PROCEEDINGS OF SCIENCE

Using Local AI with RAG for Open Educational Materials Generation for Physics Education

Dmitriy Beznosko [a,*], **Tatiana Krivosheev** [a] **and Alexander Iakovlev** [b]

a College of STEM, School of Sciences, Clayton State University,
2000 Clayton State Blvd, Morrow, GA 30260 USA,
b Upper School Science Department, Woodward Academy,
1662 Rugby Avenue, College Park, GA 30337
E-mail: dmitriybeznosko@clayton.edu

AI nowadays is a new tool that seems to be everywhere, including education. It brings convenience but also concerns - the two largest concerns with using AI by instructors are the safety of students' data and the AI lack of the specific knowledge needed in a specific class. The data safety can be addressed by using a locally run large language models model using Ollama framework, a free tool that gives user the ability to run LLM locally on your system, so no data is transmitted. The approach of using the Retrieval Augmented Generation adds the user data to the as a context for the generation of the educational materials as well as homework and quiz assignments. Brief background, installation and configuration highlights with performance and use examples are presented in this work.

The European Physical Society Conference on High Energy Physics (EPS-HEP2025)
7-11 July 2025
Marseille, France

*Speaker

ISSN 1824-8039 . Published by SISSA Medialab.

PoS(EPS-HEP2025)618

1. Introduction

As Large Language Models (LLM, commonly also called AI) are becoming integrated into all areas of life, education is to follow the trend. Students are using AI and it's now time for the instructors to use AI to make their life easier. The two main concerns that exist for the AI use by instructors are the lack of expertise in a specific topic and the concern for the privacy of the data.

A general LLM is designed to perform well in multiple areas from translation of text, writing code to reasoning and helping with cooking recipes. However, instructors often need a model that can help with specific tasks, such as generation of quiz and homework questions, training sets for students on a given chapter or topic, creation of grading rubrics and similar.

2. AI Framework on a Local PC

There are several frameworks available to run the AI models locally, with the Ollama [1] framework being the most used by the home user. The details of the installation of this framework and running a chosen model are detailed in [2]. Running model locally removes privacy concerns.

2.1 Using RAG with User Data

RAG is short for the Retrieval Augmented Generation. It is an approach to enhancing a small model in the area needed by providing it with user materials without retraining the model itself. The details on the installation and use of 'the Golden RAGtreiver' Verba that is available for download from https://github.com/weaviate/ is detailed in [2]. Here, being limited in number of pages, we will focus on using Verba for the generation of course materials.

First, the process of adding user context should be completed. For the test, a chapter on Newton's Laws was provided to the local model using the Vebra web interface and has been successfully imported as shown in Figure 1.

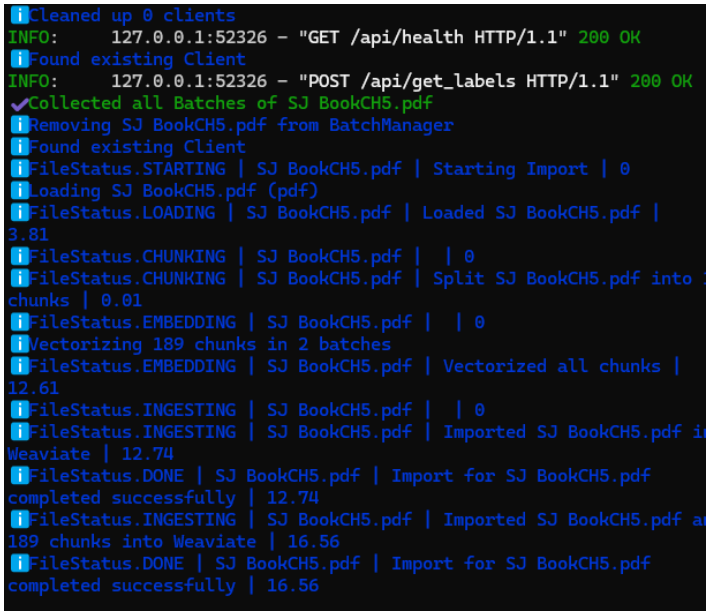

Figure 1. A process of importing the textbook chapter by Verba.

