Intelligent Educational Agent for Education Support Using Long Language Models Through Langchain

Chapter · February 2024

DOI: 10.1007/978-3-031-54235-0_24

7 authors, including:

Diego Quisi
Universidad Politécnica Salesiana (UPS)
22 PUBLICATIONS **81** CITATIONS
SEE PROFILE

Juan Pablo Salgado-Guerrero
45 PUBLICATIONS **58** CITATIONS
SEE PROFILE

Tracy Cárdenas
Universidad Politécnica Salesiana (UPS)
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Jorge Galán-Mena
University of Vigo
17 PUBLICATIONS **38** CITATIONS
SEE PROFILE

INTELLIGENT EDUCATIONAL AGENT FOR EDUCATION SUPPORT USING LONG LANGUAGE MODELS THROUGH LANGCHAIN.

Diego Quisi-Peralta[2], Juan Salgado-Guerrero[1], Pedro Neira-Maldonado[1], Jordan Murillo-Valarezo[1], Tracy Cárdenas-Arichábala[1], Jorge Galan-Mena[1], Daniel Pulla-Sanchez[1]

[1] Grupo de Investigación en Transformación Digital y Ciencia de Datos,
Universidad Politécnica Salesiana,
Cuenca, 010102, Azuay, Ecuador

[2] atlanTTic Research Centre for Information and Communication Technologies, Department of Telematics Engineering, University of Vigo, 36310 Vigo, Spain

diego.quisi.peralta@uvigo.es,
jsalgado@ups.edu.ec,pneiram1@est.ups.edu.ec,
jmurillov1@outlook.com, tcardenas@ups.edu.ec, jgalanm@ups.edu.ec
dpulla@ups.edu.ec

Abstract. This paper explores the development of an Intelligent Educational Agent (IEA) with a focus on enhancing the learning experience for university students. In an era where online education is on the rise, there is a growing demand for personalized learning tools. IEAs, powered by artificial intelligence, offer a solution by providing tailored support, explanations, answers to queries, and content adaptation. This study leverages advanced AI technologies, including the LangChain framework and the GPT-3.5 Turbo model from OpenAI, to create an adaptive educational assistant. LangChain facilitates Natural Language Processing and information analysis, while GPT-3.5 Turbo ensures context-aware responses through prompt-tuning. The research methodology involves defining functional requirements, implementing the LangChain framework for NLP, integrating the OpenAI API, and establishing an architecture with three main actors: students, teachers, and tutors/assistants. Results indicate the IEA's ability to generate precise multiple-choice tests and comprehensive academic plans. The system exhibits contextual understanding and resource generation capabilities. In conclusion, despite challenges like data quality and infrastructure requirements, developing an IEA for content adaptation based on large language models shows great promise. It has the potential to revolutionize education by providing personalized learning experiences and generating educational resources. Collaboration among education experts, developers, and researchers is crucial to fully harness this transformative potential.

Keywords: Agent, Education, Innovation, Large Language Models, Natural Language Processing.

1 Introduction

Online education is experiencing unprecedented growth, and with this transformation, significant challenges arise in meeting the needs of students. As more institutions and organizations transition to online education, students need tools that can adapt to their individual needs and learning styles, given the considerable breadth of content available on the internet. An intelligent educational support agent can provide a personalized learning experience that meets the individual needs of students and helps them achieve their educational goals [5]. In light of these considerations, this article aims to address the following key research questions: What role can intelligent educational support agents play in improving the online learning experience for students? How can the use of large language models, such as GPT-3, contribute to the development of intelligent educational agents for personalized support? What challenges and limitations are associated with implementing large language models in educational environments, particularly in resource-constrained settings?

In the field of education, advances in technology have opened up new possibilities for enhancing the learning experience. Intelligent educational agents (IEAs) are systems that use artificial intelligence techniques to provide support and assistance to students in their learning process. These agents can provide explanations, answer questions, offer feedback, and adapt educational content to the individual needs of students [6].

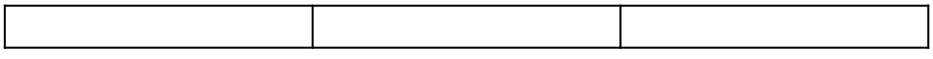
Another need that can be addressed with an intelligent educational support agent is the need for more effective interaction among students. The agent can help students navigate the platform and respond to their questions and concerns in a timely and effective manner [8] [7].

