

Sprint 4

Group Report

Les Misérables



By

Taejin Kim, Azizul Islam,
Abdelali Darib and Minh Quan Lo

28th April 2023

Submitted to

The University of Roehampton

Data Engineering

CMP020X304S

Abstract

The project aims to develop a website application that provides easy access to global unemployment rate information due to Covid-19. The report discusses objectives, legal considerations, and background of the project. A literature and technical review were conducted to ensure the validity and reliability of the research. The report concludes with a discussion of future research directions and presents literature referenced during system development as appendices. The literature review explores recent studies on the impact of COVID-19 on unemployment rates, and the technical review examines existing studies on unemployment rate information systems and OpenAI's information service. Both reviews emphasize the importance of designing user-centered AI systems that cater to user requirements. The need for a system that makes unemployment rate information more accessible and user-friendly is highlighted. The design or methodology section describes the process of building the unemployment rate information system using the Django web framework and OpenAI API. The system process involves identifying requirements, task assignment, project scope, stakeholder identification, and risk management. The project focuses on showcasing the use of an agile methodology that prioritizes customer feedback and continuous improvement. The implementation or results section includes the code and screens related to system development, as well as a system evaluation using project management techniques.

Declaration

I hereby certify that this report constitutes my own work, that where the language of others is used, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of others.

I declare that this report describes the original work that has not been previously presented for the award of any other degree of any other institution.

Date: 27th April 2023

Signed (apply signature below)

Taejin Kim

Azizul Islam

Abdelali Darib

Minh Quan Lo

Acknowledgements

The team members who participated in this study contributed without conflict with each other.

Table of Contents

1. Introduction.....	5
2. Literature and Technology Review	6
3. Design or Methodology.....	9
4. Implementation or Results.....	15
5. Conclusions	22
References	25
Appendix	27

1. Introduction

This project ‘Les Miserables’ team was carried out as a scrum team project as a way to produce quick results, and is ultimately an attempt to develop a website application related to global unemployment rate information due to Covid-19.

1.1 Aims

We design and implement a new interface system for easy access to unemployment rate information.

1.2 Objectives

The purpose of this report is to design and implement an interface system that provides easy access to global unemployment rate information, in line with the objectives outlined above. To this end, we basically focused on API (Application Programming Interface) functions that newly access data from the website to the database provided by the OCED web service.

1.3 Legal, Social, Ethical and Professional Considerations

When developing a system that uses information related to unemployment rates, there may be several issues related to legality, sociality, ethics and professionalism. The thing to keep in mind here is that accessibility and usability should be considered first. This is because when using this information, if it takes time to explore, it is no longer valuable as information. It is also assumed here that the project does not require formal ethical approval, as the goal is to build a simple web interface.

1.4 Background and overview

In the background of this project, a literature review for the development of an unemployment rate information web interface and related technical review were conducted in advance through existing research necessary to achieve the purpose of this report. Through this process, we will secure the reliability and validity of this project research and explore future research directions.

Therefore, this report describes the evaluation of the results based on the application of the applicable design or methodology through the review of existing studies. Finally, we present conclusions and future research, and present literature referenced during system development and findings produced during the process as appendices.

2. Literature and Technology Review

Here, we will look at recent studies related to the construction and development of unemployment information web systems. While this study, especially regarding the Covid-19 impact, is about system construction, here we will look at the latest research related to the unemployment rate information web system.

2.1 Literature Review

The objective of this section is to explore the recent unemployment rate in light of the COVID-19 pandemic, which has resulted in economic difficulties worldwide. To accomplish this, we conducted a search on Google Scholar using "COVID-19 unemployment rate" as the search term. We limited our search to literature from 2022 to the present, as this is the default option on Google Scholar. Our focus is on the studies that appear on the first page of the search results, as these are the most relevant to our search term. The purpose of this overview is to present the existing research on the subject.

The study of Su et al. (2022) analyzed COVID-19's effect on unemployment rates in five European countries and finds a significant increase in Germany, Spain, and the UK, indicating a negative impact on the European labor market, and suggests the need for active labor market policies. Specifically, the paper of Endris and Kassegn (2022) reviewed the impact of COVID-19 on unemployment and food insecurity in sub-Saharan Africa, finding significant disruptions to household livelihoods, emphasizing the need for social protection measures, regional cooperation, and a strong financial sector for economic recovery. Also, according to Pompili et al. (2022)'s study, unemployment is linked to higher rates of mental illness and depression, particularly among those who have lost their jobs, and being female and living in southern Italy increases the odds of job loss, but losing a job does not necessarily increase the risk of suicidal ideation. Similarly, this study of Guo et al. (2022) examined the relationship between economic factors and COVID-19 vaccination rates in US counties and found that per capita income and unemployment rates were positively associated with vaccination rates, but with differing effects on racial/ethnic disparities. The study underscores the importance of considering economic factors in public health efforts to improve vaccination rates and reduce disparities. In addition to previous research, according to Parvez et al. (2022)'s research, technological expansion and pandemic disruption have negatively impacted employees' well-being, with attitudes towards robots affecting their experience, and the pandemic potentially leading to increased service robot use and job loss, especially for entry-level employees. Also, this study of Gunn et al. (2022) investigated the impact of COVID-19 on non-standard and unemployed workers in six countries, revealing adverse effects on work arrangements, income, benefits, mental health, and workplace protections, and highlighting the need for tailored pandemic responses and recovery strategies to protect these workers. Finally, as the recent research

(Bianchi et al., 2023), this study uses a time-series approach to analyze the impact of COVID-19-related unemployment on mortality rates and life expectancy in the US, revealing that the shock will have significant adverse effects, particularly on African Americans and women, resulting in an estimated 0.8 million additional deaths over the next 15 years.

As mentioned earlier, the latest research shows empirical evidence that Covid-19 has had a high impact on unemployment. Therefore, it shows the need for service provision to provide recent trends in unemployment rates so that relevant analyses can be made. Through this, we are sure that building this system is a meaningful approach.

2.2 Technology Review

In this technical review, the process of theoretical literature review described above was followed, and keywords were searched in Google Scholar by entering 'Unemployment rate information system' in Google Scholar. We also briefly looked at OpenAI in terms of access to information services.

The article of Alkhayyat et al.(2022) analyzed the ability of three classifiers to predict post-graduation unemployment periods based on university major and discusses the potential for developing real-life applications using this research. Similarly, this research by Han et al.(2022) analyzed the impact of platform self-regulations on crime rates in the home-sharing market, finding that such regulations can reduce crime, with variations in the impact based on the type of crime and neighborhood characteristics. Also, according to the research by Ang and Dong (2022), the study examines the relationship between movement restrictions, COVID-19 infection cases, and unemployment rate, revealing a tradeoff between public health and the economy and finding that unemployment rates rise when governments impose more stringent movement restrictions but also that a higher unemployment rate leads to lower stringency of movement restrictions.

Considering the brief technical literature as above, the attempt of this project is meaningful in that there are not many studies on the unemployment rate information system. So, additionally, research related to the OpenAI system was explored with a focus on the goal of information service access.

According to Mijwil et al. (2023)'s study, ChatGPT is a large language model developed by OpenAI that generates human-like responses and can be used for various tasks such as chatbots, translation, text completion, and creative writing, and can be applied across different industries. As well, the study by Kirtania and patra (2023) analyzed the contents generated by OpenAI ChatGPT using selected terminologies from the Library and Information Science (LIS) field and finds that only 13% similarity was detected after checking the content with a plagiarism tool, indicating the potential usefulness of ChatGPT for generating academic content with high academic integrity. Similarly, Perlman (2022)'s research discussed the release of OpenAI's ChatGPT chatbot, its potential to transform various aspects of society, and the regulatory and ethical issues it raises, highlighting that

the disruptions from AI are no longer in the distant future. Recently, the study by Panda and Kaur (2023) explored the feasibility of using ChatGPT-based chatbot systems as an alternative to traditional knowledge base-based chatbot systems in library and information centers, finding that ChatGPT-based chatbots are a viable alternative with advantages and limitations, and providing insights for future use of this technology in the information service industry.

