ACTA DIDACTICA NAPOCENSIA
Volume 18, Number 1, 2025 - DOI: 10.24193/adn.18.1.12

QUIZ GENERATOR USING LARGE LANGUAGE MODELS. PROMPT ENGINEERING EXAMPLE

Nicolae FLORIAN

Abstract: In this paper we will explore an application, made by us, that can generate physics grid tests using artificial intelligence. The application analyzes the response from the large language model in the required format recognized by the application and writes it to a file that will be accepted by the test builder application. The application creates the files required for a test by parsing the prompt from the user in a request to an Application Programming Interface. Tests generated by the application can be used in student assessment at physics discipline, by level groups (easy, medium and hard).

Key words: assessment tests, tests generator, artificial intelligence, large language models

1. Introduction

In recent years, the integration of artificial intelligence (AI) into education – particularly in mathematics and science instruction – has garnered increasing attention from researchers and educators alike, thus numerous studies have been published on the use of AI in the teaching and learning of mathematics and science, as well as in the development of assessment tests. Studies such as those by Sato (2024) and Yeadon & Hardy (2023) highlight how large language models (LLMs) can offer new ways to explain complex physics concepts and personalize instruction for students. Research by Chang et al. (2024) and Kasneci et al. (2023) further underscores the rapid evolution of these models, emphasizing their capacity to support both formative and summative assessments. Al-Kamzari & Alias (2025) explores how AI-driven tools can optimize the design of educational technology systems, while Holmes et al. (2023) demonstrate the precision with which AI can operate in specialized domains like radiation oncology physics. As these technologies mature, their application in tailored student evaluation – as discussed by Zhang et al. (2024) and Jiang & Jiang (2024) – opens promising paths for adaptive testing and deep learning experiences.

The development of an application that leverages artificial intelligence to generate physics grid tests represents a significant advancement in educational technology. By utilizing a LLM through an Application Programming Interface (API) (Birhane et al. 2023; Bitzenbauer, 2023; LM Studio Documentation, 2024; Yeadon & Hardy, 2003), the application can interpret user prompts and produce the necessary files for a physics test. This process involves the application sending a request to the LLM, which then returns a response in a specific format that the application can recognize and process. In the context of generating physics tests, the accuracy would depend on several factors, including the quality of the input prompts, the sophistication of the LLM, and the parameters set for the test generation. For instance, if the application is designed to cover a wide range of topics and complexities within physics, and the LLM is fine-tuned with a comprehensive understanding of the subject matter, the resulting tests are likely to be accurate reflections of the intended assessment criteria.

Once the LLM has generated the content based on the user's prompt, the application evaluates the response to ensure it meets the predefined criteria for the test builder application. If the response is deemed suitable, the application proceeds to write it to a file. This file is structured in such a way that it is compatible with the test builder application, ensuring a seamless integration into the test generation process.

Received July 2025.

Cite as: Florian, N. (2025). Quiz generator using large language models. Prompt engineering example. *Acta Didactica Napocensia,* 18(1), 141-147, https://doi.org/10.24193/adn.18.1.12

This innovative approach not only streamlines the creation of physics tests but also introduces a level of customization and adaptability previously unattainable with traditional methods. Educators can input specific parameters or topics they wish to assess, and the application will tailor the test content accordingly. This results in a more efficient use of educators' time and resources, allowing them to focus on other critical aspects of teaching.

Furthermore, the application's ability to generate tests using AI opens up possibilities for more dynamic and interactive assessments. For instance, the application could incorporate varying levels of difficulty based on the students' performance, creating a more personalized learning experience. Additionally, the use of AI in test generation can help identify and fill knowledge gaps, as the LLM can be programmed to cover a wide range of topics and complexities within the field of physics.

2. Development of the application for generation tests in Physics

This test method is probably the most discussed of the assessment strategy used in physics (Florian, 2004; Stoenescu & Constantinescu, 1999; Stoenescu & Florian, 2015). The proposed test generator offers the possibility of objective and differentiated assessment of students by level group.

The implications of such technology extend beyond the classroom as well. As AI continues to evolve, its integration into educational tools like this application could revolutionize the way we approach learning and assessment. The potential for AI to assist in creating more engaging, comprehensive, and challenging educational materials is vast, and this application is a step towards realizing that potential.

