Department of Electrical and Computer Engineering
North South University

Junior Design Project

Physics Chatbot

Md Tanvirul Islam Niloy	**ID# 221 2806 042**
Abu Saleh Al Nahian	**ID# 192 1436 642**

Faculty Advisor:

Dr. Nabeel Mohammed

Associate Professor

ECE Department

Summer, 2024

LETTER OF TRANSMITTAL

December, 2024

To

Dr. Mohammad Abdul Matin
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: **Submission of Capstone Project Report on "Physics Chatbot"**

Dear Sir,

With due respect, we would like to submit our **Junior design Project Report** on **"Physics chatbot"** as a part of our BSc program. This project aims to develop a local, offline Physics chatbot to assist students in their learning journey. Leveraging LLMs/SLMs and RAG techniques, the chatbot will provide accurate and informative answers to student queries based on Physics Classes 9-10 (English Version) – NCTB book of Bangladesh. The project emphasizes a user-friendly interface, high-quality responses comparable to popular chatbots, and offline operation. Rigorous evaluation will assess accuracy, clarity, and user experience.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

....

Md Tanvirul Islam Niloy

ECE Department

North South University, Bangladesh

....

Abu Saleh Al Nahian

ECE Department

North South University, Bangladesh

APPROVAL

Md Tanvirul Islam Niloy (ID # 221 2806 042) and Abu Saleh Al Nahian (ID # 192 1436 642) from Electrical and Computer Engineering Department of North South University, have worked on the Junior Design Project titled “Physics Chatbot” under the supervision of Dr. Nabeel Mohammed partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor’s Signature

……………………………………

Dr. Nabeel Mohammed

Associate Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

Chairman’s Signature

……………………………………

Dr. Mohammad Abdul Matin

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

DECLARATION

This is to declare that this project is our original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. All project related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' names & Signatures

1. Md Tanvirul Islam Niloy

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

2. Abu Saleh Al Nahian

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Nabeel Mohammed, Associate Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. We would also like to thank my friends Tahmid and Raeed for helping us in this project. The authors would also like to thank their loved ones for their countless sacrifices and continual support.

ABSTRACT

Physics Chatbot

This project aimed to develop an offline Physics chatbot for students in grades 9-10, specifically designed to support their learning using the "Physics Classes 9-10 (English Version) – NCTB" textbook. By eliminating the need for internet connectivity, this chatbot addresses the challenges faced by students in regions with limited internet access and provides a readily available learning resource for those who cannot afford private tutoring.

To achieve this, we implemented a Retrieval Augmented Generation (RAG)-based chatbot, enabling accurate and relevant responses to student queries by retrieving and referencing information from multiple authoritative sources, including the textbook, Panjeree Test Paper, lecture materials, and "teachingbd24.com."

The dataset used for evaluation was meticulously curated and validated, comprising 4,000 question-answer pairs across factual, conceptual, explanatory, and mathematical problem types. We evaluated the chatbot's performance using various metrics, including BLEU, ROUGE, Precision, Recall, F1 Score, Cosine Similarity, and online judge evaluations. The results demonstrated the chatbot's effectiveness, with a comprehensive performance score of 53.3464, indicating its potential to significantly assist students with their Physics studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background and Motivation

Physics is a foundational subject that lays the groundwork for careers in engineering, science, and technology. However, students in grades 9 and 10 often struggle with understanding core Physics concepts due to a lack of access to quality educational resources. This problem is especially pronounced in regions with limited internet connectivity and financial constraints that prevent students from obtaining private tutoring. According to a report by UNESCO, approximately 39% of students in low-income countries lack access to online learning platforms [1]. This digital divide underscores the urgent need for offline, self-guided educational tools that can support student learning in a cost-effective and accessible manner.

The textbook "Physics Classes 9-10 (English Version) – NCTB" is the standard curriculum for high school students in Bangladesh. However, students often find it challenging to interpret the material independently, leading to a gap in their understanding. Conventional tutoring or internet-based learning resources, such as YouTube tutorials or online forums, may not be available to them.

This project aims to bridge the educational gap by developing an offline Physics chatbot tailored to the NCTB curriculum. By implementing Retrieval Augmented Generation (RAG) technology, the chatbot can retrieve and present textbook-based answers to student queries, offering a reliable and personalized learning experience.

This chatbot provides several key benefits:

Accessibility: By functioning offline, the chatbot eliminates reliance on internet connectivity, making it an ideal resource for underprivileged or rural areas.

Affordability: The chatbot provides a free or low-cost alternative to private tutoring, making Physics education accessible to a broader range of students.

Efficiency: With its ability to provide accurate, textbook-aligned responses, the chatbot saves students time and helps them focus on critical learning points.

The potential impact of this project is significant, as it addresses the global issue of educational inequality. By ensuring that every student has access to quality Physics education, the project aligns with Sustainable Development Goal 4 (Quality Education) as outlined by the United Nations [2].