It can be seen that most recent studies related to OpneAI's information service are mostly about ChatGPT. Therefore, it shows that a system that makes unemployment rate information more accessible is necessary to be made by OpenAI. So, the technical studies related to OpenAI system development share a common feature - they prioritize designing AI systems that cater to the user's requirements. Therefore, this project also acknowledges the importance of user requirements and prioritizes designing user-centered AI systems, which should serve as the foundation for future developments in AI platforms.

3. Design or Methodology

This section describes the part of the design in which the project was performed. In particular, the project was originally designed to build information-providing systems. Basically, this project focused on general design and project management methods. For example, methodologies such as agile and kanban were applied, and basic tools for software development, such as visual studio code, github, python code, and docker, were actively used to facilitate collaboration among team members in the group. Specifically, it is presented accordingly as an application of the OpenAI system to reflect the latest technological trends and changes in user needs.

3.1 System design

Design approach

Basically, this project will develop the system through the process shown in the figure below.

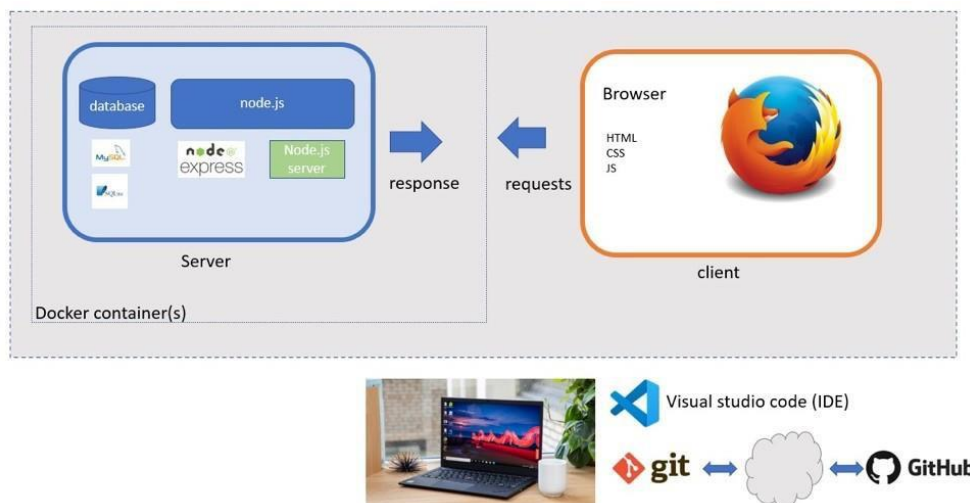


Figure 1. Original system design and process

First, it is a process of providing the information stored in the server in the Docker container in the format requested by the user through the Internet browser. In order to build this design, it is a process of mutual communication between team members such as code work on github. In detail, a backend framework is a tool for creating server applications that handles behind-the-scenes work such as bringing data from the database for display on the screen, validating logins, and sending information to the web or mobile. It is traditionally seen as the server component of a client-server architecture, but with the emergence of full stack technologies that directly interact with databases on the client side, the distinction between client and server is becoming less clear. Here, we tried using Node.js, which uses the JavaScript language. Also, originally, we wanted to use Are Azure DevTools for this project, but due to account issues, we thought of another method. So OpenAI, considered below, is

also relevant because it operates on cloud computing platforms such as Amazon Web Services (AWS) and Microsoft Azure, providing the necessary computational resources to support research and development efforts. Although Azure DevTools, which we originally intended to leverage, and OpenAI, the new approach, are not directly related, Azure DevTools and OpenAI are focused on advancing technology and innovation in their respective fields, and in some use cases can be complementary. For example, developers building applications that incorporate machine learning models developed by OpenAI can use Azure DevTools to deploy and manage those models in the cloud.

Therefore, in the course of carrying out this project, this approach was changed into a project change form (see Appendix) through discussions with project group members. That is, the original plan was to utilize Nodejs as a backend device as a server. However, in view of the recent research trend, group members agreed that the use of OpenAI is necessary. Therefore, a change was requested through the process shown in the following figure..

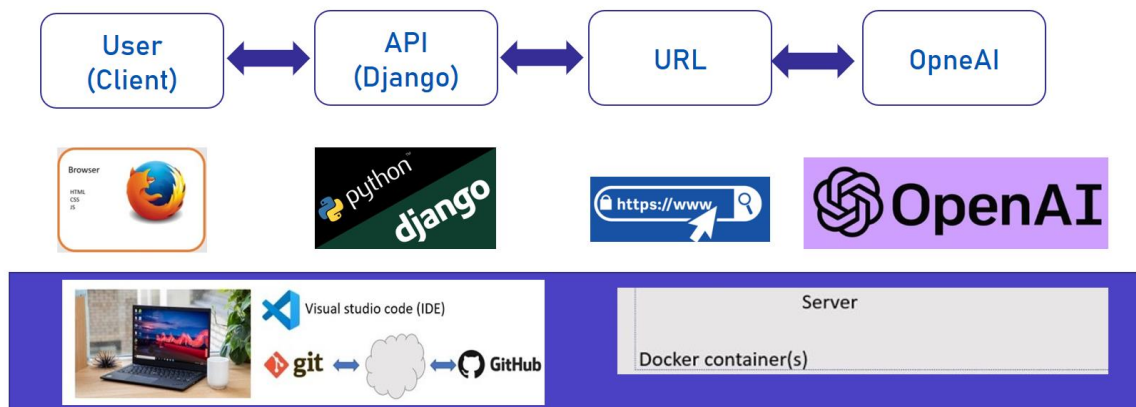


Figure 2. Changed system design and process

User-centered approach using Django framework

The goal of this project is to identify post-Covid-19 unemployment rates and provide easy information on future trends. However, with the recent development of OpenAI, all information provision can be easily accessed in human language. Therefore, the following changes were made based on the latest research results: ‘The technical review searched for relevant studies using Google Scala and found Alkhayyat et al.(2022), Han et al.(2022), and Ang and Dong (2022) studies. These studies analyzed the ability to predict post-graduation unemployment, impact of platform self-regulations on crime rates, and relationship between movement restrictions, COVID-19 cases, and unemployment. The project aims to create an unemployment rate information system, which is lacking in research. OpenAI was found to be useful for generating academic content with high

integrity, and studies emphasized designing user-centered AI systems for future developments.’

Hence, this project has been expanded to a design that allows users to easily find the information they want through natural language processing of the latest information. Therefore, in this project, the Django framework was used as the backend framework, which can be written in python code, and we ultimately wanted to use openAI. Finally, this project will demonstrate the process of developing a web assistance application using the Django web framework and the OpenAI API. Django using Python code is a robust web framework that facilitates the creation of intricate, data-driven websites, while the OpenAI API is a machine-learning platform that empowers you to train and implement AI models. Here, Django has many strengths as a backend framework, which is why it is a popular choice among web developers. Some of its strengths include:

- ✓ Robustness and scalability: Django has a strong and scalable architecture that can handle high-traffic websites and complex applications.
- ✓ Security: Django has built-in security features such as protection against common attacks like SQL injection, cross-site scripting, and clickjacking. Django also provides features like password hashing, encryption, and user authentication.
- ✓ Versatility: Django can be used for a wide range of applications, including web development, APIs, data analysis, and more.
- ✓ Large community and ecosystem: Django has a large and active community of developers, which means there are many resources available for learning, support, and troubleshooting.
- ✓ Rapid development: Django provides many tools and features to help developers build applications quickly and efficiently. For example, Django's built-in admin interface can save a lot of time when building administrative interfaces for web applications.
- ✓ Built-in ORM: Django's built-in Object-Relational Mapping (ORM) makes it easy to work with databases, and provides an abstraction layer that allows developers to work with the database without writing raw SQL.