In this context, one of the promising approaches in the development of IEAs is the use of large language models, such as GPT-3. These models are capable of generating coherent and relevant text, making them suitable for educational content generation tasks. However, these models have a significant limitation: their size and computational complexity make it challenging to implement them on devices with limited resources or in environments with limited internet connectivity [9].

On the other hand, while Intelligent Educational Agents based on large language models have demonstrated their effectiveness in content generation, we still face difficulties in precise and personalized adaptation of such content. Among the problems is the lack of the ability to fully understand the specific preferences, skills, and learning styles of each student. Although large language models can process large amounts of data and generate relevant text, they often lack the ability to adapt content accurately and effectively to the individual requirements of each student [10].

2 Related Works

The development of Intelligent Educational Agents (IEAs) based on large language models has made significant advances in recent years. These advances have been



achieved thanks to progress in natural language processing, machine learning, and primarily generative artificial intelligence. Baidoo-Anu and Owusu[1] have researched the benefits and opportunities that can arise from the use of AI, considering the benefits in the teaching and learning process. The main tool currently being studied, due to its widespread use, has been ChatGPT.
In this context, the ability of large language models to generate high-quality educational content has been demonstrated. For example, systems have been developed that can generate questions and answers in various educational domains, providing students with additional study material and practice [11].

Furthermore, progress has been made in adapting and personalizing educational content. Approaches have been proposed that allow content to be tailored to individual learning styles, skills, and prior knowledge of students. This allows students to receive more tailored support for their specific needs, thereby enhancing their learning experience [12].
Another significant advancement has been achieved in improving conversational interaction between IEAs and students. They are capable of understanding and responding to complex questions and inquiries more accurately, providing students with a higher level of engagement and interaction during the educational process [13].
In this context, Kraus and Webersinke [2] have analyzed the use of Large Language Models (LLMs) and how integrating different sources of information can enable a chatbot to interpret contexts and be more accurate in responses [13].
Additionally, Xu and Ouyang[3], through a conceptual framework, propose ways to implement Artificial Intelligence in Education. It is important to note that, as it is still an emerging field, there is still a need for a deeper understanding of the roles AI can adopt in the educational process.
Finally, it is important to highlight that there are still challenges in the development of IEAs based on large language models. These challenges include aspects related to data privacy, ethics in the use of artificial intelligence in education, and the need for ongoing human supervision and feedback to ensure the quality of generated educational content [14].

3 Methodology

In the field of education, artificial intelligence (AI) has proven to be a promising tool for enhancing the learning experience of students. In this research, the focus is on implementing an academic assistant from the perspective of a university student, leveraging AI capabilities to process information. The functional requirements of the system were established, taking into account limitations in natural language processing and the ability to remember previously stored contexts. This was done based on an experimental methodology with the following steps:

3.1 Solution Proposals

In order to address these limitations, LangChain framework has been implemented, which allows Natural Language Processing and analysis of information established through documents and images, storing it in a vector database, in this case, Qdrant.
As the main search engine, the OpenAI API has been implemented, which allows the use of the GPT-3.5 Turbo model. Through prompt-tuning, the chatbot can understand the context of the information so that, following rules, an efficient response to user queries can be obtained. To implement these frameworks along with the knowledge database, the following scheme is proposed (Fig. 1):

