adopt to the changing environment. To build an artificial neuron first we have to understand how biological neuron works. Information is transmitted to the brain from sensory organs. Neuron gains information from dendrites. There are many dendrites in a single neuron. Information is transmitted through an axon by using protons, electrons and chemical reactions. Finally, information is transmitted to the other neuron or to the brain. Each neuron forms a large number of connections with other neurons. The number of these

connections are between 5 to 200 thousand. When a human learns something new, new connections are formed. For now, programming a biological neuron is impossible. It should be something simpler. So, in computers artificial neuron models are used.

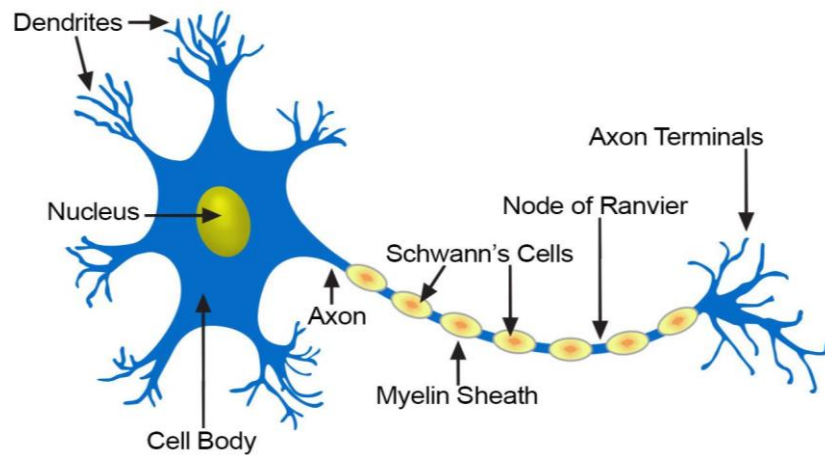


Figure 1 – The structure of a biological neuron [4].

Artificial neurons are having something similar to biological. It consists of incoming data, weights of each incoming data or inputs, sum, activation function (figure 2).

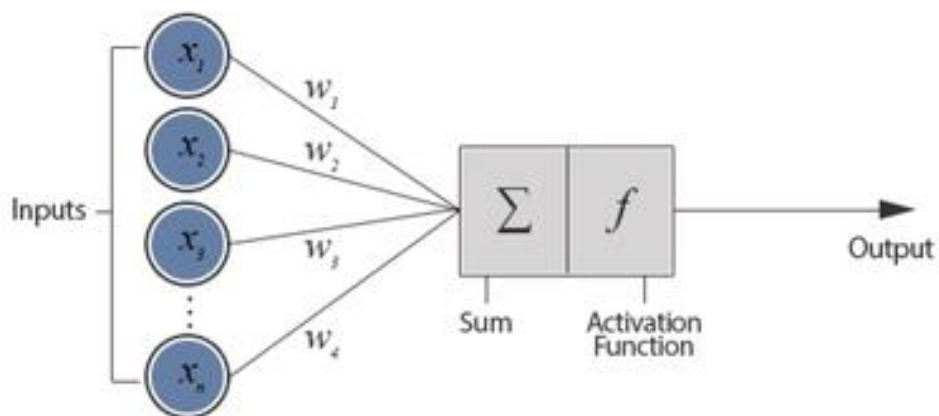


Figure 2 – Artificial neuron [5]: x_1, x_2, \dots, x_n are inputs, w_1, w_2, \dots, w_n are weights.

Inputs can either be an activation of other neurons or information from outside. If we draw analogy with biological neurons it can be an activation from other neuron or information from sensory organs. Weights w_1, w_2, \dots, w_n show that how important the current input for the neuron. Next, the sum of all weights is counted and its value

passed to the activation function. Finally, its processed by activation function and some output is formed. This output is called output activation of a neuron.

The simplest model of a neuron is perceptron. It is a bridge to understanding more complex models. Perceptron was first introduced by at the Cornell Aeronautics Laboratory by Frank Rosenblatt in 1958 [6]. He believed that, using his model humanity will build an artificial intelligent machine. Later in the New York Times an article was published, which stated: “a perceptron model will allow US military to build robots of new generation which will walk, talk and even realize the fact of self-existence. Unfortunately, this all ended as quickly as started. M.Minsky and S. Papert had a research on perceptron’s. They showed that this model has many restrictions and low efficiency [7]. Because of high expectations, people thought that all the neural network models have same deficiencies. It let to the historical event known as AI (artificial intelligence) winter, which had place from 1974 to 1980. In this period funding were reduced in the field of AI. Despite these facts the perceptron model is still in use. For example, it is used when much predictors are needed.

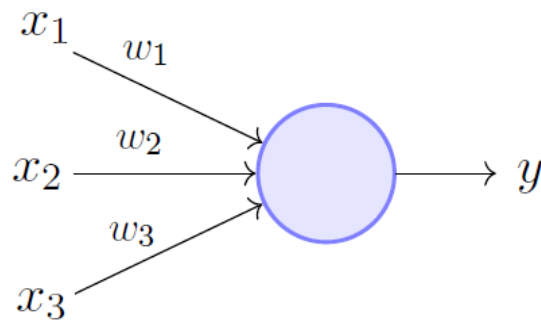


Figure 3 – Perceptron model: x_1, x_2, x_3 are inputs, w_1, w_2, w_3 are inputs, y is output

Perceptron behavior can be described by the following function:

$$f(x, w, b) = \begin{cases} 1 & \text{if } \sum_{i=1}^n w_i x_i + b > 0 \\ 0 & \text{if } \sum_{i=1}^n w_i x_i + b \leq 0 \end{cases} \quad (1)$$

where w – is vector of weight, x – input’s vector and b – bias and $f(x, w, b)$ is perceptron’s output activation.

1 LITERATURE REVIEW

1.1 Introduction to a research area

Today large companies look for ways to introduce intellectual assistant that is capable to carry out various routine office tasks. These are some examples of these tasks: staff recruitment, training of new employees, answer to questions, paperwork, planning of time and resources [16]. Intellectual assistants based on machine learning (ML) allow computer programs to learn from data.

A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E [4].

Data from which computers learn is called training set and each training example is a training sample.

Machine learning is great for next problems:

- Problems for which existing tools and solutions are not optimized and needs a lot of code to be written and rules to set up. In these kind of problems ML could allow us to write simpler code and get better results.
- Complicated problems, for which no solution found yet, in perspective ML could be a solution.
- Problems on fluctuating environment. Using ML an adaptive to new data system which can learn during work, could be built.
- Problems with big data processing. Large amount of data could be optimized and processed using ML.

It is a fact that companies around the globe are using ML in their business, because it allows them with less investment to gain more profit and spend less time for the same task.

Chatbots are one of the representatives of modern business solutions. They are getting more popular these days, because they are able to work for 24 hours 7 days a week.

Reports provided by statisa.com [9], show that the number of users of messaging applications exceeded 2.2 billion. This number will be increasing constantly. The active distribution of instant messengers as a tool for easy and quick messaging began in 2015.

People today prefer using messengers, as they are more efficient and less time consuming for business processes and everyday tasks.

Most messenger's API's are providing a chatbot solution. The first chatbot was introduced to the world in the 1960s by MIT professor Joseph Weizenbaum. It was called ELIZA. It could breed a psychologists' speech by finding out keywords or either say "okay" and transfer the conversation to other direction if it didn't know the answer.

Chatbots today are very different from ELIZA. They can solve more tasks, use natural language processing (NLP) to understand commands and use ML to improve themselves. There are four types of chatbots:

1. Click navigation chatbots. These kind of chatbots work with preloaded «if — else» rules, created through a pre-designed conversation.
2. Keyword based chatbots. These chatbots use a huge list of keywords database. They respond to users' question using search-keywords algorithm, which means if the keyword is in question the corresponding answer will be given.
3. Natural language understanding chatbots. These chatbots are trained on relevant data using neural networks and they are able to understand natural language. They form answers by creating their own sentences whose style will be similar to the training set.
4. Aware chatbots. These are bots like JARVIS or FRIDAY from Marvel movies. For now, they exist only in movies.

Most of the existing chatbots have one common deficiency: they cannot be installed on a company's server, since they use messenger applications API, so that they work only with that messenger.

Here we build a natural language understanding chatbots. There are some challenges building a Russian language chatbot:

More than 50 % of the internet content is in English language while Russian segment is less than 10%. Giants like Google, Microsoft and IBM are presenting developer's tools mostly in English. That is why the accuracy of chatbots in English are higher than in any other languages.

The other problem is Acronyms. Acronym is a word or name formed as an abbreviation from the initial components of a phrase or a word, usually individual letters. NLTK library already contains list of acronyms.

1.2 Corporate tasks analysis and development tools

Corporate employees are facing a lot of tasks to solve. For making such a regular task like making a travel order, an employee should write about it to a business travels' office, they should find tickets and hotels and coordinate with top-brass, after that they reserve the ticket and the hotel, and send confirmation to the employee. All of these processes could be automated by a bot, which could do this task a lot faster. Other example is, if some references or inquiries needed, a chatbot could send it to printer and send message to the working stuff that a stamp is needed. All of this will save a lot time and finally it brings good profit to organizations.

To create a chatbot existing tools were examined. There are lots of tools created to build chatbots [6-10]:

Morph.ai [6] is a platform that allows us to create a business chatbot which uses WhatsApp, Facebook messenger or a website as a body of the bot. Such bots understand natural language. Data from such bot cannot be stored locally in organizations' server, which can bring to corporate information leak.

Flowxo position itself as a zero coding chatbot, which allows us to build chatbots as fast as it is possible. It also can be integrated with Gmail, LinkedIn and JIRA. It has the same flaw as a previous chatbot [7].

Telegram.API [13] allows us to build a telegram chatbot. It has lots of pluses, but the deficiency is that it uses Telegram as a bot platform.

API.AI [14] allows us to build a learning chatbot which can be adopted with Android, iOS, Node.js and HTML. But the main problem is in free version there are lots of limits such as the number of requests or messages. Also, the company cannot be sure about confidential information if it is safe.

Python [15] is a simple powerful programming language and it has lots of libraries which allow us to create a machine learning and natural language understanding chatbot, these tools are free to use. Python allows us to connect own database to the user which can be stored on the company's servers.

For the magister's work we have chosen Python as a programming language and work on Ubuntu OS:

- NumPy is an open-source module for Python that provides general mathematical and numerical operations in the form of pre-compiled, fast functions. They are combined in high-level packages. They provide functionality that can be compared with the functionality of MATLAB. NumPy provides basic methods working with large matrices and arrays.

- SciPy expands NumPy capabilities. It is also an open-source software for mathematics, science, and engineering, but it is more user-friendly and has more functionality for numerical optimization and integration. NumPy and SciPy are easy to install on all popular operating systems and they are free of charge [16].

- NLTK is a library which was created in 2001 for computational linguistics course in University of Pennsylvania. Now it is the most popular tool used for natural language processing especially in the field of machine learning. It contains all important functions for text preprocessing [19].

- TensorFlow is developed by Google. It contains popular algorithms of machine learning and deep learning. It is well documented and it is easy to use. It uses computational algebra and optimization methods.

- Six is a Python library that helps to work Python 2 code in Python 3.

- Recurrent neural network (RNN) is used if current data depends on previous ones. As we know the context of sentences depends on all words, so we have to use RNN.

– Seq2seq model. It consists of encoder and decoder. Encoder analyses users' input data, decoder creates answers similar to the training set.

For the chatbot we construct a neural network. In machine learning we do not use exact programming where we have to set all rules. We use training set to train a model. During training new rules would be generated automatically [3].

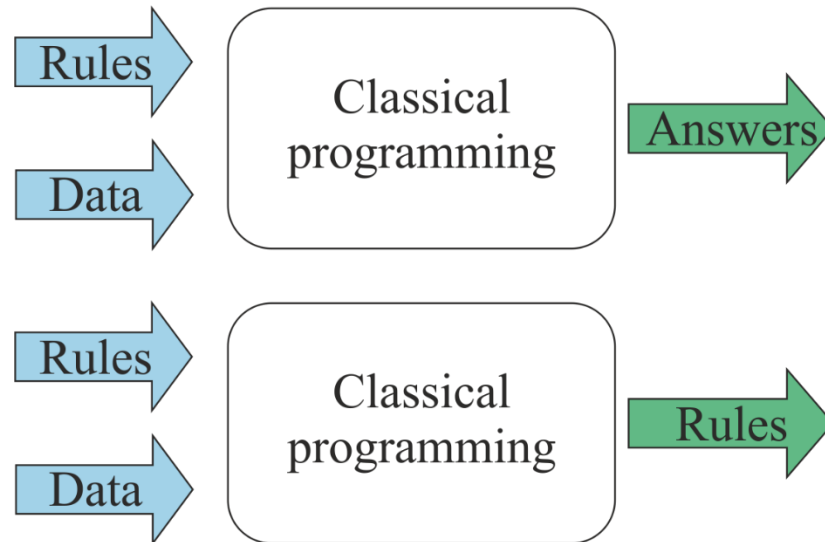


Figure 4. Machine learning concept

There are two ways to do it:

1. To build a neural network for each task. In our case these tasks are to interpret the language, to track the set of a conversation and to generate a response. Each of these systems will work separately to do its own task. It is unnecessarily complex to build and very time consuming to train.

2. The better type of building a neural network is called «End-to-End» Systems. These are chatbots that use one system that is trained on one dataset. They make no assumptions about the use case or the structure of a dialogue. It should be just trained on a relevant data. After training it will be able to chat with us about these data using natural language.

We use the second approach. In Figure 5 [21] one can see this approach.

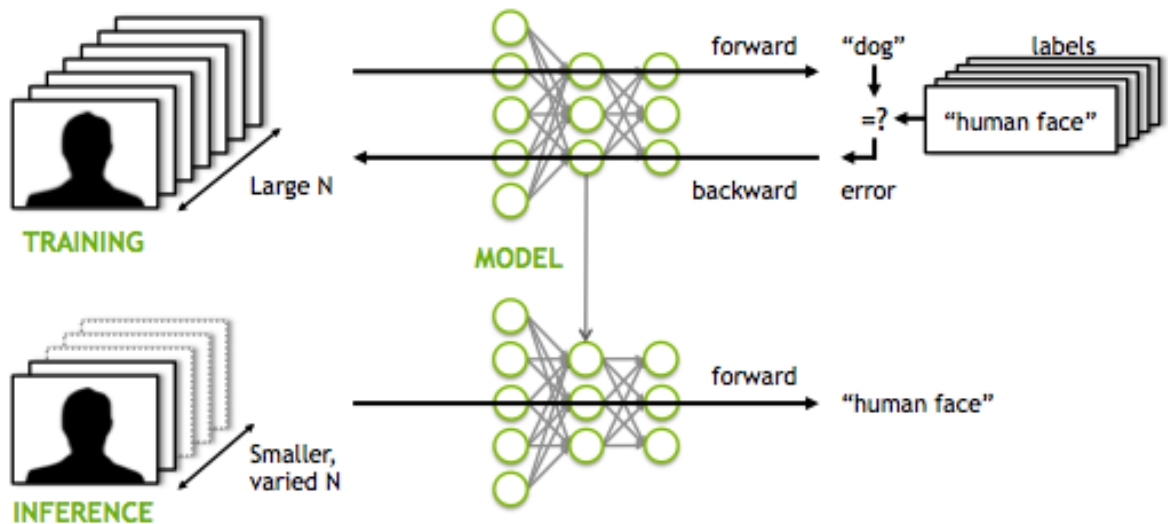


Figure5 – End-to-End Systems [18]

1.3 Dataset Preprocessing

First of all, we have to find and download a dataset. The dataset should be in forms of dialogues in Russian language. There are a lot of data on the internet in English. As we mentioned before, it is quite problematic to find data in Russian language. Luckily there is a dataset from Yandex.Toloka [22]. It consists of 10,000 dialogues.

In the second step we prepare dataset for machine learning. We should clear it and convert it to a vector. For this we use NLTK and pymorphy2 Python libraries. Data preparation consists of following steps:

- a) We use NLTK library for tokenization. We have to choose between whitespace tokenizer, punctuation tokenizer and TreeBankWord tokenizer. We have chosen whitespace tokenizer, because tests show higher accuracy in Russian language with it.
- b) For token normalization we use pymorphy2. It allows us to bring the word to the root form, which is known as lemma.
- c) Converting words to vectors by counting its Term Frequency – Inverse Term Frequency (TF-IDF) value. In this step we convert words to vectors. It allows us to

perform operations faster and consume less space. TF-IDF value is counted using formulas below:

$$TF = \frac{f_{f,d}}{\sum f_{t,d}} \quad (1)$$

$$IDF = \log \frac{N}{|\{d \in D: t \in d\}|} \quad (2)$$

$$TFIDF(t, d, D) = TF \cdot IDF \quad (3)$$

where TF is term frequency of the word

$f_{f,d}$ is frequency for term in document d ;

$\sum f_{t,d}$ is total number of words in the document;

IDF is inverse term frequency of the word;

D is number of documents in the collection;

$|\{d \in D: t \in d\}|$ is number of documents in D , which contains t .

