

Exploring how Natural Language Processing Techniques can be used for Personalized Learning

K. D. Samarasinghe (Reg. No.: MS22907370)

A THESIS

SUBMITTED TO SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

2025/01/27

Prof. Nuwan Kodagoda

Approved for MSc. Research Project:

.

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign:

Kavisha Samarasinghe

Date:

ABSTRACT

Exploring how Natural Language Processing Techniques can be used for Personalized Learning

Kavisha Samarasinghe

MSc. in Information Technology Specializing in Enterprise Application Development Supervisor: Prof. Nuwan Kodagoda

December 2024

Artificial intelligence (AI) has had a profound impact on many industries, and education is no exception. Large Language Models (LLMs), including GPT-3 and GPT-4, stand out among AI-driven technologies as ground-breaking instruments that have the power to revolutionize conventional learning settings. These models allow for a realistic, conversational interface between students and instructional information because they were trained on massive amounts of text data. They produce logical, contextually appropriate answers to a range of queries by examining linguistic patterns, giving students the opportunity to participate in individualized learning experiences. Though LLMs offer chances for dynamic learning, their effectiveness is constrained by the static nature of their knowledge, which is dependent on the training data. Tasks requiring current or specialized knowledge are made more difficult by this constraint. To close this disparity,

In particular, this thesis investigates the use of RAG and LLM models for personalized learning in educational environments that prioritize tailored instruction and flexible learning pathways. The limitations of conventional educational systems can be overcome by incorporating these AI-driven technologies, giving students access to a more flexible, personalized, and interactive learning environment. Essentially, personalized learning is adjusting the pace, approach, and substance of training to meet the individual requirements and preferences of each student. When LLMs and RAG are used together, the system can comprehend the demands of the learner and adjust in real-time to provide feedback, fresh knowledge, and questions for critical thought. This makes for an interactive learning process.

Personalized learning is based on the necessity of flexibility. Because different learners have different comprehension levels, learning styles, and rates of advancement, it is critical to have a flexible system that can change in real time to meet the demands of each unique user. Because of their extensive language comprehension skills, LLMs are excellent at offering this flexibility. LLMs can help learners learn by having adaptive discussions in which they provide factual answers to basic queries, in-depth explanations of difficult subjects, or even the introduction of new ideas. When studying Sri Lanka's history in the 1980s, for example, a student may begin by asking general questions about significant occurrences and then go further into topics like the country's political climate or economic shifts at that time. A very participatory and interesting learning

environment is made possible by the LLM's capacity to understand and answer such questions in a conversational manner.

But personalized learning uses LLMs for more than just conversation. The capacity of LLMs to give learners rapid feedback is one of its main advantages. This is particularly crucial in educational environments as pupils frequently need their misconceptions cleared up or corrected. In order to make sure the subject is understood, LLMs can spot areas where a learner might be having difficulty and provide thorough explanations or alternative explanations. Additionally, by emulating the Socratic technique of guided questioning, LLMs can motivate students to consider other viewpoints and reflect critically on their own responses. To encourage a deeper cognitive engagement with the material, the LLM could, for instance, ask the student to consider the reasons and effects of particular events rather than just providing a historical answer.

Because LLMs are trained on static data, they are intrinsically constrained even if they offer a strong framework for individualized learning. As a result, when it comes to current events or specialized issues, their expertise is limited to what they learned during their training, which may result in inaccurate or obsolete information. This restriction may make it impossible for the LLM to give precise information about certain historical events, people, or policies. Retrieval-Augmented Generation (RAG) models are useful in this situation.

RAG models provide a link between real-time information retrieval and LLMs. Through the integration of a retrieval mechanism that combs through databases or outside sources, RAG models enhance the generative process by adding highly relevant and current data. Because of this dual structure, the learning platform can deliver timely, factually accurate knowledge in addition to responses that are coherent and appropriate for the given situation. For instance, the LLM may provide a generic response based on its prior knowledge if a student asking about a particular political incident in Sri Lanka's history during the 80s asks about that event. In contrast, the RAG component enriches the learning process by retrieving recent documents, articles, or academic papers to offer a more accurate and fact-based response.

Enhancing the range and depth of information that an educational chatbot or system may offer requires the integration of RAG models. RAG models make sure the system can serve both broad learners and individuals looking for more specialized or up-to-date information by drawing from a dynamic knowledge base. This is especially helpful in subjects like history, where having access to original sources, research papers, and other academic materials may greatly improve the caliber of instruction. Furthermore, by directing students toward more resources, the retrieval element of RAG models can support reinforcement of learning by allowing them to delve further into a subject and carry on learning after their first engagement with the system.

