# Autonomous Question and Answer System Based on ChatGPT With Large Language Model

Jun Wang[1]

1 Arizona State University

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

## Abstract

Chat-GPT has become very popular in recent years. But there is a problem. Chat-GPT does not ask questions to the users. Therefore, Chat-GPT looks like a machine, not a human. However, users sometimes do not want a single answer. They want real things like food, cars, products, etc. Therefore, our system will ask users questions several times until they get what they really want. In this project, we not only resort to Chat-GPT3.5 to find questions, but also resort to traditional programming skills or databases to solve these problems. OpenAI's Chat-GPT3.5 will play the main role in this project. Furthermore, Java and Spring-Boot will be used in this project. These are mature frameworks for enterprise systems(de Oliveira, C. E., Turnquist, G. L., & Antonov, A., 2018). Finally, the MySQL database is used in this project. It provides comprehensive and reliable SQL database services. The data stored in MySQL instances can generate very large data sets(bin Uzayr, S., 2022). Python3 and some machine learning frameworks such as NumPy, Pandas, TensorFlow and PyTorch are used in this project to analyze user behavior(Liu, Y. H., 2020). In the future, the dataset can be integrated into webtext2 as a basic data element to connect the model training.

**Jun Wang**

*80Hou International Software LLC*

*Arizona State University*

## Introduction

70 years ago, Alan Turing made assumptions about how to tell whether an AI is successful or not. The Turing test is a research method in artificial intelligence (AI) to determine whether a computer can think like a human. As a result, humans can talk to AI. If humans cannot distinguish between humans and AI, who they are talking to. So, this means that AI is like a human(Saha, D., Brooker, P., Mair, M., & Reeves, S., 2023). But even today, humans have powerful computers, graphics processors, advanced algorithms and intelligent, trained models. Humans can still easily judge who they are talking to, a human or an AI. Therefore, we believe that AI is a reasonably mature technology. Today, Chat-GPT has made

significant progress. Chat GPT not only provides the user with an accurate answer, but can also provide them with very specific services, such as generating code for the programmer (Ray, P. P., 2023). As a result, more and more people are using Chat-GPT instead of Google. In today's world, we cannot yet provide users with complete AI products due to the limitations of technology. Therefore, the current AI and machine learning will combine some technologies from the Internet era to serve users as much as possible. Based on this situation, the project chooses Java, Spring-Boot, MySQL and Python to combine the API of OpenAI, which can provide the maximum utilization of current technologies for users.

## Related work

The project does not only require AI or machine learning skills such as Python, Scikit-Learn or TensorFlow. Skills other than AI such as Java, Spring-Boot and AWS must also be included in the project.

### Python

Python has a rich ecosystem of libraries and frameworks specifically designed for machine learning and data science. Libraries such as NumPy, Pandas, SciPy, Matplotlib and especially scikit-learn provide robust tools for data manipulation, analysis, visualization and model building. Python is the primary language for many popular deep learning frameworks such as TensorFlow, PyTorch and Keras. These frameworks provide efficient implementations of deep learning algorithms and enable easy experimentation with complex neural network architectures(Liu, Y. H., 2020).

### Java and Spring-Boot

In this project we have to build an enterprise system. Java and Spring-Boot form the core of our system. Java is known for its robustness and reliability and is therefore suitable for the development of complex and mission-critical applications. Its strong typing system, memory management and exception handling contribute to the creation of stable and maintainable software(de Oliveira, C. E., Turnquist, G. L., & Antonov, A., 2018). Java's scalability has proven itself and makes it a preferred choice for developing applications that need to handle high loads and scale horizontally. Java applications can run efficiently on a variety of platforms and architectures and are therefore suitable for use in different environments. Spring Boot, a microservices-focused framework based on the Spring Framework, has gained popularity for the development of microservices-based applications. It simplifies the setup and configuration of microservices and allows developers to focus on the business logic rather than the infrastructure(de Oliveira, C. E., Turnquist, G. L., & Antonov, A., 2018).

### AWS

Amazon Web Services (AWS) is one of the leading cloud computing platforms offering a wide range of services to individuals, businesses and governments. AWS offers elastic scaling, allowing users to quickly scale their infrastructure up or down as needed. This scalability is particularly useful for applications with fluctuating traffic patterns or unpredictable

workloads. AWS places great emphasis on security and compliance, offering a wide range of tools and features to help users protect their data and infrastructure. These include identity and access management (IAM), encryption, network security, monitoring and compliance certifications. Based on these situations, we deploy the project in AWS and then provide services to users(Gupta, B., Mittal, P., & Mufti, T, 2021).

## Research Questions

In the projects, we have three questions that need to be answered.

### Chat-GPT won't ask questions to users

When the user asks the first question, the Chat-GPT will answer it, and then the dialog process will be paused until the user asks another question. There is one problem: if we care about the users, we also have to care about what they need. Sometimes they need products and services. So, they ask questions to Chat-GPT. An example: A user wants to buy a car. The user asks the question where I can buy a car in San Jose. And then Chat-GPT only gives the user an answer. That ends the dialog. But if you go to a real car dealership, the agent will ask the user several questions, such as what kind of car do you like? How many budgets do you have? Or how many family members do you have? There are a series of questions that combine what you need. On the surface, Chat-GPT gives you wonderful answers, but right now Chat-GPT lacks features that focus on the user's needs. This is because Chat-GPT cannot ask questions to users (Zhong, Qihuang, Liang Ding, Juhua Liu, Bo Du, and Dacheng Tao, 2023). To solve this project, we build the application using traditional applications such as Java, Spring-Boot and MySQL. We use Java programming logic to store the user's question. And we can find the answer in our database.

