Asyk: An Educational Chatbot for Kazakh Traditional Games

Module Code	**:**	CT098-3-M - Artificial Intelligence (112023-VTK)
Intake Code	**:**	APDMF2310AI
Lecturer Name	**:**	Assoc Prof Ts Dr Vinothini Kasinathan
Hand in Date	**:**	1st of March, 2024
Student Name	**:**	Nurlan Amanov
Student Number	**:**	TP077526

Table of Contents

ABSTRACT

The main purpose of this paper is to demonstrate the work of the Asyk chatbot developed in the Python programming language using the NLTK, PyTorch and Tkinter libraries. Before that we will tell the story of the development of Artificial Intelligence and chatbots themselves, what chatbots are, what kinds of chatbots exist and what chatbots are popular nowadays. Further, to delve into the very work of the chatbot "Asyk" and understand its functionality, the principle of its work and interaction with the user. With the creation of the application, it is important to know the opinion about it as well. With the general opinion and evaluation of its work, set a course for its improvement and future integrations that will allow it to be competitive among other chatbots.

1. INTRODUCTION

1.1. Brief History

Nowadays, the field of Artificial Intelligence (AI) has gained tremendous development and popularity. Big companies are in a race to develop the best gadget that will have a huge impact on society and get huge demand. Although the field of Artificial Intelligence (AI) has gained attention only now, its history dates to the times of antiquity, where myths told about fabulous creatures that have artificial intelligence, just like humans. Thus, these are only the first mentions of artificial intelligence, where philosophers described the human brain as a mechanism, but these are only distant mentions, because the official history of AI dates to the mid-twentieth century. The beginnings of AI begin with the invention of the digital computer in the 1940s, where a machine based on abstract mathematical reasoning inspired scientists. The field of AI research took shape in the summer of 1956 at a seminar at Dartmouth College. Over time, the field of AI ran into problems and such machines proved to be very difficult to develop, which served to cut off funding. In the article "There Was No 'First AI Winter'" by Haigh (2023) it says that "The late 1980s are universally seen as the beginning of the 'AI Winter,' in which faith and funding for AI dwindled dramatically." However, with subsequent improvements in equipment, the creation of new techniques and so on has fueled the development of AI. Today, there are a huge number of examples of AI use cases: content generation, analytics, detection, but one of the most popular is chatbots.

1.2. Chatbot History

A chatbot is a computer program that simulates a conversation between a human and a user. The most popular examples of chatbots use natural language processing (NLP), which allows the chatbot to understand requests from the user and generate a response, a popular example being ChatGPT. However, the very first chatbot was created back in 1966 and it was called ELIZA. ELIZA used pattern recognition to understand the request it received and respond with a predefined pattern, however the chatbot failed the Turing Test. Turing Test is a method that helps to evaluate a machine's ability to demonstrate intelligence like a human. The subsequent chatbots are PARRY and ALICE, but of the two, ALICE was the first chatbot to pass the Turing Test. ALICE (Artificial Linguistic Internet Computer Examiner), created in the late

1990s, was the first chatbot to pass the test. ALICE used a combination of rule-based system and NLP. A rule-based chatbot works based on predefined rules, thus giving a response based on its rules. The dialogue with ALICE is available on the Internet by typing "ALICE Chatbot" into any search engine (Figure 1).

