**Project Overview**

Middle school mathematics can be tough for lots of students, and sometimes the reason is inherent in each individual's uniqueness in the way they learn or the specific portions they have trouble with. For the classrooms in this genre, traditional teaching sometimes fall short when it comes to accommodating the individual needs of the students, which in turn causes unequal chances in learning. Utilizing advancements in Large Language Models, we can develop a personalized tutoring system that is responsive to the strengths and weaknesses of each learner during the tutoring process.

Proposed project intends to create an LLM-based platform for delivering math content that includes lessons, practice problems, and instant feedback, all within a conversational environment that is engaging and interactive. Besides, the students may put forward follow-up questions on any skill they wants to deepen his or her understanding, and the computer will reduce the questions’ difficulty level depending on the student’s performance. Aside from being an automotive technology, LLMs have advanced language understanding capabilities which are useful in-molding the explanations more comprehensive, locating the specific areas of the problem, and experiencing the personalized tutoring that would be beneficial for the student-study process.

**Project Objectives and Contributions**

The goal is an interactive tutoring system that greatly adjusts maths content for middle school level towards each learner, which is conversational and inspires them to stay engaged with the learning material. The superior, compared to the existing automated tutoring systems in the same field, will be the one using the highly-tuned LLM model to perform real-time feedback loops where it identifies the dreaded subtle misconceptions and targets them with very specific explanations or even exercises.

The project is distinguished by its use of LLMs for language generation to create a finer dialogue between student and educator that boosts the learning process. Our method aspires to surpass standard multiple-choice or fill-in-the-blank questions as it generates hints and step-by-step solutions for personalized learning, which comes from comprehending the math topics, instead of memorizing procedures.

**Literature Review**

Over the years, automated tutoring systems have been deeply studied, with research studies listing early applications on rule matching or pattern methods (Cognitive Tutor system). Though these methods demonstrate some positive outcomes, they need large and complex knowledge bases or fixed scripts, which may take time to get updated and may also offer less flexibility if students spread out from expected paths. (Adaptive Learning in Mathematics) To take a recent example, language models based on neural networks are among the newest and most innovative advances.

A very big one was the progress that was made when the researchers applied transformer architectures like the BERT and GPT, which led to the beginning of the state-of-the-art LLMs, which we now have. In this respect, educational studies showed that LLMs were suitable for question generation, answer assessment, and personalized pathways creation. (GPT-3 and Few-Shot Learning for Math) Still, most of the state-of-the-art tutoring systems fall short of being able to hold a robust and contextually aware dialogue, which makes the human teacher an irreplaceable asset in education.

The literature regarding adaptive learning mostly emphasizes the customization of instructional techniques based on different learner outcomes, where either item response theory or Bayesian approaches are employed as tracking mechanisms. (Adaptive Learning in Mathematics) The system can become more adaptive by integrating the strategies already found and adapting conversation strategies to LLMs. Moreover, the ethical aspects of AI-based education and the risk of incorporating bias in the content have also been discussed in some of the recent papers. This being done, we will ensure that there are measures to maintain the accuracy, accessibility, and inclusivity of our online tutoring platform.

Finally, there exists a range of open-source student problem-solving logs and mathematics question banks, which could also be used for further training and evaluation of more complex AI tutors and models. Although these models are available, they are not commonly used together with the power of advanced LLMs. Hence, we are at the fine crossroad being served on personalized tutoring, cutting edge NLP techniques, and practical applications in the context of education.

**Methodology**

This approach uses the model I specifically trained on textbooks for the middle school math curriculum, along with an adaptation of feedback, too. The first stage of getting those materials that we have to have will be with sourcing textbooks, exercises, and question and answer sets from authentic sources. Through the objective of tuning such a pre-trained transformer model to the data, which are shorter expository texts, stepwise solutions of problems, and adaptive follow-up questions, will be addressed.

We will then build a system that watches over the deal and helps monitor the good old trend of improvement of students as time goes by. The platform is structured to keep record of the sessions each time the student is in touch with it, such as asking a problem or giving an answer to a problem. This will then be data fed into a personalized dilemma difficulty regulator, which is specifically meant to help the system to pick or even create problems that will be parallel to the level of the student.