In the context of generating physics tests, the accuracy would depend on several factors, including the quality of the input prompts, the sophistication of the LLM, and the parameters set for the test generation. For instance, if the application is designed to cover a wide range of topics and complexities within physics, and the LLM is fine-tuned with a comprehensive understanding of the subject matter, the resulting tests are likely to be accurate reflections of the intended assessment criteria.

In this work, for the use of the test generator in evaluating students, files with school programs (in PDF format) can be introduced into LLMs, as well as digital school textbooks. Depending on the targeted curriculum and its related content, tests can be differentiated, by chapters, by introducing the level of difficulty. At the same time, summative tests can be created. In our study, we focused on tests that provide an objective assessment.

For this application to work it needs an LM Studio or Ollama server that can run any modern computer. LM Studio has a friendly user interface, a variety of models available, can tell if the model can be loaded on GPU and most importantly it can open a server so other applications can use the models. The server can be accessed over the LAN, so internet connection is needed to run. On the other hand, Ollama runs in the terminal/command line and it opens automatically the server. For this use case, LM Studio was used (LM Studio Documentation, 2024).

The AI server configuration for starting the application is shown in Figure 1.

This project was written in python and generates the prompt based on user’s input. The prompt was optimized for the best results, so the user interface is comprehensive and easy to use. The application was built to g Generate a custom quiz using a local LLM endpoint.

The Quiz Generator application can be used to generate the quiz, differentiated by level groups (easy, medium and hard), and for this the steps shown in Figure 2 must be followed (Step 1. Write the topic of interest; Step 2. Select a level of difficulty; Step 3. Select the number of questions and click Submit).

For this application to work it needs an LM Studio or Ollama server that can run any modern computer LM Studio.

This application was written in Python programming language and generates the prompt based on user’s input. To ensure a user-friendly interface we are going to use the Gradio library (Gradio Documentation, 2024). The prompt was optimized for the best results, so the user interface is comprehensive and easy to use.

The application porpuse in this study was built to Generate a custom quiz using a local LLM endpoint. The model used in this project was Phi3 from Microsoft (Microsoft Azure, 2024).