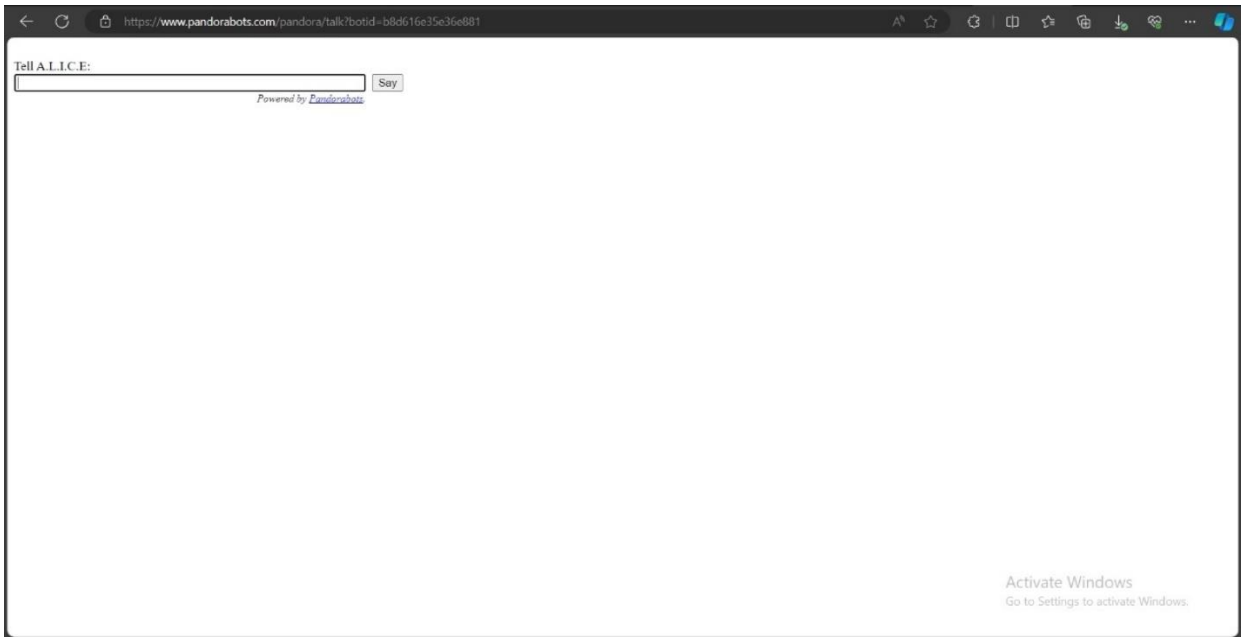

Figure 1. ALICE Chatbot Website

According to Shenoy (2024) on Yellow.ai, there are 6 types of chatbots, such as:

1. Button-Based chatbots.
2. Rules-Based chatbots.
3. Machine Learning-Based chatbot.
4. Hybrid Chatbots.
5. Voice-Enabled chatbots.

1.3. Button-Based

Button-Based Chatbots are considered the simplest chatbots, where the user communicates with the chatbot via predefined buttons or menus. Such chatbots are useful for simple queries but are not suitable for more complex ones. Such chatbots can be found on any

online shops where the user can find the information, he or she is interested in by simply navigating through a menu (Figure 2).

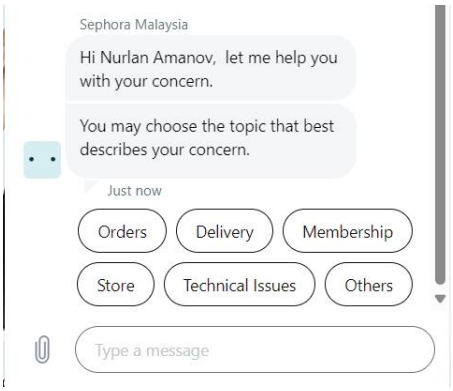

Figure 2. Button-Based Chatbot

1.4. Rule-Based Chatbot

Rule-Based chatbot uses "if-then" logic, thus the conversation is based on predefined rules. Such chatbots act as interactive FAQs and are useful for structured interactions but lack adaptability. A simple principle of Rule-Based chatbot operation is illustrated in Figure 3.

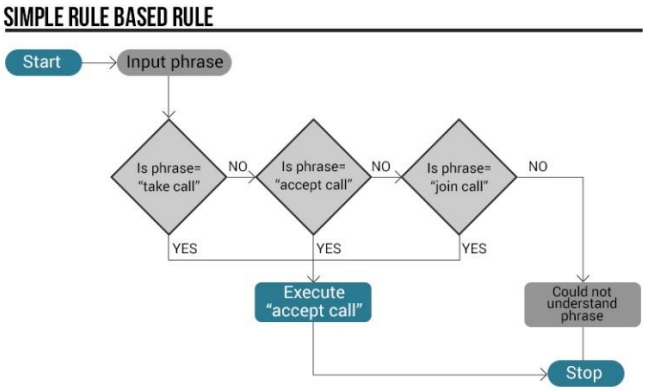

Figure 3. Simple Rule-Based Rule. https://www.cosoit.com/difference-between-rule-based-and-ai-based-chatbot

1.5. Machine Learning-Based Chatbot

Machine Learning-Based Chatbots are chatbots that use machine learning techniques to generate a response to a user's query. Unlike Rule-Based chatbots that follow rules, Machine

Learning-Based Chatbots learn from data and improve their performance. An example would be the popular ChatGPT chatbot (Figure 4).