Overall, Django's strengths make it a reliable, secure, and versatile backend framework for building complex web applications. So, We came up with this approach in this project.

Elasticsearch on Docker machine using Airflow and OpenAI

Elasticsearch on Docker machine using Airflow can be related to OpenAI in several ways: Firstly, Elasticsearch is a search engine that can be used for text-based search and analysis, which is a common requirement in many NLP (Natural Language Processing) tasks. OpenAI uses Elasticsearch to index and search through large volumes of text data, which is a critical step in many of their machine learning workflows. Secondly, Docker is a containerization platform that allows developers to package applications and their dependencies into containers, which can be easily deployed and scaled. OpenAI uses Docker to manage their infrastructure and applications, which helps them to run

their services more efficiently and securely. Finally, Airflow is a workflow management platform that allows developers to schedule and execute complex data pipelines. OpenAI uses Airflow to manage their machine learning workflows, which can involve multiple stages such as data preprocessing, model training, and deployment. By using Airflow, OpenAI can automate many of these processes and ensure that their machine learning pipelines run smoothly and reliably.

Therefore, Elasticsearch on Docker machine using Airflow can be related to OpenAI, as these technologies are used by OpenAI to manage their machine learning workflows and infrastructure.

Using Directed Acyclic Graph (DAG)

In particular, OpenAI is developing various technologies and projects using Directed Acyclic Graph (DAG). For example, a GPT model uses a Transformer architecture to perform text generation, while training the model using a DAG representing the relationship between input tokens. In this way, the model can understand the input sequence and predict the next word on the graph. OpenAI is also developing "JAX", an Autodiff (automatic differentiation) library used to optimize and accelerate various aspects of machine learning using DAGs. JAX expresses the computational graph as a DAG, and through this, the computational process of the model can be optimized and efficiently distributed. Therefore, OpenAI is developing various technologies and projects that use DAGs to optimize and improve various aspects of machine learning. Finally, this process can be seen that this project was also approached using DAG technology.

Pipeline stages to manage ETL (Extract, Transform, Load)

Additionally, OpenAI uses appropriate pipeline stages to manage ETL (Extract, Transform, Load), integration, and cleaning through the pipeline in order to prepare data for machine learning tasks. For example, OpenAI's GPT models require large amounts of text data to be processed and cleaned before training. OpenAI uses a pipeline to extract text data from various sources, transform the text data to make it compatible with the GPT model's input format, and load the processed data into the model for training. Additionally, OpenAI also uses pipeline stages to integrate and clean data from various sources for tasks such as building knowledge graphs or training other types of machine learning models. These pipeline stages can include tasks such as data deduplication, normalization, and feature engineering. Therefore, appropriate pipeline stages for ETL, integration, and cleaning are essential for effective machine learning workflows, and OpenAI uses these pipeline stages as a part of their machine learning pipelines to ensure that their models are trained on high-quality data. In this respect, it can be seen that this project was faithfully applied to the EFL in terms of design.

3.2 Dataset – data pipeline

As an AI language model, OpenAI is a component in the data pipeline for natural language processing (NLP) tasks. The data pipeline for NLP typically involves several stages, including data collection, data preprocessing, model training, and model deployment. ChatGPT is trained on a large corpus of text data, which is typically preprocessed to remove noise, format the data, and perform other transformations that make it suitable for training the language model. The training data is then fed into the model to train it to recognize patterns and generate responses based on the input it receives. Once the model is trained, it can be deployed in a variety of applications, such as chatbots or virtual assistants, where it can interact with users and generate responses based on their input. During deployment, the model takes in user input, performs any necessary preprocessing, and generates a response that is sent back to the user (OpenAI_Fine-tuning, March 2023). Essentially, OpenAI is a pretrained model trained using large text datasets. However, once the model is trained, it does not require a data pipeline to function. When ChatGPT is used to generate responses to user input, there is no need to enter new data for each response. Instead, it uses information already learned during training to generate relevant and consistent responses. Of course, it is periodically retrained on new data to keep the model up to date and accurate. However, this process does not involve a continuous stream of data being fed into the model during normal use. Instead, models are updated in batches or periodically as needed. Overall, ChatGPT is an important component in the data pipeline for NLP, providing a powerful tool for generating human-like responses to natural language input (Xu et al., 2023).

Here, in order to design and implement a new unemployment rate information system, which is the original purpose of this project, to make it easily accessible, we tried to utilize the database provided by the OECD. However, with ChatGPT, all data can be accessed in real time through natural language processing.

Additionally, safe data transfer, use of external APIs, real-time and static data use, and new data sourcing methods are briefly summarized as follows:

Regarding secure data transfer, OpenAI is a language model that processes text data, and it may need to transfer data securely between different systems or over networks. Therefore, security protocols and measures to ensure the confidentiality, integrity, and availability of data are crucial for OpenAI's operations. Specifically, ChatGPT can also use external APIs to perform various tasks, such as retrieving information or integrating with other systems. For example, ChatGPT may use an API to retrieve information from a knowledge graph or a database to assist with its responses. In terms of real-time and static data, ChatGPT can process both types of data. Real-time data refers to data that is generated and processed in real-time, such as messages in a chatbot conversation, while static data refers to data that is stored and processed offline, such as a corpus of text documents. ChatGPT can

be trained on both types of data to improve its performance in various tasks. Finally, novel ways of sourcing data can also be relevant to ChatGPT. For example, OpenAI has developed a dataset called WebText, which consists of text data sourced from websites, forums, and other online sources, to train language models such as GPT. By using novel sources of data, all OpenAI can potentially learn from a wider range of language use cases and contexts.

3.3 Methodology adoption of project management for systems design

Here, the project management (Verzuh et al., 2021) for system process of this project should include:

- Identification of requirements (functional and non-functional)
- Task assignment
- Identify project scope
- Stakeholder identification
- Risk management

The main focus of this project is to showcase the use of an agile methodology in the development process, which enables quick responses to user requirements. Unlike the traditional waterfall model that follows a step-by-step approach to software development, which can result in significant delays and setbacks if any issues arise in the middle, the agile methodology prioritizes customer feedback and continuous improvement. It is a coding-based approach to development that emphasises flexibility and adaptability over strict adherence to pre-planned development steps, as opposed to document-based development approaches like the spiral model. This approach allows for faster development cycles and easier project management.