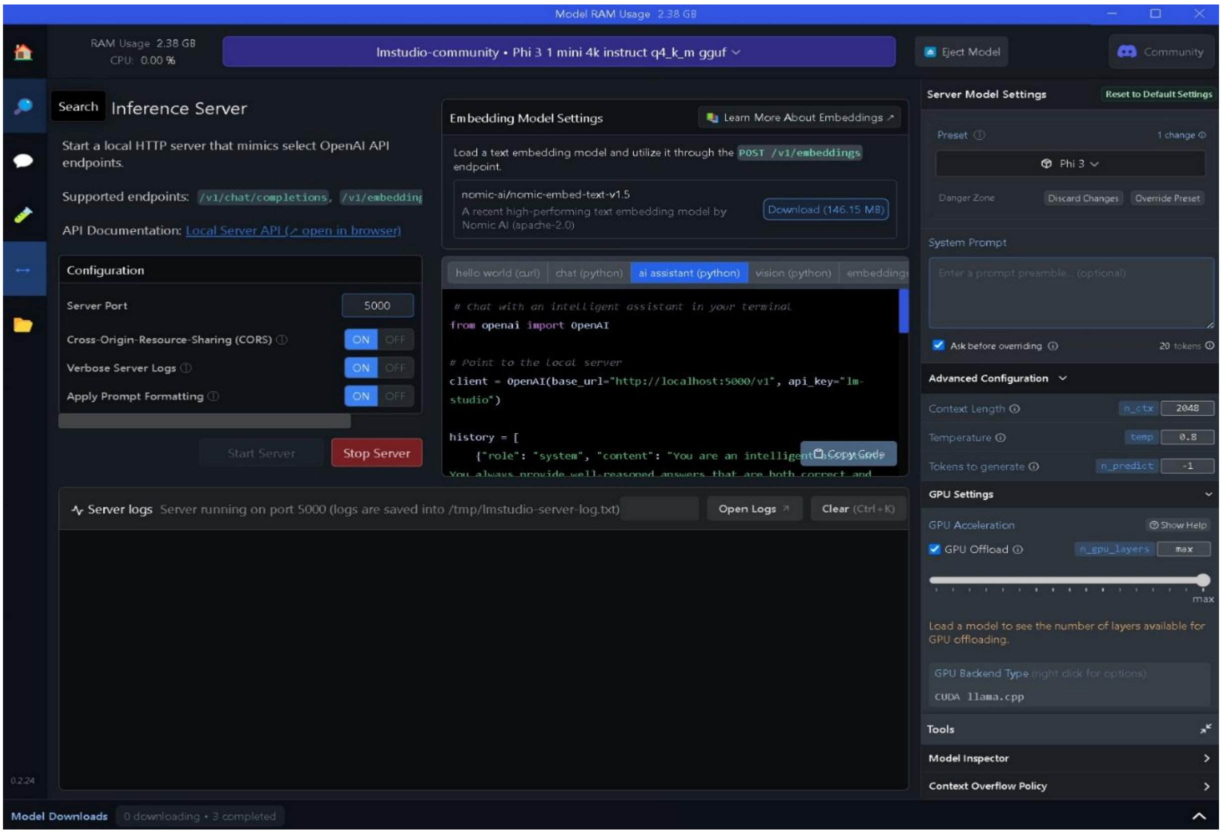

Figure 1. *The quiz generator interface allows you to configure the quiz according to the physics topic, diffusion level and number of questions*

Figure 2. *The AI server configuration*

This application is customizable, allowing the user to use smaller or bigger models, depending on the hardware resources.

Figure 3 shows the interface of the application containing the test generated for the light diffraction theme, medium level, with AI using LLM endpoint. It can be seen that the answers have also been generated. In order to apply the test to the students, the content of the items can be selected and then copied to an editable file (e.g. *.docx) to be printed and given to the students to solve.

The answers from the test generator can be used for assessment/self-assessment. It should be noted here that the duration of test generation varies depending on the device used from a few seconds to several minutes.