### Risk question

This is a problem study of the ethical concerns surrounding ChatGPT and other advanced language models (LLMs). They argue that while LLMs have impacted various industries, including texting and summarization software, they can create social biases and toxicity that pose moral and societal risks(Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J.,... & Kazienko, P., 2023). Based on these situations, Chat-GPT will not give you an answer. Or it gives users very vague answers or tips like I am an AI model, I cannot answer such questions. However, users found that some normal questions are still considered risky questions. The authors investigated the ethical dimensions of ChatGPT, including bias, reliability, toxicity, and robustness, through experimental evaluations on several sample datasets. We found that the current benchmarks cannot adequately address a significant portion of the ethical concerns, so we provided additional case studies to shed light on them. Therefore, the additional system can be considered a good choice to address this issue.

### Outdated information

Currently Chat-GPT uses webtext2 as the dataset. Webtext2 is not real-time text. Therefore, the model does not have access to real-time data or the ability to update the knowledge base beyond the last training cut-off date in January 2022(Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H.,... & Hu, X, 2023). Although Chat-GPT strives to provide accurate and relevant information based on the data on which Chat-GPT has been trained, the information provided by ChatGPT may be outdated or no longer accurate due to changes that have occurred since the model was last updated. Therefore, users must verify the information with current sources, especially for time-sensitive or rapidly changing topics. Our system can provide real-time data to users. Some real-time APIs can be integrated into the Java service logic. This allows it to handle some responses in real time that ChatGPT cannot.

## Process and Methodology

### Step1: UI

In the project, the user interface is similar to that of any chat software such as WhatsApp or WeChat. In the chat application. Users send their messages on the right side. Your friend’s reply to their message is on the right-hand side. Message senders and recipients have different colored message fields. Autonomy's QA system still uses this kind of user interface. The only difference is that the message sent by autonomy AI. In our project, AI combines the model of Chat-GPT with the Java ecosystem, which provides an additional intelligent system to enforce the model of Chat-GPT.