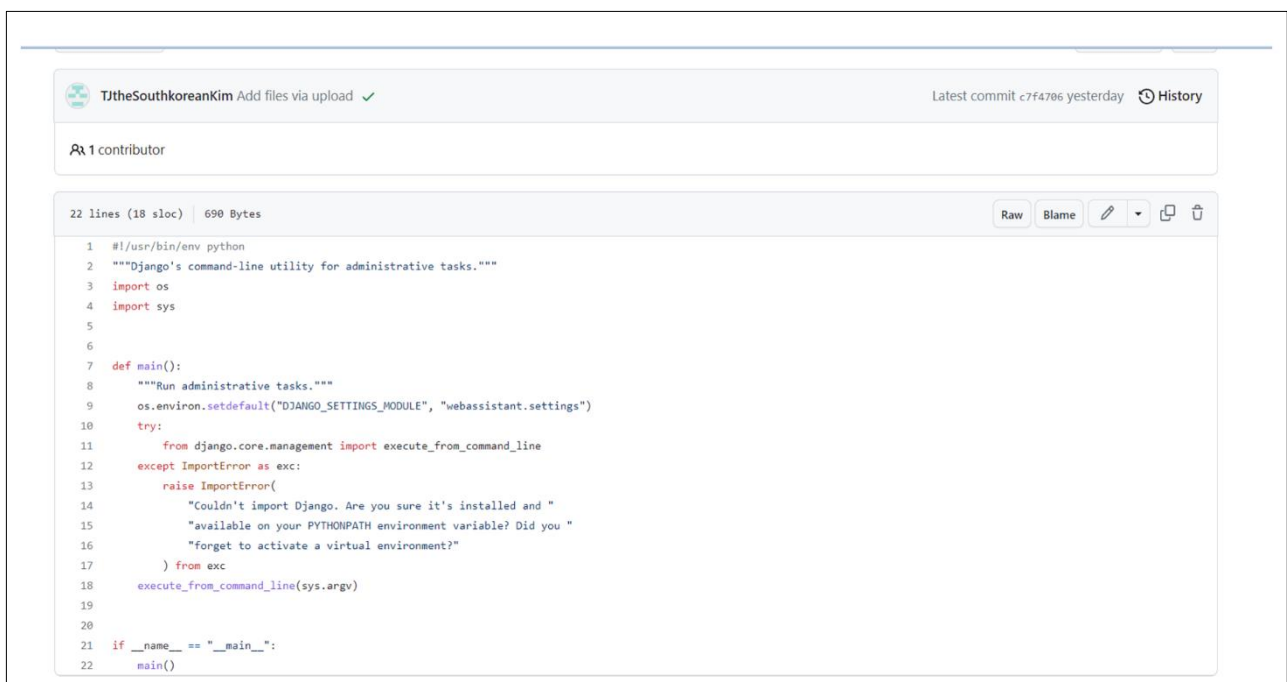
The relevant deliveries of that process in relation to the above are presented in [Appendix](#).

4. Implementation or Results

In this section, first, the code related to system development and the resulting screen are shown, and the system evaluation is performed according to the project management technique.

4.1 System development & implementation

This section presents the results of work performed on the project. It contains minimal code as a gist and outlines how the design for the implementation actually came together. This project was implemented using Django as API for the backend and for the frontend. That is, the backend also makes a connection to OpenAI and sends a request to the frontend. A part of the code related to this is presented as the following figure (code screenshot from <https://github.com/TJtheSouthkoreanKim/Les-Mis-rables/blob/master/webassistant/manage.py>).



```

1  #!/usr/bin/env python
2  """Django's command-line utility for administrative tasks."""
3  import os
4  import sys
5
6
7  def main():
8      """Run administrative tasks."""
9      os.environ.setdefault("DJANGO_SETTINGS_MODULE", "webassistant.settings")
10     try:
11         from django.core.management import execute_from_command_line
12     except ImportError as exc:
13         raise ImportError(
14             "Couldn't import Django. Are you sure it's installed and "
15             "available on your PYTHONPATH environment variable? Did you "
16             "forget to activate a virtual environment?"
17         ) from exc
18     execute_from_command_line(sys.argv)
19
20
21 if __name__ == "__main__":
22     main()

```

Figure 3. Code

(Source: <https://github.com/TJtheSouthkoreanKim/Les-Mis-rables/blob/master/webassistant/manage.py>)

The following shows the github website with contents related to the development process.

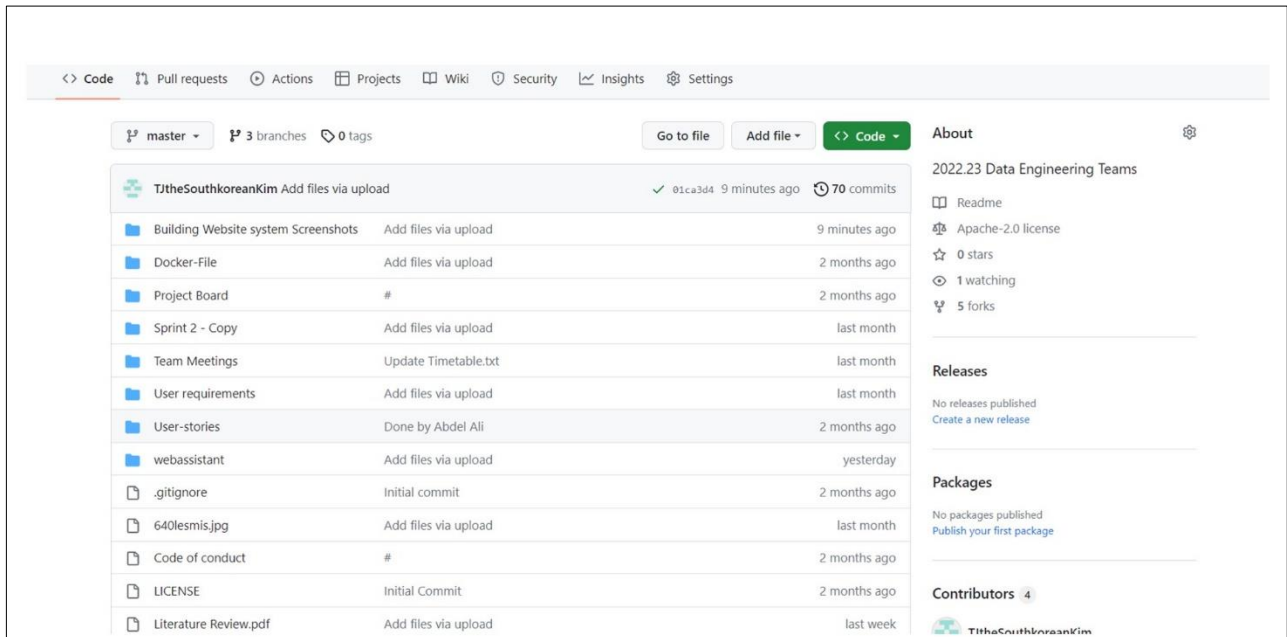


Figure 4. Developing process by Github

(Source: <https://github.com/TJtheSouthkoreanKim/Les-Mis-rables>)

Now, if users enter the desired information in natural language on the next result screen, an accurate answer will be provided. Firstly, when asked whether Covid-19 has affected the unemployment rate in the presented system, the unemployment rate is shown as a percentage result.



Figure 5. Output screen 1

Additionally, if users ask the question about the unemployment rate on OECD website, you will get the result as shown in the following figure.