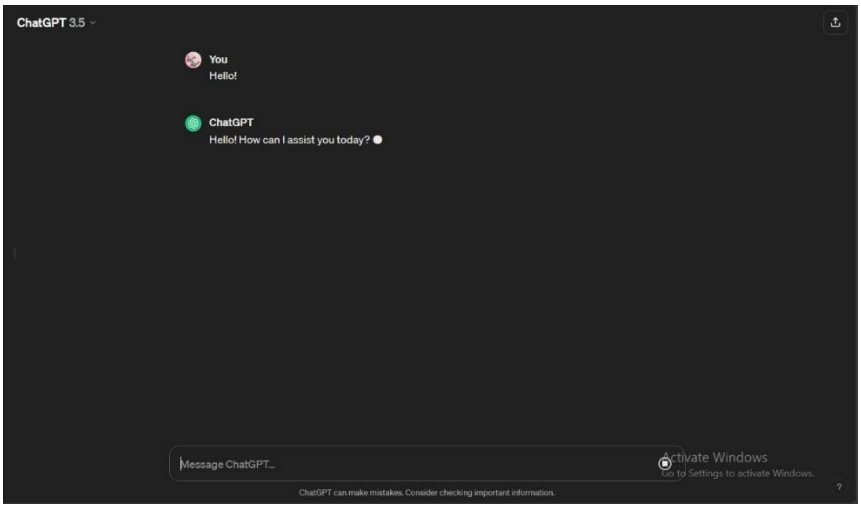

Figure 4. ChatGPT. https://chat.openai.com/

1.6. Hybrid Chatbot

Hybrid chatbot is a combination of Rule-Based and Machine Learning-Based chatbots, thus using the best of both, Hybrid tries to provide a comfortable interaction between the chatbot and the user. This is convenient because such chatbots use predefined rules for specific tasks and leave the rest to machine learning. An example of such chatbots is shown in Figure 5.

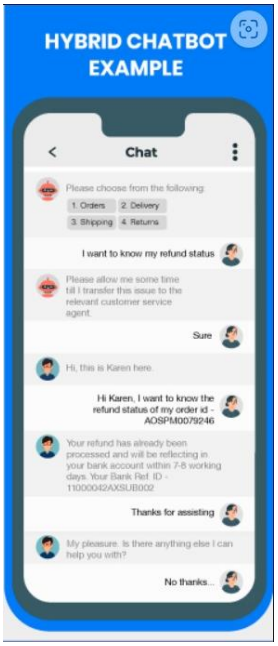

Figure 5. Hybrid Chatbot. *https://www.goodfirms.co/help-desk-software/blog/the-dominance-of-chatbots-in-customer-service-is-never-going-to-end*

1.7. Voice-Enabled Chatbot

Voice-Enabled chatbots are designed to interact with the user through spoken language using their voice. The main feature is the use of Automatic Speech Recognition (ASR) and NLP technology. Examples are the popular Siri, Alexa and Cortana (Figure 6).

Figure 6. Apple Siri. *https://economictimes.indiatimes.com/magazines/panache/now-just-say-siri-apple-planning-to-drop-hey-siri-as-wake-word-for-smart-assistant/articleshow/95347493.cms?from=mdr*

Having gained a basic knowledge of chatbots this paper introduces a new chatbot having a specific theme of Kazakh national games - Asyk.

2. MATERIALS AND METHODS

«Asyk» is an educational chatbot built in the Python programming language using the Natural Language Toolkit (NLTK), PyTorch and Tkinter libraries. We will use article “Contextual Chatbots with Tensorflow” published in Chatbots Magazine in 2017 and video “Chat Bot With PyTorch” by Patrick Loeber published on YouTube in 2020 as our references to create chatbot.

NLTK is a comprehensive library used for working with human language data. NLTK includes various modules, but for us we are interested in tokenization and stemming. The reason we are interested in these algorithms is that we use the Bag of Words (BoW) technique. BoW is the simplest and most common way of representing text for NLP. The whole process is as follows:

1. By tokenization, we split the text into individual tokens.

   Original text: "I love natural language processing."

   Tokenized words: ["I", "love", "natural", "language", "processing"]

2. Stemming reduces words to their root, which will help group them together in the future.

   Stemmed words: ["I", "love", "natur", "languag", "process"]

3. After the dictionary is created where all unique words with their own unique index will be placed.

   Vocabulary: ["I", "love", "natur", "languag", "process"]

4. For each sample, a count is kept of how many times each word from the dictionary appears in the text. This creates a list of mentions of a particular word from the numbers of that.

   Original text: "I love natural language processing."

   Stemmed words: ["I", "love", "natur", "languag", "process"]

   Bag of Words Vector: [1, 1, 1, 1, 1]

5. These counts are organized into a vector that represents the text in terms of word counts.

As a result, BoW is used as input data for Feed Forward Neural Network, where passing through the process the probability of any class is given (Figure 7). The number of classes is unlimited and if, for example, the class "greeting" was given, the response will be according to this class.