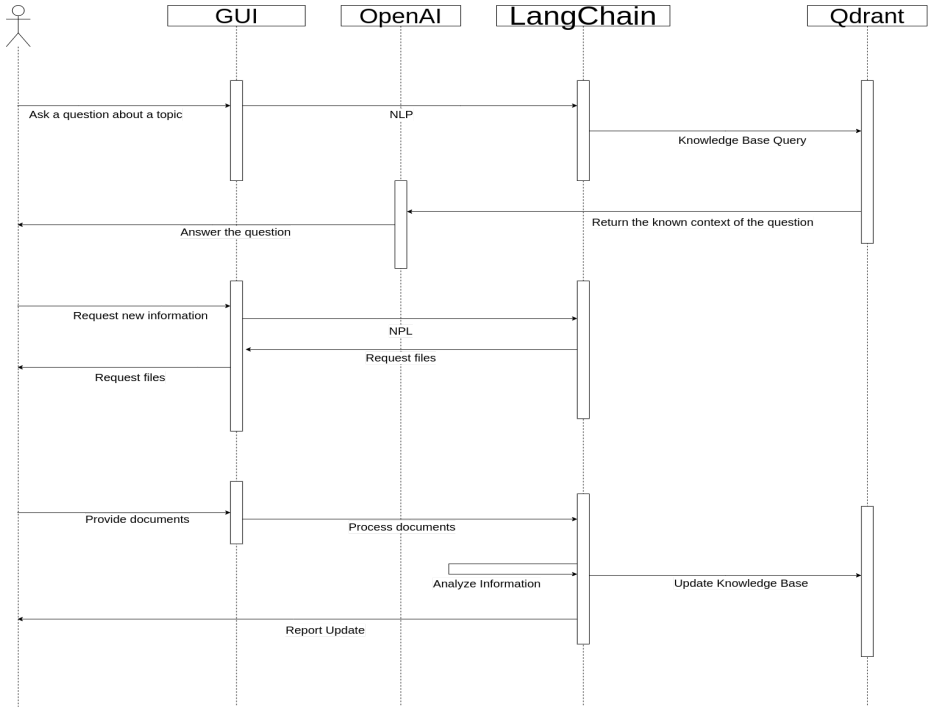


Fig 1. Sequence Diagram for Interaction with the Educational Agent.

Subsequently, the architecture of the intelligent educational agent is presented, defining the necessary components for content adaptation based on large language models, as visualized in Fig 2

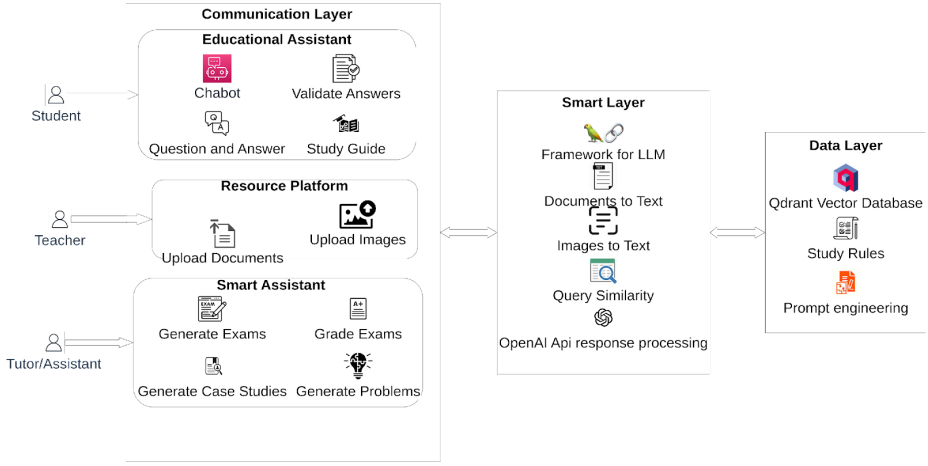


Fig 2. Architecture Diagram of the Proposed Solution.

To do this, the software is coded according to the presented architecture, identifying three main actors in the system: the student, the teacher, and a tutor/assistant who interact with the virtual assistant to upload information and generate new educational resources. It is composed of three layers, which are described below:

Communication Layer

- Educational Assistant
  - The Chatbot is the main component for the "Student" user, as the user will interact directly with the bot to complement other functionalities and directly consume the OpenAI API.
  - The Response Validation Service is a support for the user, as it allows checking answers to exams and tests. It retrieves contexts of the subject matter previously uploaded to the knowledge database.
  - The question and answer service is used to study the subjects uploaded within the knowledge base. The main actor here is the Chatbot, which responds to questions based on their context.
- Resource Platform
  - The Document Upload service allows teachers to upload information for their subject so that the chatbot can analyze it and add it to its knowledge database.
- Intelligent Assistant
  - The Exam Generation service allows Tutors/Assistants to generate a set of questions and answers related to the previously loaded context.
  - The Exam Grading service allows Tutors/Assistants to grade the questions answered by students and assign a grade based on the number of correct answers.

Intelligent Layer

- The framework for Long Language Models (LLM) is LangChain, which allows for the understanding of the original text and enables adapting the response for subsequent evaluation through prompt-tuning.
- For the Document to Text service, LangChain enables the processing of PDF documents, transcribing the information from them to be subsequently attached to the knowledge database.