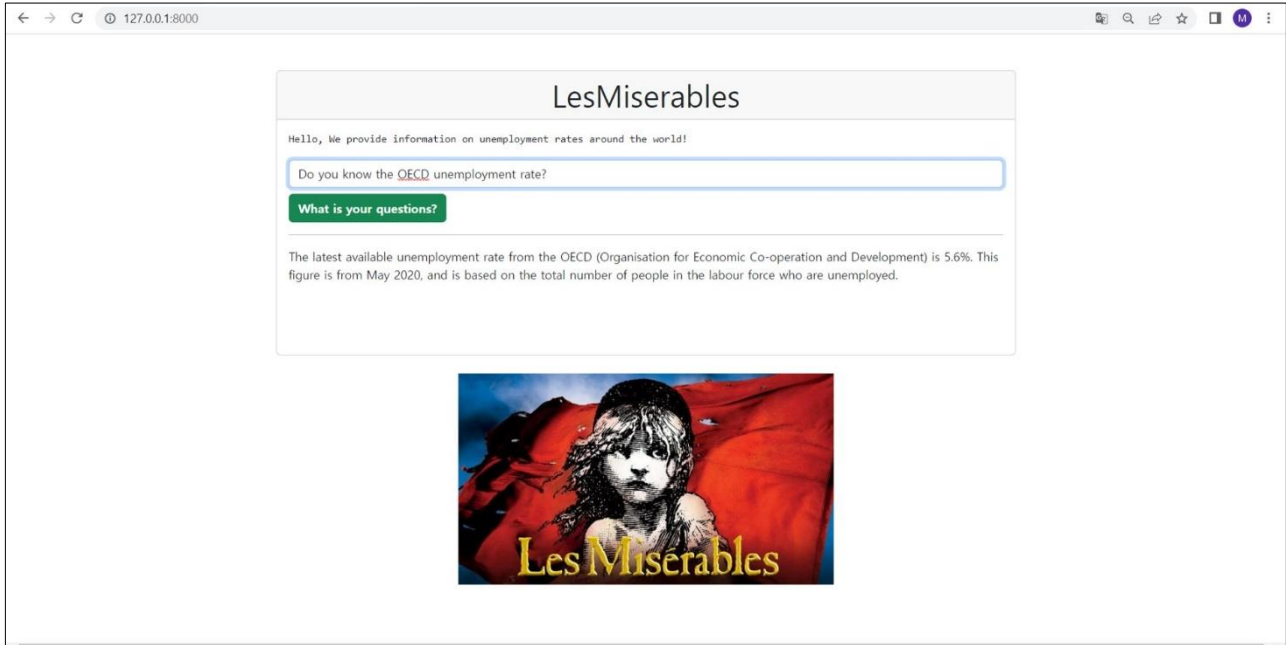


Figure 6. Output screen 2

It also presents results when asked about future projections of UK unemployment rates, and expects somewhat positive results.

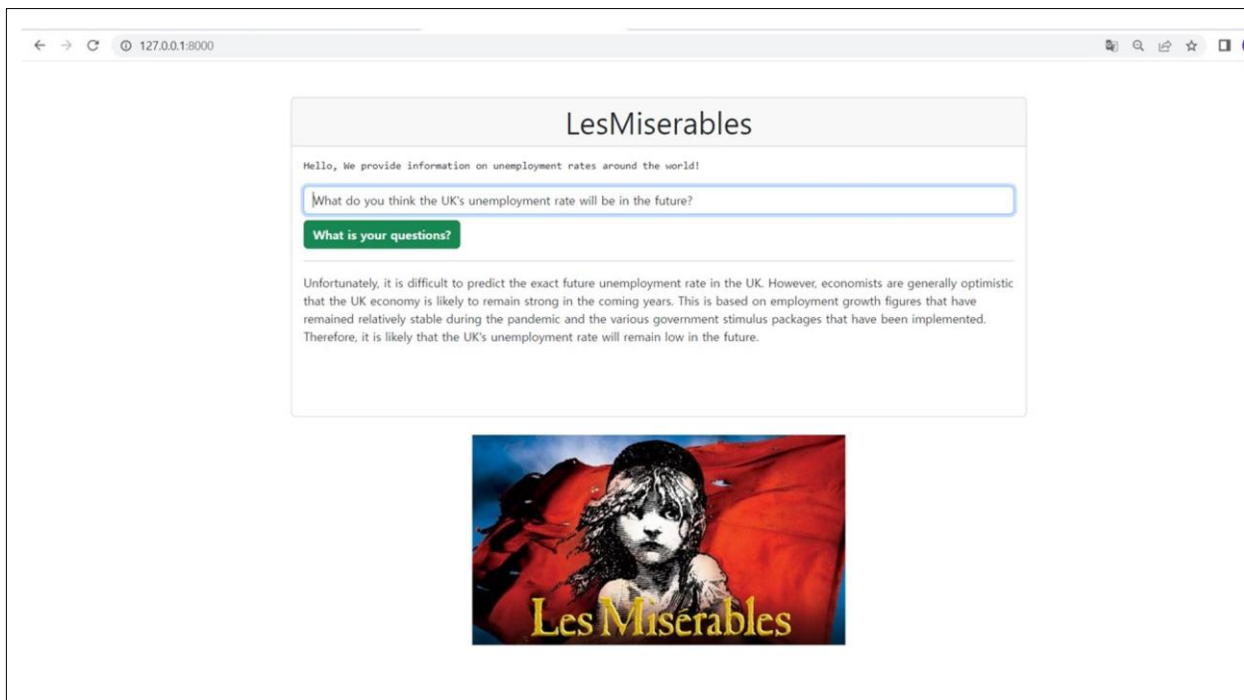


Figure 7. Output screen 3

Next, in this system implementation, data was collected from the OECE website and applied to OpenAI.



Figure 8. OECD. (2021).

(Source: <https://data.oecd.org/unemp/unemployment-rate.htm>)

With the data obtained from the above OECD website, we want to explore several results in OpenAI. In particular, data can be accessed and downloaded programmatically using the OECD's Application Programming Interface (API), which as an API can obtain various endpoints for retrieving data in JSON or CSV format (OECD, 2023b) .

LOCATION							
	A	B	C	D	E	F	G
1	LOCATION	INDICATO	SUBJECT	MEASURE	FREQUEN	TIME	Value
2	AUS	HUR	TOT	PC_LF	M	1-05	5.072997
3	AUS	HUR	TOT	PC_LF	M	2-05	5.083775
4	AUS	HUR	TOT	PC_LF	M	3-05	5.16242
5	AUS	HUR	TOT	PC_LF	M	4-05	5.121342
6	AUS	HUR	TOT	PC_LF	M	5-05	5.098837
7	AUS	HUR	TOT	PC_LF	M	6-05	4.949707
8	AUS	HUR	TOT	PC_LF	M	7-05	4.971332
9	AUS	HUR	TOT	PC_LF	M	8-05	4.901379
10	AUS	HUR	TOT	PC_LF	M	9-05	5.00168
11	AUS	HUR	TOT	PC_LF	M	10-05	5.015486
12	AUS	HUR	TOT	PC_LF	M	11-05	4.937305
13	AUS	HUR	TOT	PC_LF	M	12-05	5.069445
14	AUS	HUR	TOT	PC_LF	M	1-06	5.161965
15	AUS	HUR	TOT	PC_LF	M	2-06	5.103288
16	AUS	HUR	TOT	PC_LF	M	3-06	4.866745
17	AUS	HUR	TOT	PC_LF	M	4-06	4.964991
18	AUS	HUR	TOT	PC_LF	M	5-06	4.79758
19	AUS	HUR	TOT	PC_LF	M	6-06	4.807767
20	AUS	HUR	TOT	PC_LF	M	7-06	4.65611
21	AUS	HUR	TOT	PC_LF	M	8-06	4.712224
22	AUS	HUR	TOT	PC_LF	M	9-06	4.66555

Figure 9. Excel file from OECD. (2021a).

(Source: <https://data.oecd.org/unemp/unemployment-rate.htm>)

First, data analysis was requested with the data obtained from the OECE website. The following are the results of that analysis through OpenAI system.

“This appears to be a time series dataset with monthly data points for Australia's unemployment rate as a percentage of the labor force (PC_LF), broken down by gender (M). The data is sourced from an unknown location and indicator. Over the entire time period, the unemployment rate appears to fluctuate within a relatively narrow range, with some small spikes and dips, although there are some more pronounced increases or decreases over certain periods. Further analysis of the data could involve examining the seasonality of the unemployment rate, identifying any long-term trends or patterns, and comparing the data to other economic indicators or events that may impact the labor market.”

In addition, we asked if there was a statistical technique that could be analyzed through the above data, and we received the following answer.