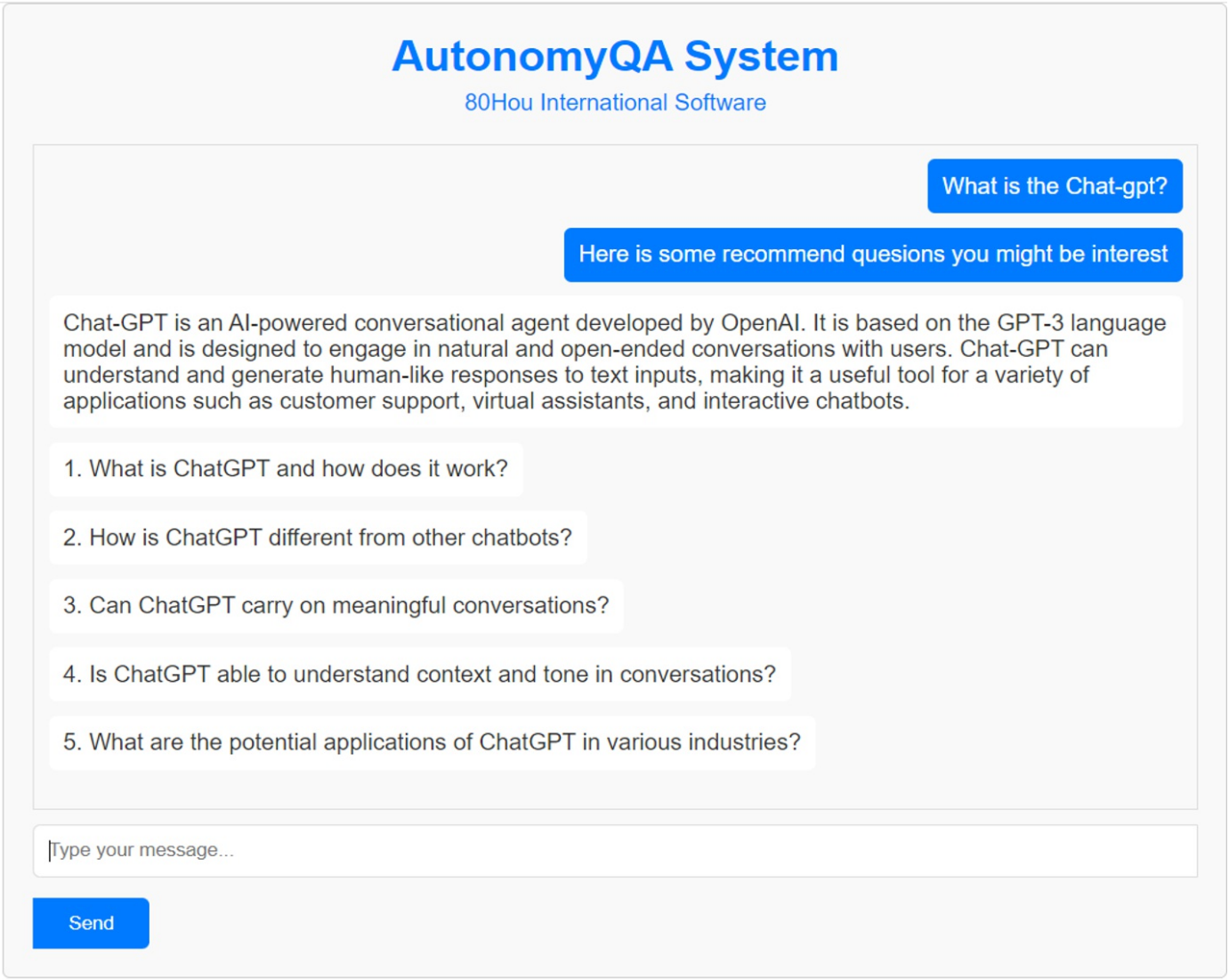


**Figure 1.**

## Step 2: Build system based on Java

A Java Spring-Boot application was created to enforce the Chat-GPT model. We still use the 3-Ties layer such as Controller, Service, and Dao to handle the user's request. In the controller layer, there is a function that receives prompts from the front page, e.g. Html and JavaScript.

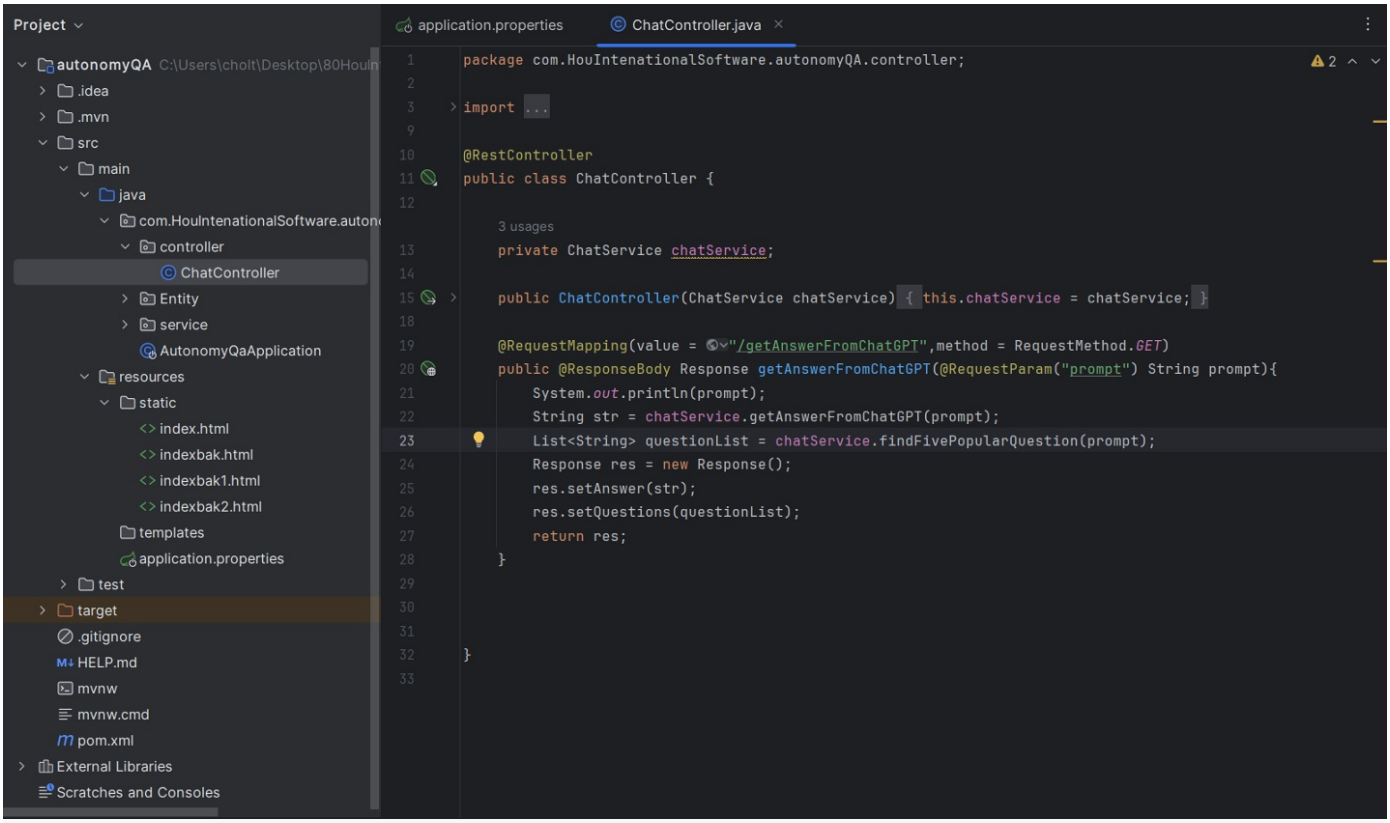


**Figure 2.**

In the services we deal with some business logic. Here, there are two main functions getAnswerfromChatGPT (String prompt). In this function Java calls the API of OpenAI and then gets a result. Another function is findFivePopularQuestion (String prompt, String answer). Based on this function, our system returns five questions to answer the users' queries.