The Socratic method is one of the most effective ways to apply LLMs and RAG models in individualized learning. A well-known instructional approach is the Socratic method of inquiry, which entails posing open-ended questions that promote introspection and more in-depth thought. The approach can support active learning, where students are more involved and take charge of their education, by encouraging them to consider their responses carefully. For example, the LLM may ask follow-up questions to encourage the student to delve deeper into the subject matter, instead than giving a direct response to a query from the student. This method promotes the growth of critical thinking abilities in addition to reiterating the students' comprehension of the material.

Beyond technological considerations, LLMs and RAG models are being implemented for individualized learning. The user experience must be prioritized in order to guarantee the efficacy of these solutions. With the use of these technologies, an interactive chatbot or virtual tutor may be created that allows students to easily ask questions, get helpful answers, and participate in insightful conversations. Furthermore, learning routes should be dynamically adjusted by the system based on the student's performance and development, which will allow it to modify the level of difficulty, recommend new subjects, or provide remediation as needed. The provision of an effective learning experience that is customized for each individual depends on this adaptability.

Ultimately, even if the application of RAG models and LLMs has great potential for individualized learning, there are certain issues that need to be resolved. For example, it's crucial to make sure the data retrieved is precise, dependable, and suitable for the student's needs. This necessitates the meticulous selection of outside information sources and the creation of algorithms capable of determining the reliability of content retrieval. Furthermore, both RAG and LLM models have large computational requirements, thus preserving high-quality outputs while maximizing efficiency is essential.

In conclusion, the integration of LLMs and RAG models into personalized learning platforms represents a significant step forward in the evolution of education. By combining the language understanding and generation capabilities of LLMs with the real-time retrieval capabilities of RAG models, educators can create adaptive, engaging, and highly effective learning environments. The addition of the Socratic method further enhances these platforms by encouraging critical thinking and active engagement. As AI continues to evolve, the potential for creating even more personalized, responsive, and impactful learning experiences will grow, ultimately transforming how we learn and interact with educational content.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Nuwan Kodagoda, for his guidance, support and encouragement throughout the entire thesis process. Without his guidance this thesis would not have been possible.

I would also like to thank the members of the thesis committee (Dr. Kalpani Manathunga, Dr. Junius Anjana and Mr. Thusithanjana Thilakarthna) for their valuable feedback and suggestions. Their insights helped me to improve my thesis significantly.

In addition, I would like to thank my fellow graduate students in the Enterprise Applications Development program for their support and friendship. We helped each other to learn and grow, and I am grateful for their companionship.

Thank you to all of you for your support. I could not have done this without you.

Kavisha Dineth Samarasinghe

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENT	vi
TABLE OF CONTENTS	. vii
List of Figures	x
Chapter 1 Introduction	1
1.1 What is personalized learning?	1
1.2 Using Large Language Models for Personalized Learning	2
1.3 The Role of Large Language Models in Personalized Learning	3
1.4 What is Socratic learning?	5
1.4.1 The Socratic Circle	5
1.4.2 The Funnel of Inquiry	5
1.4.3 The Socratic Process Flowchart	6
1.5 Thesis statement	7
Chapter 2 Literature Review	9
2.1 Personalized learning	9
2.1.1 Key Aspects of Personalized Learning and Technology	. 11
2.1.2 Use Cases of Personalized Learning Technology	. 12
2.2 NLP techniques in education	. 12
2.2.1 The Role of NLP in Education	. 14
2.2.2 Synthesis and Implications	. 15
2.3 LLM and RAG Model	. 16
2.3.1 HistoryQuest: Arabic Question Answering in Egyptian History with LLM Fine-Tuning and Transformer Models	16
2.3.2 Document Knowledge Graph to Enhance Question Answering with Retrieval Augmented	
Generation	. 17
2.3.3 An Automated System for Question Generation and Answer Evaluation	. 18
2.3.4 Science Exam Question Answering based on Retrieval-Augmented Generation	. 19
2.3.5 Retrieval Augmented Generation (RAG) Based Restaurant Chatbot with AI Testability	20
2.3.6 A Comparative Analysis of Large Language Models with Retrieval-Augmented Generation based Question Answering System	21
2.3.7 A Multi-Source Retrieval Question Answering Framework Based on RAG	22
2.3.8 Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Seman Search and Hybrid Query-Based Retrievers	ntic 23
2.3.9 Enhancing Supply Chain Efficiency Through Retrieve-Augmented Generation Approach in Large Language Models	24
2.3.10 Ask-EDA: A Design Assistant Empowered by LLM, Hybrid RAG and Abbreviation De- hallucination	25