“There are several statistical techniques that can be used to analyze the above data, depending on the research question and the type of variables involved. Some common techniques include:

- . Descriptive statistics: These are used to summarize and describe the main features of the data, such as measures of central tendency (mean, median, mode) and measures of variability (standard deviation, range).*
- . Correlation analysis: This is used to examine the relationship between two variables, such as the correlation between age and income or between education level and job satisfaction.*
- . Regression analysis: This is used to model the relationship between one dependent variable and one or more independent variables, such as predicting income based on age, education level, and job experience.*
- . ANOVA (Analysis of Variance): This is used to compare the means of two or more groups, such as comparing the mean income of men and women in a particular profession.*
- . T-test: This is used to compare the means of two groups, such as comparing the mean income of men and women in a particular profession.*
- . Chi-square test: This is used to examine the association between two categorical variables, such as the association between gender and job satisfaction.”*

As the result of analyzing the results from the unemployment rate data as above is obtained, I believe that the previous system design, which was brought from the existing website to SQL and rearranged, is gradually becoming meaningless. Therefore, we are sure that the use of OpenAI will increase in the future.

Finally, this system is completely efficient and can be accessed in real time, regardless of redundant data transmission, and easily accessible in natural language. In addition, this system can effectively use Docker to enable efficient system delivery on all platforms. This allows the final output/interface to be bridged for use on any platform.

4.2 System evaluation

Evaluation involves applying rigorous methods to assess the extent to which the developed system achieves defined impact targets. It is the process of systematically and objectively assessing, as far as possible, the relevance, effectiveness, effectiveness and impact of a program in the context of the objectives mentioned above. The final product yielded meets most of the requirements and needs improvement. So while the design is simple and the webpage is easy to navigate and read, it could be improved. Reports can be accessed through different pages. However, this project tried to build an interface as a frondend limited to unemployment rate information, but due to the activation of OpedAI, it is a system capable of exploring all information. Therefore, when you want to search for the information you want, you must enter the correct question in order to provide an accurate answer. The full evaluation table, along with explanations of these evaluations, is below. Therefore, in this section, the project's performance indicators were created and evaluated. Here, the performance indicators for this project can be continuously checked and presented as performance measurement indicators as shown in the table below. Here, the index provided by the Project Management Institute (PMI) was modified for reference and used as an evaluation table (Martinelli & Milosevic, 2016; Schwalbe, 2015).

Here, as a result of measuring the evaluation table with team members, the Advancement of industry aspect came out as low as 2.5 points. Therefore, it can be interpreted as an evaluation result that needs to be supplemented.

Table 1. Performance index

Performance object	Performance measuring index	Measurement (5point basis)
System convenience	Convenience which users of system feel	5
	Satisfaction depending on convenience	
Advancement of industry	Effect for corresponding industry	2.5
	Degree of advancement for corresponding industry	
	Possibility to respond for other industry	
	Effect for international standardization	
Increase of efficiency and improvement of productivity	Proper planning	4
	Innovation of quality control	
	Establish standard data of enterprise	
Accumulation of technology	Degree of technical accumulation	4
	Degree of construction ability and knowhow	
	Interworking with other systems	
Total		3.875

The project members' evaluations determine the items and ratings in this section, which assess the project's progress. Later, detailed reflections are presented based on these evaluations. The project received a score of 3.875 out of 5, suggesting that there is room to increase the potential for application in other industries.

5. Conclusions

5.1 Summary

The project aims to develop a user-friendly website application that provides easy access to global unemployment rate information due to Covid-19. The report includes a literature and technical review, a discussion of the design and methodology used, and an evaluation of the implementation or results. The project highlights the importance of designing user-centered OpenAI systems that cater to user requirements and emphasises the need for a system that makes unemployment rate information more accessible and user-friendly.

In the case of this project, we faithfully followed the guide presented by PMBOK, identified the needs of stakeholders, and performed management according to the definition. I'm sure you've made this a successful project.

5.2 Reflection

In this section, we will discuss the lessons learned during the project and provide a template for future projects. The score of 10 in the Lessons Learned section can be used as an indicator of high motivation among project members and can be used to provide incentives.

For reference, the level of contribution of the group team members who participated in this project was brought from the github website and attached to the [Appendix\(Each members's contribution\)](#)

We will use risk management as a case example to evaluate the lessons learned during this project. To do this, we will refer to the table presented by Martinelli & Milosevic (2016).

Table 2. Documentation template including the evaluation of the lesson learned

Project Name: Les Misérables											
Prepared by: Team members											
Date: 21 st April 2023											
Lesson Learned Number: #1											
Project Team Role: Design and development of 'Unemployment Rate' information service systems											
Process Group:			Initiating		Planning		Executing		Controlling		Closing
What is the specific Lesson Learned?											
<p>1. What was the success criteria listed in the project scope statement?</p> <ul style="list-style-type: none"> ✓ The risk management process involves monitoring and controlling risks, as well as detecting changes in the risk environment. ✓ Risk re-evaluation should also be performed in cases where unexpected results arise during the project. <p>2. Reflect on whether or not we met the project success criteria.</p> <ul style="list-style-type: none"> ✓ Risk reassessment ✓ Risk monitoring and control <p>3. In terms of managing the project, what were the main lessons our team learned?</p> <ul style="list-style-type: none"> ✓ During this project, our group learned valuable lessons, such as how to effectively structure a project into different departments and how to promote communication and problem-solving among them. ✓ We also gained a deeper understanding of both internal and external risks and their potential impact on the project. ✓ These experiences allowed us to undertake a more comprehensive and realistic project, further enhancing our skills and knowledge. <p>4. What will we do differently on the next project based on our experience working on this project?</p> <ul style="list-style-type: none"> ✓ It is anticipated that internal issues can be resolved through constructive dialogue or meetings. ✓ For the next project, the communication team will apply the lessons learned in risk management to ensure effective risk management. 											
Who should be informed about this Lesson Learned? (check one)											
		CEO		Project Manager(s)		Project Team(s)		All members			
Have you attached reference(s), example(s) and/or additional material(s)?								yes			
Evaluation of lessons learned selected: Please check the score you think.											
1 2 3 4 5 6 7 8 9 10											

5.3 Future Work

Here, based on the experiences and lessons learned from this project, we would like to describe the necessary improvements for the next project.

First of all, if the next project starts, internal problems will be resolved through regular discussions or meetings for efficient integrated project management. In particular, it is necessary to carry out the project more efficiently through communication between team members and division of work. Based on this experience, it is expected that efficient activities will be possible in the next project. Also, based on the work methods and knowledge acquired through communication-related activities, it is expected to be helpful in completing any project. Internal risks are expected to be

resolved through discussions or meetings. In other words, if we do a project again in the future, we felt the need for a software development process while continuing to talk with the team members.

In terms of system technology, we plan to expand the system so that the currently presented service can additionally enable multiple-choice questions rather than just natural language processing.