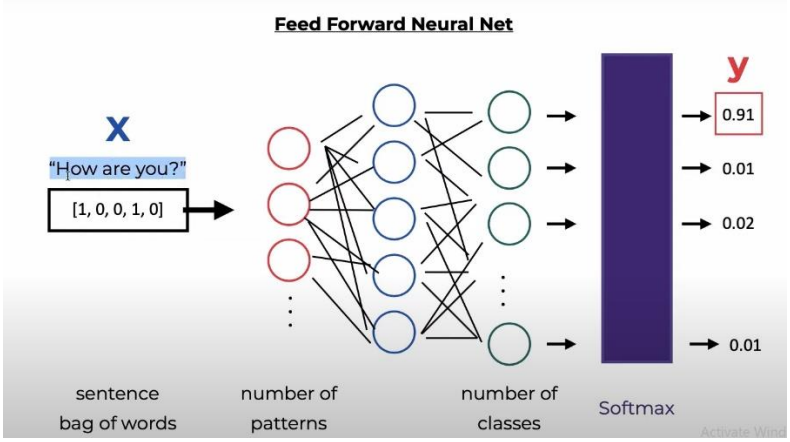

Figure 7. Feed Forward Neural Network.

https://www.youtube.com/watch?v=RpWeNzfSUHw&t=910s

PyTorch is a machine learning library developed by Facebook. The Asyk chatbot system is used to create, train, and utilize neural networks. PyTorch is needed to create the model architecture, load pre-trained weights, and perform forward propagation for predictions.

Tkinter is a standard Graphical User Interface (GUI) toolkit. It is the most widely used library for creating simple and complex GUI applications. For more comfortable interaction between the user and the bot a primitive GUI was created in Tkinter.

3. CHATBOT SYSTEM

As was stated earlier, the interface of the program was created by using the Tkinter library. The program interface was created to make the interaction between the chatbot and the user more comfortable. The interface consists of a window that greets the user with the message "Say 'hello' to Asyk!" and a line where you type your message to the chatbot. The "Send" button or the "Enter" key is used to send the message itself. The full interface is shown in Figure 8.