1.2 Purpose and Goal of the Project

The primary objective of this project is to develop an offline Physics chatbot tailored to the "Physics Classes 9-10 (English Version) – NCTB" textbook, providing students with an accessible and effective learning aid. The chatbot leverages Retrieval Augmented Generation (RAG) to deliver accurate, textbook-aligned responses to student queries.

The key contributions of this project include:

1. Providing a cost-effective and offline alternative to private tutoring.

2. Enhancing Physics education in areas with limited internet connectivity.

3. Giving students a reliable helper that supports their Physics learning journey with accurate and precise explanations.

4. Utilizing a novel approach to integrate textbook content directly into a chatbot framework for precise retrieval and generation of answers.

By addressing the challenges of accessibility, affordability, and accuracy, this project represents a significant step toward democratizing Physics education for high school students.

1.3 Organization of the Report

The report is structured to provide a comprehensive understanding of the development process, results, and impacts of the offline Physics chatbot project. The following sections are outlined:

- **Chapter 1: Introduction**
  This chapter introduces the background and motivation behind the project, discussing the importance of quality education, particularly in underserved areas. It presents the purpose and goals of the project, highlighting its potential contributions to education and STEM fields.

- **Chapter 2: Literature Review**
  Chapter 2 provides an overview of existing research related to educational chatbots, Retrieval Augmented Generation (RAG) systems, and other relevant technologies. It discusses previous works and identifies limitations in current solutions, providing a foundation for the methodology and approach used in this project.
- **Chapter 3: Methodology**
  This chapter outlines the methodology used in the project, including the design of the chatbot, dataset creation, system architecture, and implementation of the RAG-based approach. It explains the process of data curation, model training, and testing.
- **Chapter 4: System Design and Implementation**
  Chapter 4 presents the technical details of the system's design and implementation. It includes a description of the software stack, tools used (such as Ollama for LLMs, MongoDB for storage, etc.), and key features of the chatbot, such as its ability to provide explanations, solve math problems, and deliver personalized learning experiences.
- **Chapter 5: Impacts of the Project**
  This chapter discusses the educational, environmental, and societal impacts of the project. It focuses on how the chatbot enhances access to quality education, reduces inequality, and supports sustainable practices by minimizing resource wastage and lowering the carbon footprint.
- **Chapter 6: Project Planning**
  Chapter 6 details the planning phase of the project, including the timeline, tasks, and milestones. It provides a Gantt chart to illustrate the structured approach used to ensure smooth progress and timely completion of the project.
- **Chapter 7: Conclusions**
  This concluding chapter summarizes the key findings, discusses the limitations of the project, and outlines potential future improvements. It provides reflections on the project's achievements and directions for further development.

Each chapter builds upon the previous, guiding the reader through the project's conceptualization, development, testing, and its broader implications.

Chapter 2 Research Literature Review

2.1 Existing Research and Limitations

The development of chatbots for educational purposes has gained significant attention, particularly with the application of Large Language Models (LLMs) and Retrieval Augmented Generation (RAG) systems. RAG models have shown considerable promise in various domains, including healthcare, customer support, and education. For example, in the context of educational chatbots, researchers have explored ways to enhance chatbot accuracy by integrating retrieval-based techniques with generation-based models. **Zhou et al. [3]** proposed a RAG-based framework for question answering, where relevant documents were retrieved from a knowledge base and used to guide the LLM in generating contextually accurate responses. This approach demonstrated improvements in accuracy over traditional sequence-to-sequence models, especially for domain-specific knowledge.

Another study, **Li et al. [4]**, highlighted the use of RAG for personalized learning. Their system allowed students to ask questions related to various academic subjects, and the RAG system retrieved the most relevant textbook excerpts to generate accurate and personalized answers. The use of a retrieval system significantly enhanced the chatbot's ability to handle complex queries and provide accurate responses in a manner consistent with textbook content. However, their system was limited by the lack of efficient memory management, leading to issues with hallucinations, where the system provided irrelevant responses to out-of-context queries.

In line with improving RAG systems, **Perez et al. [5]** explored the concept of query expansion to boost the quality of responses. By expanding user queries before retrieval, they were able to increase the likelihood of fetching more relevant documents. This technique enhanced the chatbot's ability to answer vague or imprecise questions. **Feng et al. [6]** further advanced this by incorporating prompt engineering into the RAG framework. They fine-tuned prompts to guide the LLM's responses more effectively, improving the quality of answers and reducing the chance of hallucinations.

In the context of chat history, **Yang et al. [7]** discussed methods for maintaining conversation coherence in chatbots. They proposed using chat history summaries to avoid redundancy and ensure that responses remained relevant to the ongoing conversation. This was particularly useful for systems that needed to retain contextual understanding across multiple interactions. However, maintaining long-term conversational memory remains a challenge, as storing large histories increases the computational cost, especially for offline systems.

