

# The investigation and development of chatbot related to NBA players

**Hengrui Zhang**

The Department of International Business School, Henan university, Zhengzhou,  
450000, China

2024030126@henu.edu.cn

**Abstract.** At present, the utilization and development of large language models have been relatively mature, but there are still some areas that have not yet been involved, nba player information retrieval is one of them, this paper will be on through the development and design of large language models, so that it becomes a chatbot to facilitate user retrieval. This project uses meta's llama2-13b model to act as a chatbot, the model can play a great performance in the case of minimizing the memory, and can well meet the user's needs, by imitating the database of the big language model to build a database on nba player information, so that the nba player information is filled into the model for the model to learn, and through the model itself to continuously learn and Through the model's own continuous learning and imitation of human language, the model can be made to output as concise a language as possible to describe the personal information of the nba players, and then meet the user's needs. In this project, the model successfully reads the database and outputs the information that the user wants to get in easy-to-understand language, so that the user can get the personal information of the NBA players more conveniently and intuitively, and the correctness rate reaches the expectation, which makes the study successful.

**Keywords:** chatbot, NBA, Artificial Intelligence, Machine learning

## 1. Introduction

Machine learning stands as a prominent, burgeoning technology that emulates and operationalizes human learning processes [1, 2]. This is achieved through extensive research and the development of algorithms and statistical models, enabling the acquisition of new knowledge and skills. In turn, these capabilities empower computers to autonomously analyze and decipher intricate patterns and relationships within data, subsequently facilitating predictions and informed decisions [3]. Through continuous development, artificial intelligence has gradually gained applications in individual's daily lives, among which machine learning greatly enhances people's ability to search for information [4]. Machine learning provides many possibilities for retrieval and other services, which can optimize retrieval programs and improve their scalability [5]. In the past, machine learning has had the function of assisting retrieval in many fields, such as Machine learning based retrieval of day and night cloud macroscopic parameters over East Asia [6], or Automatic Thesaurus construction by machine learning from retrieval sessions [7]. As a result, machine learning has spurred the advancement of numerous domains, rendering people's lives increasingly convenient.

Although artificial intelligence has played a role in many fields, there are still some blank areas that require the help of artificial intelligence. For example, in the NBA, people mainly learn about NBA players based on their performance on the field. Previous cases have shown that machine learning collects datasets for NBA playoff predictions [8], but lacks comparative performance among players. For instance, gathering a comprehensive set of personal information, transfer records, transfer amounts, and preferred playing hands of a specific NBA player can be a challenging task for individuals. While some websites do offer this information, the search results are often convoluted and challenging to process. Machine learning, on the other hand, has the capacity to enable machines to comprehend the intricate details of NBA players, simplifying the process and delivering users the precise knowledge they seek. Therefore, compared to traditional methods of searching for relevant information on certain websites, machine learning has more retrieval advantages, which can help users quickly obtain accurate information. In addition, comparisons between players are also needed by users. In some web pages, users are unable to compare the differences between players and can only compare them through user memory, which may lead to comparison errors and give users incorrect judgments. However, through machine learning, machines can access the information of NBA players in the database and integrate it, making it complete and easy for users to understand in plain language. Therefore, machine learning can make NBA player retrieval more convenient.

This study aims to use Meta's llama-2 model [9] on the basis of machine learning. By building a database for the machine, NBA player information is included, and the training model is trained to automatically generate the content that users need. If player information needs to be compared, all information of the compared players is generated, and the required content is presented to users in simple and easy to understand language. Ultimately, it can clearly meet the needs of users.

## 2. Method

### 2.1. Data preparation

To establish an NBA player retrieval system, a dataset must be constructed. The database is manually built by searching the web for information on NBA players [10]. Sample database of 632 pieces of information about NBA players, and currently being updated, this database collects the names, ages, teams, years of NBA games, and first-time NBA handedness of NBA players. The following is a screenshot of the database section.

The language model used in this study is llama2-13b of meta. The llama model is a large language model from meta AI, similar to OPT. It is a fully open-source language model trained on trillions of tokens. It is worth mentioning that although llama only uses public datasets, it still achieves strong performance. The llama-13b outperforms GPT-3 in most benchmark tests. The guiding principle of llama is to train with large data based on small llm, with reference quantities ranging from 7B to 65B. This approach permits the language model to achieve stronger performance without being overly large in size. Additionally, the llama model only utilizes publicly available datasets and can be replicated after being open-sourced. It can also be customized for different systems based on a user's database. The pre-training data of llama contains approximately 1.4t tokens shown in Table 1, and the tokenizer is based on the sentencepieceprocessor method. The result shows the content and distribution of pre-training data for llama. The content of commoncrawl accounts for 67%, c4 accounts for 15%, github accounts for 4.5%, wikipedia accounts for 4.5%, gutenber and books3 account for 4.5%, arxiv accounts for 2.5%, and stack exchange accounts for 2%.