Data Layer

- Study rules are the norms to which the chatbot is subject. These are the rules that ensure the bot speaks a single language and can respond to questions with answers containing context-specific information.
- Prompt Engineering allows for efficient control of the responses provided by the Long Language Models.

3.2 Tools and Technologies

A thorough evaluation of various tools and technologies was carried out in order to address the challenges associated with implementing an AI-based academic assistant. The main purpose was to identify solutions capable of managing information, analyzing a wide variety of data (such as documents and images), and retaining previously stored contexts. In the following sections, the selection process is detailed, and the main features of the selected tools are described:

LangChain: The use of agents from Large Language Models (LLM) in LangChain offers several advantages, such as their ability to process significant volumes of information, understand context, and quickly adapt to new tasks. These agents can effectively tackle complex challenges and provide accurate solutions that meet the specific needs of users[15].
Qdrant: For efficient storage and retrieval of information, Qdrant was selected as the vector database. Qdrant is a database specifically designed to store and query vectors, which is particularly useful in the context of the academic assistant, where data is represented as vectors in a multidimensional space. Qdrant offers high search speed and allows for complex queries based on similarity and proximity [16].

3.3 Rule Design and Response Logic

Prompt engineering is a discipline used to efficiently utilize language models (LM) as it helps in better employing Large Language Models (LLM). It is also used in the fields of artificial intelligence and natural language processing to design and optimize commands or instructions given to language models. This discipline seeks to find the most suitable and effective formulation to obtain results from the LM [17]. Based on this, the academic assistant was implemented using the Knowledge-Generated Prompt technique, as shown in Fig 3:

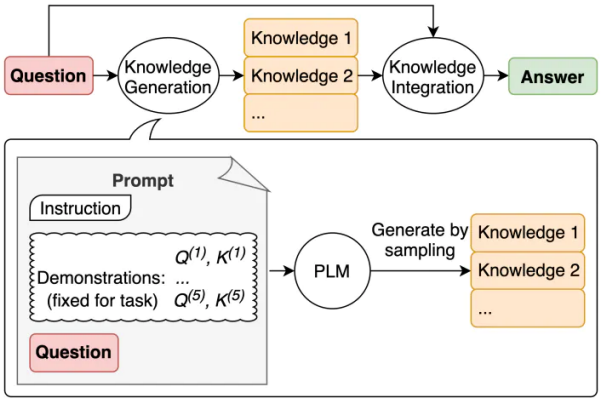


Fig 3. Knowledge-Generated Prompt Technique.[4]

In this process, the question-response flow is understood, and prior to formulating the question, context is provided to the LLM to achieve a standardized format for the response. It's important to note that this context should be related to the question. For this purpose, LangChain was implemented with its modules to build indexes that enable an understanding of the context of the question asked. At the same time, it retrieves vectors stored within Qdrant that contain the necessary information to formulate a prompt with the essential background to obtain an efficient and focused response.

4 Results

To fully understand the performance of the educational assistant and its impact on the educational field, we conducted a comprehensive evaluation based on specific tests. Here, we present and exemplify two use cases:

Multiple-Choice Test Generation: In this test, we evaluated the assistant's ability to generate a multiple-choice test related to the subject of Information Security, as shown in Figure 4. The results highlighted the following aspects:

- **Accuracy of Answers:** The assistant demonstrated a remarkable ability to generate precise multiple-choice questions and answers, reflecting its aptitude for accurately addressing the topics presented.
- **Contextual Understanding:** The assistant appropriately maintained the context of the Information Security subject, responding coherently to requests related to the topics, demonstrating its contextual understanding.
- **Resource Generation Capability:** In this evaluation stage, we did not explore the assistant's ability to produce other types of resources, as we focused on generating multiple-choice tests.