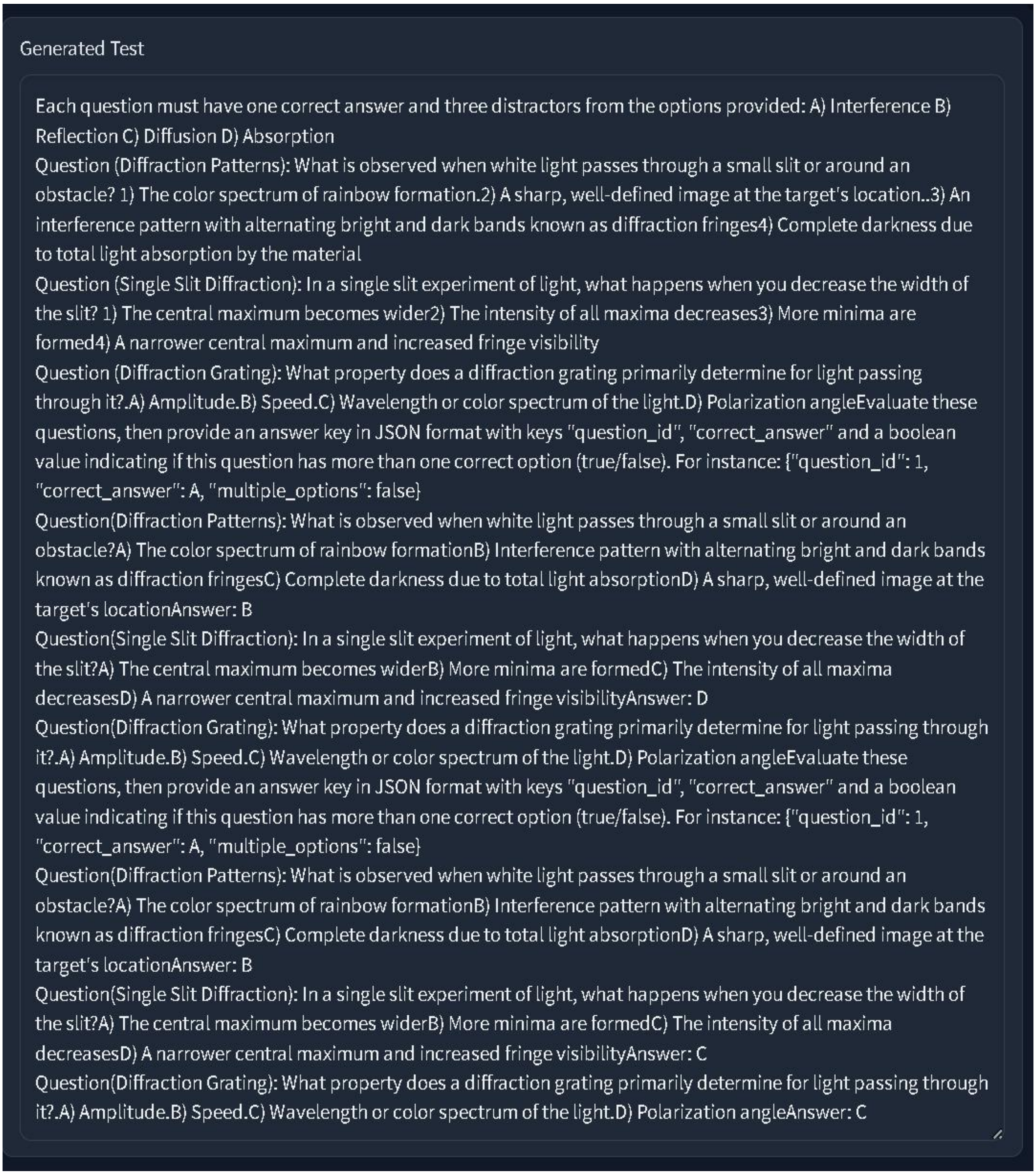

Figure 3. *Presentation of the AI generated quiz using the Quiz Generator application*

In this study 5 LLMs were tested: phi3, mistral, llama2, gema, qwen, each with its variants: from 1.4 billion to 8 billion parameters. For generating the tests, it is recommended to insert the school curriculum, file in PDF format. The user of the test generator can ask the application to take into account certain contents and competences from the loaded syllabus, as well as select the level of difficulty and the number of questions. Test validation is provided within the model used by the application. It is recommended that the model has a larger number of parameters so that the probability of the model to be wrong is as low as possible.

3. Pedagogical considerations on integrating the test generating application into education

The integration of AI into educational tools like test generation applications could lead to more dynamic and interactive assessments. By potentially incorporating varying levels of difficulty based on students' performance, AI-generated tests could offer a more personalized learning experience. This adaptability, combined with the AI's ability to cover a broad spectrum of physics topics, could result in tests that are not only accurate but also highly relevant and challenging for students.

Moreover, the application's ability to analyze the LLM's response and format it for the test builder application plays a crucial role in maintaining the accuracy of the tests. If the application can effectively interpret and evaluate the LLM's output against the required standards for the test, it can ensure that the generated files are of high quality and accurately represent the physics concepts intended to be tested.

To ensure the accuracy of AI-generated physics tests and mitigate potential inaccuracies, several safeguards can be implemented. These include rigorous validation and testing of the AI model, continuous learning and updating of the model with new data, and the incorporation of expert review in the test generation process.

Validation and testing involve subjecting the AI to a wide range of scenarios and comparing its outputs against established physics problems and their solutions. This helps in identifying any biases or errors in the model's understanding of physics concepts. Continuous learning is another crucial safeguard; as the AI is exposed to more data over time, it can refine its algorithms and improve its accuracy in generating test questions.

The validation of tests generated by LLMs is very important and involves several complementary steps. First, a didactic and curricular review conducted by education specialists is necessary to ensure the coherence of the items with the learning objectives. Subsequently, the tests need to be tried out on real samples of students to highlight any ambiguities or dysfunctions. Another important aspect is the cultural and linguistic adaptation of the content, so that the formulations are relevant and accessible in the targeted educational context. Ideally, this process can be complemented by a rigorous psychometric analysis, which evaluates the test's performance in terms of reliability, validity, and item discrimination capacity.

With the proposed test generator, tests can also be made for the assessment of students and according to their individual and group particularities, significantly facilitating their self-instruction, self-testing and self-assessment.

The use of LLM models in the generation of educational tests offers significant innovative potential due to their speed of generation, customizability and extensive thematic coverage. These tools can support teachers in developing items, scales and rationales, streamlining formative and summative assessment processes. However, in the absence of qualified pedagogical intervention, there is a risk of producing items that are not aligned with the curriculum, ambiguous or culturally inappropriate, and of standardizing assessment. Therefore, the integration of these technologies requires careful validation, contextual adaptation and critical filtering in order to maintain educational relevance and fairness of assessment.