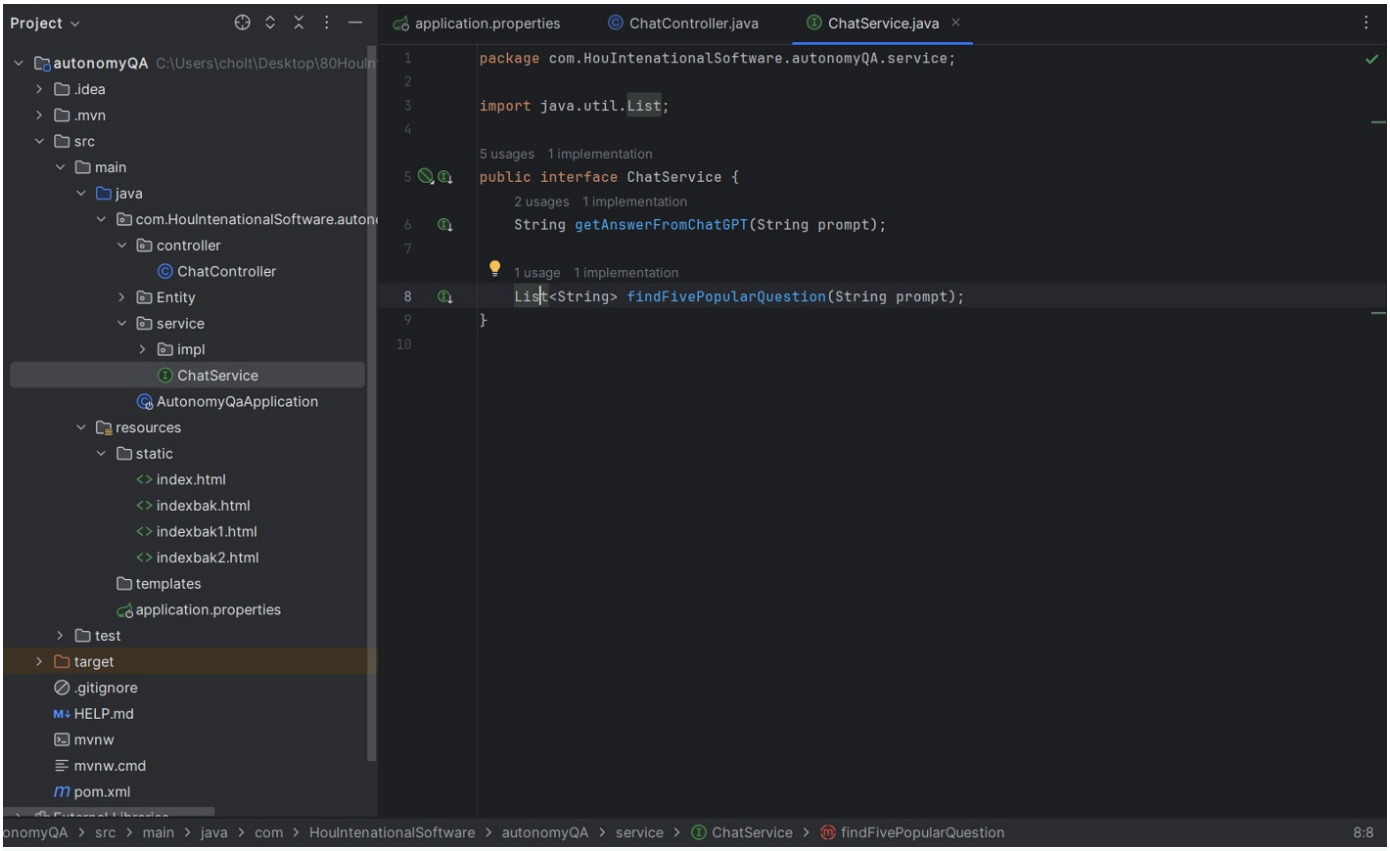


**Figure 3.**

The OpenAI API is still called via the traditional channels. The system must obtain permission from OpenAI, such as URL, apikey and model. The project uses Java to generate the HTTP request and response that go through the OpenAI API.