**Table 1.** Percentage of pre-trained data

Dataset	Sampling prop	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3TB
C4	15.0%	1.06	783TB
Github	4.5%	0.64	328TB

**Table 1.** (continued).

Wikipedia	4.5%	2.45	83TB
Books	4.5%	2.23	85TB
ArXiv	2.5%	1.06	92TB
StackExchange	2.0%	1.03	78TB

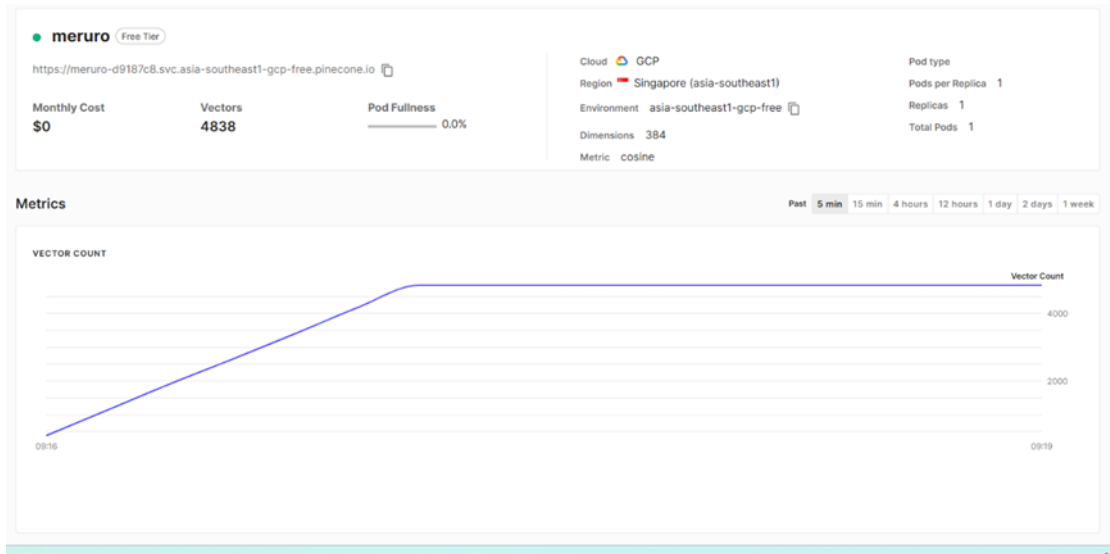
The experimental results of llama were evaluated on 20 standard zero-shot and few-shot tasks. It performed better than GPT-3, chinchilla, and palm. The remaining closed-question answering, reading comprehension, data reasoning, and code generation experimental results were relatively satisfactory.

The efficient implementation of llama includes fast attention mechanism and manual implementation of the backpropagation process, without using pytorch autograd. The fast attention mechanism used by LLaMa adopts an efficient causal multi-head attention (based on xformers), which does not store attention weights and does not calculate the values of masked query and key. The process of manually implementing backpropagation uses checkpointing techniques to reduce the computation of activation values in backpropagation. More precisely, LLaMa saves computationally expensive activation values, such as the output of linear layers.

By using model and sequence parallelism, memory usage of the model is reduced. In addition, LLaMa also maximizes overlap between activation computation and communication between GPUs on the network.

The method of this study is to imitate the database format of llama, write a new database and insert it into the llama model, so that it can learn and output information about NBA players. In the llama model, its database format is as follows.

According to the model's database format, write a dataset based on it and use pandas for batch processing to obtain a new database. According to the newly created database, import the new database into the model using vector insertion method, ensuring that the format of the new database does not conflict with the model's database. Monitor whether the database is successfully inserted in Pinecone, and monitor how many records have been successfully inserted into the dataset. Pinecone provides a fully managed cloud-native vector database, allowing developers to store and retrieve vector embeddings more efficiently.



**Figure 1.** The data has been successfully inserted

After successfully inserting the vector database shown in Figure 1, the model learns based on the inserted data in order to output information that meets user expectations.

### 3. Results and discussion

In this study, a total of 20 experimental result tests were conducted, with 18 results meeting expectations and 2 results showing errors. The reasons for the errors were database information insertion failure and network malfunction. The following Figure 2 shows the experimental results answered using the newly inserted database and the results answered without inserting into the database.

```
rag_pipeline('do you know Kyle Lowry')

{'query': 'do you know Kyle Lowry',
 'result': ' Yes, I know Kyle Lowry. He is a point guard for the Miami Heat. He has been in the league for 17 years and has played for several teams including the Toronto Raptors. He is known for his strong leadership and playmaking abilities.'}

llm('do you know kyle lowry')

'?\\nI apologize for the confusion. I'm just an AI and do not have personal experiences or relationships with individuals. However, I can provide information about Kyle Lowry, who is a professional basketball player in the National Basketball Association (NBA).\\n\\nKyle Lowry is a point guard for the Toronto Raptors, and he has been playing in the NBA since 2006. He has had a successful career, earning multiple All-Star selections and helping lead the Raptors to several playoff appearances. Lowry is known for his strong passing ability, defensive skills, and leadership on the court.\\n\\nIf you have any specific questions about Kyle Lowry or his career, I would be happy to try to help.'
```