The higher value of TF-IDF is better, which means the word with highest TF-IDF value is the main word in the user's query. In Python, to count TF-IDF value sklearn and pandas libraries are used.

The problem with TF-IDF value is if the dataset is huge it will be a big matrix to store and perform operations with. We can apply some restrictions to it, because if we look into the matrix, we will see that most of values are zeros. We will delete unnecessary values to make the chatbot learn and work more quickly.

Thirdly, we build our seq2seq model which consists of encoder and decoder. With encoder machine will understand what user wrote and follow the context of dialogue. Using decoder, a chatbot replies to a user by building its own sentences, which are similar to the training set.

Also, inside the model we split data into training and test set. Normally, it is 75% given to learning part and 25% to tests. Using test data, we can count accuracy of a chatbot by the formula below:

$$Acc = \frac{E \cdot 100}{Q} \quad (4)$$

1.4 Purpose and tasks research

So, in the result of the review of publications about chatbots development process, it can be concluded.

First, we specified and analyzed four types of chatbots. We looked into the history of chatbots development and chose the natural language understanding chatbot to develop.

Secondly, we analyzed technologies for chatbot development, data preprocessing and defined chatbot features for Russian language.

These results imply the need for creating a chatbot prototype with natural (Russian) language processing. The purpose of the research is to develop a cross-platform chatbot that can be integrated with corporate networks to improve productivity. For achieving the proposed purpose, we have such tasks:

- a) choice suitable stack technologies for creating a prototype chatbot;
- b) learn the chatbot on a training dataset;
- c) test the chatbot on formal requests including jargon words and orthographical mistakes.

2 THEORETICAL ASPECTS

2.1 Neural network

A neural network is a sequence of neurons interconnected by synapses. The structure of the neural network came into the world of programming directly from biology. Due to this structure, the machine acquires the ability to analyze and even memorize different information. Neural networks are also capable of not only analyzing incoming information, but also reproducing it from their memory.

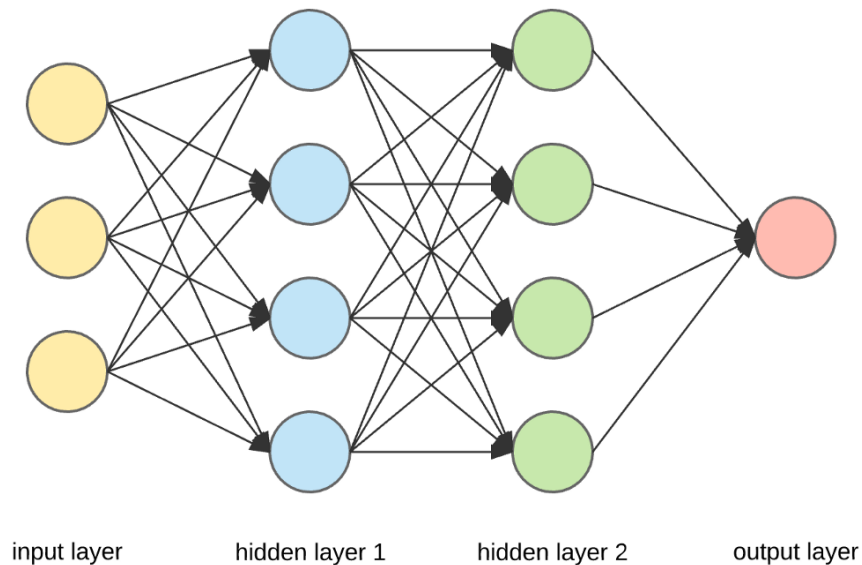


Figure6 – Neural network [8]

We use neural networks to build a chatbot, but first let's understand what is chatbot.

2.2 About chatbots

According to statista.com [9], last year the number of users of messaging applications exceeded 2.2 billion. This number will be increasing constantly.

But sending messages is not only a people's prerogative. Live chat buddies are easy to replace with chatbots. According to the Humanity in the Machine report, the vast majority of users are ready to receive business services with their help, because

this guarantees a quick and clear answer. A little less than a third of Americans called online communication the best way to contact the seller.

We can see in XXI century that a program can compete with human. Let us see from where chatbots came and what are they able to do now.

Chatbots are not new. First chatbot was presented to the world in 1966 by Joseph Weizenbaum in MIT. He was called "Eliza" and talked like a therapist: asked again, asked for more information and etc. Some at the first contact did not even realize that their interlocutor was a machine.

Chatbots today are very different from ELIZA. They can solve more tasks, use natural language processing (NLP) to understand commands and use ML to improve themselves.

Chatbots can be divided into 4 groups:

1. Click navigation chatbots. These kind of chatbots work with preloaded «if — else» logic, created through a pre-designed conversation.
2. Keyword based chatbots. These chatbots use a huge list of keywords database. They answer to the user question if the keyword is in question.
3. Natural language understanding chatbots. These types of chatbots use neural networks to understand natural language and to form answer to the users or do their tasks.
4. Aware chatbots. This kind of chatbots are not developed yet, since for now it is impossible. It is bots like JARVIS or FRIDAY from Marvel movies.

Existing chatbots have one common deficiency: they cannot be installed on a company's server, since they use Messenger's API, so that they work only with that messenger [10].

In this work we will build the third type of chatbot that will be easy to use in any corporate systems. During this work we will cover the following steps:

- a) analyze the existing tools for developing our chatbot;
- b) cover theoretical part;
- c) develop a console chatbot prototype;
- d) test chatbot by chatting and giving tasks.

2.3 Relevance of work

There is a great demand for chatbots. We can see that many companies are already using them and some companies are offering cases and hackathons to build a chatbot that satisfies their needs. Regular mobile phone users are also need a chatbot, because statistics show that most of the apps that are downloaded to user's smartphone is used once or twice. If you have a chatbot all these apps can exist inside it and most of apps can be deleted. We build a corporate chatbot that can be installed on any corporate local or global networks.

2.4 Analyze the existing tools for developing chatbot

To create a chatbot existing tools were examined. There are lots of tools created to build chatbots [11-15].

Morph.ai [11] is a platform that allows to create a business chatbot which uses WhatsApp, Facebook messenger or a website as a body of the bot. Such bots understand natural language. Data from such bot cannot be store on local network of the organization which can bring to corporate information leak.

Flowxo [12] position itself as a zero coding chatbot, which allows to build chatbots as fast as it is possible. It also can be integrated with Gmail, LinkedIn and JIRA. It has the same flaw as previous chatbot.

Telegram.API [13] allows to build a Telegram chatbot. It has lots of pluses, but the deficiency is that it uses Telegram as a bot platform.

API.AI [14] allows to build a learning chatbot which can be adopted with Android, iOS, Node.js and HTML. But the main problem is in free version there're lots of limits such as the number of requests or messages. Also, the company cannot be sure about confidential information if it is safe.

Python [15] is a simple yet powerful programming language and it has lots of libraries which allow to create a machine learning and natural language

understanding chatbot these tools are free to use. Python allows to connect own database to the user which can be stored on the company's servers.

In this work we will use Python programming language.

2.5 Development tools and libraries

To develop a chatbot we will use next libraries:

- NumPy,
- SciPy,
- NLTK,
- TensorFlow,
- Six.

NumPy is an open-source module for python that provides general mathematical and numerical operations in the form of pre-compiled, fast functions. They are combined in high-level packages. They provide functionality that can be compared with the functionality of MATLAB. NumPy provides basic methods working with large matrices and arrays.

The main feature of NumPy is the array object. Arrays functionality is very similar to objects called lists, the difference is that elements of the array must have the same data type, like int or float. Arrays perform numerical operation more efficient and faster than lists.

SciPy expands NumPy capabilities. It is also an open-source software for mathematics, science, and engineering, but it is more user-friendly and has more functionality for numerical optimization and integration. NumPy and SciPy are easy to install on all popular operating systems and they are free of charge. They are both are easy to use, but powerful packages [16].

NLTK is a library which was created in 2001 for computational linguistics course in University of Pennsylvania. Now it is the most popular tool used for natural language processing especially in the field of machine learning. It contains all important functions for text preprocessing [19].

TensorFlow is developed by Google. It contains popular algorithms of machine learning and deep learning. It is well documented and it is easy to use. It uses computational algebra and optimization methods.

TensorFlow has the following features:

- It uses multidimensional arrays which are known as tensors for mathematical calculations;
- It supports machine learning and deep learning methods;
- It contains data sets scaling algorithms;
- It uses GPU for computing.

Six is a Python library that helps to work Python 2 code in Python 3.

There are two ways to build a chatbot using neural network:

1. We create a system that consist of three deep learning neural network. One for understanding natural language, another one to trace conversation context or so called Long-Short Term Memory (LSTM) and the third one is to generate answers. All neural networks will be trained separately, which means more time will be elapsed, more computer resources and 3 different datasets are needed. Indeed, it is unnecessary complex work to do.

2. There is simpler way to build a chatbot. It is to build so called “end-to-end” system (Figure 7). These kind of chatbots are trained on only one dataset and is needed only one network to build which would proceed input in natural language, trace the context and generate the answer.

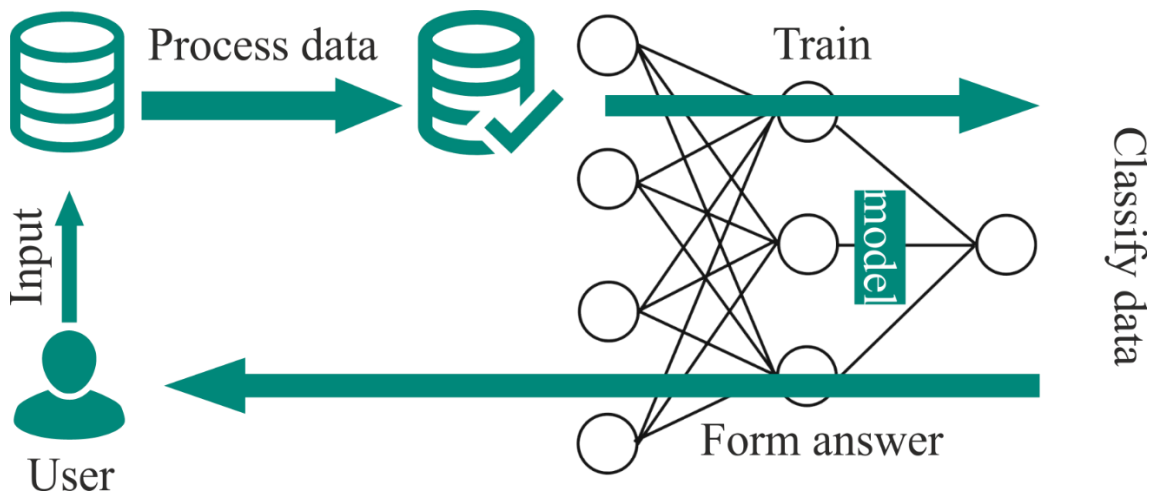


Figure 7 – End-to-end system [18]

The simplest type of neural network is one directional, which means it proceeds data from input layer all the way to the output layer. Only data with the fixed size can be imported into this kind of network. It could either be a number or an image. For example, it can be a labeled dataset of temperatures. The output of this neural network can be the prediction of the given word as hot or cold.

As we know conversation cannot be a fixed size. It is a sequence that consists of words. To proceed this kind of data we need a neural network that can except sequences as an input. This kind of neural network is recurrent neural network.

Recurrent Neural Networks are mostly used for natural language processing. First, they count random sentences based on how often they appear in the texts in all documents of the corpus. This would be a measure of semantic and grammatical correctness. This kind of models are used in translation. Finally, this language models generate new text. For example, model trained on Lord of the ring dataset will generate text looking like that movie.

RNN uses information consistently, which means it will follow the context of the sentences. In traditional neural networks all data is analyzed independently, which will not give the best result. Different words translated without context could be meaningless. To predict the next word of the sentences we have to use RNN, because it allows to analyze previous words. Recurrent neural networks perform the same task again and again to each element until the sentence ends, the output layer depends on previous objects. RNNs are also known as the networks that have “memory”.

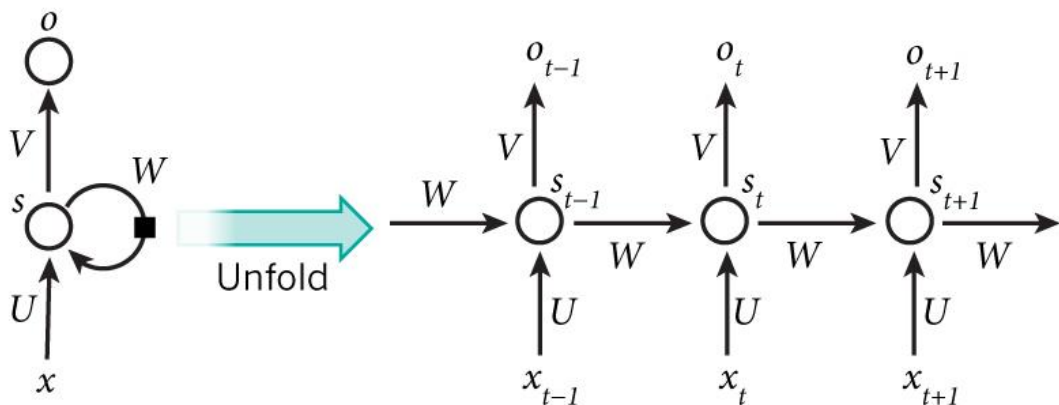


Figure 8 – Recurrent neural network unfolded. x is input, s is hidden state step (or memory), o – output, W, U, V – matrices.

In Figure 8 we can see an unfold of the RNN. Unfolded means we show all the network action. For example, if our sentences have 6 words it will be 6 layers, a layer for each word. Formulas that used in the RNNs:

x_t – input in step t,

s_t – it is hidden state in step t, we can say that it is memory. s_t as a function depends on previous states and current state of input x_t : $s_t = f(Ux_t + Ws_{t-1})$. Function $f()$ is usually non-linear, for example, $\tan h$. s_{-1} is used to count the first hidden state, usually initiated with vector of zeros.

o_t – output layer on step t. It could be the layer of varieties if we want to predict a word.

We will use RNNs in TensorFlow. Our steps will be:

1. Download the relevant dataset to train the chatbot.
2. Create a training model.
3. Train the model using the training model
4. Test it by chatting.

2.6 Dataset

Finding dataset is the important and one of difficult part of the research, because there are only few datasets in Russian language on the Internet. There is an open dataset from Yandex.toloka to train chatbots which consist of 10 thousand dialogs [22].

2.7 A training model

In our end-to-end model we use encoder and decoder, where encoder is used from one side of the conversation, which means it will process natural language and the decoder is from the other side of conversation, which will generate answers. First, we have to split our data into two training sets. One for encoding and the other one for the decoding.

2.8 Natural language processing

Computers are capable of efficiently and quickly processing any structured data. In the natural language, there is no strict structuring and typing. This leads to the fact that it is difficult for computers to process incoming information if it does not comply with strict rules.

Each language can distinguish its own rules and structures, but this is not enough for quality processing. Moreover, if we consider the texts compiled by ordinary Internet users, you can find a huge number of errors, both syntactic and spelling.

All this leads to problems of natural language processing (NLP). A separate area in artificial intelligence with the same name NLP deals with these problems.

For full text processing it is necessary to conduct its full analysis. At the same time, not all theoretical developments have yet been implemented in practice, which leads to certain difficulties. The main problem is the complexity of the parsing of the text. The difficulty lies in the fact that natural language serves not only for the formal processing of information, but also for the living imagination. Often, the analysis of voluminous texts is a difficult task not only for computers, but also for humans, since most dialogues have some kind of real or fictitious foundation.

For a complete analysis of the text, the finished system must analyze the text in terms of syntax (sentence structure), semantics (concepts used in the text) and pragmatics (the correct use of concepts and the purposes of their use). After analysis, the system should generate its answer, suitable for inference.

