**Syllabi Chat: an LLM Chatbot to Query a Database of Syllabus using RAG**

Applied Project, Section 202

Zhihang Zhu

NetID: zz5018

NYU ID: N18397241

**Table of Contents**

[**Client Information 3**](#_37jc1d1910gc)

[1.1 Company Name 3](#_9uc6uxe5gqoh)

[1.2 Company Location 3](#_2g4akbed6i1j)

[1.3 Project Sponsor and Location 3](#_dcxdkg2whtkc)

[1.4 Business Description 3](#_2tam2dvfrpxo)

[1.5 Relationship to the Client 3](#_45mulq46uh5w)

[**Project Information 4**](#_99detxv5o6ac)

[2.1 Project Goal 4](#_d96dlnfaz9r6)

[2.2 Benefit of The Project 4](#_vo9d0zdwqpmt)

[2.3 Project Brief Description 4](#_xg2x7b5uyvpw)

[2.4 Objectives and Metrics 5](#_m7yuz2kpsj67)

[2.5 My Role in the Project 6](#_91gnteyxg4ub)

[2.6 Proposed Duration 7](#_7l4up651vvwl)

[2.7 Resources Required 7](#_m6fhijgvmcmm)

[**Learning Anticipations 8**](#_q9i64bwlp1nc)

[3.1 Area of MASY Study Covered 8](#_nvr9hiwv4a1t)

[3.2 Learning Outcomes 8](#_ycpwlsitg8bx)

[3.3 Anticipated Project Finding 8](#_4vsyicq5gl2h)

[3.4 Academic Advisor 8](#_aews6sd58oxu)

# Client Information

## 1.1 Company Name

The Digital Forge

NYU School of Professional Studies and the Management and Systems program (MASY), is a New York-based learning institution.

## 1.2 Company Location

NYU School of Professional Studies is at 12 West 43rd Street, NY, NY.

## 1.3 Project Sponsor and Location

Dr. Andres Fortino (agf249@nyu.edu) can be reached over virtual conference calls as per project requirements.

## 1.4 Business Description

New York University (NYU) is a private research university based in New York City. The MASY degree is based on a unique curriculum that provides students with experiential learning opportunities to develop strong management and leadership skills and gain a comprehensive knowledge of current information technologies.

## 1.5 Relationship to the Client

The Client’s relationship with the Project Manager will be that of an independent contractor, and nothing in this sponsorship is intended to or should be construed to, create a partnership, agency, joint venture, or employment relationship.

# Project Information

## 2.1 Project Goal

Develop a chatbot for students and faculties to query course syllabi with high efficiency and accuracy.

## 2.2 Benefit of The Project

The project aims to develop a chatbot that can query a large database of course syllabi using Retrieval-Augmented Generation (RAG) techniques. I have decided to name the chatbot as Syllabi Chat. This tool will enhance the current internal system for syllabi queries, which is limited by the number of documents that can be uploaded to a commercially available Large Language Model (LLM).

The result of the project can significantly enhance the query efficiency of the syllabus database. Syllabi Chat will go beyond simple string matching in query data. When querying the information for a specific or related course, clients will only need to enter their description and thoughts into this tool, and the background system of Sylla Chat will provide relatively accurate recommendations based on these users’ ambiguous descriptions. This means the user experience will be very convenient and the given query results would be quite accurate.

## 2.3 Project Brief Description

The purpose of this project is to create a chatbot that effectively queries a large database of course syllabi using RAG techniques. Recent advancements in LLMs and information retrieval methods have opened new possibilities for document querying, allowing for more precise and contextually relevant responses. This project seeks to explore these possibilities by developing a prototype chatbot that can overcome the limitations of the current system and provide enhanced access to course information.

To begin, the project will involve collecting and preprocessing syllabi data from the Management and Technology programs (MSPM and MASY) over the past five years. This comprehensive dataset will form the basis for augmenting the LLM's knowledge base using the RAG approach. The chatbot will be designed to receive and interpret user queries, providing relevant information from the syllabi database.

The chatbot's development will include the creation of prompts that enable the LLM to generate accurate and useful responses. A user-friendly interface will be developed to facilitate easy interaction for both faculty and students. The effectiveness of this chatbot will be tested through a limited technology trial, comparing its responses to those generated by traditional search methods. This evaluation will focus on factors such as response relevance, accuracy, and comprehensiveness.