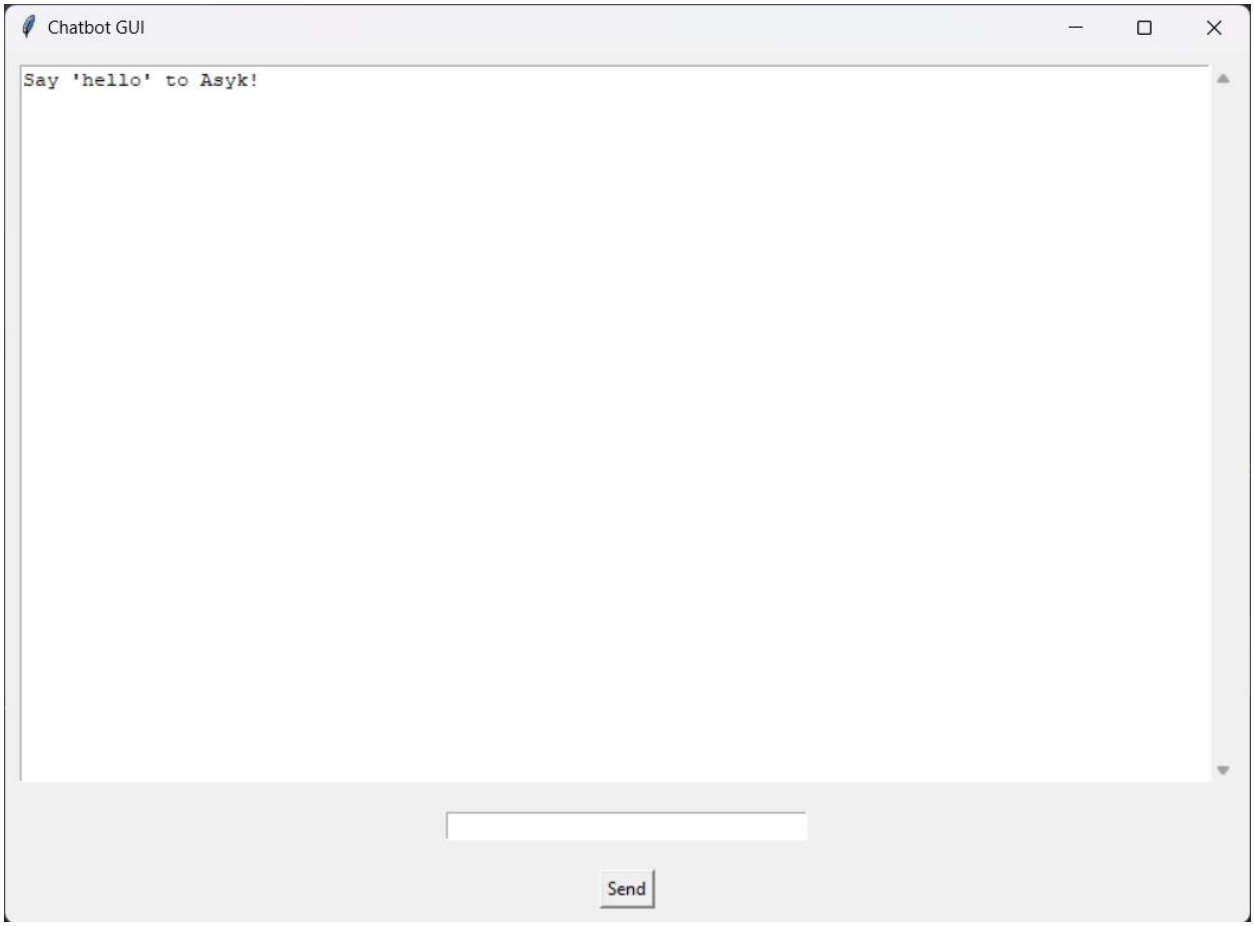

Figure 8. Interface of «Asyk» Chatbot

The program was created in the Python language environment. For its functionality 4 .py files (chatbot.py, nltk_utils.py, train.py, model.py) and one .json file (intents.json) are used. The .json file serves as a repository for the data used for machine learning of the program. It contains information about the name of a particular group, its pattern, and the answer to the query. At the time of program development and writing this document, the .json file contains the following

groups of files: "greetings", "thanks", "goodbye", "identity", "information", "tymak", "altybakan", "kyz kuu", "kokpar", "audarspak", "baiga", "tyime alu", "alaman baiga", "aksuyek", "oramal", "asau mastek".

The groups "tymak", "altybakan", "kyz kuu", "kokpar", "audarspak", "baiga", "tyime alu", "alaman baiga", "aksuyek", "oramal", "asau mastek" are responsible for information about the Kazakh national games themselves. The "greetings" group is responsible for greetings, the "goodbye" group is responsible for saying goodbye, the "thanks" group is responsible for responding to the user's gratitude, the "identity" group contains information about the purpose of the bot and the knowledge that it holds, and the "information" group shares generalized information about the Kazakh national games. Figure 9 shows what the file itself looks like. To fill up train data we will use information from Wikipedia and Welcome.kz (2019) websites.

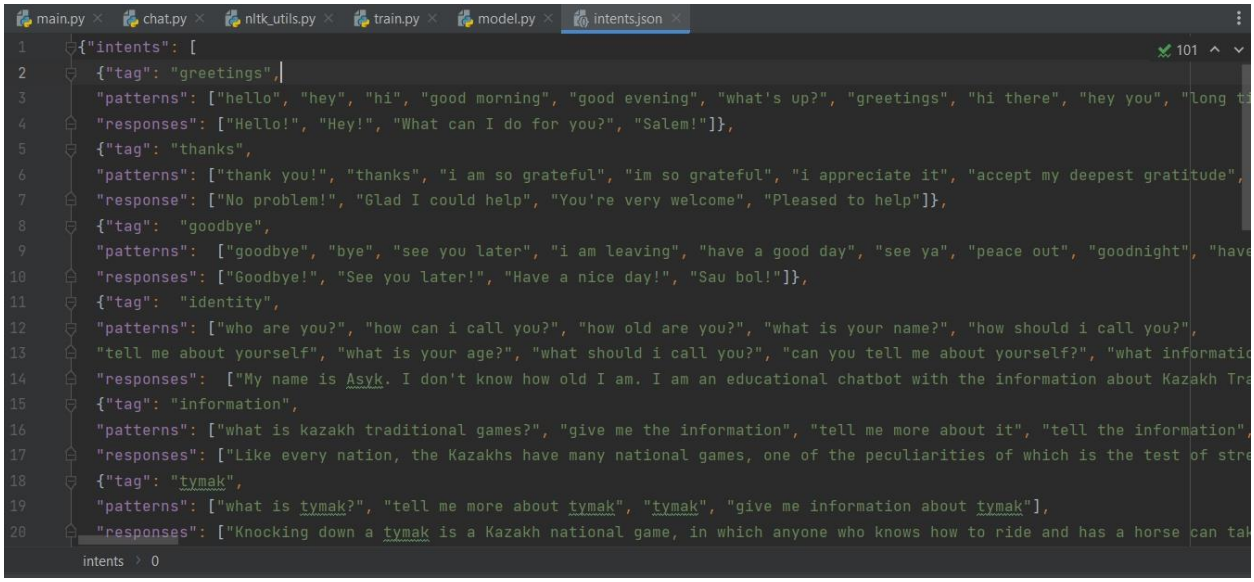

Figure 9. JSON file

Of the four .py files, the most important one is chatbot.py, which is responsible for launching the chatbot itself. The other three support its functionality. As it was said earlier three libraries were used, namely NLTK, PyTorch and Tkinter.