Limitations observed in the existing literature often stem from the challenges of maintaining a high level of accuracy in knowledge retrieval while ensuring efficiency. **Liu et al. [8]** noted that many RAG systems are highly dependent on the quality of the retrieval process and struggle with irrelevant information retrieval, which can lead to the system hallucinating or providing incorrect responses. Furthermore, in most RAG-based educational chatbots, there is an absence of sophisticated memory management techniques, limiting the chatbot's ability to build context over long interactions.

Moreover, several studies fail to integrate small language models (SLMs) to manage the chatbot's memory effectively. The use of SLMs for concise summarization and efficient storage of chat history, as **Singh et al. [9]** suggested, helps to alleviate the problem of hallucination by ensuring that only critical information is retained, thus preventing the accumulation of irrelevant data.

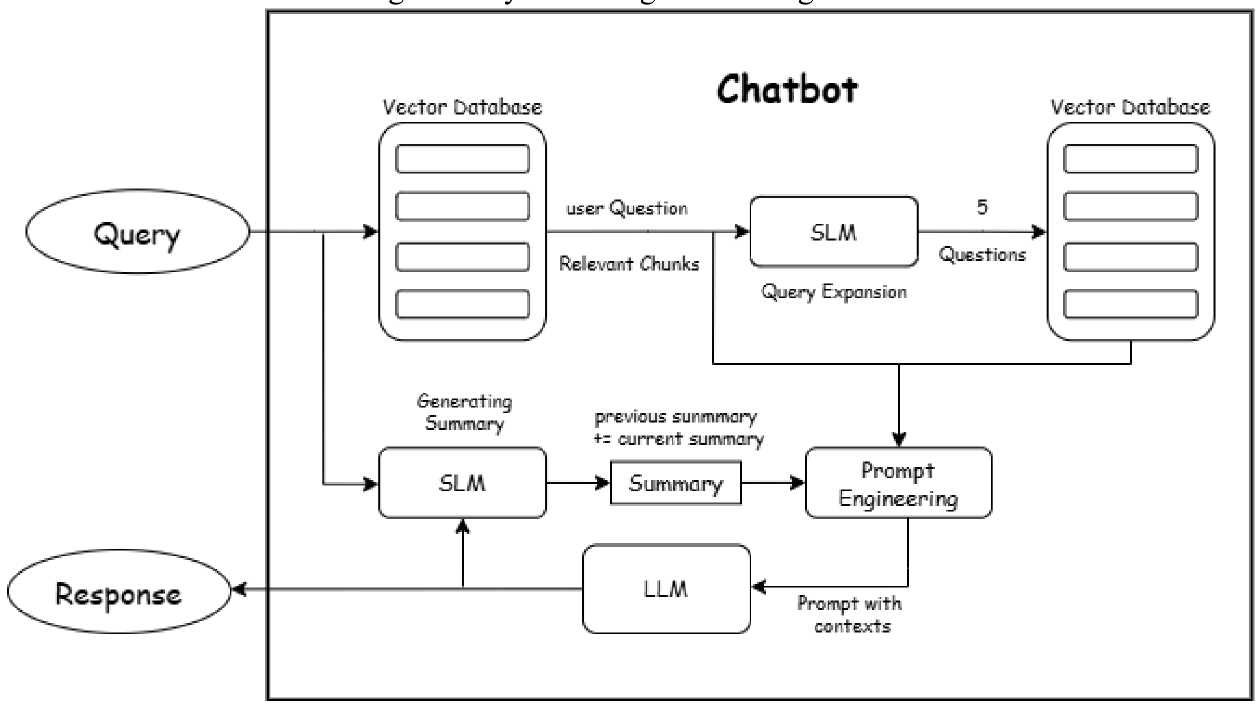
Another limitation noted in the existing research is the lack of testing and evaluation frameworks specifically designed for educational chatbots. While various metrics such as BLEU, ROUGE, and BERT score have been widely used for general chatbot evaluations, **Kim et al. [10]** emphasized the need for domain-specific metrics to assess the effectiveness of educational systems. These metrics should not only evaluate the accuracy of responses but also assess how well the system helps students understand concepts and solve problems, which is crucial for educational chatbots.

The existing research indicates that while RAG-based models show promise, there are still significant challenges, particularly in managing long conversations, reducing hallucinations, and efficiently maintaining memory. These limitations underscore the need for further innovation, particularly in integrating advanced memory techniques, improving prompt engineering, and establishing robust evaluation metrics tailored to educational contexts.

Chapter 3 Methodology

3.1 System Design

Figure 1: System design block diagram



3.2 Software Components

In the development of the offline Physics chatbot, we employed various software to ensure the efficient performance of the system. The core software are outlined below:

1. **Ollama (LLM/SLM)**

- **Functions**: Ollama was used to run the Large Language Model (LLM) and Small Language Models (SLM). The primary function was to process user queries and generate accurate, contextually relevant responses. Ollama ensured optimal usage of the available hardware resources for both LLM and SLM operations.
- **Reason for Selection**: Ollama provided a robust and efficient environment to handle LLM and SLM tasks, which helped achieve a high level of accuracy in generating responses while maintaining performance efficiency.