2.3.11 Comparative Analysis of RAG, Fine-Tuning, and Prompt Engineering in Chatbot Development	26
2.3.12 Enhancing the RAG Retrieval Engine Through Multi-Encoder Fusion	27
2.3.13 Investigating on the External Knowledge in RAG for Zero-Shot Cross-Language Transfe	r 27
2.3.14 RAG: An efficient reliability analysis of logic circuits on graphics processing units	28
2.3.15 Large Language Models in Education: A Systematic Review	29
2.3.16 Unlocking the Potential of Generative AI in Large Language Models	30
2.3.17 The Language Model Revolution: LLM and SLM Analysis	31
2.3.18 Exploring the Application of Large Language Models in Spoken Language Understandin Tasks	g 32
2.3.19 Exploring the Application of Large Language Models in Spoken Language Understandin Tasks	g 33
2.3.20 Integrated Method of Deep Learning and Large Language Model in Speech Recognition.	34
2.3.21 Structuring Natural Language Requirements with Large Language Models	35
2.3.22 Integrated Method of Deep Learning and Large Language Model in Speech Recognition.	36
2.3.23 Research on the Architecture Design Based on Large Language Model	37
2.3.24 The implementation solution for automatic visualization of tabular data in relational datal based on large language models	bases 38
2.3.25 A Large Language Model-based Approach for automatically Optimizing BIM	39
Chapter 3 Research Methodology	40
3.1 Research design	40
3.1.1 High-Level Design	40
3.2 Technologies Used	43
3.2.1 Python	43
3.2.2 Gradio	43
3.2.3 Huggingface	47
3.3 Implementation	49
3.3.1 Contextual Setup	49
3.3.2 Question Generation with Hugging Face Models	50
3.3.3 Reference Answers and Hints	53
3.3.4 User Interaction Flow	53
3.3.5 Verify Answer	54
3.3.6 Feedback and Next Steps	56
3.3.7 Summary of used Models used.	56
3.4 Data collection methods	57
3.4.1 Populating the Knowledge Bank	57
3.4.2 User Competency Level Determination	58
3.5 Ethical considerations	59

3.5.1 Data Privacy and Security	59
3.5.2 Algorithmic Transparency and Fairness	60
3.5.3 Informed Consent and User Autonomy	60
3.5.4 Inclusivity and Accessibility	61
Chapter 4 Results and Discussion	62
4.1 Implementation Outputs	62
4.2 Quantitative Results – User Evaluation	65
4.2.1 User Contentment	65
4.2.2 Learning Improvement	65
4.2.3 Average Answer Accuracy After Hints	66
4.2.4 Time Spent Engaging Users	66
4.2.5 Learning Effectiveness Scores	67
Chapter 5 Interpretation of the results	68
5.1 Understanding User Engagement and Learning Outcomes	68
5.2 Evaluating the Questions Generation Engine	68
5.2.1 User/Student Management and Knowledge Bank Effectiveness	70
5.2.1 Assessing Ethical Compliance and Data Privacy	70
5.2.1 Long-term Impact and Scalability	70
5.3 Limitations of the study	71
Chapter 6 Conclusion	75
6.1 Summary of the main findings	75
6.2 Restatement of the thesis statement	77
6.3 Concluding remarks	79
Chapter 7 References	80
Chapter 8 Appendix	85
8.1 Project File Structure	85
8.2 Code Details	86
8.2.1 chatbot_main.py	86
8.2.2 question_answers_engine.py	87
8.2.3 verify_answer.py	88
8.2.4 context.txt.	88
8.3 Key Technical Challenges and Solutions	89
8.4 Future Extensions	89

List of Figures

Figure 3.1.1 : High-level Diagram	40
Figure 4.1.1 : Initilization of Question and Answering Dialouge	62
Figure 4.1.2 : Providing Correct answer but context is in different order.	63
Figure 4.1.3 : Counter-Questioning for wrong answers with hint	64
Figure 7.1.1.1 : Gradio Interface Setup:	86
Figure 7.1.1.2 : Response Function	86
Figure 7.1.2.1 : Question Generatio	87
Figure 7.1.2.2: Hint Generation with Pegasus	87
Figure 7.1.3.1: Semantic Similarity with Sentence Transformers	88
Figure 7.1.3.2 : Fuzzy Matching and Named Entity Recognition	88