

# Design of Intelligent Customer Service Knowledge Base for Medical Insurance Based on Foundation-scale Models

Yixuan Lyu<sup>1,\*</sup>

<sup>1</sup>Northeast Yucai Foreign Language School, Shenyang, Liaoning, 110179, China

\*Corresponding author's e-mail: lvyx060614@hotmail.com

---

**Abstract:** In this paper, we propose a method for constructing an intelligent customer service knowledge base system for medical insurance based on natural language processing (NLP) models. This method utilizes a local vector database to construct the knowledge base system and retrieves industry knowledge by calling the vector database through foundation-scale models. The use of this technology ensures the security of business data and reduces platform training costs. The intelligent knowledge base of medical insurance constructed using this method can automatically handle user's inquiries, complaints, and general business transactions, effectively improving the efficiency and intelligence level of customer service.

**Keywords:** Medical insurance, knowledge base, vector database, natural language processing (NLP), foundation model.

---

## 1. Introduction

With the continuous development of the medical industry and the constantly change in the medical insurance policies, improving service capabilities has become an urgent need for industry reform. Intelligent customer service is the key to improving service capabilities. The medical insurance intelligent customer service system uses information technology to handle service requests automatically and accurately, improve the efficiency and quality of medical insurance services, save costs, and improve the service experience of the people.

Foundation models are machine deep learning technologies that have rapidly developed in recent years, which have obvious advantages in natural language processing and the generation of multimodal content. As an intelligent service platform in the medical insurance field, the medical insurance intelligent customer service system can better serve customers by using the natural language processing and content production capabilities of foundation models. However, because of the high deployment and training costs of foundation model platforms, and considering of the privacy and security of industry data, how to use foundation models effectively have become the current challenges in industry applications.

This paper describes a method for constructing a foundation model l industry application based on a local vector database. By using a locally deployed vector database to store medical insurance industry policies, domain knowledge, and business information, a domain knowledge base is formed[1]. Then foundation models are called to retrieve knowledge and generate content, realizing automatic customer inquiry services. This method not only utilizes the general capabilities of foundation models but also solves the problems of industry data security and high deployment and training costs of l foundation models. It is currently a widely promotable way to build foundation model industry applications.

## 2. Objectives and Key Technologies

### 2.1. Objectives

To construct an intelligent customer service knowledge base that supports channels such as telephones, service machines, and mobile apps, that can automatically reply to and handle service requests based on customer voice, including:

Answering users' questions about insurance policies: Users can ask various questions about insurance policies to the intelligent customer service through voice or text, such as payment standards, insurance benefits, complaints and reports, and other areas.

Querying medical insurance information: provide users with medical insurance information inquiry services, such as users can query their own insurance participation records, payment records, treatment records, and other information through intelligent customer service.

Handling medical insurance business: assist users in handling various medical insurance businesses, such as handling insurance registration, payment, benefit application, complaints and reports, and other business.

Providing voice recognition and synthesis functions: provide voice recognition and voice synthesis functions. Users can ask questions online by voice, and intelligent customer service can respond in a voice way as well.

Providing AI-powered chatbots for customer service: utilize technologies such as natural language processing and machine learning to mimic human conversation and engage with users, offering a range of services.

### 2.2. Key technology

The key technologies necessary for developing an intelligent customer service system encompass natural language processing, machine learning, and data analysis mining[2]. Natural language processing constitutes the foundational element of the system, converting diverse user inputs, including speech and text, into computationally processable formats and automatically synthesizing the

output comprehensively for users[3]. Machine learning technology can be employed to enhance the accuracy and intelligence of the intelligent customer service system by training it on historical conversation data, optimizing product performance and enhancing the overall user experience[4].

In recent years, research on foundation models has significantly boosted the effectiveness of natural language processing, representing significant progress in the field of artificial intelligence. Notably, platforms such as ChatGPT4 as well as a range of domestic large-scale model platforms provide highly effective natural language processing capabilities. A vector database is a specialized database that stores and processes vector data. Foundation models can acquire knowledge of the business domain through vector databases, and in this article, we used a vector database to construct a local business database, promoting faster response efficiency of foundation models and ensuring data security.