A complete system, with the possibility of full analysis, should have the following modules: Graphemic analysis is a system of primary analysis natural text, preparing information for subsequent processing (morphological and syntactic analysis). Typically, graphemic analysis breaks down the source text into various components: highlighting paragraphs and headings, dividing text into words, highlighting the surname of the middle name (if the name and middle name are

written with initials), email addresses, highlighting sentences, highlighting steady turns and etc.

Morphological analysis - allows us to determine the normal form of words. This is necessary in order to process the text based on the normal form, and not different word forms.

Pre-syntactic analysis - is used to combine individual lexical units into one syntactic or, conversely, its division into several. Variable inextricable phrases are combined into one syntactic unit (for example, “beat the buck” It is especially necessary to deal with word division, for example, in the German language, where several arbitrary interconnected words can be combined into one complex “on the fly”, and it is not possible to place all such combinations into a morphological analysis.

Parsing is the hardest part of text analysis. Here it is necessary to determine the roles of words and their relationship with each other. The result of this step is a set of trees showing such connections. The task is complicated by a huge number of alternative options that appear during the parsing, related to both the ambiguity of the input data (the same word form can be obtained from different normal forms) and the ambiguity of the parsing rules themselves.

Post syntactic analysis serves two purposes. On the one hand, it is necessary to clarify the meaning embedded in the words and expressed using various means of the language: prepositions, prefixes or affixes that create one or another word form. On the other hand, the same thought can be expressed by different constructions of the language. Semantic analysis analyzes the text “in meaning”. On the one hand, semantic analysis clarifies the connections that post-syntactic analysis could not clarify, since many roles are expressed not only by means of the language, but also taking into account the meaning of the word. On the other hand, semantic analysis allows you to filter out some word meanings or even whole parsing variants as “semantically disconnected” [17].

3 PROGRAMMING A CHATBOT

3.1 Text analysis steps

All the queries from user is processed with NLTK library. Firstly we “Tokenize” incoming data. It is the process of dividing the text into words, sentences or a paragraph. This units in tokenization known as tokens. We can use TreeBankWord, punctuation or whitespace tokenizer, which are built in to NLTK library. Tests show that whitespace tokenizer is the best for the Russian language. For higher accuracy we first use sentence tokenizer, then whitespace tokenizer.

Next, we should normalize our tokens. For this aim we can either use lemmatization or stemming. Stemming is finding the root form of the word by removing or replacing suffixes or prefixes. The root form of the word called stem. Lemmatization is the same process, but it uses vocabulary and morphological analysis. It returns the dictionary form of the word which is known as lemma. For some tasks it is good to use lemmatization, but for some stemming. NLTK library shows not good results for stemming and lemmatization for non-English languages. Its accuracy in Russian language is around 15%, which is unacceptable. For this reason, we should look for the analog of NLTK. Another library for natural language processing which shows higher accuracy is pymorphy2.

The pymorphy2 library is text analyzer which can:

1. Bring the word to the normal form;
2. Put the word in the right form in the sentences;
3. Return grammatical information about the word.

Tests shown the accuracy about 61%. It is not good enough, but it is best of the available libraries for natural language processing for Russian language.

After this we should find main words in the user’s query. It could be done with counting TF-IDF (term frequency-inverse document frequency) value. It is the count of numerical statistic value of the word in a sentence, document collection and language corpus at all. It can be counted as follows [20]:

$$TF = \frac{f_{f,d}}{\sum f_{t,d}} \quad (2)$$

$$IDF = \log \frac{N}{|\{d \in D: t \in d\}|} \quad (3)$$

$$TFIDF(t, d, D) = TF \cdot IDF \quad (4)$$

where TF is term frequency of the word

$f_{f,d}$ – frequency for term in document d ;

$\sum f_{t,d}$ – total number of words in the document;

IDF – inverse document frequency of the word;

D – number of documents in the collection;

$|\{d \in D: t \in d\}|$ – number of documents in D , which contains t .

The higher value of TF-IDF is better, which means the word with highest TF-IDF value is the main word in the user's query. In Python to count TF-IDF value sklearn and pandas libraries are used. A code fragment is in listing 1.

Listing 1 – Counting TF-IDF value in Python programming language

```
>from sklearn.feature_extraction.text importTfidfVectorizer
>import pandas as pd
>text=input().split()
>tfidf=TfidfVectorizer(min_df=2, max_df=0.5,ngram_range(1,2))
>features=tfidf.fit.transform(texts)
>pd.DataFrame(features.todense(),columns=tfidf.get_feature_names())
```

Counting TF-IDF value directly will slow down chatbots work. For example, if we have an input from user and as a result, we have matrix of words whose TF-IDF values would be searched and compared with the corpus TF-IDF values which consists of 25 000 rows and 75 000 columns. That is a huge matrix and if we look to its values 99.8% of them are zeros. As we can see if we go this way it will be an irrational usage of resources. It is better to try to predict TF-IDF value using machine

learning. We could also put some restrictions to read only non-zero values. The machine learning will predict the probability in a given diapason using logistic regression. Responses would be grouped by TF-IDF values.

3.2 Building a model

As text is preprocessed and ready, we can build our seq2seq model.

One of the most popular architectures in machine translation is the sequence to sequence (Seq2Seq) models [3]. Such models consist of two recurrent networks: an encoder and a decoder. The encoder builds a representation of the input word sequence. Further, the resulting representation (the last output and the value of the network cell) is copied to the decoder. According to the received representation, the decoder tries to restore the target sequence of words. In machine translation tasks, input and output sequences are sentences in different languages. In question-answer and dialogue systems - a question and an answer.

To convert words to input vectors, the so-called embedding matrix is used. The number of rows in this matrix is equal to the size of the dictionary, and the number of columns is equal to the size of the LSTM cell. Each line corresponds to a vector representation of the corresponding word. Each word before being fed to the input of the LSTM network is replaced with the corresponding row of the representation matrix.

The decoder receives a special <GO> symbol at the first step of the input, then at each step the word generated in the previous iteration is fed. The response generation continues until a special word is generated - the end of line marker <EOL>. During training, the target symbol is transmitted as the generated symbol to the next measure, and the distribution on the predicted symbols is transferred to the loss function.

During the prediction, it is required to find the most probable proposal from the point of view of the model. It is impossible to do this directly, since the model allows you to calculate only the best word with fixed previous ones. The trade-off

between greedy word picking and exhaustive search is Vem Search. When using this method, a small number of the best candidates are selected at each iteration, and the remaining hypotheses are discarded.

After applying Beam Search, the network often begins to respond using the most common answers in the sample: **yes, no, I don't know**. To combat this, two models can be trained [4]: a predictive answer on a question and a question on an answer. The sentences generated by the beam search of the first model are rearranged according to a convex combination of the logarithms of the likelihood of the two models:

$$\lambda \log P (A | Q) + (1 - \lambda) \log P (Q | A) \quad (5)$$

where Q is the question;

A is the answer;

$\lambda \in [0, 1]$ is a hyperparameter.

Now frequency responses will have a low probability, since it is rarely possible to reconstruct a question from them. To generate long sentences, a penalty is added to the ranking function, encouraging the generation of a large number of words. Usually, $\gamma |A|$, where $|A|$ is the number of generated words, and γ is the hyperparameter.

3.3 Building and testing a chatbot

First of all, we have to install Anaconda. Anaconda is the programming environment for Python which contains all necessary tools to get started.

Anaconda allows to create different environment for each project. Since we use Ubuntu Linux operation system, we create environment using terminal by using the following command:

```
>conda create -n chatbot python=3.5 anaconda
```

which means create environment with name “chatbot” using anaconda and install on it Python version 3.5.

Next, we have to activate the environment:

```
>conda activate chatbot
```

We have to install all the needed libraries next. For this we will use pip command. There are 2 package managers in Python pip and conda. They both work the same way.

```
>pip install tensorflow==0.12.1
```

```
>pip install numpy
```

We have to open **Jupyter notebook** to start programming by typing the same in the terminal:

```
>jupyter notebook
```

Now we are ready to start programming on Python. Firstly, we have to import all libraries:

```
> import tensorflow as tf
```

```
> import numpy as np
```

```
> import time
```

```
> import re
```

```
>import nltk
```

Next, we have to prepare our data to work with seq2seq model. As we had analyzed before we have to open data, clean it, tokenize it, then normalize tokens and count TF-IDF values.

To clean data we need to remove unnecessary symbols, so we will write a function for it:

```
> def clear_text(text):
```

```
>     text = text.lower()
```

```
>     text = re.sub (r "[-()\"]\`~<>${#!?|}]", "", text)
```

```
>     return text
```

Now we have a function that cleans text, so we would use it:

1) We clean questions:

```
>clear_questions = []
```

```
> for question in questions:
>     clear_questions.append(clear_text(question))
```

2) We clean answers

```
> clear_answers = []
> for answer in answers:
>     clear_answers.append(clear_text(answer))
```

After all of this operation we have to make it easier for the model to learn from dataset. For this we have to clean less frequent words. After this we have to add some tokens to the end of dictionaries.

```
> tokens = ['<PAD>', '<EOS>', '<OUT>', '<SOS>']
> for token in tokens:
>     words_of_questions_to_int[token]=len(words_of_questions_to_int)+1
> for token in tokens:
> words_of_answers_to_int[token]=len(words_of_answers_to_int)+1
```

As we mentioned before these tokens will help the model to understand the beginning and the end of sentences.

Next, we have to build our seq2seq model. As we know it consists of encoder and decoder. It will be a translation model (Figure 9). In the Figure we can see the translation model from English to French, where a user enters any data and encoder processes and understands it step by step. Then decoder gives the translation. Instead of English to French translation we will put Russian to Russian, which means both encoder and decoder will work with the same language. We will also have to follow the dialog -context. That's why we used the LSTM (Long-short term memory).

First, we build Encoder with LSTM included as follows:

```
> def encoder (input_rnn, size_rnn, layers_num, prob, seq_len):
>     lstm = tf.contrib.rnn.BasicLSTMCell(size_rnn)
>     lstm_dropout = tf.contrib.DropoutWrapper(lstm, prob_input = prob)
>     cell_encoder = tf.contrib.rnn.MultiRNNCell
([lstm_dropout]*layers_num)
```



```

> output_encoder, state_encoder = tf.nn.bidirectional_dynamic_rnn
(cell_fw = cell_encoder, cell_bw = cell_encoder, seq_len = seq_len, inputs =
input_rnn, dtype = tf.float32)
> return state_encoder

```

Next, we build decoder:

```

> def decoder(inbuilt_decoder_input, inbuilt_decoder_matrix,
state_encoder, words_num, seq_len, size_rnn, layers_num, w2int, prob, batch_size):
> with tf.variable_scope('decoding') as scope_decoding:
> lstm = tf.contrib.rnn.BasicLSTMCell(size_rnn)
> dropout_lstm = tf.contrib.rnn.DropoutWrapper(lstm, input_prob=prob)
> cell_decoder = tf.contrib.rnn.MultiRNNCell([dropout_lstm] *
layers_num)
> weights = tf.truncated_normal_initializer(stddev=0.1)
> biases = tf.zeros_initializer()
> output = lambda x: tf.contrib.layers.fully_connected(x, num_words,
None, scope=scope_decoding, weights_initializer=weights, biases_initializer=biases)
> predict_training = decode_training_set(state_encoder, cell_decoder,
inbuilt_decoder_input, seq_len, scope_decoding, output_function, prob, batch_size)
> decoding_scope.reuse_variables()
> test_predictions = decode_test_set(state_encoder, cell_decoder,
inbuilt_decoder_matrix, w2int['<SOS>'], w2int['<EOS>'], seq_len - 1, words_num,
scope_decoding, output_function, prob, batch_size)
> return training_predictions, test_predictions

```

Now we run out the model. After training the model we can finally test our chatbot. The result can be seen in Figure 10. Testing results showed that the probability of recognizing formalized user requests reaches 90% and higher.

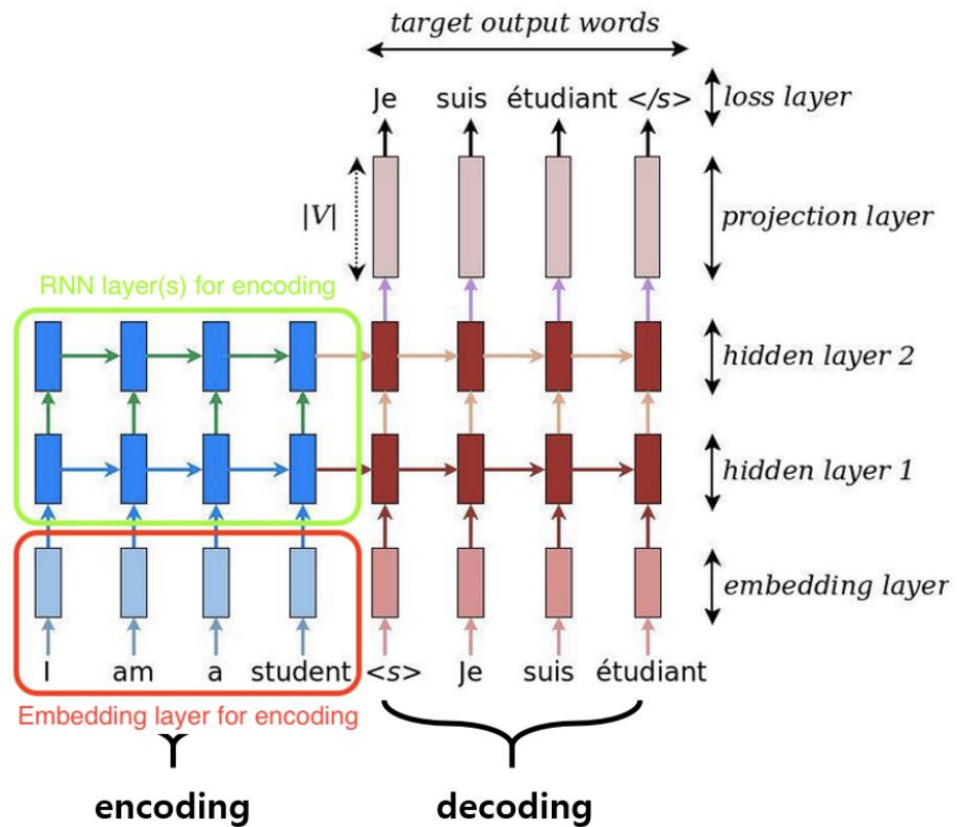


Figure 9 – Seq2Seq model

```

File Edit View Insert Cell Kernel Help
+ ✂ 📄 📄 ⬆ ⬇ ▶ Run 🔄 ⏪ Code 🗨
In [18]: 1 response('привет')
        Рад вас снова видеть!

In [19]: 1 classify('в какое время вы работаете?')
Out[19]: [('hours', 0.99916303)]

In [20]: 1 response('в какое время вы работаете?')
        Мы работает каждый день с 9:00 до 21:00

In [21]: 1 classify('Вы принимаете карты?')
Out[21]: [('payments', 0.99686974)]

In [23]: 1 response('Вы принимаете карты?')
        Мы принимаем VISA, Mastercard и МИР

In [24]: 1 classify('пока')
Out[24]: [('goodbye', 0.9846563)]

In [25]: 1 response('пока')
        Хорошего дня!

```

Figure10 – Illustrative example of a chatbot. Lines 19, 21, 24 probability of a question belonging to a particular topic

Since mobile applications are popular these days, chatbot can be integrated into a mobile app (Figure11). Python can be used for web programming. We created back-end part in this work. Front-end part of application can be created using web frameworks for Python like Django or Flask and the mobile app can be created using android studio for Android smartphones or XCode for apple devices.

Also, we created a website widget (Figure12) and webpage for our chatbot. Widget can be informative for people visiting companies' websites and can give needed information to the visitor or connect to the operator if needed. The widget prototype was created using **Django (Flask)** software.