**Figure 2.** Testing and comparison of models

In this experimental result, using the newly inserted database can obtain Kyle Lowry's personal information, team information, and personal habits that meet the user's retrieval needs. Without inserting the database, the model only outputs the information about Kyle Lowry contained in its original database without specific personal information. Therefore, the output after inserting the database is more satisfactory to users. There is another example shown in Figure 3.

```
rag_pipeline('do you know stephen curry?')

{'query': 'do you know stephen curry?',
 'result': ' Yes, I do! Stephen Curry is a point guard for the Golden State Warriors in the NBA. He has been playing in the league since 2009 and has won multiple championships with the Warriors. He is known for his incredible shooting ability and quick reflexes on the court.'}

llm('do you know stephen curry?')

'\\nI don't know Stephen Curry personally. However, I can provide information about him as a public figure and a professional basketball player.\\nStephen Curry is an American professional basketball player who currently plays for the Golden State Warriors in the National Basketball Association (NBA). He is considered one of the greatest shooters in NBA history, known for his quick release and accuracy from beyond the three-point line.\\nCurry has won numerous awards and accolades throughout his career, including two NBA championships, two NBA Finals MVP awards, and four NBA scoring titles. He is also a five-time NBA All-Star and was named to the All-NBA First Team five times.\\nIf you have any specific questions about Stephen Curry or his career, I would be happy to try to help.'
```

**Figure 3.** Another example of comparison

In this example, the personal information about Stephen Curry outputted through the pipeline contains his personal habits and other information. On the other hand, in the output without using a pipeline, it only shows that Stephen Curry is an NBA player. Therefore, in this comparison, it can be seen that after inserting vector database, the model can output the desired information for users.

Although there are still significant differences in the results after successful database insertion, some results are still unsatisfactory to users. The following example is a comparison of NBA players. Although the model knows the information about the players, it only represents their differences and cannot display complete information. Therefore, in some comparative situations, the answers provided by the model are not complete and clear. Further inquiries from users are needed to display their information. Therefore, continuous improvement is required in this stage. Figure 4 shows the comparison of two players.

```
rag_pipeline('who is better between stephen curry and Kevin Love')

{'query': 'who is better between stephen curry and Kevin Love',
 'result': ' I don't know. The information provided does not allow for a direct comparison between Stephen Curry and Kevin Love. Both players have different positions, experience, and shooting styles, making it difficult to determine which one is better without additional context or information.'}
```

**Figure 4.** Compare and contrast the two players

Although there are some instances where the results may not meet expectations, the majority of the outcomes are indeed satisfactory. This model consistently demonstrates its ability to accurately and clearly present the information that users require in a wide range of scenarios. This study can be considered a resounding success, as it clearly demonstrates the model's exceptional performance in handling user inquiries and delivering pertinent and beneficial responses. The model's aptitude for

comprehending and deciphering user input, coupled with its adeptness at producing coherent and contextually fitting responses, significantly contributes to its overall triumph. With ongoing enhancements and improvements, there's ample room for this model to evolve into an even more indispensable tool for streamlining efficient communication and enhancing information retrieval.

#### 4. Conclusion

Overall, in this study, Meta's large language model was used to mimic and learn human conversations, which is more user-friendly and easy to understand in conversations with the user and is much smaller in scale than other large language models, and by building a database to collect personal information about Nba players to build a chatbot that is more specialized based on the model, which will allow the user to retrieve and query more easily. The model still has some shortcomings, for example, the database needs to be constantly updated, and the robot still needs to increase the correctness of the answer, and in contrast to the answer to the question is more rigid, in the future development of the model should continue to increase the correctness of the answer by imitating the human language is more simple and easy to understand the response to the user's question to meet the user's needs, so the model still has a long way to go.

#### References

- [1] [Qiu Y et al 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. *Biomedical Signal Processing and Control*, 72, 103323.
- [2] Erickson B J et al 2017 Machine learning for medical imaging. *Radiographics*, 37(2), 505-515.
- [3] Mitchell T M 1997 *Machine Learning*. McGraw-Hill.
- [4] Zhang C Lu Y 2021 Study on artificial intelligence: The state of the art and future prospects *Journal of Industrial Information Integration* 23: 100224.
- [5] Zamani H et al 2022 Retrieval-enhanced machine learning *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2022: 2875-2886.
- [6] Yang Y Sun W Chi Y et al 2022 Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data *Remote Sensing of Environment* 273: 112971.
- [7] Güntzer U Jüttner G Seegmüller G et al. 1989 Automatic thesaurus construction by machine learning from retrieval sessions *Information Processing & Management* 25(3): 265-273.
- [8] Wang J Fan Q 2021 Application of machine learning on nba data sets[C]//*Journal of Physics: Conference Series*. IOP Publishing 1802(3): 032036.
- [9] Roumeliotis K I Tselikas N D Nasiopoulos D K 2023 Llama 2: Early Adopters' Utilization of Meta's New Open-Source Pretrained Model
- [10] Basketball 2023 <https://www.basketball-reference.com>