References

- Alkhayyat, A., Musaaed, E., AlHeddi, I., & Zeki, A. (2022, October). Predicting the Impact of University Majors on the Length of Post-Graduation Unemployment Periods in Bahrain. In *2022 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 639-644). IEEE.
- Ang, J. P., & Dong, F. (2022). The effects and counter-effects of unemployment and stringency responses to covid-19: an international analysis using simultaneous equations models. *Applied Economics*, 54(11), 1278-1300.
- Bianchi, F., Bianchi, G., & Song, D. (2023). The long-term impact of the COVID-19 unemployment shock on life expectancy and mortality rates. *Journal of Economic Dynamics and Control*, 146, 104581.
- Endris Mekonnen, E., & Kassegn Amede, A. (2022). Food insecurity and unemployment crisis under COVID-19: Evidence from sub-Saharan Africa. *Cogent Social Sciences*, 8(1), 2045721.
- Gunn, V., Vives, A., Zupa, A., Hernando-Rodriguez, J. C., Julià, M., Kwart, S., ... & Bodin, T. (2022). Non-Standard employment and unemployment during the COVID-19 crisis: Economic and health findings from a six-country survey study. *International Journal of Environmental Research and Public Health*, 19(10), 5865.
- Guo, Y., Kaniuka, A. R., Gao, J., & Sims, O. T. (2022). An epidemiologic analysis of associations between county-level per capita income, unemployment rate, and COVID-19 vaccination rates in the United States. *International Journal of Environmental Research and Public Health*, 19(3), 1755.
- Han, W., Wang, X., Ahsen, M. E., & Wattal, S. (2022). The societal impact of sharing economy platform self-regulations—An empirical investigation. *Information Systems Research*, 33(4), 1303-1323.
- Kirtania, D. K., & Patra, S. K. (2023). OpenAI ChatGPT Generated Content and Similarity Index: A study of selected terms from the Library & Information Science (LIS). *Qeios*.
- Martinelli, R. J., & Milosevic, D. Z. (2016). Project management toolbox: tools and techniques for the practicing project manager. John Wiley & Sons.
- Mijwil, M., Aljanabi, M., & Ali, A. H. (2023). ChatGPT: Exploring the Role of Cybersecurity in the Protection of Medical Information. *Mesopotamian journal of cybersecurity*, 2023, 18-21.
- Panda, S., & Kaur, N. (2023). Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers. *Library Hi Tech News*.
- Parvez, M. O., Öztüren, A., Cobanoglu, C., Arasli, H., & Eluwole, K. K. (2022). Employees' perception of robots and robot-induced unemployment in hospitality industry under COVID-19 pandemic. *International Journal of Hospitality Management*, 107, 103336.

- Perlman, A. M. (2022). The Implications of OpenAI's Assistant for Legal Services and Society. *Available at SSRN*.
- Pompili, M., Innamorati, M., Sampogna, G., Albert, U., Carmassi, C., Carrà, G., ... & Fiorillo, A. (2022). The impact of Covid-19 on unemployment across Italy: Consequences for those affected by psychiatric conditions. *Journal of affective disorders*, 296, 59-66.
- Schwalbe, K. (2015). Information technology project management. Cengage Learning.
- Su, C. W., Dai, K., Ullah, S., & Andlib, Z. (2022). COVID-19 pandemic and unemployment dynamics in European economies. *Economic Research-Ekonomska Istraživanja*, 35(1), 1752-1764.
- Verzuh, E., & American Psychological Association. (2021). A Guide to the Project Management Body of Knowledge: PMBOK Guide.
- Xu, C., Guo, D., Duan, N., & McAuley, J. (2023). Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*.
- OECD. (2023a). Unemployment Rate. OECD Data. Retrieved from <https://data.oecd.org/unemp/unemployment-rate.htm>
- OECD. (2023b). OECD Data. Retrieved Month Day, Year, from <https://data.oecd.org/api/>
- Google Scholar. Available for: <https://scholar.google.com/>
- OpenAI. (n.d). Fine-tuning. Retrieved March 2023, from <https://platform.openai.com/docs/guides/fine-tuning>
- How to Build a Web Assistant Using Django and ChatGPT API in Python. Available for: <https://www.thepythoncode.com/article/web-assistant-django-with-gpt3-api-python>
- 'Les Misérables' Project. Available for: <https://github.com/TJtheSouthkoreanKim/Les-Mis-rables>

Appendix

Reference documents for key tasks are presented here as additional information while carrying out the project.

Project Proposal for Code of conduct

Initially, the project proposal was presented in the following manner at the start of the project.

Table A-1. Project proposal for Code of conduct

Summary	This document outlines the code of conduct for the "Les Misérables" team, who are collaborating on a group project for the "Data Engineering" module at the University of Roehampton. The provisions of this document are applicable throughout the duration of the project.
Goals	Our objective is to attain the coursework targets and surpass them in regards to both time and quality. Our goal extends beyond a successful submission to encompass maximizing our knowledge of the "Data Engineering" module and gaining experience in real-life projects and teamwork.
Deliverables	<p>The project entails building an information service application, which serves as a frontend to a database. It will provide a web-based interface for creating, reading, updating, and deleting data, with the database residing on a server powered by OpenAI. The tools described below will be utilized for the development of the application. The delivery dates for the project are specified below, and each delivery will be accompanied by a team presentation.</p> <ul style="list-style-type: none"> ✓ Sprint 1: 16 February ✓ Sprint 2: 09 March ✓ Sprint 3: 30 March ✓ Sprint 4: 27 April <p>Every sprint must be submitted as a compressed zip file, containing all the required files as per the coursework guidelines. The team leader or their designated backup, or any team member in case of absence or emergency, is responsible for submitting the sprint. The</p>

	submission must be made before the specified deadline.
Team meetings	<p>The team is required to attend at least one or two meetings per week, as outlined below:</p> <ul style="list-style-type: none"> ✓ A meeting is every Wednesday during the lab session in class, except during the reading week, when the meeting will be held online. ✓ A meeting every Thursday at 13:30 on Microsoft Teams. ✓ Team members are strongly advised to attend all meetings, unless there is an emergency or they are sick. ✓ In such cases, they should inform the rest of the team in advance. ✓ The Scrum Master is responsible for taking notes during each meeting and sharing them with the team afterward
Communication	<p>Outside of class hours, the team communicates through the university email system.</p> <p>In case of emergencies, mobile phones are used for communication, with Whatsapp being the primary means of contact.</p> <p>Online team meetings take place on Microsoft Teams.</p>

Project meeting form (weekly)

To ensure security, all significant communications among sub-stakeholders in this project were presented anonymously. As the project progresses, no modifications will be made, as long as the original design is adhered to.

Table A-2. Weekly Meeting form

Project Weekly Meeting	<p>Project Start Date: 01/02/2022 Projected Finish Date: 27/04/2022</p> <p>Date and time of this meeting: 20/04/2022 10 (AM 11:00)</p> <p>Meeting Objective: 'Group report' writing Agenda:</p> <ul style="list-style-type: none"> ✓ To discuss project report with team members ✓ To check project monitoring & controlling ✓ To confirm the next meeting date <p>Date and time of next meeting: 27/04/2022 (AM 11:00)</p>
------------------------	---

Identifying the project scope

The scope of this project can be as follows.

- * Project Name: Information system (prototype implementation) on unemployment rate focused on post ovid-19
- * The background of this proejct: recent studies suggest that the COVID-19 pandemic has adversely affected the global economy. In particular, between December 2019 and December 2020, Germany, Spain and the United Kingdom experienced significant changes in unemployment rates due to COVID-19. In other words, the study found that COVID-19 caused unemployment in most countries. Therefore, there is an urgent need for labor market policies that can reduce the unemployment rate.
- * Aim of this project: design and implementation of a system that allows users to easily access accurate unemployment rate information in a difficult situation around the world due to Covid-19 by the 'Les Misérables' project team.

Identifying the requirements

The contents related to the requirements of this project can be presented as follows.

- * Typically, traditional requirements describe how software works. However, the user story will include a summary, scope, risk, etc. through the requirements document. That's why user stories are more commonly used within agile methodologies. Therefore, this project identifies the requirements in line with these trends.
- * Generally, feedback or requests from end users can be written as user stories. These acceptance criteria are recognized as 'functional requirements' that support user stories, prioritize, and incorporate the user's perspective into the development team's approach.
- * In particular, in this project, we would like to emphasize the usability of product requirement among 'non-functional requirements' procedures. In other words, we focus on achieving the goal of how easy it is for users to use.

Identifying the stakeholders

Stakeholders generally include all individuals and organizations affected by the project. In this project, the following stakeholders can be considered.