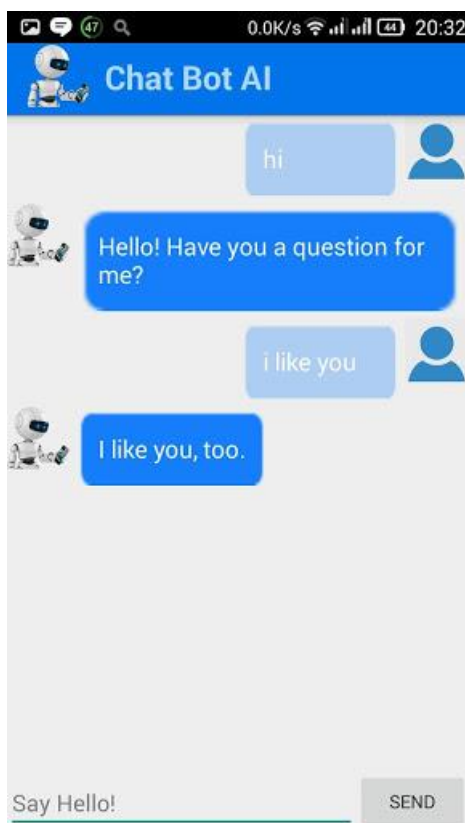


Figure11 – Chatbot mobile application

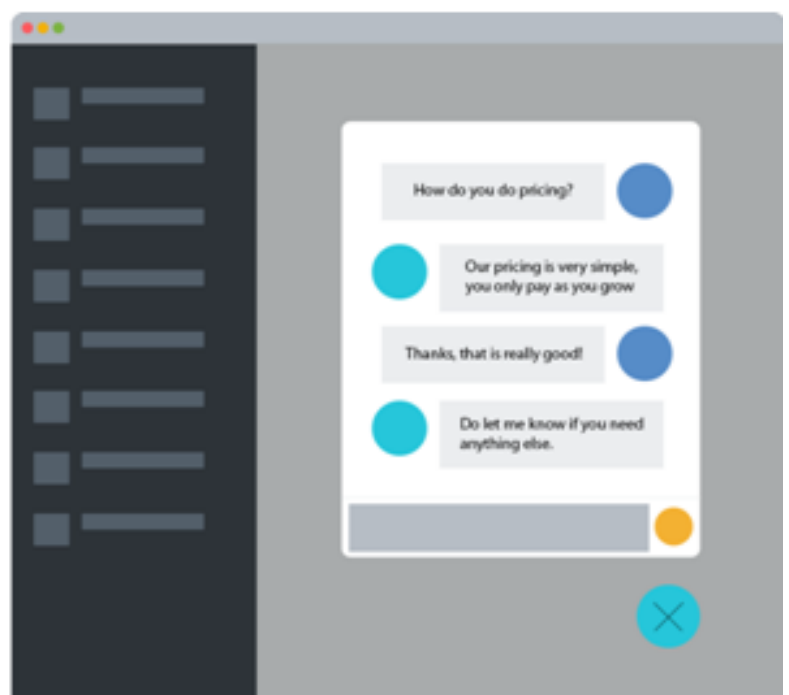


Figure12 – Chatbot widget for a website

4 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

В этом разделе указаны вредные факторы негативно влияющие на организм человека, такие как электромагнитное излучение, неоптимальный микроклимат помещения, недостаточность освещения и электрический ток. Также указан характер вредного воздействия данных факторов на организм человека и последствия их длительного и чрезмерного воздействия.

Также рассмотрены вопросы, связанные с организацией рабочего места и условий, в которых будет реализовываться проект, описанный в данной работе, а именно, обработка данных полученных в полученных на рабочем месте, в соответствии с нормами производственной санитарии, техники безопасности и охраны труда и окружающей среды.

Условия труда оказывают прямое воздействие на здоровье человека и его состояние в процессе работы, что обуславливает их отвечать всем требованиям безопасности и санитарно-гигиеническим требованиям. Все факторы, воздействующие на работника в процессе осуществления трудовой деятельности, принято разделять на два типа: вредные и опасные производственные факторы.

Рабочим местом является аудитория 427а 10-ого корпуса Томского политехнического университета по адресу проспект Ленина 2, рабочей зоной является стол с компьютером.

В рамках данной работы разработан чат-бот с обработкой естественного языка. Он может быть использован в предприятиях для облегчения рутинных задач, таких как: отправка писем по электронной почте, электронный документооборот, поиск в интернете и тд.

4.1 Правовые и организационные вопросы обеспечения безопасности

4.1.1 Эргономические требования к рабочему месту

В процессе работы, все используемые предметы должны находиться в зоне досягаемости. Оптимальное размещение предметов труда и документации в зонах досягаемости приведено на рисунке 13.

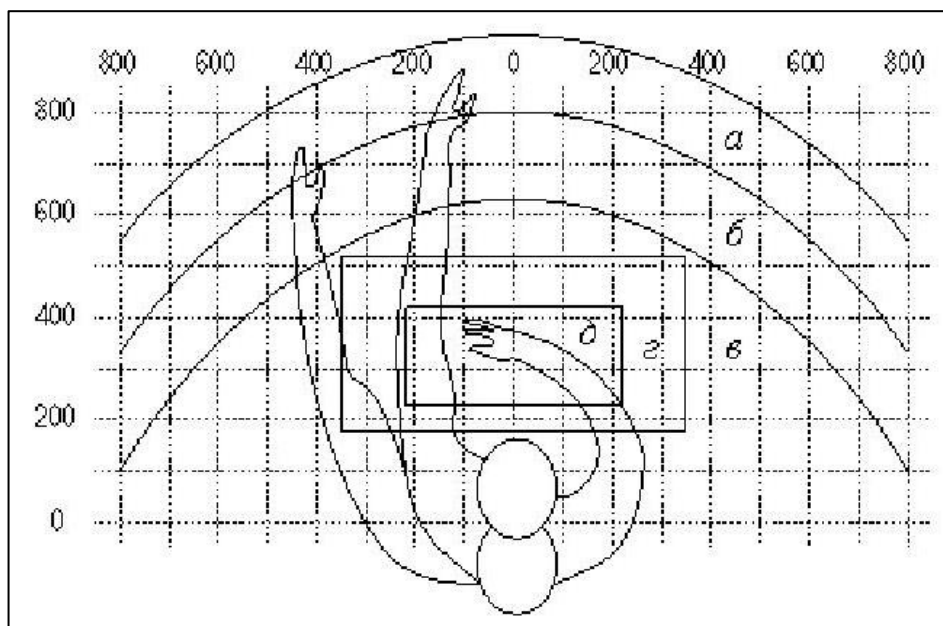


Рисунок 13 –Зона досягаемости.

- а – зона максимальной досягаемости;
- б – зона досягаемости пальцев при вытянутой руке;
- в – зона легкой досягаемости ладони;
- г – оптимальное пространство для грубой ручной работы;
- д – оптимальное пространство для тонкой ручной работы [23].

4.1.2 Особенности законодательного регулирования проектных решений

В соответствии с государственными стандартами и правовыми нормами обеспечения безопасности предусмотрена рациональная организация труда в течение смены, которая предусматривает:

- длительность рабочей смены не более 8 часов;
- установление двух регламентируемых перерывов (не менее 20 минут после 1-2 часов работы, не менее 30 минут после 2 часов работы);
- обеденный перерыв не менее 40 минут.

Обязательно предусмотрен предварительный медосмотр при приеме на работу и периодические медосмотры.

Каждый сотрудник должен пройти инструктаж по технике безопасности перед приемом на работу и в дальнейшем, должен быть пройден инструктаж по электробезопасности и охране труда.

Предприятие обеспечивает рабочий персонал всеми необходимыми средствами индивидуальной защиты. Таким образом, по итогу проделанной работы можно сделать вывод о том, что рабочее место инженера соответствует нормам, ГОСТам и СанПиН. По степени физической тяжести работа инженера относится к категории лёгких работ.

4.2 Производственная безопасность

При работе с электрическими приборами возникают различные вредные факторы, которые негативно воздействуют на организм человека. Например, в результате постоянной физической или моральной нагрузки человек испытывает сильное напряжение, в результате чего возникает хронический стресс. В свою очередь, стресс способен повлиять на жизнедеятельность организма, то есть, ухудшить работу органов дыхания, сердечно-сосудистой системы, пищеварения, желез внутренней секреции. Данная проблема часто рассматривается в рамках психофизиологии.

Также, при работе с электрическими приборами возникают вредные и опасные факторы, их перечень приведен в таблице 1.

Таблица 1–Опасные и вредные факторы при работе с электроприборами

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные документы
	Разработка	Изготовление	Эксплуатация	
Отклонение показателей микроклимата в помещении	-	-	+	СП 2.2.4.548-96 [24]
Недостаток естественного освещения	-	-	+	СП 52.13330.2016 [25]
Недостаточная освещенность рабочей зоны	-	-	+	
Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека	-	-	+	СанПин 2.2.4.3359-16 [26] ГОСТ 12.1.019-2017 [27] ГОСТ 12.1.004-91 ССБТ [28]
Повышенная напряженность электромагнитного поля	-	-	+	ГОСТ 12.1.010-76 ССБТ [29]

4.2.1 Отклонение показателей микроклимата в помещении

Микроклимат является важной характеристикой производственных помещений. В организме человека происходит непрерывное выделение тепла. Одновременно с процессами выделения тепла происходит непрерывная теплоотдача в окружающую среду. Равновесие между выделением тепла и теплоотдачей регулируется процессами терморегуляции, т.е. способностью организма поддерживать постоянство теплообмена с сохранением постоянной температуры тела. Отдача тепла происходит различными видами: излучением, конвекцией, испарение влаги.

Нарушение теплового баланса в условиях высокой температуры может привести к перегреву тела, и как следствие к тепловым ударам с потерей сознания. В условиях низкой температуры воздуха возможно переохлаждение организма, могут возникнуть простудные болезни, радикулит, бронхит и другие заболевания. Показателями, которые характеризуют микроклимат рабочей зоны, являются:

- температура воздуха, °С;
- относительная влажность воздуха, %;
- скорость движения воздуха, м/с.

Оптимальные значения этих характеристик зависят от сезона (холодный, тёплый), а также от категории физической тяжести работы. Для математика и программиста она является лёгкой (1а), так как работа проводится сидя, без систематических физических нагрузок.

Оптимальные показатели микроклимата рабочей зоны, согласно СанПиН 2.2.4.548 [30], представлены в таблице 5, допустимые – в таблице 2, допустимые в таблице 3.

Таблица 2 – Оптимальные показатели микроклимата

Период года	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Теплый	23-25	40-60	0,1
Холодный	22-24	40-60	0,1

Таблица 3 – Допустимые показатели микроклимата

Период года	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движение воздуха, м/с	
	Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин		Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин
Теплый	21,0-22,9	24,1-25,0	15-75	0,1	0,2
Холодный	20,0-21,9	25,1-28,0	15-75	0,1	0,1

Мероприятия по доведению микроклиматических показателей до нормативных значений включаются в комплексные планы предприятий по охране труда. Это такие мероприятия, как:

- механизация и автоматизация производственных процессов, дистанционное управление ими;
- применение технологических процессов и оборудования, исключающих образование вредных веществ или попадания их в рабочую зону;
- установка систем вентиляции, кондиционирования, отопления.

К мероприятиям по оздоровлению воздушной среды в помещении относятся правильная организация вентиляции и кондиционирования воздуха, отопление помещений. Вентиляция может осуществляться естественным и

механическим путём. В зимнее время в помещении необходимо предусмотреть систему отопления.

По степени физической тяжести работа программиста относится к категории лёгких работ. Таким образом, делаем вывод о том, что рабочее место программиста соответствует нормам показателей микроклимата так, как есть соответствие температурным показателям в помещении в холодный период – 23,3 °С, и в теплый период – 24,6 °С. Относительная влажность воздуха составляет 58%, что соответствует диапазону.

4.2.2 Недостаток естественного освещения

Помещения с постоянным пребыванием людей должны иметь естественное освещение. Без естественного освещения допускается проектировать помещения с временным пребыванием людей, помещения, которые определены соответствующими сводами правил и стандартами организаций на проектирование зданий и сооружений, а также помещения, размещение которых разрешено в подвальных этажах зданий и сооружений.

В учебных и учебно-производственных помещениях общеобразовательных организаций, интернатов, профессиональных образовательных организаций при одностороннем боковом освещении нормируемое значение коэффициента естественной освещенности (КЕО) должно быть обеспечено в расчетной точке, расположенной на пересечении вертикальной плоскости характерного разреза помещения и условной рабочей поверхности на расстоянии 1,2 м от стены, наиболее удаленной от световых проемов и иметь значение не более 0,5% [31].

Равномерность естественного освещения не нормируется для производственных помещений с боковым освещением.

В аудитории имеется одно окно с одной стороны, чего достаточно для этого помещения, так как в аудитории имеется искусственное освещение.

4.2.3 Недостаточная освещенность рабочей зоны

Производственное освещение является неотъемлемым элементом условий трудовой деятельности человека. При правильно организованном освещении рабочего места обеспечивается сохранность зрения человека и нормальное состояние его нервной системы, а также безопасность в процессе производства. Производительность труда и качество выпускаемой продукции находятся в прямой зависимости от освещения.

При неудовлетворительном освещении ощущается усталость глаз и переутомление, что приводит к снижению работоспособности. В ряде случаев это может привести к головным болям. Головные боли также могут быть вызваны пульсацией освещения, что в основном является результатом использования электромагнитных пускорегулирующих аппаратов (ПРА) для газоразрядных ламп, работающих на частоте 50 Гц.

Для характеристики естественного освещения используется коэффициент естественной освещенности (КЕО). Величины КЕО для различных помещений лежат в пределах 0,1-12%.

Освещённость на рабочем месте должна соответствовать характеру зрительной работы; равномерное распределение яркости на рабочей поверхности и отсутствие резких теней; величина освещения постоянна во времени (отсутствие пульсации светового потока); оптимальная направленность светового потока и оптимальный спектральный состав; все элементы осветительных установок должны быть долговечны, взрыво-, пожаро-, электробезопасны.

Работа с приборами относится к зрительным работам средней точности для помещений жилых и общественных зданий. Согласно СП 52.13330.2016 [25], такие помещения должны удовлетворять требованиям, указанным в таблице 4.

На рабочем месте соблюдаются необходимые нормы освещенности согласно ГОСТ Р 55710-2013[32]. «Освещение рабочих мест внутри зданий.

Нормы и методы измерений». В административных зданиях, а в частности в помещениях для чтения, письма и обработки данных эксплуатационная освещенность должна составлять 500 ЛК, что соответствует показателям.

Таблица 4 – Требования к освещению помещений жилых и общественных зданий при зрительной работе средней точности

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Разряд зрительной работы	Подряд зрительной работы	Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %	Искусственное освещение		Естественное освещение, КЕО ен, %, при	
					Освещенность на рабочей поверхности и от системы общего освещения, ЛК	Коэффициент пульсации освещенности Кп, %, не более	Верхнем или комбинированном	Боковом
Высокой точности	Более 0,5	В	1	Не менее 70	150	20	2,0	0,5
			2	Менее 70	100	20	2,0	0,5

4.2.4 Повышенное значение напряжения в электрической цепи, замыкание которой может пройти через тело человека

Измерительные приборы должны подключаться к сети питания, имеющие защитное заземление. Соединять и разъединять вилки, розетки электрических соединений допускается только при выключенном сетевом выключателе.

Основными причинами поражения человека электрическим током могут быть следующие:

- непосредственное прикосновение к токоведущим частям, оказавшимся под напряжением;
- соприкосновение с конструктивными частями, оказавшимися под напряжением.

Электрический ток, проходя через организм человека, оказывает тепловое (ожоги, нагрев сосудов), механическое (разрыв тканей, сосудов при

судорожных сокращениях мышц), химическое (электролиз крови), биологическое (раздражение и возбуждение живой ткани) или комбинированное воздействие.

Основными средствами и способами защиты от поражения электрическим током являются:

- недоступность токоведущих частей для случайного прикосновения;
- защитное заземление, зануление или отключение;
- вывешивание предупреждающих надписей;
- контроль за состоянием изоляции электрических установок;
- использование дополнительных средств защиты.

Требования электробезопасности электроустановок производственного и бытового назначения на стадиях проектирования, изготовления, монтажа, наладки, испытаний и эксплуатации, а также технические способы и средства защиты, обеспечивающие электробезопасность электроустановок различного назначения приведены в ГОСТ Р 12.1.019-2017 [27].

4.2.5 Повышенный уровень электромагнитных излучений

В качестве источника электромагнитного излучения в данной работе рассматриваются средства измерения и электрические приборы.

Степень и характер воздействия ЭМП на организм человека зависят:

- от интенсивности излучения;
- частоты колебаний;
- поверхности тела облучаемого;
- индивидуальных особенностей организма;
- режима облучения (непрерывный или прерывистый)

продолжительности воздействия;

- комбинированного действия других факторов производственной среды.

В диапазонах промышленной частоты, радиочастот, инфракрасного и частично ультрафиолетового света электромагнитные поля оказывают тепловое воздействие. Перегревание отдельных тканей и органов ведёт к их заболеваниям, а повышение температуры тела на 1 °С и выше вообще не допустимо.

Влияние электромагнитных излучений заключается не только в их тепловом воздействии. Микропроцессы, протекающие в организме под действием излучений, заключаются в поляризации макромолекул тканей и ориентации их параллельно электрическим силовым линиям, что может приводить к изменению свойств молекул; особенно для человеческого организма важна поляризация молекул воды.