4. Conclusion

In conclusion, the application described in this paper is a testament to the transformative power of AI in education. By automating the test generation process and providing a customizable platform for educators, it enhances the teaching and learning experience. As this technology continues to develop, it will undoubtedly uncover new avenues for innovation in educational assessment and beyond.

While the exact accuracy of AI-generated physics tests would ultimately depend on the specific application and its implementation, the current trends and research in AI and physics suggest a high potential for accuracy. As AI continues to advance and as more data becomes available to train these models, it's likely that the accuracy of such tests will only improve, offering valuable tools for educators and students alike.

References

Al Kamzari, F. & Alias, N. (2025). A systematic literature review of artificial intelligence (AI) in secondary school physics: applications, benefits, and challenges. *Interactive Learning Environments*, 1-18. https://doi. org/10.1080/10494820.2025.2508323.

Birhane, A., Kasirzadeh, A., Leslie, D. et al. (2023). Science in the age of large language models. *Nat Rev Phys* 5, 277–280. https://doi.org/10.1038/s42254-023-00581-4.

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology, 15*(3), ep430

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1-45.

Florian, G. (2004). *Tratarea diferențiată a elevilor la fizică*. Else.

Gradio Documentation. (2024). https://www.gradio.app/docs.

Holmes, J., Liu, Z., Zhang, L., Ding, Y., Sio, T. T., McGee, L. A., Ashman, J. B., Li, X., Liu, T., Shen, J., & Liu, W. (2023). Evaluating large language models on a highly-specialized topic, radiation oncology physics. *Frontiers in Oncology*, *13*, 1219326.

Jiang, Z. & Jiang, M. (2024). *Beyond answers: Large language model-powered tutoring system in physics education for deep learning and precise understanding*. arXiv preprint arXiv:2406.10934.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274.

LM Studio Documentation. (2024). https://lmstudio.ai/docs.

Microsoft Azure. (2024). *Introducing Phi-3: Redefining what's possible with SLMs*. https://azure.microsoft.com/en-us/blog/introducing-phi-3-redefining-whats-possible-with-slms.

Polverini, G. & Gregorcic, B. (2024). How understanding large language models can inform the use of chatgpt in physics education. *European Journal of Physics*, *45*(2), 025701.

Sato, K. (2024) Exploring the Educational Landscape of AI: Large Language Models' Approaches to Explaining Conservation of Momentum in Physics. arXiv preprint arXiv:2407.05308.

Shengcai, L., Caishun, C., Xinghua, Q., Ke, T., & Yew-Soon, O. (2023). *Large language models as evolutionary optimizers*. arXiv:2310.19046 (2023).

Stoenescu, G., & Constantinescu, R. (1999). *Teaching physics*. Sitech.

Stoenescu, G., & Florian, G. (2015). *Didactica fizicii.* Sitech.

Tian, Y., Zicheng, X., Yuanzhi, L. & Zeyuan, A.-Z. (2024). *Physics of Language Models: Part 2.1, Grade-School Math and the Hidden Reasoning Process*. arXiv preprint arXiv:2407.20311.

Zhang, H., Da, J., Lee, D., Robinson, V., Wu, C., Song, W., Zhao, T., Raja, P., Zhuang, C., Slack, D., Lyu, Q., Hendryx, S., Kaplan, R., Lunati, M., & Yue, S. (2024). *A careful examination of large language model performance on grade school arithmetic*. arXiv preprint arXiv:2405.00332, (2024).

Yeadon, W., Hardy, T (2023). *The impact of AI in physics education: A comprehensive review from gcse to university levels*. arXiv preprint arXiv:2309.05163 (2023).

Author

Nicolae FLORIAN, National University of Science and Technology Politehnica of Bucharest, Faculty of Automatic Control and Computers, Department of Systems Engineering, Independence Square, 060042, Bucharest, Romania. E-mail: florian_nick@windowslive.com

Acknowledgement

The author (**FN**) acknowledges that this work was carried out within the doctoral program at National University of Science and Technology Politehnica of Bucharest, Faculty of Automatic Control and Computers, Department of Systems Engineering, Independence Square, 060042, Bucharest, Romania.