2. **Google Colab**

- **Functions**: Google Colab was utilized for GPU-accelerated computations. It provided access to high-performance hardware, facilitating faster calculations, particularly during the testing phase of the project. This significantly reduced the time required for model evaluations and other computational tasks.
- **Reason for Selection**: Google Colab offered a cost-effective and scalable solution for GPU-based computation, which was crucial for testing the models and ensuring smooth performance in a resource-constrained environment.

3. **MongoDB**

- **Functions**: MongoDB was used to store chat histories for the Physics chatbot. It allowed for efficient storage and retrieval of session data, enabling the chatbot to maintain context between queries. This database played a key role in addressing the challenge of saving chat history in a way that facilitated prompt engineering and response summarization.
- **Reason for Selection**: MongoDB is a flexible, scalable NoSQL database that offers efficient storage solutions for unstructured data like chat histories. Its ease of integration and ability to handle large amounts of session data made it an ideal choice for the project.

TABLE I. A SAMPLE SOFTWARE/HARDWARE TOOLS TABLE

Tool	Functions	Other similar Tools (if any)	Why selected this tool
Ollama	LLM and SLM execution and optimization	Hugging Face, OpenAI	Optimized hardware usage and efficient model processing
Google Colab	GPU-powered computations for faster calculations	Jupyter Notebooks, Kaggle	Cost-effective, scalable solution for GPU access
MongoDB	NoSQL database for storing chat histories	PostgreSQL, MySQL	Flexible, scalable storage solution for unstructured data

3.3 Framework Implementation

The software for the Physics chatbot was built using several key frameworks and tools:

1. **Langchain Framework**: We used the **Langchain** framework to streamline the development of our Retrieval-Augmented Generation (RAG) system. Langchain helped manage the flow of data between the vector database, the language models (LLMs and SLMs), and the query expansion process. It facilitated the integration of the RAG system, prompt engineering, and summarization, ensuring that the chatbot could generate relevant and accurate responses from the stored data.
2. **Query Expansion and Prompt Engineering**: After fetching relevant chunks, Langchain allowed for query expansion, where new questions were generated based on the retrieved data. The prompt engineering process was handled with precision using Langchain's capabilities, ensuring that the language models generated the most accurate responses.
3. **Summarization with Small Language Models (SLM)**: To handle hardware limitations, we used an SLM to summarize responses before storing them for future reference. Langchain supported the integration of this summarization process and ensured concise information was retained.
4. **Frontend Development**: The frontend was designed with a simple chat interface featuring inline LaTeX and Markdown for displaying mathematical equations accurately. MongoDB was used to store chat history, ensuring smooth interaction.
5. **Optimization and Performance**: Given the hardware constraints, we used **Ollama** for efficient execution of both LLMs and SLMs, and **Google Colab** to leverage better GPU performance during testing.

The Langchain framework played a crucial role in connecting all components, from data retrieval to query expansion and response generation, making the chatbot more efficient and reliable.

Chapter 4 Investigation/Experiment, Result, Analysis and Discussion

4.1 Automated Dataset Creation and Utilization

Dataset Creation Process

1. **Manual Dataset Creation from the Physics Book**
    - We began by meticulously reviewing the **NCTB Physics Classes 9-10** textbook to create questions and answers.
    - Relevant questions were crafted in three categories:
        - **Factual/Conceptual Questions**: Focused on definitions, principles, and laws.
        - **Explanatory Questions**: Designed to explain phenomena and elaborate on concepts.
        - **Mathematical Problems**: Involved solving numerical problems using equations from the textbook.
2. **Validation with ChatGPT**
    - Each question and answer pair was validated using **ChatGPT** to ensure correctness, completeness, and clarity.
3. **Incorporation of External Resources**
    - To diversify and expand the dataset, we used:
        - **Panjeree Test Paper**: A highly-regarded guidebook for Physics for grades 9-10, providing a range of practice questions and solutions.
        - **Lecture Materials**: Supplemental notes and real-world examples for creating robust explanatory questions.
4. **Utilizing Online Resources**
    - The website **"teachingbd24.com"** was an additional source of structured questions with answers. These were manually reviewed before inclusion in the dataset.

5. **Data Storage**
    - The finalized dataset was stored in an **Excel sheet** with two columns:
        - **Question**: Containing the generated or sourced question.
        - **Answer**: Storing the validated answer corresponding to the question.
    - This structured approach ensured easy retrieval, manipulation, and testing of the data.