Длительное и систематическое воздействие на человека полей ВЧ и УВЧ вызывает:

- повышенную утомляемость;
- головную боль;
- сонливость;
- гипертонию;
- боли в области сердца.

Длительное и систематическое воздействие на человека полей СВЧ вызывает, кроме того:

- изменения в крови;
- катаракту (помутнение хрусталика глаза);
- нервно-психические заболевания.

Согласно СанПиН 2.2.4.3359-16 [26], предельно допустимые уровни энергетических экспозиций электромагнитного поля диапазона частот ≥ 30 кГц - 300 ГГц, не должны превышать значений, приведенных в таблице 5.

Для понижения уровня напряженности электромагнитного поля следует использовать мониторы с пониженным уровнем излучения, устанавливать защитные экраны, устранять неисправности. А для профилактики

компьютерного зрительного синдрома, улучшения визуальных показателей видеомониторов, повышения работоспособности, снижения зрительного утомления возможно применение защитных очков со спектральными фильтрами, разрешенных Минздравом России для работы с персональным компьютером.

Таблица 5–Предельно допустимые уровни энергетической экспозиции

Диапазоны частот	Предельно допустимые уровни энергетической экспозиции		
	По электрической составляющей, (В/М) 2 х ч	По магнитной составляющей, (А/М) 2 х ч	По плотности потока энергии (мкВт/см) 2 хч
30 кГц-3 МГц	20 000,0	200,0	-
3-30 МГц	7 000,0	-	-
30-50 МГц	800,0	0,72	-
50-300 МГц	800,0	-	-
300 МГц-300ГГц	-	-	200,0

На рабочем месте программиста используются мониторы с пониженным уровнем излучения. Таким образом можно сделать вывод о том, что рабочее место соответствует нормам СанПиН 2.2.4.3359-16 [26].

4.3 Экологическая безопасность

В настоящее время, когда встает проблема рационального использования природных ресурсов, охраны окружающей среды, уделяется большое внимание организации разумного воздействия на природу. Необходимо совершенствовать технологические процессы с целью сохранения окружающей среды от вредных выбросов.

В связи с тем, что основным средством работы являются средства измерения и электрические приборы, серьезной проблемой является электропотребление. Это влечет за собой общий рост объема потребляемой электроэнергии. Для удовлетворения потребности в электроэнергии, приходится увеличивать мощность и количество электростанций. Соответственно, рост энергопотребления приводит к таким экологическим

нарушениям, как глобальное потепление климата, загрязнение атмосферы и водного бассейна Земли вредными и ядовитыми веществами, опасность аварий в ядерных реакторах, изменение ландшафта Земли. Целесообразным является разработка и внедрение систем с малым потреблением энергии. В аудитории не ведется никакого производства. К отходам, производимым в помещении, можно отнести бытовой мусор. Сточные воды здания относятся к бытовым сточным водам. За их очистку отвечает городской водоканал. Основным видом мусора – это отходы печати, бытовой мусор (в т. ч. люминесцентные лампы), неисправное электрооборудование, коробки от техники, использованная бумага. Утилизация отходов печати вместе с бытовым мусором происходит в порядке, описанном в ФЗ от 24.06.1998 №89-ФЗ (ред. от 25.12.2018) «Об отходах производства и потребления» [33] и в распоряжении Правительства РФ от 25.07.2017 №1589-р [34] осуществляется сотрудниками лаборатории и предусматривает следующие этапы:

1. Правильное заполнение акта списания с указанием факта невозможности дальнейшей эксплуатации перечисленной в акте измерительной техники, о чем имеется акт технического осмотра;

2. Осуществление списания перечисленной в акте измерительной техники с баланса предприятия с указанием в бухгалтерском отчете, так как утилизация возможна для осуществления только после окончательного списания;

3. Непосредственно утилизация измерительной техники с полным демонтажем устройств на составляющие детали с последующей сортировкой по видам материалов и их дальнейшей передачей на перерабатывающие заводы.

4.4 Безопасность в чрезвычайных ситуациях

Чрезвычайными ситуациями в подобных помещениях могут быть пожары. Основы пожарной безопасности определены по ГОСТ 12.1.004-91 [27] и ГОСТ 12.1.010-76 [28].

Все производства по пожарной опасности подразделяются на 5 категорий: А, Б, В, Г, Д. Аудитория, в которой будет выполняться работа, относится к категории В.

Причинами пожара могут быть:

- токи короткого замыкания;
- электрические перегрузки;
- выделение тепла, искрение в местах плохих контактов при соединении проводов;
- курение в неположенных местах.

Тушение горящего электрооборудования под напряжением должно осуществляться имеющимися огнетушителями ОУ-5. Чтобы предотвратить пожар в аудитории, необходимо:

- содержать помещение в чистоте, убирать своевременно мусор. По окончании работы поводится влажная уборка всех помещений;
- работа должна проводиться только при исправном электрооборудовании;
- на видном месте должен быть вывешен план эвакуации из помещения с указанием оборудования, которое нужно эвакуировать в первую очередь;
- уходящий из помещения последним должен проверить выключены ли нагревательные приборы, электроприборы и т.д. и отключение силовой и осветительной электрической сети.

Также необходимо соблюдение организационных мероприятий:

- правильная эксплуатация приборов, установок;
- правильное содержание помещения;
- противопожарный инструктаж сотрудников аудитории;
- издание приказов по вопросам усиления ПБ;
- организация добровольных пожарных дружин, пожарно-технических комиссий;

- наличие наглядных пособий и т.п.

В случаях, когда не удастся ликвидировать пожар самостоятельно, необходимо эвакуироваться вслед за сотрудниками по плану эвакуации и ждать приезда специалистов, пожарников. При возникновении пожара должна сработать система пожаротушения, издав предупредительные сигналы, и передав на пункт пожарной станции сигнал о ЧС, в случае если система не сработала, по каким-либо причинам, необходимо самостоятельно произвести вызов пожарной службы по телефону 101, сообщить место возникновения ЧС и ожидать приезда специалистов.

4.5 Выводы по разделу «Социальная безопасность»

Таким образом можно сделать вывод по всему разделу «Социальная ответственность», что на рабочем месте показатели микроклимата, освещенности и электромагнитных излучений находятся в норме, что говорит о безопасности для человека. Анализ выявленных опасных факторов рабочего помещения показал, что электробезопасность, пожаробезопасность и экологическая безопасность находятся под контролем, и все необходимые меры для обеспечения безопасности принимаются.

5 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

В научно-исследовательских проектах важную роль играет его коммерциализация. Разработчикам и исследователям следует учитывать такие факторы как его экономическая эффективность, конкурентоспособность и экологичность, так как спонсоры и доноры обращают на это особое внимание. Рентабельность проекта зависит от её себестоимости, времени на реализацию и спрос.

В данном разделе рассматривается раздел коммерческая выгода НИОКР и расчет затрат, необходимых для её проведения.

5.1 Предпроектный анализ

5.1.1 Потенциальные потребители результатов исследования

Разработчик заинтересован в монетизации его проекта, причём проект должен обеспечить прибыль в достаточном объеме. Поэтому перед разработкой необходимо чёткое представление о том, кто является конечным потребителем, т.е. целевой рынок и сегментировать его.

Целевой рынок – сегменты рынка, на котором будет продаваться в будущем разработка. Сегмент рынка – это особым образом выделенная часть рынка, группы потребителей, обладающих определенными общими признаками. Сегментирование – это разделение покупателей на однородные группы, для каждой из которых может потребоваться определенный товар или услуга [35].

В ходе данной работы был разработан чат-бот, который позволяет автоматизировать большинство процессов для сотрудников предприятий, что позволяет существенно сократить время при выполнении повседневных задач, таких как электронный документооборот, обучение сотрудников, отправка писем с электронной почты и так далее.

Целевым рынком чат-ботов являются коммерческие организации. Критериями сегментирования в нашем случае выступают компании и требуемый для них функционал. Результат сегментирования представлен на рис. 14.

		По требуемому функционалу		
		Чат-бот для поиска информации и обработки простых запросов на естественном языке	Чат-бот с расширенным функционалом (отправка сообщений по электронной почте, заполнение документов и электронный документооборот)	Чат-бот с расширенными возможностями и гибкой настройкой всех функций.
Размер компании	Крупные			
	Средние			
	Мелкие			

Рисунок 14 – Карта сегментирования рынка чат-ботов

5.1.2 Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения

Нужно провести анализ существующих конкурирующих разработок на рынке. Так как рынки постоянно развиваются этот анализ должен быть систематическим. Анализ подобных разработок позволяет более эффективно конкурировать с оппонентами путем внесения корректировок в проект. Важна реалистичная оценка конкурентов. Для этого нужно использовать всю информацию об их разработках такие как:

- Технические характеристики разработки
- Конкурентоспособность
- Завершенность проекта

- Бюджет
- Уровень проникновения на рынок
- Финансовое положение

Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения. Анализ проведен в таблице 6, где конкурент 1 – это компания «Singularika» [36], и конкурент 2 – компания «Крок» [37].

Анализ конкурентных технических решений определяется по формуле:

$$K_i = \sum_{i=1}^n B_i B_i \quad \text{б)}$$

где K – конкурентоспособность разработки или конкурента; B_i – вес i -го показателя (в долях единицы); B_i – балл i -го показателя.

Таблица 6 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес	Баллы			Конкурентоспособность		
		B_{ϕ}	$B_{к1}$	$B_{к2}$	K_{ϕ}	$K_{к1}$	$K_{к2}$
Технические критерии оценки ресурсоэффективности							
1. Способствует увеличению скорости выполнения рутинных задач	0,15	5	4	4	0,75	0,6	0,6
2. Удобство в эксплуатации (соответствует требованиям потребителей, все функции внутри одного приложения)	0,13	5	5	5	0,65	0,65	0,65
3. Безопасность	0,10	4	4	4	0,4	0,4	0,4
4. Надежность	0,10	4	3	3	0,4	0,3	0,3
5. Простота эксплуатации (понятный интерфейс и функции)	0,09	5	4	4	0,45	0,36	0,36
Экономические критерии оценки эффективности							
6. Цена	0,10	5	3	4	0,5	0,3	0,4
7. Конкурентоспособность продукта	0,09	4	4	4	0,36	0,36	0,36
8. Предполагаемый срок эксплуатации	0,09	5	5	5	0,45	0,45	0,45
9. Срок выхода на рынок	0,08	4	5	5	0,32	0,4	0,4
10. Послепродажное обслуживание	0,07	5	3	3	0,35	0,21	0,21
Итого	1	46	40	41	4,63	4,03	4,13

Таким образом, конкурентоспособность нашей разработки составила 4,63, в то время как у «Singularika» и «Крок» этот показатель составил 4,03 и 4,13 соответственно. Результаты показывают, что данная разработка является конкурентоспособной и имеет преимущества по некоторым критериям.

5.1.3 SWOT-анализ

SWOT-анализ – это комплексный стратегический анализ научно-исследовательского проекта, которая применяется для исследования внутренних и внешних среды проекта. SWOT – это английская аббревиатура, которую можно расшифровать как Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы).

Анализ проводится в несколько этапов. На первом этапе нужно определить эти сильные и слабые стороны, возможности и угрозы относительно нашего проекта.

На втором этапе нужно создать несколько интерактивных матриц и посмотреть различные взаимосвязи на шаге 1. В интерактивной матрице ставится знак «+», если одни характеристики сильно соответствуют другим (например, соответствие сильных сторон возможностям), знак «-», если эта связь слабая и «0», если имеются сомнения. Интерактивные матрицы приведены в таблицах 7-10.

Третий этап – это делать логические выводы исходя из интерактивных таблиц. Результаты приведены в таблице 11.

Таблица 7 – Интерактивная матрица «Сильные стороны и возможности»
Таблица 8 – Интерактивная матрица «Слабые стороны и возможности»

		Сильные стороны		
		C1	C2	C3
Возможности	B1	+	+	+
	B2	+	0	+
	B3	+	+	-

		Слабые стороны		
		Сл1	Сл2	Сл3
Возможности	B1	+	+	+
	B2	-	0	+
	B3	+	-	+

Таблица 9 – Интерактивная матрица «Сильные стороны и угрозы»

		Сильные стороны		
		C1	C2	C3
Угрозы	У1	0	0	0
	У2	+	-	+
	У3	+	-	-
	У4	-	+	-

Таблица 10 – Интерактивная таблица «Слабые стороны и угрозы»

		Слабые стороны		
		C1	C2	C3
Угрозы	У1	0	+	0
	У2	+	+	+
	У3	+	+	+
	У4	+	+	+

Таблица 11 – SWOT-анализ, первый этап.

	<p>Сильные стороны (С):</p> <ol style="list-style-type: none"> Опытный научный руководитель; Продукт, реализованный на языке Python и бесплатных библиотеках машинного обучения; Чат-бот с LSTM (Longshort-termmemory – долгая краткосрочная память); 	<p>Слабые стороны (Сл):</p> <ol style="list-style-type: none"> Малый опыт работы разработчика на реальных проектах; Трудоёмкий процесс разработки и тестирования; Большие затраты времени на вычисления и обучение модели;
<p>Возможности (В):</p> <ol style="list-style-type: none"> Актуальность разработки; Ориентация на крупных потенциальных покупателей; Относительная легкость внедрения продукта; 	<p>V1C1C2C3 – позволят установить относительно невысокую цену на актуальный продукт;</p> <p>V2C1C3 – помогут удовлетворить потребности крупных покупателей.</p> <p>V3C1C2 – легко внедряемое современное решение в корпоративные сети.</p>	<p>V1Сл1Сл2Сл3 – разработка ведется на достаточно известном языке программирования Python, то на большинство вопросов уже имеются ответы в документациях библиотек, а также примеры использования.</p> <p>V2Сл3 – обычно крупные компании имеют большие сервера, которые смогут быстрее справиться с большими массивами данных.</p>
<p>Угрозы (У):</p> <ol style="list-style-type: none"> Возможный экономический кризис после пандемии COVID19; Дополнительные государственные требования при получении лицензии на продукт; Возможная уязвимость ПО Измениться поведение ПО при использовании больших массивов данных 	<p>У2С1С3 – опытный научный руководитель сможет помочь в продвижении продукта.</p> <p>У3С1 – программное обеспечение может быть проверено на наличие уязвимостей научным руководителем.</p> <p>У4С2 – так как используется инструменты с открытым исходным кодом, то на большинство вопросов, возникающих в процессе разработки, можно найти ответы в сети интернет.</p>	<p>У1Сл2 – реализовать проект в условиях пандемии ещё сложнее.</p> <p>У2Сл1Сл2Сл3 – наличие опыта работы в реальных проектах снизило бы количество времени затрачиваемую на получение лицензии на конечный продукт.</p> <p>У3Сл1Сл2Сл3 – перечисленные сложности могут способствовать появлению уязвимостей в ПО.</p> <p>У4Сл1Сл2Сл3 – вышеуказанные сложности могут способствовать изменению поведения данных.</p>

В рамках проведения SWOT-анализа были изучены сильные и слабые стороны, а также возможности и возможные угрозы, которые необходимо учесть при реализации проекта.

5.1.4 Оценка готовности проекта к коммерциализации

Независимо от того на какой стадии находится программное обеспечение (научная разработка) важно оценить готовность его к коммерциализации и выяснить уровень собственных знаний для её проведения (или завершения). На таблице 12 приведены оценки степени готовности проекта к коммерциализации, где оценка проработанности трактуется следующим образом:

- 5 – имеется положительное заключение независимого эксперта;
- 4 – выполнено качественно;
- 3 – выполнено, но под сомнением;
- 2 – проработано слабо;
- 1 – не проработано.

В таблице также приведены оценки уровня знаний у разработчика, где баллы трактуются так:

- 5 – знаком с теорией, выполняет, может консультировать;
- 4 – знаком с теорией и самостоятельно выполняет;
- 3 – знаком с теорией и практическими примерами применения;
- 2 – знаком с теорией;
- 1 – не знаком или мало знаком.

Оценка готовности научного проекта к коммерциализации (или уровень знаний) определяется по формуле:

$$B_{\text{сум}} = \sum B_i \quad (7)$$

где $B_{\text{сум}}$ – это суммарное количество баллов по каждому направлению; B_i – балл по каждому показателю.