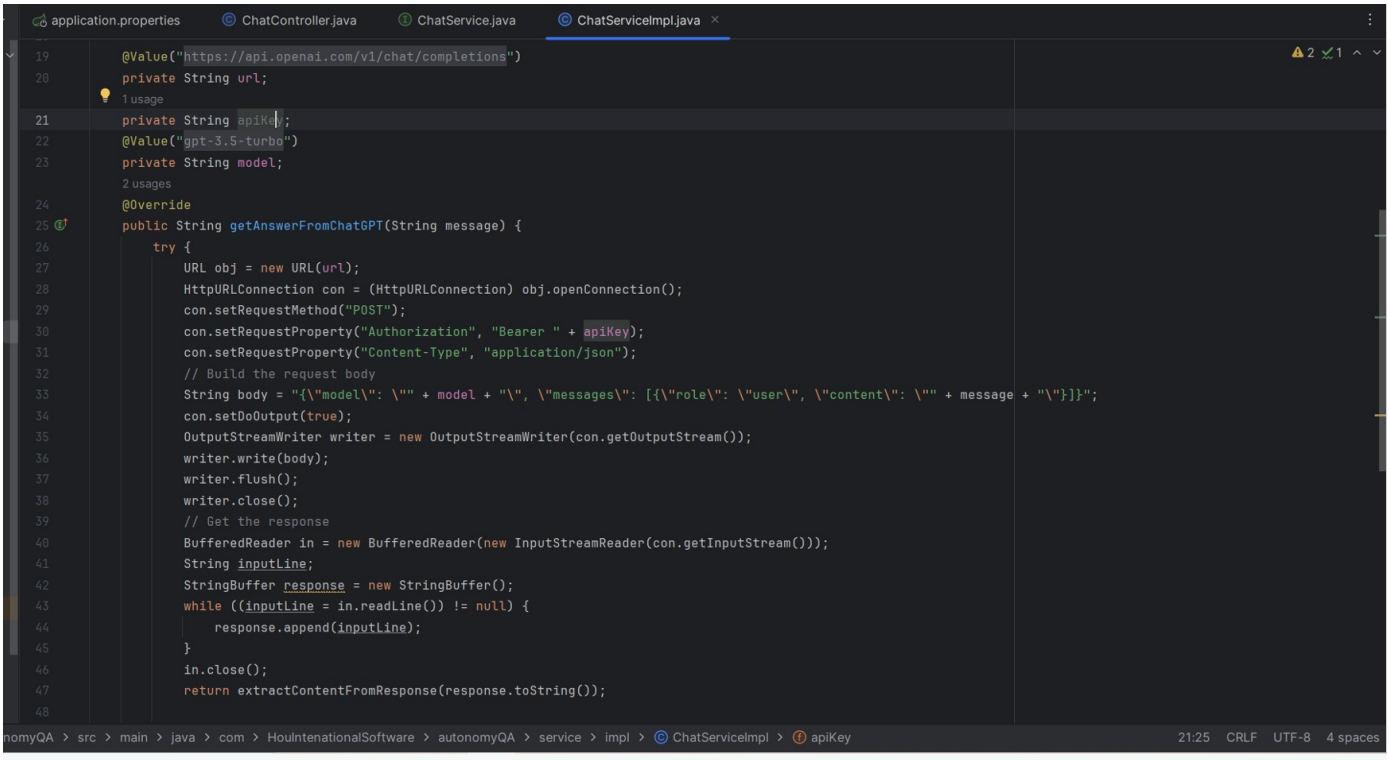

Figure 4.

### Step 3: database to build dataset

```
CREATE TABLE Users (
    UserID INT PRIMARY KEY AUTO_INCREMENT,
    Username VARCHAR(50) NOT NULL,
    Password VARCHAR(100) NOT NULL,
    Email VARCHAR(100) NOT NULL,
    Phone VARCHAR(20),
    Address VARCHAR(255),
    DateJoined DATE NOT NULL
);
CREATE TABLE QuestionAndAnswer (
    QuestionID INT PRIMARY KEY AUTO_INCREMENT,
    UserID INT NOT NULL,
    QuestionText TEXT NOT NULL,
    AnswerText TEXT NOT NULL,
    DateAsked DATE NOT NULL,
    DateAnswered DATE,
    FOREIGN KEY (UserID) REFERENCES Users(UserID)
);
```

Based on these two tables, the user dataset can be created. And then we can use Python to train the model using this dataset.

### Step 4: deploy system into AWS

There are several advantages to storing applications in AWS:

**Scalability:** With AWS, users can easily scale their application as needed. Users can dynamically adjust resources to handle increased or decreased workloads without having to make significant upfront hardware investments(Wittig, Andreas, and Michael Wittig, 2023).

**Flexibility:** AWS provides a wide range of services, including compute power, storage options, databases, machine learning, analytics and more. This flexibility allows users to choose the specific services that meet the needs of their application(Wittig, Andreas, and Michael Wittig, 2023).

**Cost efficiency:** AWS works on a pay-as-you-go model, meaning you only pay for the resources user use. This can be more cost-effective than maintaining and managing physical hardware, especially for small and medium-sized

businesses(Wittig, Andreas, and Michael Wittig, 2023).

**Global accessibility:** AWS has data centers (regions) around the world. This global infrastructure allows users to deploy their application close to them, reducing latency and improving performance for a global user base(Wittig, Andreas, and Michael Wittig, 2023).

**Reliability and availability:** AWS offers high availability and reliability through redundant and geographically distributed data centers. This ensures that users application remains available even in the event of hardware failures or other problems(Wittig, Andreas, and Michael Wittig, 2023).

**Security:** AWS offers a robust set of security features and compliance certifications. AWS invests heavily in security measures to protect data and infrastructure and provides tools to help you secure your applications(Wittig, Andreas, and Michael Wittig, 2023).