When entering queries to the chatbot, the response is generated almost instantly, however, in case the chatbot does not have an answer to the query, it will display "I cannot give you an answer, please ask again". A variant of the dialogue between the user and the chatbot is

shown in Figure 10.

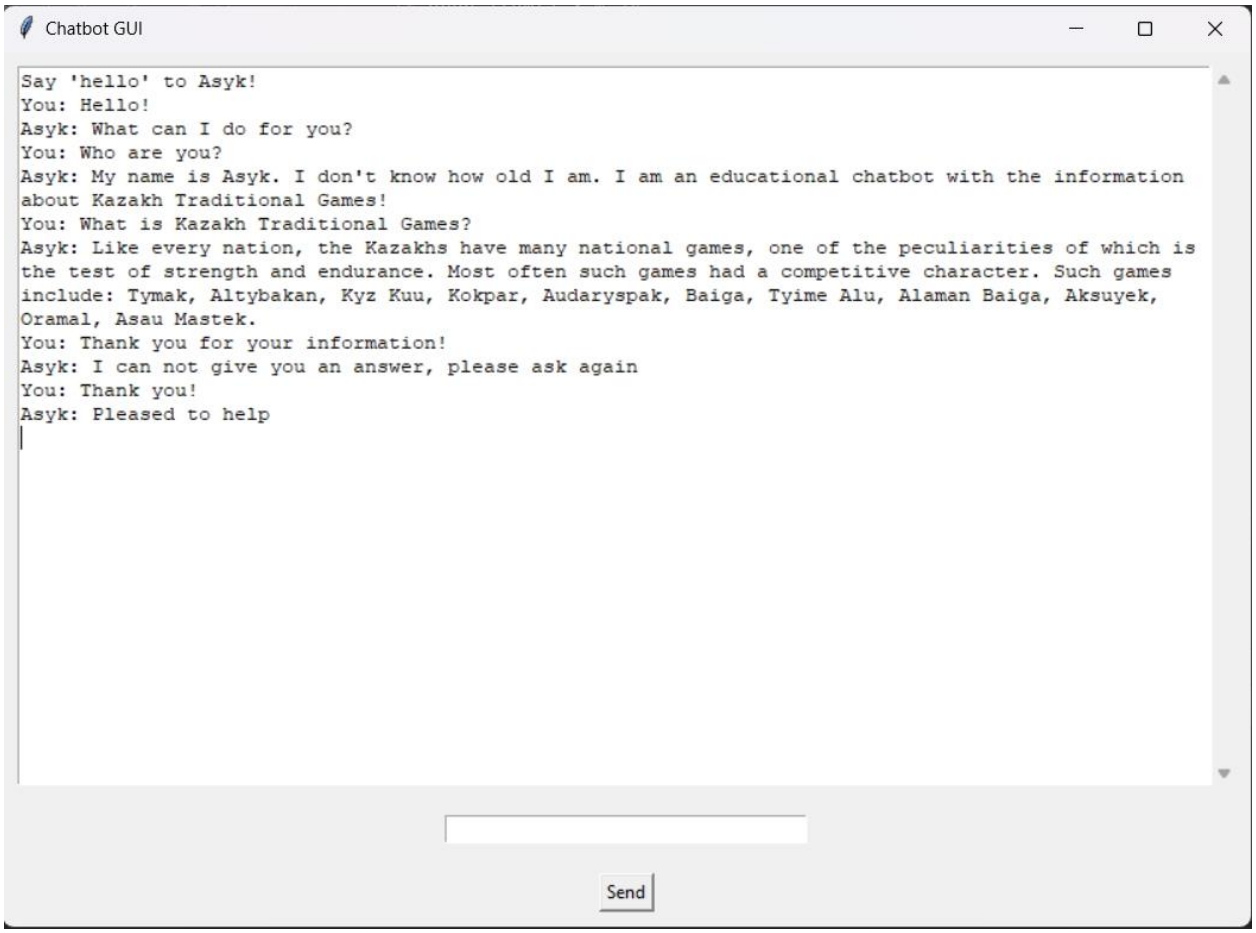

Figure 10. Dialogue Between User and Chatbot «Asyk»

4. USER ACCEPTANCE TESTING

The main time of this project was spent on its development. To have a small insight into the app itself and the opinions of the people around it, Google Forms was created with five questions. Two yes/no questions, two open-ended questions and a scale question where the app itself was rated. Ten people took part in the questionnaire, which is not enough for a complete picture and for future development it is worthwhile surveying a lot more people. Despite this, the first information about the app was gathered.