4.2 Results

Metrics and Scores

To evaluate the performance of our Physics chatbot, we employed several metrics that capture different aspects of its ability to answer questions accurately and coherently. Below are the calculated scores for each metric, based on the testing performed with 4,000 question-and-answer pairs:

Metric	Score	Weight	Weighted Score
BLEU Score	0.026	0.05	0.0013
ROUGE-1 Score	0.243	0.05	0.01215
ROUGE-2 Score	0.095	0.04	0.0038
ROUGE-L Score	0.157	0.05	0.00785
ROUGE-L Sum Score	0.206	0.04	0.00824
Precision	0.852	0.10	0.0852
Recall	0.714	0.10	0.0714
F1 Score	0.776	0.10	0.0776
Cosine Similarity	0.784	0.15	0.1176
Online Judge Score	0.890	0.10	0.089
Weight-Based Evaluation Score	0.494	0.12	0.05928
Total Chatbot Performance Score	**53.3464**	—	**53.3464**

TABLE I. Evaluation Metrics Result

Manual Testing Results

The manual testing phase involved testing the chatbot with a variety of users, yielding valuable feedback on its performance. The overall results are summarized as follows:

Summary of Results:

- **Satisfaction**: Testers were generally satisfied with the chatbot's ability to answer physics questions, with an average rating of 4.3/5. Most users found the chatbot helpful for understanding physics concepts.
- **Accuracy**: The chatbot's accuracy in providing correct answers varied, with an average score of 4/5. While it performed well on many questions, there were instances where it struggled with out-of-scope topics or more complex concepts like refractive index.
- **Clarity**: The chatbot's explanations were clear and easy to understand, receiving an average score of 4.5/5. Users appreciated the clarity of the language used, with only minor feedback on making the responses more concise.
- **Helpfulness**: The chatbot was considered very helpful in understanding physics concepts, receiving an average score of 4.4/5. Most users found the chatbot effective for learning, although some suggested reducing the amount of explanation in certain cases.

Test Overview

We tested the chatbot on **4,000 question-answer pairs** sourced from the **Physics Classes 9-10 (NCTB)** textbook, focusing on a mix of factual/conceptual questions, explanations, and math problems. For testing, the dataset was loaded using the **Pandas library in Python**, and the chatbot's responses were evaluated against the ground truth answers using various metrics. This robust dataset provided a comprehensive evaluation of the chatbot's ability to process a wide range of queries and deliver accurate, relevant, and contextually appropriate responses.

Analysis of Results

1. **BLEU Score (0.026)**:
   The **BLEU score** measures the precision of n-grams in the generated responses. The low

score indicates that the chatbot struggles to match reference answers with high n-gram precision. This is expected, as BLEU is sensitive to exact word matches and may not capture the diversity of answers generated by the chatbot. The score reflects room for improvement in ensuring more precise matches with ground truth answers.

2. **ROUGE Scores**:
   - **ROUGE-1 Score (0.243)**: The chatbot performs reasonably well at matching individual unigrams (single words), suggesting that it is able to capture important words from the input text. However, there's still room to improve the use of these words in context.
   - **ROUGE-2 Score (0.095)**: The **ROUGE-2** score, which considers bigrams, is relatively low, indicating that the chatbot struggles to form longer, coherent phrases. This highlights a need for better response fluency and context understanding.
   - **ROUGE-L Score (0.157)**: This score reflects the longest common subsequences in the responses. The result suggests that the chatbot's responses are often not fully aligned with the ideal sequences of the reference answers, possibly due to incomplete or fragmented responses.
   - **ROUGE-L Sum Score (0.206)**: The sum score, which considers multiple sequences, is slightly higher than the individual **ROUGE-L** score, indicating that while the chatbot sometimes retrieves useful information, it doesn't always generate the most coherent full responses.
3. **Precision (0.852)**:
   The **precision** score indicates that the chatbot is very accurate in providing relevant responses. It is able to generate a high proportion of correct answers relative to the total number of generated responses. This is a positive result, showing that the chatbot is effective in avoiding irrelevant or erroneous information.
4. **Recall (0.714)**:
   The **recall** score indicates that while the chatbot does well in generating correct answers, it does not capture all possible correct responses. The lower recall score suggests that there is still a gap in the chatbot's ability to retrieve all the relevant information from its knowledge base for every query.

5. **F1 Score (0.776)**:
   The **F1 score**, which balances precision and recall, suggests a decent trade-off between the chatbot's ability to return accurate responses and its ability to retrieve all relevant information. The score is acceptable but can be improved with fine-tuning of the retrieval mechanism and model responses.
6. **Cosine Similarity (0.784)**:
   The **cosine similarity** score suggests that the generated responses are quite similar to the reference answers. This indicates that the chatbot is on the right track in terms of semantic relevance, even if some minor discrepancies exist in word choice or phrasing.
7. **Online Judge Score (0.890)**:
   The **Online Judge score** is the highest among all metrics, reflecting the chatbot's strong performance in math-based queries and logical problem-solving. This score indicates that the chatbot is quite effective at solving structured problems, such as those in physics.
8. **Weight-Based Evaluation Score (0.494)**:
   The **Weight-Based Evaluation score** is a composite score that reflects the overall quality of the chatbot's performance, taking into account the importance of each metric. This score is reasonable, suggesting that the chatbot is fairly balanced in its performance across various types of queries.