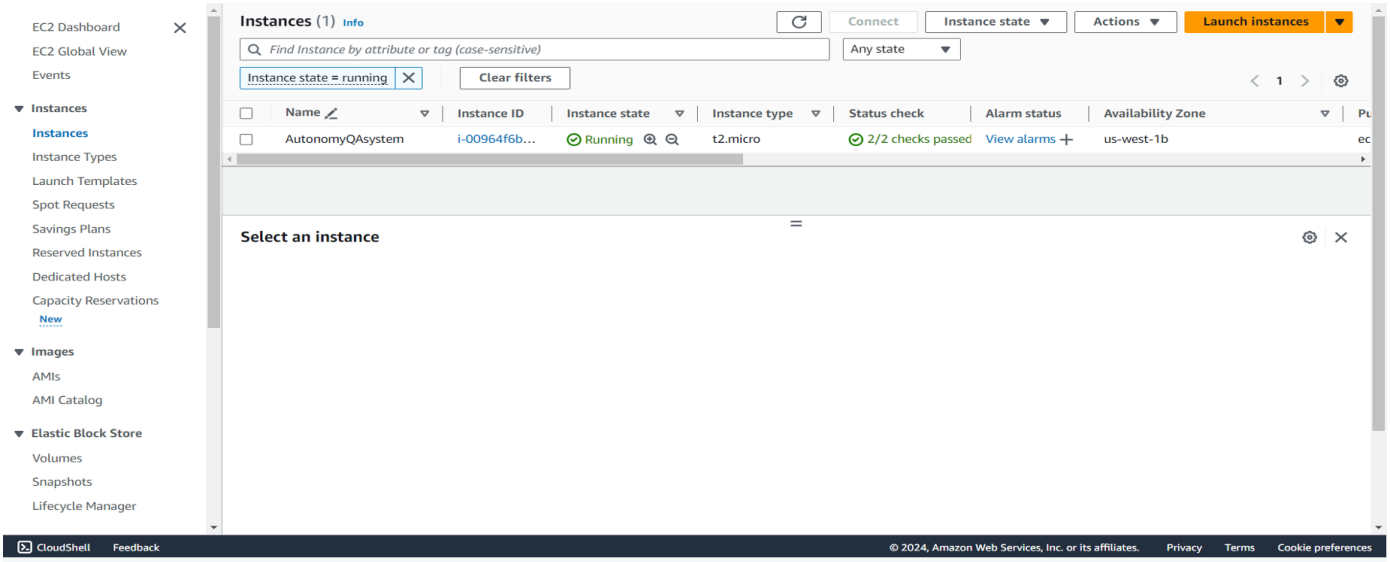


**Figure 5.**

## System architecture

In the high-level architecture, users send the query or request to the Autonomy Question Answer System. When the system has completed the validation, it calls three parts. First, the system calls Chat-GPT to get the first answer. In the second part, the system will make a recommendation. Since Chat-GPT cannot answer real-time or risk questions, the recommendation system will answer such questions. To solve outdated information, the system can call Quora or Watson Discovery API to get the latest answer(Das, D., & Semaan, B., 2020).

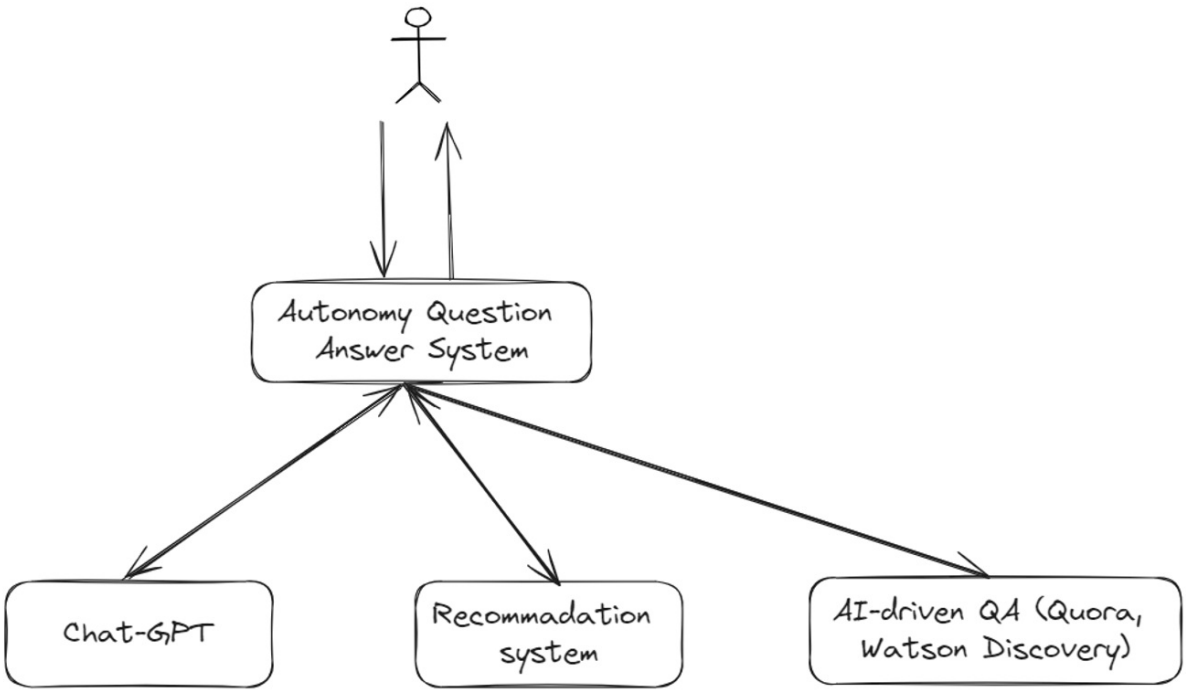


**Figure 6.**

This high-level architecture uses a distributed system to ensure high availability when many users access the system simultaneously. First, the user request goes through load balancing and is then forwarded to the Autonomy Question Answer System running on AWS EC2. The Redis cluster stores the hot data from the question and answer. If the system cannot retrieve the answer from Redis. It calls Chat-GPT4, a recommendation system and an AI-driven API. The answer and question are sent back to the user. The cluster database instance is configured and deployed behind the recommendation system to make the dataset as large as possible. The dataset is trained with another Python project.

