Conception and development of a teaching assistant using an open-source LLM augmented with specific education resources

Keywords

LLM, RAG, knowledge graph, teacher assistant, learning analytics

Context

Generative AIs (GAI) strongly influence the education context, both from the teacher and the learner perspectives. GAI assistants become a *de facto* standard for accessing resources on Internet, getting help on a specific topic, or even to solve problems automatically. It has been shown, that LLM-based chatbots can be efficiently used to stand in for the teacher, when he/she is not available, to answer questions posted by learners [Hicke2023, Liu2024]. To provide learners with course-specific support, Retrieval Augmented Generation (RAG) strategies are particularly efficient.

Subject

The main objective of the project is to go a step further towards the definition of a Teacher Assistant (TA) [Qadir2023, Kasneci2023]. The goal of the project is to conceive and implement an LLM-based TA that also integrates knowledge about the course it is deployed for. Examples of education knowledge are for instance to consider the structure of the course, prerequisite relations between notions, as well as meta-data about the resources leveraged by the RAG strategy. The internship is thus organized in three phases:

- Modeling of the education context as a knowledge graph,
- Conception of a TA relying on an open-source LLM and a RAG technology that leverages the knowledge graph,
- Implementation of the solution as a chatbot in an online course.

Duration and supervisors

The internship will last from 4 to 6 months and a continuation with a PhD is expected. The trainee will physically work at the IMT Atlantique – Brest (Lab-STICC laboratory) and will be supervised by the following team:

- Grégory SMITS (IMT Atlantique, Lab-STICC)
- Maria-Teresa SEGARRA MONTESINOS (IMT Atlantique, Lab-STICC)
- Jean-Marie GILLIOT (IMT Atlantique, Lab-STICC)

Month salary is around 600€.

Required skills

We are looking for a motivated student currently completing a Master or an engineering cursus in computational sciences and the following skills are required:

- communication: to collaborate and cooperate with the different members of the projects. Scientific reports will have to be written and a scientific publication in a learning analytics conference is targeted.
- Scientific background in data science: build a knowledge graph, manage vector DB, LLM (Llama), RAG.

To apply, send your CV and cover letter to Grégory Smits (gregory.smits@imt-atlantique.fr) **before the 30th of November 2024**.

References

[Hicke2023] Hicke, Y., Agarwal, A., Ma, Q., & Denny, P. (2023). ChaTA: Towards an Intelligent Question-Answer Teaching Assistant using Open-Source LLMs. *arXiv preprint arXiv:2311.02775*.

[Liu2024] Liu, C., Hoang, L., Stolman, A., & Wu, B. (2024, July). HiTA: A RAG-Based Educational Platform that Centers Educators in the Instructional Loop. In *International Conference on Artificial Intelligence in Education* (pp. 405-412). Cham: Springer Nature Switzerland.

[Qadir2023] Qadir, J. (2023, May). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE.

[Kasneci2023] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.