Таблица 12 – Оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1	Определен имеющийся научно-технический задел	4	4
2	Определены перспективные направления коммерциализации научно-технического задела	5	5
3	Определены отрасли и технологии (товары, услуги) для предложения на рынке	5	5
4	Определена товарная форма научно-технического задела для представления на рынок	2	2
5	Определены авторы и осуществлена охрана их прав	3	3
6	Проведена оценка стоимости интеллектуальной собственности	5	5
7	Проведены маркетинговые исследования рынков сбыта	2	2
8	Разработан бизнес-план коммерциализации научной разработки	1	1
9	Определены пути продвижения научной разработки на рынок	4	4
10	Разработана стратегия (форма) реализации научной разработки	3	3
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	2	3
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	4	4
13	Проработаны вопросы финансирования коммерциализации научной разработки	4	4
14	Имеется команда для коммерциализации научной разработки	4	4
15	Проработан механизм реализации научного проекта	4	4
Итого		52	54

Значение $B_{\text{сум}}$ позволяет говорить о том, на сколько готовы проект и разработчики к коммерциализации. Так как их значения выше среднего можно делать вывод о том, что у проекта перспективы выше среднего.

5.1.5 Методы коммерциализации результатов научно-технического исследования

При коммерциализации научно-технических разработок продавец (а это, как правило, владелец соответствующих объектов интеллектуальной собственности), преследует вполне определенную цель, которая во многом зависит от того, куда в последующем он намерен направить (использовать, вложить) полученный коммерческий эффект. Это может быть получение средств для продолжения своих научных исследований и разработок (получение финансирования, оборудования, уникальных материалов, других научно-технических разработок и т.д.), одноразовое получение финансовых ресурсов для каких-либо целей или для накопления, обеспечение постоянного притока финансовых средств, а также их различные сочетания.

Правильный выбор метода коммерциализации может существенно сократить время на продвижение товара. Для данного проекта оптимальным решением будет инжиниринг. Инжиниринг – это самостоятельный вид коммерческих операций предполагает предоставление на основе договора инжиниринга одной стороной, именуемой консультантом, другой стороне, именуемой заказчиком, комплекса или отдельных видов инженерно-технических услуг, связанных с проектированием, строительством и вводом объекта в эксплуатацию, с разработкой новых технологических процессов на предприятии заказчика, усовершенствованием имеющихся производственных процессов вплоть до внедрения изделия в производство и даже сбыта продукции [35].

5.2 Инициация проекта

На этом этапе принято выполнить определенные работы для успешного запуска проекта. Здесь следует определить цели и результаты проекта, определить рабочую группу, определить ограничения и допущения. Вся эта информация заносится в документ, называемый уставом проекта.

Уставом проекта называется документ, который определяет бизнес-потребности и текущее понимание потребностей заказчика, а также новый продукт, услугу или результат, который планируется создать.

Цели и результат проекта. Здесь необходимо определить цели и результаты проекта, а также, определяются внутренние и внешние заинтересованные стороны проекта, которые в конечном итоге повлияют на итоговый результат проекта. Информацию по заинтересованным сторонам проекта приведены в таблице 13, а цели проекта, ожидаемые результаты, критерий приемки и требования к результату проекта на таблице 14.

Таблица 13 – Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Крупные предприятия, которые обладают корпоративными сетями и манипулируют большими массивами данных.	Внедрение чат-бота за короткие сроки, автоматизация рутинных задач.

Таблица 14 – Цели и результаты проекта

Цель проекта:	Разработка чат-бота с LSTM с обработкой естественного языка встраиваемый в корпоративные сети предприятий.
Ожидаемый результат:	Разработанный для внедрения в корпоративные сети предприятий.
Критерии приемки результата проекта:	- Точность обработки запросов на естественном языке свыше 85%; - Функции взаимодействия с пользователями, в том числе: отправка писем по электронной почте, поиск в интернете и другое.
Требования к результату проекта:	Требование:
	1. Отсутствие угрозы утечки информации корпоративных пользователей;
	2. Необходимое количество времени для внедрения в корпоративные сети;
	3. Надежность ПО (отказоустойчивость, не конфликтность с другими приложениями).

Организационная структура проекта. На данном этапе необходимо определить, кто будет входить в рабочую группу данного проекта, определить роль каждого участника, а также прописать функции, выполняемые каждым из них и их трудозатраты в проекте. Эта функция представлена в таблице 15.

Таблица 15 – Рабочая группа проекта

ФИО, основное место работы, должность	Роль в проекте	Функции в проекте	Трудозатраты, час.
Семенов М.Е. доцент, к.ф.м.н, отделение экспериментальной физики ТПУ	Руководитель проекта	Координатор	28
Исмоилов Н.И., магистрант	Исполнитель по проекту	Разработчик	546

В ходе реализации научного проекта, помимо магистранта задействован руководитель проекта:

Руководитель проекта – отвечает за реализацию проекта в пределах заданных ограничений по ресурсам, координирует деятельность участников проекта. В рамках данного проекта эту роль выполняет руководитель магистерской диссертации.

Исполнитель по проекту – специалист, выполняющий отдельные работы по проекту. В рамках этого проекта исполнителем является магистрант.

Ограничения и допущения проекта. Ограничения проекта – это факторы, которые ограничивают действия членов команды, а также границы проекта, за рамки которых выходить нельзя. Эти ограничения приведены в таблице 16.

Таблица 16 – Ограничения проекта

Фактор	Ограничения/допущения
Бюджет проекта	109 051,36 руб.
Источник финансирования	НИИ ТПУ
Сроки проекта	27.01-31.05.2020
Дата утверждения плана управления проектом	27.01.2020
Дата завершения проекта	31.05.2020
Прочие ограничения и допущения	Ограниченные вычислительные мощности для проведения расчетов

5.3 Планирование управления научно-техническим проектом

Группа процессов планирования состоит из процессов, осуществляемых для определения общего содержания работ, уточнения целей и разработки последовательности действий, требуемых для достижения данных целей [35].

План управления должен состоять из следующих элементов:

- Иерархическая структура проекта
- План проекта
- Бюджет проекта
- Организационная структура проекта
- План управления коммуникациями проекта
- Реестр рисков проекта

5.3.1 Иерархическая структура проекта

В Иерархической структуре проекта структурируется и определяется все содержание проекта. На рис. 1 построена иерархическая структура научно-исследовательского проекта «Разработка чат-бота с обработкой естественного языка».

5.3.2 План проекта

В рамках планирования проекта нужно построить календарный и сетевой графики проекта.

Линейный график представлена в виде таблицы (таблица 17).

Диаграмма Ганта – это, по сути, гистограмма, которая визуализирует календарный план проекта, на котором работы по теме отображаются в виде шкалы времени. Диаграмма проекта построена на рис. 16.

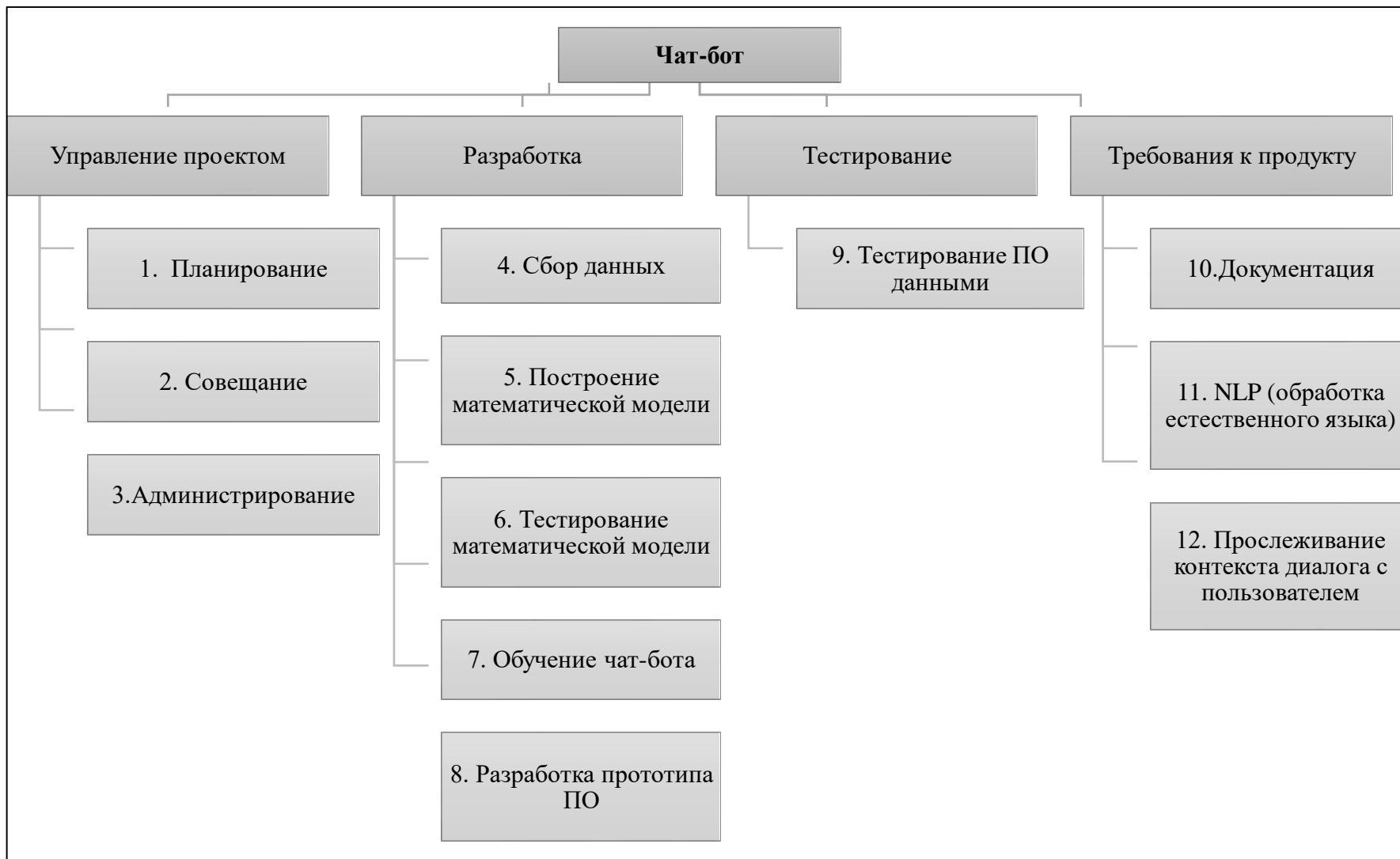


Рисунок15 – Иерархическая структура «Разработка чат-бота с обработкой естественного языка»

Таблица 17 – Календарный план проекта

Код работы (из ИСР)	Название	Длительность, дни	Дата начала работ	Дата окончания работ	Состав участников (ФИО ответственных исполнителей)
1	Планирование	5	27.01.2020	31.01.2020	Семенов М.Е. Исмоилов Н.И.
2	Совещание	1	3.02.2020	3.02.2020	Семенов М.Е. Исмоилов Н.И.
4	Сбор данных	14	4.02.2020	21.02.2020	Исмоилов Н.И.
5	Построение математической модели	14	24.02.2020	12.03.2020	Исмоилов Н.И.
6	Тестирование математической модели	7	13.03.2020	23.03.2020	Исмоилов Н.И.
7	Обучение чат-бота	21	24.03.2020	20.04.2020	Исмоилов Н.И.
8	Разработка прототипа ПО	5	21.04.2020	27.04.2020	Исмоилов Н.И.
9	Тестирование ПО с тестовыми данными	5	28.04.2020	04.05.2020	Семенов М.Е. Исмоилов Н.И.
10	Документирование	29	21.04.2020	29.05.2020	Исмоилов Н.И.
Итого:		90	27.01.2020	29.05.2020	

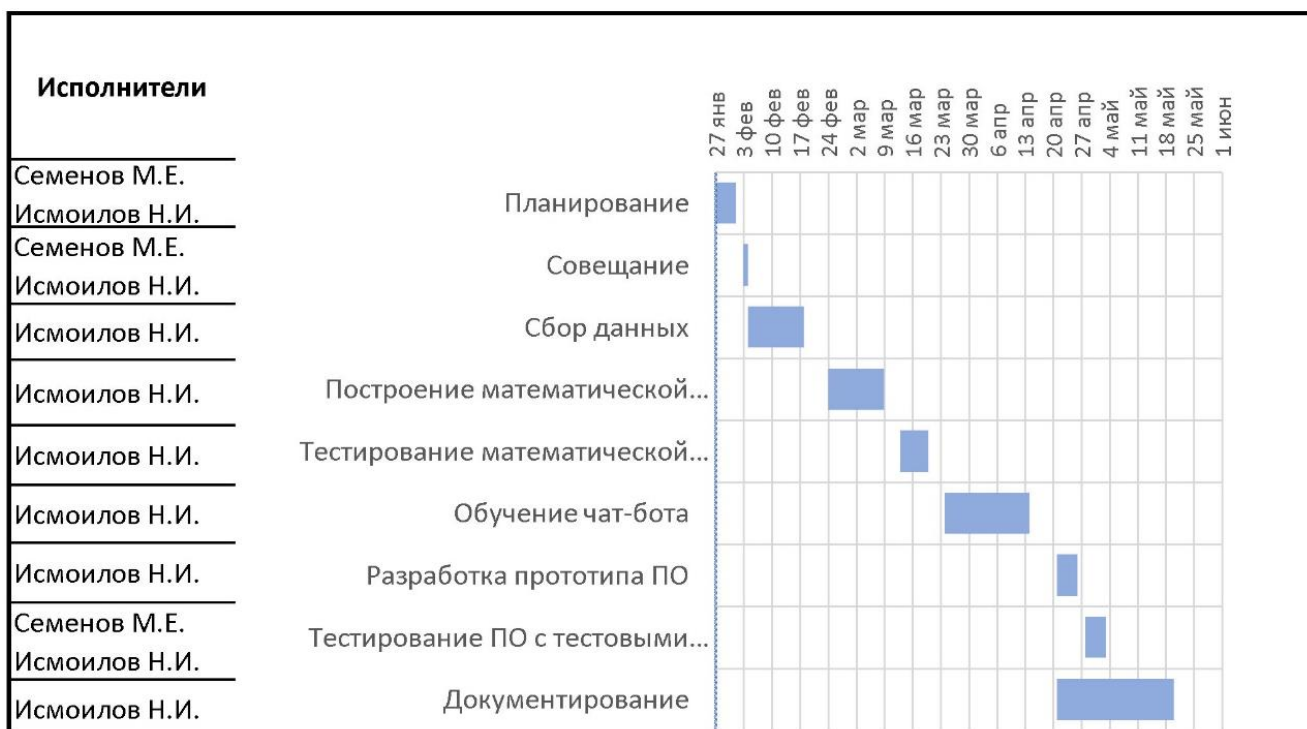


Рисунок 16 – Диаграмма Ганта

5.3.3 Бюджет научного исследования

При планировании бюджета необходимо указать полные и достоверные данные о планируемых расходах, необходимого для выполнения проекта.

Специальное оборудование для научных(экспериментальных)работ

В данную статью включают все затраты, связанные с приобретением специального оборудования (приборов, контрольно-измерительной аппаратуры, стендов, устройств и механизмов), необходимого для проведения работ по конкретной теме. Определение стоимости спецоборудования производится по действующим прейскурантам, а в ряде случаев по договорной цене. Расчет затрат на спецоборудование приведен в таблице 18.

Таблица 18– Расчет затрат по статье «Спецоборудование для научных работ»

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб.
1.	Ноутбук Dell Inspiron 5558	1	32500	32500

Основная заработная плата

Статья включает основную заработную плату работников, непосредственно занятых выполнением проекта, (включая премии, доплаты) и дополнительную заработную плату.

$$C_{зп} = Z_{осн} + Z_{доп}, \quad (8)$$

где $Z_{осн}$ – основная заработная плата; $Z_{доп}$ – дополнительная заработная плата.

Основная заработная плата ($Z_{осн}$) руководителя от университета:

$$Z_{осн} = Z_{дн} * T_{раб}, \quad (9)$$

где $Z_{осн}$ – основная заработная плата работника; $T_{раб}$ – продолжительность работ, выполняемых научно-техническим работником, раб. дн.; $Z_{дн}$ – среднедневная заработная плата работника, руб.

Средняя заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} * M}{F_{\text{д}}}, \quad (10)$$

где $Z_{\text{м}}$ – месячный должностной оклад работника, руб.; M – количество месяцев работы без отпуска в течении года: при отпуске 48 раб. дня за полгода $M = 10,4$ месяца, 6-дневная неделя; $F_{\text{д}}$ – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{ТС}} * (1 + k_{\text{пр}} + k_{\text{д}}) * k_{\text{р}}, \quad (11)$$

где $Z_{\text{ТС}}$ – заработная плата по тарифной ставке, руб.; $k_{\text{пр}}$ – премиальный коэффициент, равный 0,3 (т.е. 30% от $Z_{\text{ТС}}$); $k_{\text{д}}$ – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5 (в НИИ и на промышленных предприятиях – за расширение сфер обслуживания, за профессиональное мастерство, за вредные условия: 15-20% от $Z_{\text{ТС}}$); $k_{\text{р}}$ – районный коэффициент, равный 1,3 (для Томска).