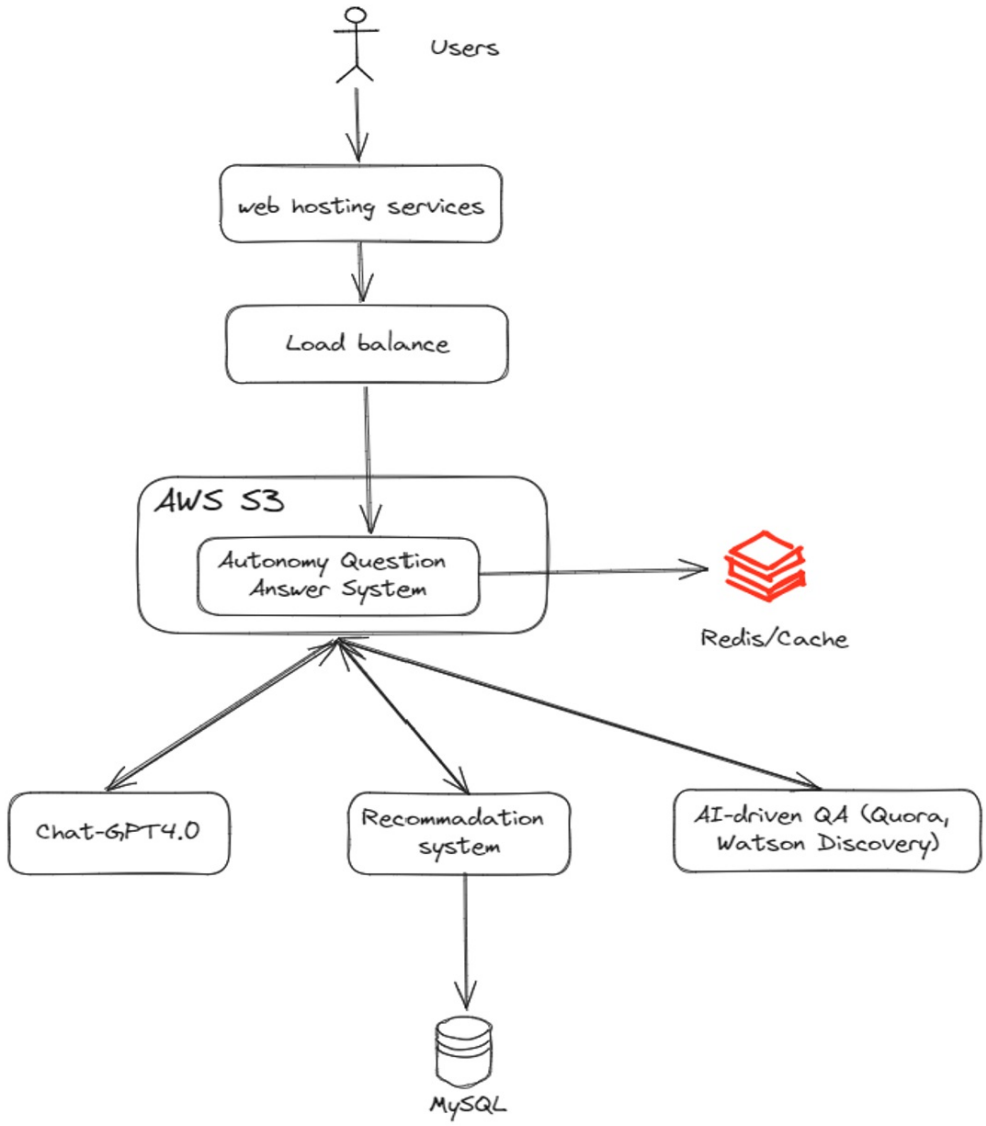


**Figure 7.**

## The Impact of Autonomy QA System Across Different Fields

In the current situation, Chat-GPT cannot customize the content of a specific area. Due to the training, the new model requires new data sets, and the maintenance and cleanup of such data sets is required at least monthly or annually. For these reasons, the training of Chat-GPT is slower. For example, Chat-GPT's model has used a dataset like webtext2 until 2024, which is based on the 2021 version. The Autonomy QA system can therefore be used in various fields such as banks, telecommunications companies and Internet companies. It will combine Chat- GPT to provide real-time and more accurate information. In addition, the system can still be used in the following areas(Roumeliotis, K. I., & Tselikas, N. D., 2023):

**Customer service and support:** The Autonomy QA System is often used in customer service chatbots to handle requests, provide assistance and solve problems. Its ability to understand natural language makes it an effective tool for engaging with customers and providing them with relevant information(Wang, J., Liang, Y., Meng, F., Sun, Z., Shi, H., Li,

Z.,... & Zhou, J., 2023).

**Research and information retrieval:** The Autonomy QA System can help researchers by summarizing articles, answering research questions and generating hypotheses. It is also used for information retrieval tasks by helping users find relevant information from large data sets or document repositories(Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S.,... & Han, W., 2023).

**Legal assistance:** The Autonomy QA System can assist you with legal research, drafting legal documents and providing legal advice. It is used by law firms, legal departments and lawyers to optimize their workflows and increase efficiency(Ray, P. P., 2023).

**Social media and marketing:** The Autonomy QA System is used in social media management tools, chatbots and marketing automation platforms to engage with users, create content and analyze social media trends(Roumeliotis, K. I., & Tselikas, N. D., 2023).

## Prospects for Autonomy QA Applications

Today, Chat-GPT spends 700,000 dollars per day, the revenues cannot compensate for the daily costs. Many investor representatives predict that OpenAI will be bankrupt in 2024. Thanks to Autonomy QA system, which can be customized, Autonomy QA system can be integrated with Amazon Order, Quora and Reddit API to generate some successful business models and get the full revenue. There are several other areas that can be integrated with the Autonomy QA system(Roumeliotis, K. I., & Tselikas, N. D., 2023).