Finally, the project will produce a comprehensive report documenting the development process and the results of the technology trial. This report will include initial findings and discuss the potential implications of using AI-assisted information retrieval in educational settings. Additionally, a draft for a technical conference paper will be prepared, outlining the project's approach, key findings, and the broader impact on improving access to and understanding of course information.

## 2.4 Objectives and Metrics

Objective #1: Design a user-friendly chatbot for course syllabi query tasks within the next 3 months (September 10, 2024 – December 5, 2024).

Measurement: Do related research and write a comprehensive literature review about LLMs and RAG. All the needed files would be stored in GitHub repository and all the add, delete, merge, pull, and modify operations would be recorded. At the final stage, combine the back-end system with the front-end query page.

Objective #2: Build up the user-friendly GUI for students and faculties after the background algorithms are developed (Expected in November, 2024).

Measurement: Delivery in HTML5 based web application by end of the semester. It would be better to have a server to support the website going live while this chatbot would mainly be based on web pages.

Objective #3: Develop efficient query algorithms using RAG and LLMs in the first two months (September 10, 2024 – November 14, 2024).

Measurement: Design embedded query algorithms for that syllabus. The detailed algorithms and their working principles would be delivered through the user manual by the end of the semester.

Objective #4: Use data sets specified by the client to test the validity of the algorithm.

Measurement: The verification of this chatbot system would be delivered by the end of this semester. Also, the detailed methodology will be included in the final presentation.

## 2.5 My Role in the Project

In this project, I will be the project manager and I will define the project scope, set schedules, develop the project plan, and coordinate the project. Besides, several necessary points should be taken into consideration:

* The progress of the project should be well managed within the timeline and reasonable plans and weekly summary should be produced during the project implementation period.
* All the final project files and supporting documentation should be delivered via an accessible GitHub repository. Every operation performed in the project should be displayed in the history document.
* Regular (Weekly) meetings with the client should be held to ensure clients were being updated with the most recent information, progress, and issues about the project. If there are any concerns or issues in the project, related solutions and modifications should be done as soon as possible.

Also, I will be the people to conduct the whole project:

* All steps in the project must be well documented as the project progresses and comes up with summary progress reports weekly. Also, every operation of the project will be included in the GitHub and I would be responsible for the repository management task.
* The final deliverable must be in working code as an installable module, as well as instructions and necessary tutorials for installation and use. If the project is working smoothly, more versions that can run on multiple platforms may be developed within 3 weeks. (I hope so)
* Ensuring the final project files and supporting documentation to be delivered via an accessible GitHub repository.

## 2.6 Proposed Duration

The project will start on September 1st, 2024, and the final version will be delivered on December 12th, 2024. Approximately 325 hours in total.

## 2.7 Resources Required

The required resources are shown in the following list:

* Necessary programming tools: Adobe Dreamweaver for Web application development, Vue.js for Front-back connection, MYSQL for database management, and Python for AI algorithms development.
* Related AI algorithms packages, developed through Python libraries.

# Learning Anticipations

## 3.1 Area of MASY Study Covered

* Project Management Methodology: For this project, I will be the project manager. I will initiate the project, plan, execute, monitor/control and close the project. Similar to what I did in PM class but on a real-world project.
* Research Process Methodology: I will need to use the skills I learned in my Research Process Methodology (RPM) class to conduct my research and make the results into a research paper format. I will also practice the research method discussed in class to conduct the market research if necessary.
* Database Management: I will use the skills that I learned in this course about data process and structure, and advanced database applications.

## 3.2 Learning Outcomes

* Necessary Project management skills. Mainly about Agile Methods.
* Programming Languages: Python
* Server management: Nginx
* Database management: MYSQL

## 3.3 Anticipated Project Finding

Potential Problems and Limitations:

* Creating a program required lots of expertise, knowledge including full stack program development skills, rigorous version control. It is difficult to always build the best performance apps right on time and meet all the requirements in one version, but sooner after the project finished, it will require frequent updates and further maintenance.
* This project will only be implemented on web applications. Other versions like mobile devices and independ applications need more time and resources to develop.

## 3.4 Academic Advisor

Treesa Ouseph

## 3.5 Expected Graduation Date

December, 2024