Therefore, this experiment will most probably disprove our premise, and as such we will run it as a pilot with either the middle school students or willing participants. The performance indicators will be tracked, namely the pre- and post-tutorial test results; there will also be a specific input from the users (the assessment of the given explanations), and hence these talks will determine the final shaping of the algorithm.

**Technical Considerations**

For the tutoring platform, we plan to fine-tune an existing converter-based LLM, possibly a GPT-style model, based on a well-organized dataset of middle school math problems and detailed solutions. We will look for a balance between utilizing the powerful pre-trained models and focusing their capabilities on specific areas of math education. Considering our own training pipeline, we will explore open-source models like GPT-Neo or DeepSeek-R1-Distill-Qwen-32B, but we may also consider commercial API like GPT-4. for high level reasoning capabilities. The cost and the model performance are the key factors in the decision.

For resource considerations, fine tuning the models like GPT will require a dedicated GPU environment to handle large-scale parametric training. We are planning to conduct a targeted fine tuning phase, updating only a subset of the model parameters like LORA, to keep the computational budget at an acceptable level. In this way, we can reduce training time as well as GPU requirements while keeping the large-scale language capabilities which is necessary for high-quality tutoring. We have 4090 GPU for local training, and we have verified that we are able to run LLM up to 32B locally. Otherwise, if a commercial API is used, the main consideration be the cost of use, ensuring that our budget can sustain the required number of queries throughout the semester.

**Timeline and Milestone**

1. **Week 1:** Finalize project requirements, determine math content scope, and collect any necessary data sets.
2. **Week 2–3:** Establish initial data processing pipeline; Start fine-tuning LLM with a minimal test set.
3. **Week 4:** Implement the adaptive assessment engine and integrate it with the reasoning capabilities of the LLM.
4. **Week 5:** Conduct Pilot test with a small group of users, collect feedback, and refine the model based on the results.
5. **Week 6:** Perform scale-out assessments, address edge cases, and complete final usability improvements.
6. **Week 7:** Compile the results, complete the documentation, prepare the final presentation.

**Expected Outcomes and Evaluation**

We expect to have a prototype of an AI-driven tutoring platform capable of interactive conversations, personalized questions, and customized feedback loops. It will be a website or mobile app where middle school students can solve problems, ask questions, and get instant, context-aware explanations.

Success will be measured by various methods. Quantitatively, we will track user performance improvements, error rates, and usage metrics such as average session duration. Qualitatively, we will collect feedback from students and professors to measure overall user satisfaction. These assessments will help us determine if the platform is truly improving learning achievements compared to traditional resources.

**Potential Challenges and Mitigation Strategies**

We must ensure that the LLM remains accurate in its responses. Math logical errors are not acceptable, even subtle mistakes can mislead students. We will incorporate a set of validation checks and fallback logic, such as verifying correct solutions with external math libraries.

Another challenge involves inappropriate content generation. We can mitigate this problem by using an established content moderation framework to filter user input and model output.

**Conclusion**

In summary, our proposal outlines an LLM-powered tutoring platform designed to improve middle school math education by providing personalized, context-aware instruction. By harnessing the adaptive capabilities of language models, it aims to address individual student needs in ways traditional methods cannot. Throughout the semester, our primary goal is to design and validate a prototype that demonstrates tangible academic gains and a user-friendly experience.

We are excited about the potential impact on learners’ outcomes and the broader educational technology community. This project aligns strongly with our academic and professional goals of using AI to enhance student's math learning experiences. We look forward to advancing the state of personalized math instruction and sharing our findings with peers and educators alike.

**Reference**

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30–43.

Doignon, J. P., & Falmagne, J. C. (1999). Knowledge Spaces. Springer.

Hendrycks, D., Burns, C., Kadavath, S., Arora, A., Basart, S., Tang, E., … Song, D. (2021). Measuring Mathematical Problem Solving With the MATH Dataset. NeurIPS.

Ling, W., Yogatama, D., Dyer, C., & Blunsom, P. (2017). Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems. ACL.

Saxton, D., Grefenstette, E., Hill, F., & Kohli, P. (2019). Analysing Mathematical Reasoning Abilities of Neural Models. ICLR

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language Models are Few-Shot Learners. NeurIPS.

Liang, C., Norouzi, M., Berant, J., Le, Q., & Lao, N. (2019). Neural Symbolic Machines: Learning Semantic Parsers on Freebase With Weak Supervision. ACL.