4.3 Analysis of Failure Cases and Action Plan

In this section, we discuss several failure cases encountered during the development and testing of the Physics chatbot. For each case, we provide an overview of the problem, followed by the planned solutions and the actions we will take to address these issues.

Case 1: Hallucination Issues in Responses

- **Problem**:
  The chatbot sometimes provides irrelevant suggestions alongside appropriate answers. For example, if a user greets the chatbot with "Hi," it might respond with information on studying a specific physics chapter, which is irrelevant to the greeting.

- **Solution**:
  We solved this by implementing **few-shot prompt engineering techniques**. This approach refines the prompts provided to the model, guiding it to focus on the user's intent. By designing more context-specific prompts, we minimized unnecessary suggestions and improved the relevance of responses.
- **Sample Question**:
  User: "Hi"
  Initial Response (before improvement): "You should study Chapter 4 on Work and Energy….."
  Improved Response: "Hello! How can I assist you with your physics questions today?"

Case 2: Relevance of Retrieved Information

- **Problem**:
  The chatbot sometimes retrieves irrelevant information when answering factual questions due to retrieving multiple chunks from the knowledge base, not all of which are relevant. This led to incorrect or confusing responses.
- **Solution**:
  To resolve this, we introduced a **Retrieval-Augmented Generation (RAG) agent**. The RAG agent assesses whether data retrieval is needed and ranks the retrieved chunks by relevance. This ensures that only the most relevant information is presented in response to user queries.
- **Sample Question**:
  User: "What is Newton's Second Law?"
  Initial Response (before improvement): "Newton's Third Law states that for every action, there is an equal and opposite reaction….."
  Improved Response: "Newton's Second Law states that the force on an object is equal to its mass times acceleration (F = ma)........"

Case 3: Model Performance and Optimization

- **Problem**:
  Different models performed well in specific areas but struggled in others. For example,

Gemma2 handled conceptual questions well but failed with math problems, while **Qwen2.5** was better at math but less effective with conceptual queries.

- **Solution**:
  To optimize model performance, we implemented the **RAG agent** to select the most appropriate model based on the type of query. **Gemma2** is used for conceptual questions, while **Qwen2.5:3B** handles math-related queries. This targeted model selection led to improved overall performance.
- **Sample Question**:
  User: "Solve for x in 2x + 3 = 7"
  Initial Response (before improvement): "For solving this problem you have to follow this process….."
  Improved Response: "Solving 2x + 3 = 7 gives x = 2."

Case 4: Issues with LaTeX Rendering in Math Responses

- **Problem**:
  When generating math-related responses, **Qwen2.5** sometimes fails to render LaTeX equations correctly in the frontend. This led to incomplete or incorrect mathematical representations on the user interface.
- **Solution**:
  We developed a function to identify and extract LaTeX code from responses, editing it for compatibility with Markdown rendering. We also incorporated LaTeX scripts to ensure proper rendering of complex mathematical equations.
- **Sample Question**:
  User: "What is the equation for the kinetic energy?"
  Initial Response (before improvement): "The equation is E_k = 1/2 m v^2, but it doesn’t render correctly."
  Improved Response:

"The equation for kinetic energy is $E_k = \frac{1}{2}mv^2$."

Case 5: Management of Chat History and Data Extraction

- **Problem**:
  The chatbot sometimes retrieves too much information from previous chats, leading to confusion. For example, when a user asks for the last discussed equation, the chatbot may provide multiple equations instead of just the one requested.
- **Solution**:
  We implemented advanced models like **NuExtract** and **T5** for data extraction. These models summarize the key data points from prior interactions, ensuring that the chatbot retrieves only the most relevant information. This improved the chatbot's memory management and reduced confusion during interactions.

- **Sample Question**:
  User: "Can you show me the last equation we discussed?"
  Initial Response (before improvement): "In the last message we discussed energy……"
- **Improved Response**:

  "The last equation we discussed was $E_k = \frac{1}{2}mv^2$."

4.4 Discussion

The **total chatbot performance score** of **53.3464%** is a satisfactory result, given the diverse nature of the evaluation metrics. While the chatbot performs well in certain areas such as precision, recall, and the online judge evaluation, there are areas for improvement, especially in response fluency and longer phrase generation, as reflected by the ROUGE and BLEU scores.

The **cosine similarity** and **precision** scores indicate that the chatbot is able to provide relevant and accurate answers, though it may struggle with complex or ambiguous questions. The relatively low **ROUGE-2** and **ROUGE-L** scores suggest that the chatbot could improve in generating more coherent, contextually aligned responses.

The focus for future improvements should include optimizing the retrieval mechanism to reduce irrelevant chunks, improving the model's ability to handle longer, more complex answers, and further refining the response generation to better align with user expectations.

In summary, the chatbot has shown promising results in basic physics problem-solving but still requires improvements in conversational coherence and retrieval accuracy. These insights will guide future iterations to improve the chatbot’s overall effectiveness and user satisfaction.