### 3. Design and Implementation

The medical insurance knowledge base is a system of medical insurance knowledge constructed by organizing and summarizing information related to medical insurance policies, processing procedures, and designated medical institutions. To ensure data security, a vector database was utilized to build a local medical insurance knowledge base for

this project. This database employs a large language model to retrieve information from the knowledge base through interface calls, creating a secure and reliable medical insurance knowledge management service system. It provides a range of comprehensive, integrated, and multi-channel support services for intelligent customer service systems.

#### 3.1. Application of vector database

The application of vector databases plays a crucial role in building industry applications based on large language models. Although foundation models can offer general answers, their knowledge depth, accuracy, and timeliness are limited in vertical fields. To address this issue, vector databases can be employed to convert enterprise knowledge base documents and data into vector representations through vector embedding. This enables the retrieval of similar knowledge through similarity search functions and interaction with foundation models, resulting in intelligent applications in proprietary fields.

Vectors can represent text and knowledge as mathematical vectors, allowing for functions such as text similarity calculation, knowledge base retrieval, and reasoning. Vector embedding is an AI-native data representation method ideal for various AI-based tools and algorithms. It can represent unstructured data or knowledge, including text, images, audio, and video, as illustrated in the figure below.

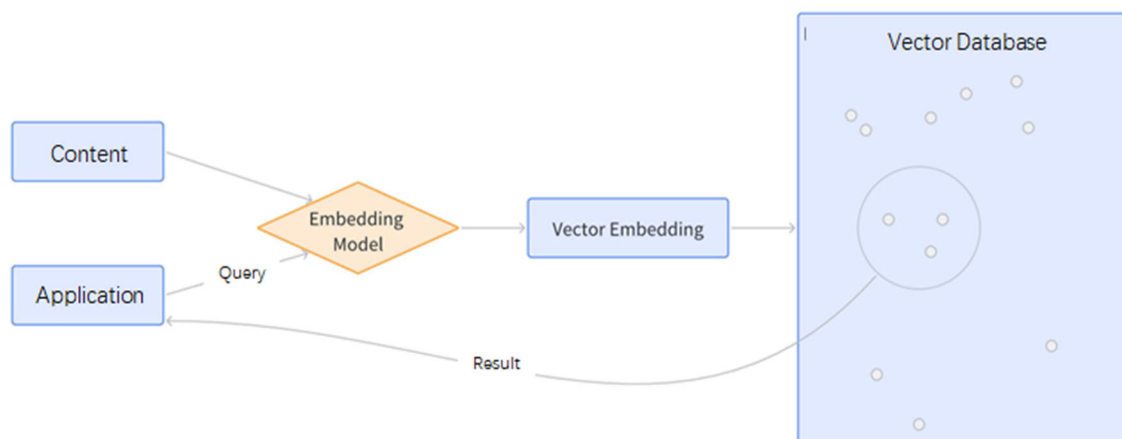


Figure 1.

In this project, the entire process of utilizing a vector database involves the following steps:

(1) Generating vectors for the content to be indexed, including text, images, and videos, using an embedding model.

(2) Adding these vectors, along with their corresponding original content, to the vector database.

(3) When an application issues a query, using the same embedding model to create vectors for the query and searching for similar vectors within the database. As previously stated, these similar vectors are associated with the original content used to create them.

This article utilizes the open-source vector database Milvus which is designed specifically for fast storage, retrieval, and analysis of foundation-scale vector data. It offers high-performance and scalable architecture, supporting various vector index algorithms and query modes. Milvus can be widely applied in fields such as recommendation systems, image searching, natural language processing, and machine learning, helping users quickly discover and analyze similar

vector data.

#### 3.2. Knowledge base design

The entire knowledge storage system is divided into three parts based on usage scenarios:

(1) Policy searches scenario

In this scenario, traditional structured data storage, Elasticsearch storage, and OSS file storage are mainly used. After being processed and extracted the text content of the original policy documents, they are stored in ES for search use. Meanwhile, the text content is stored in a structured document database for full text display. Finally, the original policy documents are saved in file storage for traceability.

(2) Policy Q&A scenario

In this scenario, vector libraries are mainly used. Similar to the first scenario, the text content needs to be extracted first. As for policy files, it needs to be sliced and segmented before being stored in the vector library. Regarding policy interpretations, questions and answers are extracted and compiled into Q&A pairs, indexed by question and stored in

the vector library. To better address policy issues, we use the summarization ability of foundation language models to condense the policy documents before they are stored, and the outcomes are similarly archived in the vector database.