**Development of chatbot solutions:** Users can create chatbot solutions based on the Autonomy QA system and sell them to companies in various industries. These chatbots can be used for customer service, lead generation, sales support and more. By charging companies for access to your chatbot solutions, you can generate revenue(Kalla, D., & Smith, N., 2023).

**Content creation services:** Offer content creation services to individuals and businesses. Use the Autonomy QA system to create articles, blog posts, social media content, product descriptions and other written content. Charge your customers for each piece of content created or offer subscription plans for ongoing content needs(Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y., 2023).

**Customized AI solutions:** Develop customized AI solutions tailored to specific business needs with the Autonomy QA system. This could include the development of AI-driven virtual assistants, personalized recommendation systems, language translation services or sentiment analysis tools. Charge your customers for the development and implementation of these solutions(Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y., 2023).

**Productivity tools and plugins:** Develop productivity tools, plugins or extensions that leverage the Autonomy QA System to streamline workflows, automate tasks or improve communication. Offer these tools as paid products or as part

of a subscription-based service for individuals and businesses(Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y., 2023).

## Limitations

Although the Autonomy QA system and similar AI language systems have made considerable progress, they still have some limitations:

**Dependence ChatGPT:** The Autonomy QA system relies on a large corpus of text data for training. The quality, variety and size of this data have a significant impact on the performance and behavior of the model. The data should be representative of the language and contexts that the model is likely to encounter during inference(Wang, J., Liang, Y., Meng, F., Sun, Z., Shi, H., Li, Z.,... & Zhou, J., 2023).

**Software libraries and frameworks:** The autonomous QA system is usually implemented using Java and frameworks such as Spring or Spring-Boot. These frameworks provide the necessary tools and libraries for the creation, training and use of AI models(Wang, J., Liang, Y., Meng, F., Sun, Z., Shi, H., Li, Z.,... & Zhou, J., 2023).

**Infrastructure and deployment environment:** The Autonomy QA system requires a suitable infrastructure and deployment environment in order to work efficiently. This includes hardware infrastructure, cloud services, containerization platforms or server configurations(Wang, J., Liang, Y., Meng, F., Sun, Z., Shi, H., Li, Z.,... & Zhou, J., 2023).

**Maintenance and updates:** The Autonomy QA System requires to be regularly maintained, updated and improved to ensure it remains effective and up to date. This may include retraining the model with new data, fine-tuning parameters or incorporating new techniques from AI research(Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H.,... & Hu, X, 2023).

**APIs and external services:** The Autonomy QA System can make use of external APIs or services for additional functions, such as language translation, sentiment analysis or content retrieval(Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y., 2023).

## Conclusion

This paper provides a solution to complement some drawbacks of using Chat-GPT, a popular AI model for QA in today's world. The Autonomy QA system not only answers the questions but can also ask users five questions related to the user's first question. When the user selects one of these five questions, the autonomy QA system can answer the question and ask five more questions until the user stops. Finally, the autonomous QA system can provide recommendations to the user, such as blogs from Quora or Reddit, products from Amazon API or videos from YouTube. All this content can simplify the whole process for the user's question. Based on these solutions, OpenAI's Chat-GPT API cannot provide all the content itself. Therefore, Java and Spring-Boot are used in the Autonomy QA system to call the API of Quora, Amazon and YouTube. The user's behavior is stored in the Autonomy QA system. The recommendation system

behind the Autonomy QA System will use a cluster of MySQL database instances to support large amounts of data. This data is used to train new models on a regular basis. All in all, the Autonomy QA System can be seen as a complementary system to the current AI model system(de Oliveira, C. E., Turnquist, G. L., & Antonov, A., 2018).

## References


- bin Uzayr, S. (2022). *Mastering MySQL for Web: A Beginner's Guide*.
- Das, D., & Semaan, B. (2020). *Quoras:A Python API for Quora Data Collection to Increase Multi-Language Social Science Research*.
- de Oliveira, C. E., Turnquist, G. L., & Antonov, A. (2018). *Developing Java Applications with Spring and Spring Boot: Practical Spring and Spring Boot solutions for building effective applications*.
- Kalla, D., & Smith, N. (2023). *Study and analysis of chat GPT and its impact on different fields of study*
- Liu, Y. H. (2020). *Python Machine Learning By Example: Build intelligent systems using Python, TensorFlow 2, PyTorch, and scikit-learn*.
- Ray, P. P. (2023). *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*.
- Roumeliotis, K. I., & Tselikas, N. D. (2023). *Chatgpt and open-ai models: A preliminary review*.
- Saha, D., Brooker, P., Mair, M., & Reeves, S. (2023). *Thinking like a machine: Alan Turing, computation and the praxeological foundations of AI*.
- Wang, J., Liang, Y., Meng, F., Sun, Z., Shi, H., Li, Z.,... & Zhou, J. (2023). *Is chatgpt a good nlg evaluator? A preliminary study*.
- Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S.,... & Han, W. (2023). *Zero-shot information extraction via chatting with chatgpt*.
- Wittig, Andreas, and Michael Wittig. (2023). *Amazon Web Services in Action: An in-depth guide to AWS*
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). *A brief overview of ChatGPT: The history, status quo and potential future development*.
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H.,... & Hu, X. (2023). *Harnessing the power of llms in practice: A survey on chatgpt and beyond*.