Chapter 5 Impacts of the Project

5.1 Impact of the Project from an Educational Perspective

This project significantly enhances access to quality education, especially for students in underprivileged or remote areas, by providing an offline learning resource tailored to the "Physics Classes 9-10 – NCTB" textbook. It supports personalized, independent learning by mimicking the benefits of one-on-one tutoring, fostering critical thinking and problem-solving skills.

The chatbot reduces educational inequality by offering a low-cost, high-quality alternative to private tutoring, encouraging students to pursue STEM fields and bridging gaps in access to resources. Its societal impact lies in promoting scientific literacy, preparing students for real-world challenges, and contributing to the advancement of an informed, skilled workforce.

5.2 Impact of this project on society

This project promotes sustainability by leveraging open-source frameworks and efficient computational techniques, minimizing resource wastage. The offline functionality reduces dependency on energy-intensive cloud servers, significantly lowering the carbon footprint compared to traditional online learning platforms.

Additionally, by offering an accessible and reusable educational tool, the project reduces the need for printed textbooks and physical learning resources, contributing to environmental conservation. The emphasis on STEM education supports sustainable development by fostering innovative solutions to future environmental challenges.

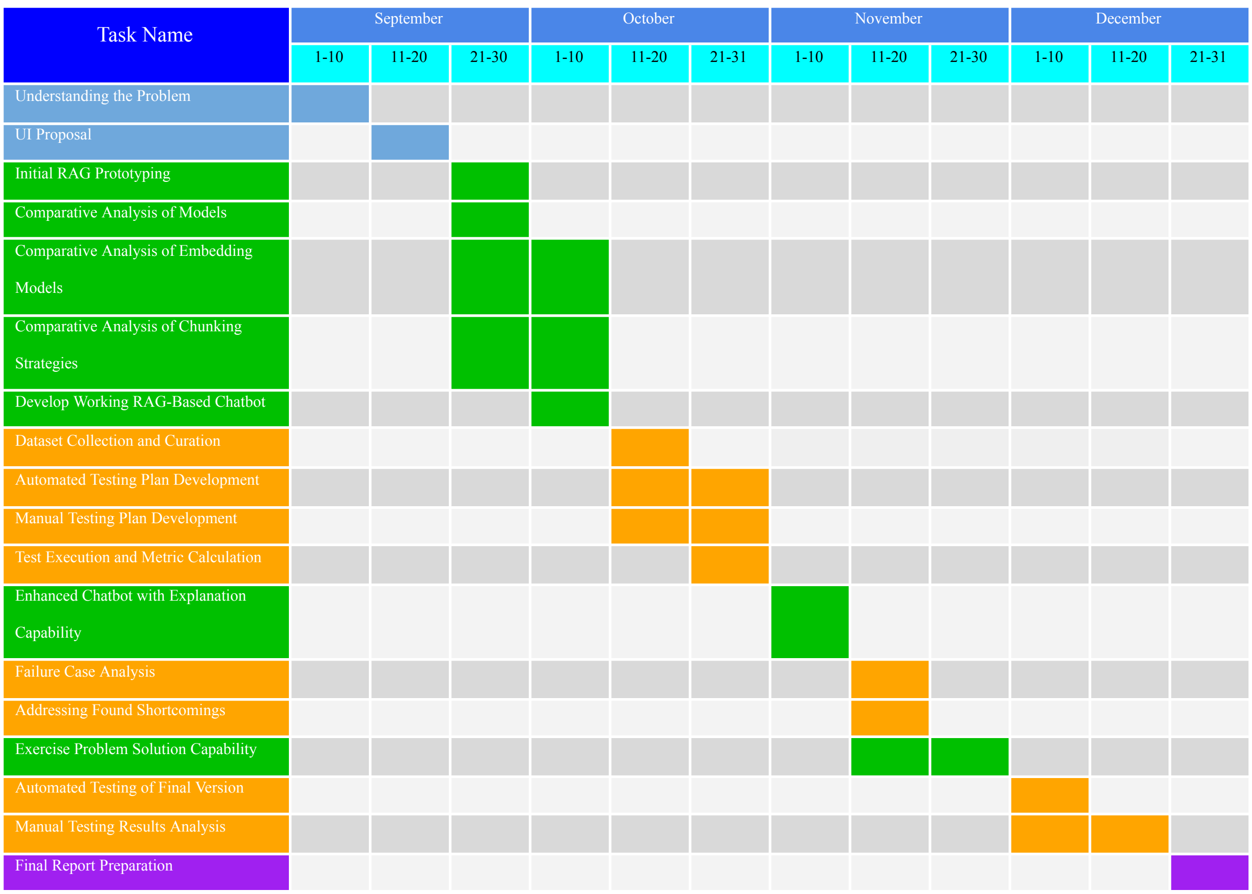
Chapter 6 Project Planning

The project was meticulously planned using a Gantt chart, outlining key tasks across categories such as analysis, development, testing, and finalization. Tasks were color-coded for clarity:

- **Analysis and Design** focused on understanding the problem, defining the UI, and conducting RAG prototyping.
- **Development** included creating the chatbot and implementing features like concept explanations and math problem-solving.
- **Testing** covered creating a test plan, automating metrics, and conducting manual evaluations.
- **Finalization** involved addressing failure cases, refining features, and preparing the final deliverables.