The first question "Do you like the idea of integrating chatbots into the education system?" aims to evaluate the very idea of integrating chatbots into education. All 10 people answered "yes" (Figure 11).

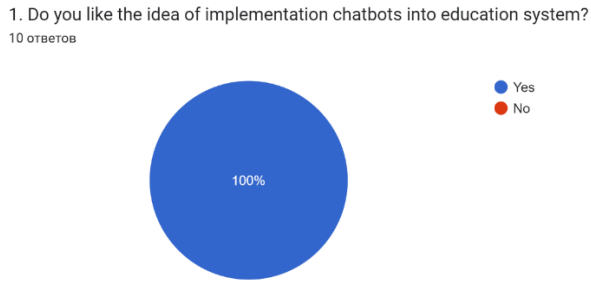

Figure 11. Question 1

The second question "Have you found the information provided by Asyk accurate and informative?" aims to find out how useful information about Kazakh culture can be. All 10 people answered "yes" (Figure 12).

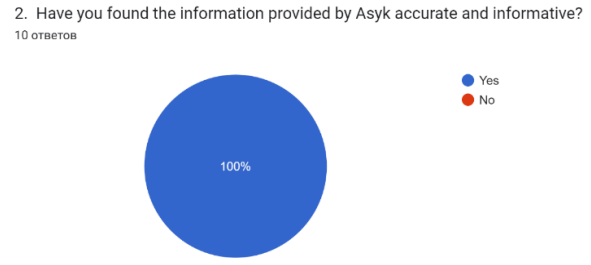

Figure 12. Question 2

The third question was open-ended and sounded like "What topics or aspects of Kazakh culture would you like to see in Asyk?". The options were: music, musical instruments, batyrs, clothes, and two answers were about completely making Asyk as a chatbot about the culture of Kazakhstan.

Figure 13. Question 3

The fourth question was to rate the performance of the program, where 1 is bad and 10 is good (Figure 14). The overall score was 5.1.

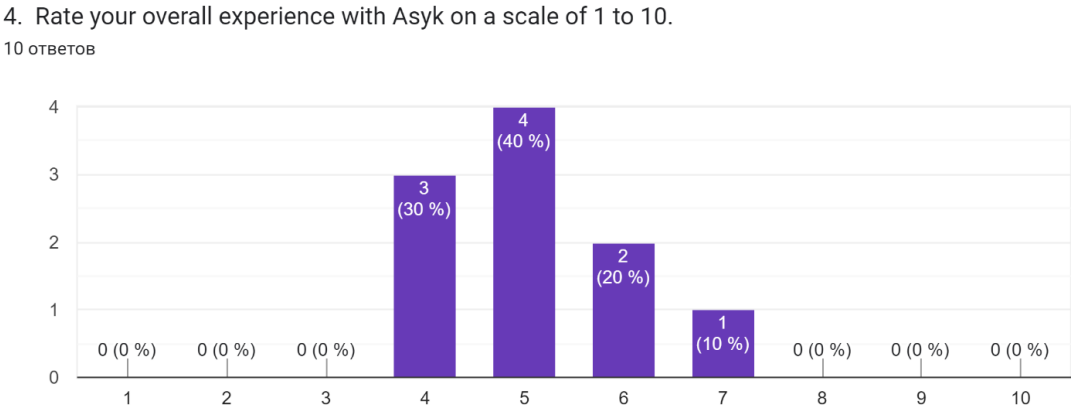

Figure 14. Question 4

The fifth question, also open-ended, was "Do you have any suggestions for improving Asyk?". The main suggestions for improvement were to improve its database, increase the scope of its knowledge, improve the interface (Figure 15).