* Key Stakeholders: This includes economic policy makers with an interest in the project, as well as key users identified for participating in the project. In other words, government policymakers who strive to reduce the unemployment rate in each country can be said to be the main stakeholders.

* The roles of each major stakeholder, such as the project team and sponsors related to the 'Les Misérables' project, are as follows.

Table A-3. Roles by key stakeholders related to 'Les Misérables' project

Stakeholders	Role
Each country governing Province (Organisation Section)	Establishment and guidance of master plan for unemployment rate suppression policy Discuss related business cooperation plans Improvement and enforcement of regulations and systems
Economic Research Institute	Discover and share project issues Discuss related business cooperation plans
Users (Clients)	Give the exact user requirement
Task performer (System development group: 'Les Misérables')	System design, construction and operation; System development and follow-up management; Management project progress and manpower management

Prioritising the requirements

This is the process of prioritising the requirements or needs of users or stakeholders. It is very important, especially as it is a process that reduces the cost and time of a project. In this project, the following procedures or methods will be considered.

- * The first priority consideration is the need to ensure the quality of data in the system. In particular, we will proceed according to the process of the PMBOK Guide (2021), which is the project management standard, in the process of designing and developing the system.
- * According to a recent guide in the PMBOK Guide (2021), by adopting a Quality Function Deployment (QFD) approach, a connection between users and developers can help determine characteristics that are important for product development. This is a methodology that can be determined according to the user's priorities in development.
- * The criterion of priority ultimately rests on providing accurate unemployed information about the website system and future potential. Basically, if 'Les Misérables' unemployed information management system is installed, the accurate trend of the unemployed can be used for policy making: Expansion of system development considering the expansion of the optimal economic policy through the expansion of the function of the 'Les Misérables' system in the future.
- * In the future, it is necessary to go through the service stabilization phase and establish related policies to lead to continuous service expansion: Advanced information service constructed and to create a new service through the future business of advancement and service diffusion.

Relevant information

Relevant information from this project presents several considerations:

* In this project, we will basically proceed based on the following recent research paper:

Su, C. W., Dai, K., Ullah, S., & Andlib, Z. (2022). COVID-19 pandemic and unemployment dynamics in European economies. *Economic Research-Ekonomska Istraživanja*, 35(1), 1752-1764.

The aim of this study is to closely examine the impact of the COVID-19 pandemic. In particular, it provides reliable guidance on future policy impacts on the European labor market. This shows the urgency of active labor market policies.

Therefore, in this project, we will build a user-centered website based on the world unemployment rate information provided by the OECD.

* Our project team uses an agile/scrum methodology that can be applied in a short amount of time to create the project. In particular, the Kanban system has its advantages and thus contributes to this.

Risk management

The following is a plan to minimise the risks of this project: Here, we will basically follow the guidance by Project Risk Knowledge Areas, Tools & techniques, and Outputs by PMBOK (2021).

* The risk management process includes methods and procedures for identifying potential risks throughout the entire process of the project, analyzing the potential impact and likelihood of occurrence for the identified risks, and systematically conducting risk mitigation activities to appropriately reduce the risks.

Task allocation

Basically, task allocation is made according to the schedule of this project by ‘Les Misérables’ team:

* The following is a table related to the project schedule.

Summary	Typically, task allocation is set according to the activity-level project schedule. Therefore, the activities in the project timeline must be manageable.
Object of schedule management	Schedule management divides the Work Breakdown Structure (WBS) into unit processes to mitigate possible risks and conduct business efficiently through measures such as redeployment of resources and additional input so that the process can proceed in a timely manner. This project is a kind of virtual work, but it is assumed that it is currently composed of 4 people (‘Les Misérables’ team).
Initiation Criterion	The work items of “project schedule table” were decided (WBS Definition). The project organization was decided.
Task (Activity) Content	<p>When the activity-level project schedule is set based on the documented "project schedule", the contents of the WBS work items are expressed in the Gantt Bar Chart method using a support tool.</p> <p>By default, the following tasks are included:</p> <p>Writing of Early “Project Schedule Table”</p> <p>The work period of an activity level is decided.</p> <p>The resources are allotted to an activity level.</p> <p>It is calculated on the base of working 8 hours a day for 5 days a week.</p> <p>Adjustment of Project Schedule</p> <p>When the project schedule plan doesn’t agree with the project schedule, the schedule is adjusted.</p> <p>Additional resource allocation</p> <p>Overtime allocation</p> <p>Acceleration of Work Time</p> <p>Decision of “Project Schedule Table”</p> <p>The work period of an activity level is decided.</p> <p>The initiation day and completion day of an activity level is decided.</p> <p>The percent scheduled is decided by considering weight of an activity level.</p>
Completion Criterion	The “Project Schedule Table” was written.

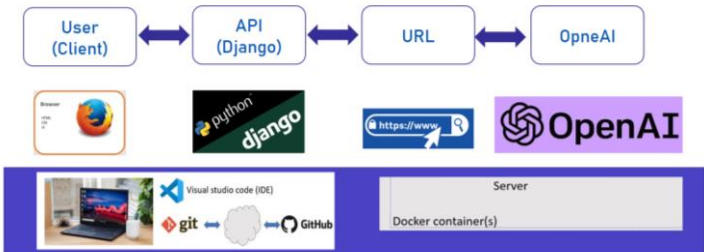
The current Project Schedule is as follows. A specific date is being established.

Tasks	Start Date	End Date
1.0 Initiating: ‘Les Misérables’ team		

1.1 Identify key Stakeholders		
1.2 Prepare business case		
1.3 Prepare project charter		
2.0 Planning		
2.1 Hold project kickoff meeting		
2.2 Prepare scope statement		
2.3 Prepare WBS		
2.4 Prepare schedule baseline		
2.5 Identify, discuss, and prioritize risks		
3.0 Executing		
3.1 System analysis		
3.1.1 User request feature		
3.1.2 Analysis of Requirement		
3.2 System design		
3.2.1 Internal Design of Components		
3.2.2 Component Realization Design		
3.2.3 User Interface Realization Design		
3.2.4 Database Design		
3.3 System construction		
3.3.1 Database Construction		
3.3.2 Component Realization		
3.3.3 User Interface Realization		
3.4 System testing		
4.0 Monitoring and Controlling		
5.0 Closing		
5.1 Prepare final project report		
5.2 Prepare final project presentation		
5.3 Lessons learned		

Change Request of project

Table A-3. Change Request form

Project Details:	
Project Name: <i>LesMiserables</i>	
Request #: 1	Date of Request: 16 th March 2023
Requested By: Taejin Kim	
Request Description:	
<p>The original plan was to utilize Nodejs as a backend device as a server. However, in view of the recent research trend, group members agreed that the use of OpenAI is necessary. Therefore, a change was requested through the process shown in the following figure.</p> 	
Reasons for this Change Request:	
<p>The goal of this project is to identify post-Covid-19 unemployment rates and provide easy information on future trends. However, with the recent development of OpenAI, all information provision can be easily accessed in human language. Therefore, the following changes were made based on the latest research results.</p> <p>‘The technical review searched for relevant studies using Google Scala and found Alkhayyat et al.(2022), Han et al.(2022), and Ang and Dong (2022) studies. These studies analyzed the ability to predict post-graduation unemployment, impact of platform self-regulations on crime rates, and relationship between movement restrictions, COVID-19 cases, and unemployment. The project aims to create an unemployment rate information system, which is lacking in research. OpenAI was found to be useful for generating academic content with high integrity, and studies emphasized designing user-centered AI systems for future developments.’</p>	
Chosen solution:	
<p>Finally, the decision was made to modify the scope and process of the project to align with current trends.</p>	

Approval Signature(s) and Date(s):

Group members all, 22th March 2023

The results of each members's contribution by Github website

Table A-4. Each members's contribution