Таблица 19 – Баланс рабочего времени

Показатели рабочего времени	Руководитель проекта	Исполнитель по проекту
Календарное число дней	365	365
Количество нерабочих дней		
– выходные дни	52	52
– праздничные дни	14	14
Потери рабочего времени		
– отпуск	48	48
Действительный годовой фонд рабочего	251	251

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{б}} * k_{\text{р}}, \quad (12)$$

где $Z_{\text{б}}$ – базовый оклад, руб.; $k_{\text{р}}$ – районный коэффициент, равный 1,3 (для Томска)

Таблица 20 – Расчет основной заработной платы

Исполнители	З _б , руб.	k _р	З _м ,руб.	З _{дн} , руб.	T _р раб. дн.	З _{осн} , руб.
Руководитель проекта	33664,00	1,3	43763,20	1813,30	11	19946,30
Исполнитель по проекту	1988,00	1,3	2584,40	108,08	90	9727,20

Дополнительная заработная плата

Дополнительная заработная плата включает оплату за непроработанное время (очередной и учебный отпуск, выполнение государственных обязанностей, выплата вознаграждений за выслугу лет и т.п.) и рассчитывается исходя из среднего значения в 10-12% от основной заработной платы, работников, непосредственно участвующих в выполнении темы:

$$Z_{\text{доп}} = K_{\text{доп}} \cdot Z_{\text{осн}}, \quad (13)$$

где $Z_{\text{доп}}$ – дополнительная заработная плата, руб.; $K_{\text{доп}}$ – коэффициент дополнительной зарплаты; $Z_{\text{осн}}$ – основная заработная плата, руб.

Таблица 21– Заработная плата исполнителей НИОКР

Заработная плата	Руководитель проекта	Исполнитель по проекту
Основная заработная плата	19946,30	9727,20
Дополнительная заработная плата (10%)	1994,63	972,72
Итого по статье С _{зп}	21940,93	9699,92

Отчисления на социальные нужды

Статья включает в себя отчисления во внебюджетные фонды.

$$C_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}), \quad (14)$$

где $k_{\text{внеб}}$ – коэффициент отчислений во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и прочее).

Величина отчислений согласно постановлению Правительства РФ от 06.11.2019 №1407 [38] составляет: взносы на ОПС 22%, взносы на ОМС 5,1% и взносы на ВНиМ 2,9% итого 30%.

Таблица 22 – Отчисления во внебюджетные фонды

	Руководитель проекта	Исполнитель по проекту
Зарплата	21940,93	10699,92
Отчисления во внебюджетные фонды	6582,28	2209,98

Прочие прямые расходы

Ноутбук потребляет 0,15 кВт/ч. В Томске 1кВт/ч стоит 3,5 рублей. На работу за компьютером было затрачено 520 часов, что означает стоимость потраченной электроэнергии $0,15 \cdot 520 \cdot 3,5 = 273$ руб.

Накладные расходы

В данную статью входят расходы на содержание аппарата управления и общехозяйственных (общеуниверситетских) служб, которые в равной степени относятся ко всем выполняемым НИОКР. По этой статье учитываются оплата труда административно-управленческого персонала, содержание зданий, оргтехники и хозяйственного инвентаря, амортизация имущества, расходы по охране труда и подготовке кадров.

Накладные расходы лаборатории 10 корпуса НИ ТПУ составляют 80-90% от суммы основной и дополнительной заработной платы, работников, непосредственно участвующих в выполнении темы. Расчет накладных расходов ведется по следующей формуле:

$$C_{\text{внеб}} = k_{\text{накл}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}), \quad (15)$$

где $k_{\text{накл}}$ – коэффициент накладных расходов. $C_{\text{накл}} = 80\% \cdot C_{\text{зп}} = 26\,112,68$.

На основании сделанных расчётов можно посчитать плановую себестоимость НИР по форме, приведенной в таблице 23.

Таблица 23 – Группировка затрат по статьям

Вид работ	Статьи						
	Специальное оборудование для научных (экспериментальных) работ	Основная заработная плата	Дополнительная заработная плата	Отчисления на социальные нужды	Прочие прямые расходы	Накладные расходы	Итого плановая себестоимость
Создание и обучение чат-бота на базе математической модели	32500	29673,50	10699,92	9792,26	273	26112,68	109051,36

5.3.4 План управления коммуникациями проектами

План управления коммуникациями отражает требования к коммуникациям со стороны участников проекта. Управление коммуникациями нашего проекта приведен в таблице 24.

Таблица 24– План управления проектами

Какая информация передается	Кто передает информацию	Кому передается информация	Когда передает информацию
Обмен информацией о текущем состоянии проекта	Исполнитель проекта	Руководителю проекта	Еженедельно
Документы и информация по проекту	Исполнитель проекта	Руководитель проекта	Ежемесячно
О выполнении контрольной точки	Исполнитель проекта	Руководитель проекта	Не позже дня контрольного события по плану управления

5.3.5 Реестр рисков проекта

Идентифицированные риски проекта — это возможные неопределенные события, которые могут возникнуть в проекте и могут повлечь за собой нежелательные эффекты.

Таблица 25 – Реестр рисков

Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска*	Способы смягчения	Условия наступления
Отсутствие спроса со стороны ключевых потребителей для приобретения продукта	Отсутствие продаж	2	4	Средний	Разработка политики продвижения	Продукт не вызывает интерес, финансовый кризис
Технологические риски, связанные с использованием неизвестных ОС на серверах заказчика	Проблемы с внедрением готовой программы в корпоративные сети	1	3	Низкий	Изучить потенциальных клиентов	Переход на отечественное ПО, оборонное предприятие
Конкуренция на рынке выше ожидаемого	Снижения спроса на продукт	2	4	Средний	Изучение альтернативных технологий для снижения стоимости продукта	Цена, Выход нового продукта на рынок

* Уровень риска может быть: высокий, средний или низкий в зависимости от вероятности наступления и степени влияния риска.

5.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

5.4.1 Оценка абсолютной эффективности исследования

В основе проектного подхода к инвестиционной деятельности предприятия лежит принцип денежных потоков (cashflow). Особенностью является его прогнозный и долгосрочный характер, поэтому в применяемом подходе к анализу учитываются фактор времени и фактор риска. Для оценки общей экономической эффективности используются следующие основные показатели:

- чистая текущая стоимость (NPV);
- индекс доходности (PI);
- внутренняя ставка доходности (IRR);

– срок окупаемости (DPP).

Чистая текущая стоимость (NPV) – это показатель экономической эффективности инвестиционного проекта, который рассчитывается путём дисконтирования (приведения к текущей стоимости, т.е. на момент инвестирования) ожидаемых денежных потоков (как доходов, так и расходов).

Расчёт NPV осуществляется по следующей формуле:

$$NPV = \sum_{t=1}^n \frac{ЧДП_{опt}}{(1+i)^t} - I_0, \quad (16)$$

где: ЧДП_{опt} – чистые денежные поступления от операционной деятельности;

I_0 – разовые инвестиции, осуществляемые в нулевом году;

t – номер шага расчета ($t = 0, 1, 2 \dots n$)

n – горизонт расчета;

i – ставка дисконтирования (желаемый уровень доходности инвестируемых средств).

Расчёт NPV позволяет судить о целесообразности инвестирования денежных средств. Если $NPV > 0$, то проект оказывается эффективным.

Расчет чистой текущей стоимости представлен в таблице 26. При расчете рентабельность проекта составляла 20 %, амортизационные отчисления 10 %. Бюджет проекта = 109052 руб.

Коэффициент дисконтирования рассчитан по формуле:

$$КД = \frac{1}{(1+i)^t}, \quad (17)$$

где: i – ставка дисконтирования, 20 %;

t – шаг расчета.

Таким образом, чистая текущая стоимость по проекту в целом составляет 109194,26 рублей, что позволяет судить о его эффективности

Индекс доходности (PI) – показатель эффективности инвестиции, представляющий собой отношение дисконтированных доходов к размеру инвестиционного капитала. Данный показатель позволяет определить

инвестиционную эффективность вложений в данный проект. Индекс доходности рассчитывается по формуле:

$$PI = \sum_{t=1}^n \frac{ЧДП_t}{(1+i)^t} / I_0, \quad (18)$$

где: ЧДП - чистый денежный поток, млн. руб.;

I_0 – начальный инвестиционный капитал, млн. руб.

Таблица 26 – Расчет чистой текущей стоимости по проекту в целом

№	Наименование показателей	Шаг расчета				
		0	1	2	3	4
1	Выручка от реализации, руб.	0	130863	130863	130863	130863
2	Итого приток, руб.	0	130863	130863	130863	130863
3	Инвестиционные издержки, руб.	-109052	0	0	0	0
4	Операционные затраты, руб.	0	0	0	0	0
5	Налогооблагаемая прибыль	0	130863	130863	130863	130863
6	Налоги 20 %, руб.	0	26172,60	26172,60	26172,60	26172,60
7	Итого отток, руб.	-109052	26172,60	26172,60	26172,60	26172,60
8	Чистая прибыль, руб.	0	104690,40	104690,40	104690,40	104690,40
9	Чистый денежный поток (ЧДП), руб.	-109052	115595,60	115595,60	115595,60	115595,60
10	Коэффициент дисконтирования (КД)	1	0,833333	0,694444	0,578704	0,482253
11	Чистый дисконтированный доход (ЧДД), руб.	-109052	96329,63	80274,67	66895,64	55746,32
12	Σ ЧДД		299246,26			
12	Итого NPV, руб.		109194,26			

$$PI = \frac{299246,26}{109052} = 2,744$$

Таким образом $PI > 1$, то проект является эффективным.

Значение ставки, при которой обращается в нуль, носит название «внутренней ставки доходности» или IRR. Формальное определение «внутренней ставки доходности» заключается в том, что это та ставка дисконтирования, при которой суммы дисконтированных притоков денежных средств равны сумме дисконтированных оттоков или $=0$. По разности между IRR и ставкой дисконтирования i можно судить о запасе экономической прочности инвестиционного проекта. Чем ближе IRR к ставке дисконтирования i , тем больше риск от инвестирования в данный проект.

$$\sum_{t=1}^n \frac{ЧДП_{опt}}{(1 + IRR)^t} = \sum_{t=0}^n \frac{I_t}{(1 + IRR)^t} \quad (19)$$

Между чистой текущей стоимостью (NPV) и ставкой дисконтирования (i) существует обратная зависимость. Эта зависимость представлена в таблице 27 и на рисунке 17.

Из таблицы и графика следует, что по мере роста ставки дисконтирования чистая текущая стоимость уменьшается, становясь отрицательной. Значение ставки, при которой NPV обращается в нуль, носит название «внутренней ставки доходности» или «внутренней нормы прибыли». Из графика получаем, что IRR составляет 1,14.

Запас экономической прочности проекта: $114\% - 20\% = 94\%$.

Как отмечалось ранее, одним из недостатков показателя простого срока окупаемости является игнорирование в процессе его расчета разной ценности денег во времени.

Этот недостаток устраняется путем определения дисконтированного срока окупаемости. То есть это время, за которое денежные средства должны совершить оборот.

Таблица 27 – Зависимость NPV от ставки дисконтирования

№	Наименование показателя	0	1	2	3	4	NPV, руб
1	Чистые денежные потоки, руб	-109052	130863	130863	130863	130863	
2	Коэффициент дисконтирования						
	0,1	1	0,90909	0,82645	0,75131	0,68301	
	0,2	1	0,83333	0,69444	0,5787	0,48225	
	0,3	1	0,76923	0,59172	0,45517	0,35013	
	0,4	1	0,71429	0,5102	0,36443	0,26031	
	0,5	1	0,66667	0,44444	0,2963	0,19753	
	0,6	1	0,625	0,39063	0,24414	0,15259	
	0,7	1	0,58824	0,34602	0,20354	0,11973	
	0,8	1	0,55556	0,30864	0,17147	0,09526	
	0,9	1	0,52632	0,27701	0,14579	0,07673	
	1	1	0,5	0,25	0,125	0,0625	
	1,1	1	0,47619	0,226757	0,10798	0,051419	
	1,2	1	0,45455	0,20661	0,09391	0,04269	
	1,3	1	0,43478	0,18904	0,08219	0,03573	
3	Дисконтированный денежный доход, руб.						
	0,1	-109052	118966	108151	98319,3	89381,2	305766,1
	0,2	-109052	109053	90877,1	75730,9	63109,1	229717,57
	0,3	-109052	100664	77433,7	59564,4	45818,8	174428,75
	0,4	-109052	93473,6	66766,8	47690,6	34064,7	132943,72
	0,5	-109052	87242	58161,3	38774,2	25849,5	100975,04
	0,6	-109052	81789,4	51118,4	31949	19968,1	75772,818
	0,7	-109052	76978,2	45281,3	26636,1	15668,3	55511,893
	0,8	-109052	72701,7	40389,8	22438,8	12466	38944,26
	0,9	-109052	68875,3	36250,1	19079	10041,6	25194,012
	1	-109052	65431,5	32715,8	16357,9	8178,94	13632,063
	1,1	-109052	62315,7	29674,1	14130,5	6728,83	3797,2435
	1,2	-109052	59483,2	27037,8	12289,9	5586,32	-4654,77
	1,3	-109052	56897	24737,8	10755,6	4676,33	-11985,33

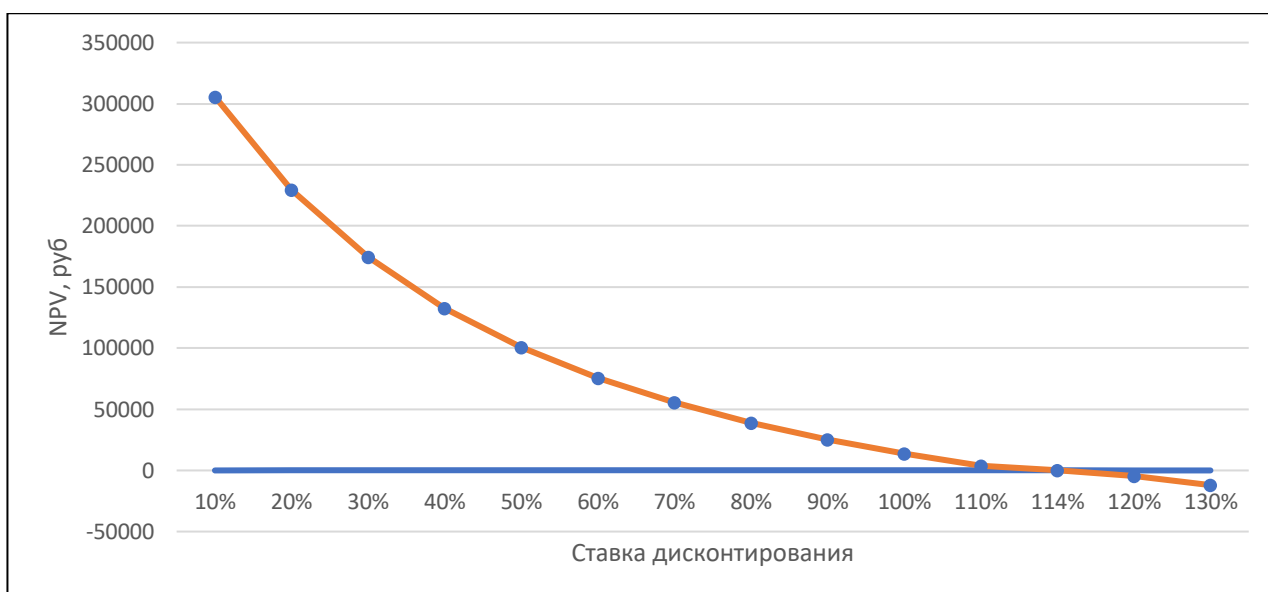


Рисунок 17 – Зависимость NPV от ставки дисконтирования

Наиболее приемлемым методом установления дисконтированного срока окупаемости является расчет кумулятивного (нарастающим итогом) денежного потока (таблица 28).