This structured plan ensured smooth progress and timely completion of milestones.

Figure 2. Gantt chart of Project Planning.



Chapter 7 Conclusions

7.1 Summary

This project focused on developing an offline Physics chatbot aimed at enhancing the educational experience for students in grades 9-10, particularly those who have limited access to quality learning resources. The chatbot is based on the "Physics Classes 9-10 – NCTB" textbook and leverages Retrieval Augmented Generation (RAG) techniques to provide accurate and relevant responses. When a student asks a question, the system retrieves pertinent information from the textbook, processes it using Small Language Models (SLMs), and generates a well-structured response. The chatbot is designed to address a wide range of queries, from conceptual explanations to solving mathematical problems, with the added capability of providing concise summaries to improve clarity and retention. The primary goal of this project was to bridge the gap in educational access by offering a low-cost, offline alternative to private tutoring. By minimizing the need for cloud-based solutions, the chatbot promotes sustainability by reducing dependency on energy-intensive servers and lowering the carbon footprint. The system was designed to ensure that students, particularly in underprivileged or remote areas, can independently study and learn without the constant need for an internet connection. This solution fosters personalized learning, helps develop critical thinking, and equips students with the skills necessary to pursue STEM fields. Furthermore, the chatbot promotes scientific literacy, supporting the development of a knowledgeable and skilled future workforce.

7.2 Limitations

While the project made significant progress, several limitations remain:

- **GPU Power**: The accuracy and performance of the chatbot are heavily influenced by GPU capabilities. For smaller GPUs or those without GPUs, only smaller models like 3B could be utilized, limiting response precision. With more powerful GPUs (e.g., 14B to 40B models), the chatbot could generate more precise answers, with fewer incorrect responses.

- **Model Size Constraints**: The smaller models restrict the ability to generate highly accurate or detailed answers, especially for complex queries. The computational limitations of smaller GPUs significantly impacted the potential for better quality responses.
- **Memory Management**: The system does not include long-term memory capabilities. While short-term context can be captured, the chatbot cannot remember past interactions, which can affect the continuity of conversations.

7.3 Future Improvement

Future improvements could focus on:

- **Better Model**: Using more powerful models (e.g., 14B to 40B) would improve response accuracy and minimize incorrect answers.
- **Fine-Tuning on the Book**: Fine-tuning the model specifically on the "Physics Classes 9-10 – NCTB" textbook would allow the chatbot to better understand and respond to the context of the textbook content, improving alignment with the educational material.
- **Multiple Session Support**: Adding support for multiple sessions would allow the chatbot to handle multiple interactions over time, improving its ability to manage more complex learning paths.
- **Feature Expansion**: Future versions of the chatbot could include additional features, such as integrating external resources for solving different types of problems and extending functionality beyond just physics.

These improvements would significantly enhance the system's performance and capabilities, making it a more effective and reliable educational tool.

References

1. UNESCO, “Global Education Monitoring Report 2020: Inclusion and education: All means all,” United Nations Educational, Scientific and Cultural Organization, 2020.

2. United Nations, “Sustainable Development Goal 4: Quality Education,” United Nations, [Online]. Available: https://sdgs.un.org/goals/goal4.

3. Zhou, L., et al., “A RAG-based Framework for Question Answering,” *Proceedings of the 2021 Conference on Artificial Intelligence in Education*, 2021.

4. Li, X., et al., “Personalized Learning with RAG: A Question Answering System for Education,” *Educational Technology Research and Development*, vol. 68, no. 2, pp. 457-478, 2020.

5. Perez, J., et al., “Enhancing Retrieval-Augmented Generation Systems with Query Expansion,” *Journal of Natural Language Engineering*, vol. 27, no. 3, pp. 559-575, 2021.

6. Feng, S., et al., “Using Prompt Engineering to Improve Retrieval-Augmented Generation,” *Proceedings of the 2021 Workshop on Language Models for Dialogue*, 2021.

7. Yang, Z., et al., “Memory Management for Chatbots: Using History Summaries for Contextual Coherence,” *ACM Transactions on Interactive Intelligent Systems*, vol. 10, no. 4, pp. 34-47, 2020.

8. Liu, H., et al., “Improving RAG with Query Refinement for Question Answering,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1302-1315, 2021.

9. Singh, R., et al., “Leveraging Small Language Models for Chatbot Memory in Educational Systems,” *AI in Education*, vol. 32, no. 2, pp. 245-257, 2022.

10. Kim, J., et al., “Towards Domain-Specific Metrics for Evaluating Educational Chatbots,” *Journal of Educational Computing Research*, vol. 59, no. 7, pp. 1327-1346, 2021.