The following prompt was then issued in the Verba chat window: "Write 3 problems with solutions on the topic of Newton's laws for the college-level physics quiz with biological content. Format output as LaTeX".

There are few caveats what should be noted here. First, different models may understand a statement 'college level' differently, so the user may want to be more specific here, using phrasing such as 'calculus-based' or 'algebra-based' questions. Also, different models have varying success with LaTeX output, as a rule, larger models do better, and the models that were marked for writing code handle LaTeX output better – thus, a bit of manual clean-up may be needed. The output in Figure 2 is generated using Overleaf.com site.

Newton's Laws in Biological Context: Quiz Questions

Sample quiz with solutions

Questions and Solutions

Question 1. A cheetah accelerates from rest to a speed of 30 m/s in 3 s. If the cheetah has a mass of 50 kg, what is the average net force exerted by its muscles during this acceleration?

Solution:

Using Newton's Second Law:

$$F = ma$$

Acceleration:

$$a = \frac{\Delta v}{\Delta t} = \frac{30}{3} = 10\,\mathrm{m/s^2}$$

Force:

$$F = 50 \times 10 = 500\,\mathrm{N}$$

Question 2. A human sprinter pushes against the ground with a force of 800 N at the start of a race. If the sprinter's mass is 70 kg, what is their initial acceleration?

Solution:

Using Newton's Second Law:

$$a = \frac{F}{m} = \frac{800}{70} \approx 11.43\,\mathrm{m/s^2}$$

Question 3. A frog jumps vertically upward with an initial velocity of 5 m/s. What is the net force acting on the frog at the peak of its jump?

Solution:

At the peak of the jump, velocity is zero, but gravity still acts:

$$F_{\mathrm{net}} = -mg$$

Assuming mass $m = 0.5\,\mathrm{kg}$:

$$F_{\mathrm{net}} = -0.5 \times 9.8 = -4.9\,\mathrm{N}$$

(The negative sign indicates downward direction.)

Figure 2. A sample LaTeX formatted quiz generated using Verba.

Additionally, note that some models (granite3.2-vision:2b, llama3.2-vision:11b, etc.) can recognize images and handwriting, thus providing valuable help in grading hand-written solutions and similar submissions.

3. Conclusion

With the latest technological advances, an LLM can be used on personal computers. The key characteristics are the amount of video (or shared with system) memory available and the discrete video card. These models have small database size and don't have much information initially. The process known as RAG enhances LLM responses by providing the context from the user-supplied files.

LLM can provide teachers with help on various course materials, such as quizzes and study guides, in recognizing handwriting and summarizing students' responses; it is useful in the labs as well [3] for recognizing hand drawn plots and other graphing responses. All this is possible because the smaller and distilled LLM models can secure and privately run on a local PC, eliminating the worry about the privacy of the data and materials.

4. Acknowledgement

Work is sponsored by ALG (Affordable Learning Georgia) round 26 grant RG017.

5. References

[1] Ollama Inc, 435 Hamilton Ave, Palo Alto, CA 94301, US

[2] Dmitriy Beznosko, Tatiana Krivosheev, Alexander Iakovlev. Approach to Physics Education Using Local AI with RAG for Open Educational Materials Generation. In proceedings of 39th International Cosmic Ray Conference, PoS(ICRC2025)1238, 2025, DOI: https://doi.org/10.22323/1.501.1238

[3] Dmitriy Beznosko, Tatiana Krivosheev, Alexander Iakovlev. *Ultimate makeover for an introductory physics lab*. In proceedings of 38th International Cosmic Ray Conference **444** 1597, 2023/7/25, PoS(ICRC2023)1597, DOI: https://doi.org/10.22323/1.444.1597