(3) Providing language data for model training  
The training data for large models are generally in JSON format, such as:

```
{
  "instruction": "rewrite the sentence based on today's date",
  "input": "I want to see the trend of China's GDP over the past decade",
  "output": "I want to see the trend of China's GDP from May"
}
```

It is divided into three parts:(1) task prompt phrases (instruction section), which refers to what type of task is to be performed and will be trained separately for these tasks later;(2) the input section, which is the question asked; and (3) the output section, which is the expected result.

### 3.3. Calling the knowledge base with a foundation model

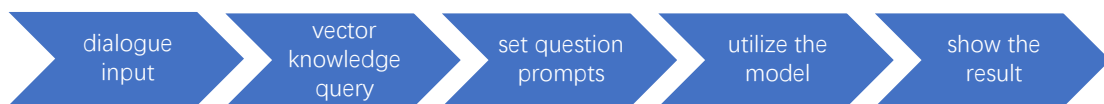


Figure 2.

After constructing the vector database, the following process is used to call the local vector database using a foundation model:

- (1) Dialogue input
- (2) Directly query the vector library, returning content
- (3) Construct question prompts
- (4) Call the model again and return the result

Based on the known information below, provide professional and concise answers to user questions. If the answer cannot be obtained from the provided content, respond with, "The content provided in the knowledge base is insufficient to answer this question." Fabrication is strictly prohibited.

## 4. Significance and Value of This Project

This article describes a foundation model application mode - building a knowledge base through a local vector database and calling a foundation model through it to implement an intelligent customer service knowledge base, which has certain significance for exploring the industry applications of foundation models. This implementation mode can not only reduce the technical difficulty and high cost brought by local deployment of foundation models but also provide efficiency for data schemes, protect data privacy and security, and avoid complex foundation model training processes. Specifically, it has the following significance:

**Reducing the cost of foundation model deployment and training:** Using a local vector library to call a foundation model can avoid local deployment and minimize the need for training, which can effectively reduce the system construction cost and technical difficulty.

**Protecting data privacy and security:** Constructing a knowledge base through a local vector library can ensure that the related knowledge does not become the general capability of the foundation model, which can guarantee data security and better protect data privacy.

**Improving data access speed:** Storing data on local devices, avoiding network latency and improving data access speed while reducing data transmission costs.

**Supporting offline calculation:** For scenarios requiring offline calculation, storing data in a local vector library makes it convenient for offline calculation.

Using a local vector library is currently a recommended industry implementation mode by multiple foundation model manufacturers. Through this project, we have built an intelligent customer service system that utilizes a foundation model to call a local knowledge base on medical insurance, achieving good application results. This has certain reference value for the industry applications of foundation models in vertical fields.

## 5. Summary

This article describes the construction method of foundation model applications based on local vector databases, which represents a practical exploration of the practical application of foundation models in the industry. Vector databases support vector storage, fast searching, indexing, sorting, and similarity calculations. Given that the training and inference processes of foundation models entail numerous vector computations, vector databases provide outstanding support for application systems based on foundation models.

By assembling the outstanding capabilities of foundation models in natural language processing and knowledge bases with local vector databases, industrial applications based on foundation models can effectively address problems related to the privacy, security, and data access efficiency of industrial data. This approach solves technical issues and mitigates high costs associated with deployment and model training, and holds significant potential for promotion.

## References

- [1] Zhang, H., Ding, Y., Zhang, Y., Feng, S. (2021). Design and Implementation of an E-commerce Intelligent Customer Service System Based on Deep Neural Networks Software Engineering, 05: 33-37.
- [2] Wang, D., Wang, W., Wang, S., Fu, J., Zhu, F. (2017) A review of natural language understanding methods for limited domain question answering systems. Computer Science, 44 (8): 1-841.
- [3] Bao, J. (2014). Research on Knowledge Based Automatic Question Answering. Thesis of Harbin Institute of Technology.
- [4] Lu, J., Li, S. (2019) Design of an intelligent customer service robot question answering system based on a knowledge base. Computer Science and Applications, 9 (11): 2098-2104.