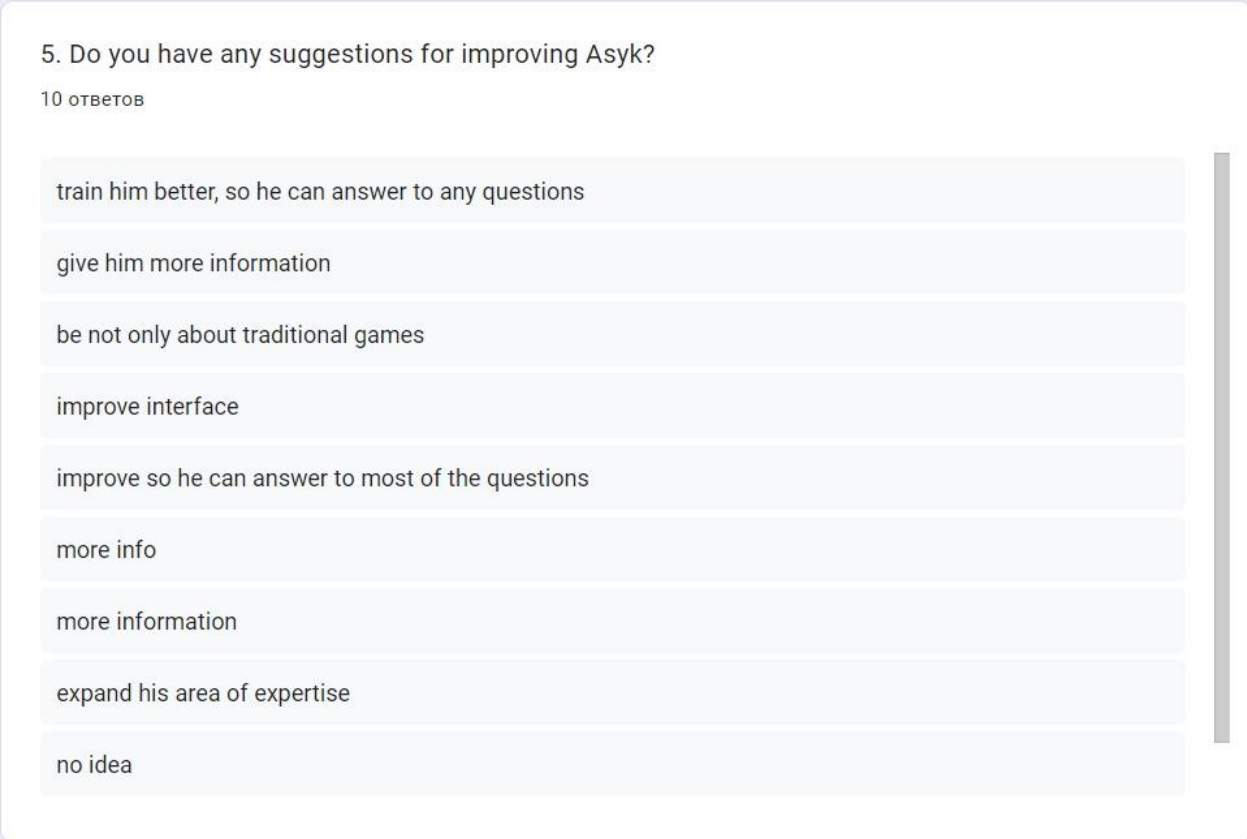

Figure 15. Question 5

5. FUTURE ENHANCEMENT AND CONCLUSION

The development of AI has made a huge contribution to the development of many areas in people's lives and one of these useful applications is chatbots that can provide quick help for a user who has a question on a particular topic. The development of AI related applications always has the potential to get a place in any field, that's why the chatbot "Asyk" can get its development in the field of education.

However, such conclusions require a lot more testing of the app and its evaluation. Conducting surveys not among 10 people but much more, but before that the chatbot requires improvement of its database on which it is trained. The more queries the chatbot can cover, the greater its efficiency. With the expansion of its database there is the possibility of expanding its scope of expertise, thus transforming into a chatbot not only Kazakh national games, but also the culture itself. The interface is simple and requires improvement for a more comfortable experience. With more time the chatbot has the potential to be a useful product that will help many people.

REFERENCE

Alina, Dzh. (2019). National games in Kazakhstan. https://welcome.kz/ru/kazakhstan/national-games

gk_ (2017). Contextual Chatbots with Tensorflow. Chatbots Magazine. https://chatbotsmagazine.com/contextual-chat-bots-with-tensorflow-4391749d0077

Haigh, T. (2023). There Was No'First AI Winter'. *Communications of the ACM*, *66*(12), 35-39.

Loeber, P. (2020). Chat Bot With PyTorch. YouTube. https://www.youtube.com/watch?v=RpWeNzfSUHw&t=910s

Shenoy, A. (2024). 6 types of chatbots – How to choose the best for your business?. https://yellow.ai/blog/types-of-chatbots/

“Sport in Kazakhstan” (2018) Wikipedia. https://en.wikipedia.org/wiki/Sport_in_Kazakhstan