Таблица 28 – Дисконтированный срок окупаемости

№	Наименование показателя	Шаг расчета				
		0	1	2	3	4
1	Дисконтированный денежный доход ($i=0,20$), руб.	-109052	96329,63	80274,67	66895,64	55746,32
2	То же нарастающим итогом, руб.	-109052	-12722,37	67552,30	134447,94	109194,26
3	Дисконтированный срок окупаемости	$PP_{\text{диск}} = 1 + (12722,37 / 80274,67) = 1,16$ года				

Социальная эффективность научного проекта (таблица 29) учитывает социально-экономические последствия осуществления научного проекта для общества в целом или отдельных категорий населения или групп лиц, в том числе как непосредственные результаты проекта, так и «внешние» результаты в смежных секторах экономики: социальные, экологические и иные внеэкономические эффекты.

Таблица 29 - Критерий социальной эффективности

ДО	ПОСЛЕ
Трата большого количества времени для выполнения рутинных задач, что ведет к избыточной трате электроэнергии	Экономия электроэнергии за счет использования чат-бота для рутинных задач, так как тратиться меньше времени за компьютером для решения тех же задач.

5.4.2 Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трех (или более) вариантов исполнения научного исследования. Для этого наибольший интегральный показатель реализации технической задачи принимается за базу расчета (как знаменатель), с которым соотносятся финансовые значения по всем вариантам исполнения.

Интегральный финансовый показатель разработки определяется по следующей формуле:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}, \quad (20)$$

где: $I_{\text{финр}}^{\text{исп.}i}$ – интегральный финансовый показатель разработки;

Φ_{pi} – стоимость i -го варианта исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное увеличение бюджета затрат разработки в размах (значение больше единицы), либо соответствующее численное удешевление стоимости разработки в размах (значение меньше единицы, но больше нуля).

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить по следующей формуле:

$$I_{pi} = \sum a_i \cdot b_i, \quad (21)$$

где: I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i^a, b_i^p – бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности приведен в форме таблицы (таблице 30).

Таблица 30 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ ПО	Весовой коэф. параметра	Текущий проект	Аналог 1	Аналог 2
1. Способствует увеличению скорости выполнения рутинных задач	0,23	5	4	4
2. Удобство в эксплуатации (соответствует требованиям потребителей, все функции внутри одного приложения)	0,21	5	5	5
3. Безопасность	0,19	4	4	4
4. Надежность	0,19	4	3	3
5. Простота эксплуатации (понятный интерфейс и функции)	0,18	5	4	4
Итого	1	23	20	20

$$I_m^p = 5 \cdot 0,23 + 5 \cdot 0,21 + 4 \cdot 0,19 + 4 \cdot 0,19 + 5 \cdot 0,18 = 4,62$$

$$I_1^A = 4 \cdot 0,23 + 5 \cdot 0,21 + 4 \cdot 0,19 + 4 \cdot 0,19 + 5 \cdot 0,18 = 4,39$$

$$I_2^A = 4 \cdot 0,23 + 5 \cdot 0,21 + 4 \cdot 0,19 + 4 \cdot 0,19 + 5 \cdot 0,18 = 4,39$$

Интегральный показатель эффективности разработки $I_{финр}^p$ и аналога $I_{финр}^a$ определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{финр}^p = \frac{I_m^p}{I_\phi^p}; I_{финр}^a = \frac{I_m^a}{I_\phi^a} \quad (22)$$

Сравнение интегрального показателя эффективности текущего проекта и аналогов позволит определить сравнительную эффективность проекта. Сравнительная эффективность проекта определяется по формуле:

$$\mathcal{E}_{\text{ср}} = \frac{I_{\text{финр}}^p}{I_{\text{финр}}^a}, \quad (23)$$

где: $\mathcal{E}_{\text{ср}}$ – сравнительная эффективность проекта;

$I_{\text{финр}}^p$ – интегральный показатель разработки;

$I_{\text{финр}}^a$ – интегральный технико-экономический показатель аналога.

Сравнительная эффективность разработки по сравнению с аналогами представлена в таблице 31.

Таблица 31 – Сравнительная эффективность разработки

№ п/п	Показатели	Разработка	Аналог 1	Аналог 2
1	Интегральный финансовый показатель разработки	0,9	0,97	1
2	Интегральный показатель ресурсоэффективности разработки	4,62	4,39	4,39
3	Интегральный показатель эффективности	5,133	4,526	4,39
4	Сравнительная эффективность вариантов исполнения	1	0,88	0,855

Вывод: Сравнение значений интегральных показателей эффективности позволяет понять, что разработанный вариант проведения проекта является наиболее эффективным при решении поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

В ходе выполнения раздела финансового менеджмента рассчитан бюджет научного исследования, определена чистая текущая стоимость, (NPV), равная 109194,26руб.; индекс доходности $PI=2,744$; внутренняя ставка доходности $IRR=1,14$, срок окупаемости $PP_{\text{дск}}=1,16$ года, тем самым инвестиционный проект можно считать выгодным и экономически целесообразным.

CONCLUSIONS

In this work, the goal was achieved and the following tasks were successively solved.

First of all, we analyzed existing chatbots and found out that mainly they are developed under some messengers' API. It is much easier to go this way, but the drawback is companies doesn't want to use messengers' API. They think it is not safe and could let to the outflow of information. That is why we used programming language without messengers' API and that's why our development is relevant.

Secondly, we analyzed development tools and chosen next technology stack:

- a) Python programming language, because it is mostly used in the field of machine learning. It means lots of libraries and technologies are built for developers, which makes it is the process of development;
- b) NumPy for working with arrays and matrices;
- c) SciPy – to expand NumPy's functionality;
- d) NLTK (Natural Language Toolkit) – to prepare dataset for machine learning;
- e) pymorphy2 – used for text preprocessing in combination with NLTK to achieve higher accuracy;
- f) TensorFlow – which uses tensors (n-dimensional arrays of basic data types) to create a neural network with given parameters;
- g) Six to make program files compatible between Python versions.

In third place, we prepared data for machine learning and applied current data conversation steps:

- a) tokenization;
- b) token normalization;
- c) converting words to vectors by counting its Term Frequency – Inverse Term Frequency (TF-IDF) value.

For tokenization we used white space tokenizer from the NLTK library. For token normalization we converted words into lemmas by using the pymorphy2 library.

Finally, we counted TF-IDF values which allowed us to convert words into vectors to perform operations faster.

Fourth, we implemented a chatbot as a console application with a simple graphical user interface. The developed chatbot represents “end-to-end” system. In this system we had to train only one model which at first understand what user wrote and then respond to him/her in text mode. Also, it could follow the context of dialogue using LSTM (Long-Short Term Memory). To train a chatbot we used a dataset from Yandex.Toloka. RNN (Recurrent neural network) is used to build seq2seq model. We propose to use the recurrent neural network because in our case the next result depends on the previous context. The seq2seq model uses encoder and decoder to process an input text and to answer respectively. Encoder used to understand what user wrote and to follow the context and decoder used to form answers.

Finally, we provided a set of tests for the implemented chatbot by live dialogue with it. During testing, we used slang in text messages, made spelling errors, but the bot recognized and corrected them well. The results of communication have been presented in figure 7.

Testing results showed that the probability of recognizing formalized user requests reaches 90% and higher.

Also, in this work, we assessed the risks and dangers to health and the environment, and considered ways to minimize them. We found out that the workplace microclimate, lighting and electromagnetic radiation are normal, which indicates safety for humans. An analysis of the identified hazardous factors of the working room showed that electrical, fire and environmental safety are under control and all necessary measures to ensure safety are being taken.

After that, we analyzed financial part of the project. In this section, we calculated the budget of the project, the net present value (NPV), profitability index (PI), the internal rate of return (IRR) and the payback period. Based on these indicators we concluded that this investment project is profitable and economically feasible.

In general, we can conclude that the proposed chatbot architecture and its software implementation are cross-platform and can be used on PCs, smartphones and are easily integrated into corporate networks and at the same time require minimal time and human resources for support and further development. Also, the project development process is safe for health and it is economically feasible.

Список публикаций студента

1. Г. С. Чурсин, **Н. И. Исмоилов**, А. В. Доржиев Встраиваемый чат-бот в готовые корпоративные системы // Перспективы развития фундаментальных наук: сборник научных трудов XVI Международной конференции студентов, аспирантов и молодых ученых, г. Томск, 23-26 апреля 2019 г. : в 7 т. — Томск : Изд-во ТПУ, 2019. — Т. 3 : Математика. — [С. 92-94].
Режимдоступа:http://earchive.tpu.ru/bitstream/11683/55932/1/conference_tpu-2019-C21_V3_p92-94.pdf
2. **Ismoilov Nurullo**. Russian Language Neural Net Chatbot with Natural Language Processing [Electronic resource] / **N. Ismoilov**, M.E. Semenov // 14 International Forum on Strategic Technology (IFOST-2019), October 14-17, 2019, Tomsk, Russia: [Proceedings]. - Tomsk: TPU Publishing House, 2019. - [С. 135-138]. –Заглавие с экрана. – Свободный доступ из сети Интернет. Режим доступа: [conference_tpu-2019-C127_p135-138.pdf](http://conference.tpu-2019-C127_p135-138.pdf)

REFERENCES

- [1] Experimentally induced visual projections into auditory thalamus and cortex/ Sur, M., P.E. Garraghty and A.W. Roe // Science 242 1988.P.1437–1441.
- [2] Tactile sensory substitution studies / Bach-y-Rita P // Annals of the New York Academy of Sciences -2004. P. 83–91, 2004.
- [3] Brainport technologies– helping people with disabilities live a better life [electronic resource]/ Wicab, Inc. Middleton, WI, USA / URL: <https://www.wicab.com/>, free access – Lang. Eng. Access date: 25.04.2020
- [4] Ideas for inspiration [electronic resource] / Pinterest, Inc. at 984 Brannan Street, San Francisco, CA 94107, USA / URL: <https://www.pinterest.com/>, free access – Lang. Eng, Ru. Access date: 25.04.2020
- [5] Get smarter about what matters to you [electronic resource]: Biological Neurons versus Artificial Network / A Medium Corporation – Author: Swathi Arulgroup Nagendra; Publication date: 25.10.2018 / URL:<https://medium.com/@swaramk/biological-neural-network-vs-artificial-neural-network-69b4fa2b093b>, free access – Lang. Eng. Access date 26.04.2020
- [6] The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain / FrankRosenblatt // Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6 – 1958, P386-408
- [7] Perceptrons: an introduction to computational geometry / Marvin Minsky, Seymour A. Papert // The MIT Press – 1987, P. 308-311
- [8] Butler analytics – Big data and machine learning in use [electronic resource]: Neural network data mining explained / Butler analytics, Inc. / URL: <https://www.butleranalytics.com/neural-network-data-mining-explained/>, free access – Lang. Eng. Access date 26.04.2020
- [9] Statista – The statistics portal for market data, market research [electronic resource] :Number of mobile phone messaging app users worldwide from 2018 to 2022 (in billions)/ Statista, Inc. / Author: J. Clement, Publication date:

8.10.2019 / URL: <https://www.statista.com/statistics/483255/number-of-mobile-messaging-users-worldwide/>, free access – Lang. Eng. Access date 01.05.2020

[10] Russian Language Neural Net Chatbot with Natural Language Processing [electronic resource] / N. Ismoilov, M.E. Semenov // 14 International Forum on Strategic Technology (IFOST-2019), October 14-17, 2019, Tomsk, Russia – Tomsk: TPU Publishing House – 2019 // URL: http://earchive.tpu.ru/bitstream/11683/57459/1/conference_tpu-2019-C127_p135-138.pdf, free access – Lang. Eng. Access date: 01.05.2020

[11] Morph.ai – Having meaningful conversations with your users [electronic resource] / Scupids Tech Private Limited / URL: <https://morph.ai/>, free access – Lang. En. Access date: 28.04. 2020

[12] Flow XO – Create a chatbot with zero coding skills required [electronic resource] / Flow XO LLC / URL: <https://flowxo.com/>, free access – Lang. En. Access date: 28.04.2020

[13] Telegram a new era of messaging [electronic resource]: Telegram APIs / Telegram, Inc. / URL: <https://core.telegram.org/api>, free access – Lang. En. Access date: 28.04.2020

[14] Build natural and rich conversational experiences [e-resource] / Google, Inc. / URL: <https://dialogflow.com/>, free access – Lang. En. Access date: 28.04.2020

[15] Head first Python / P. Barry // O'Reilly Media Inc.-2016–624 p.

[16] Find, install and publish Python packages with the Python package index [electronic resource]: SciPy/ Python software foundation; Python community / URL: <https://pypi.org/project/scipy/>, free access – Lang. En., Ru., Esp., Fr. Access date: 02.05.2020

[17] Встраиваемый чат-бот в готовые корпоративные системы: магистерская диссертация [e-resource] / Г. С. Чурсин // Национальный исследовательский Томский политехнический университет (ТПУ) // URL: <http://earchive.tpu.ru/bitstream/11683/54985/1/TPU735276.pdf>, free access – Lang. Ru., Access date: 10.05.2020

[18] Wondererd.net – A nerds home on the net [electronic resource] / Wondererd.net / Deep Learning in Virtual Data Center Management – 2016 / URL: <https://www.wondererd.net/blog/deep-learning-in-virtual-data-center-management/>, free access – Lang. En., Access date: 03.05.2020

[19] Natural language processing with Python / S. Bird, E. Klein, E. Loper // O'Reilly Media Inc. – 2009 - 463 p.

[20] Coursera | Online Courses & Credentials From Top Educators [electronic resource]: Online course: Natural Language processing – National Research University Higher School of Economics / Coursera, Inc. / URL: <https://www.coursera.org/learn/language-processing/home/>, Access for registered users, Lang. En., Access date: 08.05.2020

[21] Sequence to sequence learning with neural networks / Ilya Sutskever, Oriol Vinyals, Quoc V Le // In Advances in neural information processing systems – 2014, P. 3104-3112.

[22] Открытые датасеты [electronic resource] / Яндекс / URL: <https://toloka.ai/ru/datasets>, free access, Lang. Ru., Access date: 06.05.2020

[23] ГОСТ 12.2.032 -78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования. М.: Изд-во стандартов, 1986. – 9 с.

[24] Санитарно-эпидемиологические правила и нормативы: СанПиН 2.2.4.548–96 Гигиенические требования к микроклимату производственных помещений. М.: Минздрав России, 1997. – 20 с.

[25] Свод правил: СП 52.13330.2016. Естественное и искусственное освещение. М.: Минрегион России, 2011. – 74 с.

[26] Санитарно-эпидемиологические правила и нормативы: СанПиН 2.2.4.3359-16. Санитарно-эпидемиологические требования к физическим факторам на рабочих местах. М.: Минздрав России, 2003. – 39 с.

[27] ГОСТ 12.1.019-2017 ССБТ Электробезопасность. Общие требования и номенклатура видов защиты. – М.: Изд-во стандартов, 2010. – 28 с.

[28] ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования. – М.: Изд-во стандартов, 2006. – 67 с.

[29] ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования. – М.: Изд-во стандартов, 2003. – 6 с.

[30] Санитарные правила и нормы СанПиН 2.2.4.548-96 "Гигиенические требования к микроклимату производственных помещений".

[31] Санитарные правила и нормы СанПиН 2.2.1/2.1.1.1278-03. Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий – 34с.

[32] ГОСТ Р 55710-2013 Освещение рабочих мест внутри зданий– 13с.

[33] ФЗ от 24.06.1998 №89-ФЗ (ред. от 25.12.2018) «Об отходах производства и потребления»

[34] Распоряжение Правительства РФ от 25.07.2017 №1589-р

[35] Финансовый менеджмент, ресурсоэффективность и ресурсосбережение: учебно-методическое пособие / Н.А. Гаврикова, Л.Р. Тухватулина, И.Г. Видяев, Г.Н. Серикова, Н.В. Шаповалова; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2014. – 73 с.

[36] Singularika | Разработка чат ботов для бизнеса [электронный ресурс] / Singularika / URL: <https://chatbots.singularika.com/>, свободный доступ, Яз.рус., Дата обращения: 20.04.2020

[37] Ваш личный корпоративный помощник – БОТ [электронный ресурс] / ЗАО «КРОК инкорпорейтед» / URL: <https://bot.croc.ru/>, свободный доступ, Яз. рус., Дата обращения: 20.04.2020

[38] Постановление Правительства РФ от 6 ноября 2019 г. N 1407 “О предельной величине базы для исчисления страховых взносов на обязательное социальное страхование на случай временной нетрудоспособности и в связи с материнством и на обязательное пенсионное страхование с 1 января 2020 г.”
Доступ: <https://www.garant.ru/products/ipo/prime/doc/72878832/>