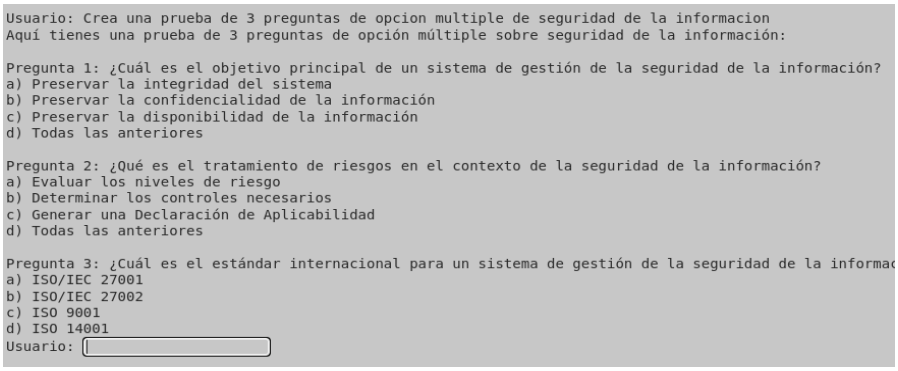


Fig 4. Test Generation Service.

Academic Planning Generation: In the second test, we evaluated the assistant's ability to create a detailed academic plan for the subject of Information Security, and the result generated is shown in Fig 5. The following aspects stood out:

- **Accuracy of Responses:** The plan generated by the assistant was highly accurate and covered the essential aspects of the subject comprehensively, demonstrating its ability to provide a coherent plan.
- **Contextual Understanding:** The assistant maintained appropriate context in the conversation, identifying key elements of the subject and its structure, supporting its ability to understand the academic environment.
- **Resource Generation Capability:** The plan included recommendations for additional resources, highlighting the assistant's ability to provide complementary study materials and enrich the learning experience.

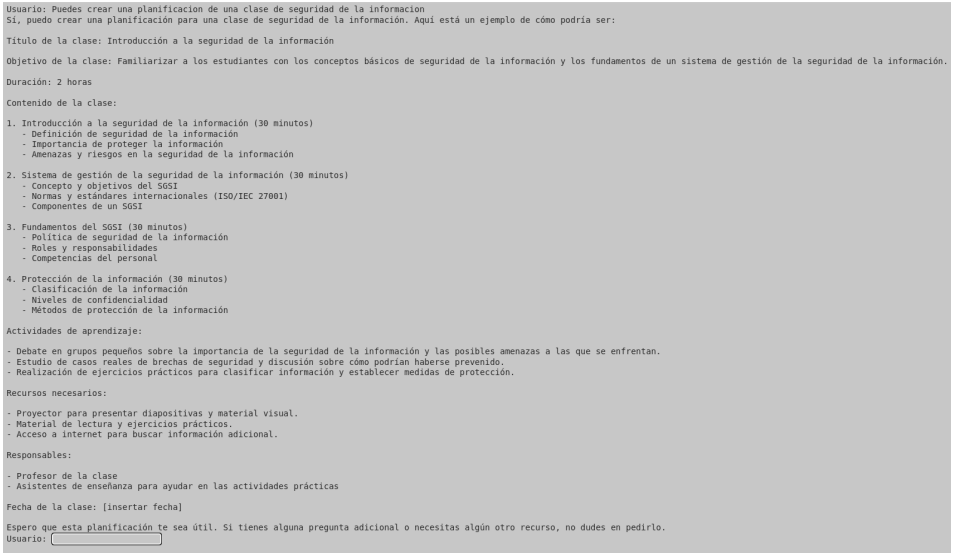


Fig 5. Academic planning generation service.

Laboratory tests were conducted using synthetic data and simulations to validate the operability of the educational agent. The synthetic data verified the system's ability to process educational information, while the functional tests evaluated its effectiveness in providing real-time educational support. These processes played a fundamental role in ensuring that the system met the performance and reliability standards required for its application in an educational environment. To this end, a process of uploading the subject content information and on the basis of questions and answers validating the results of the educational agent was carried out. In this context, information from the Information Systems subject was used.

5 Discussion

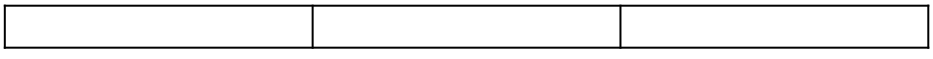
The use of large language models to adapt educational content has the potential to personalize the learning experience of each student more effectively through a chatbot. These models have the ability to understand and analyze extensive texts, capturing context and semantics more accurately, allowing the intelligent educational agent to tailor content to the specific needs of each student, providing more detailed explanations, relevant examples, and personalized recommendations.

Furthermore, content adaptation based on large language models can address the diversity of learning styles and knowledge levels of students. By understanding the context and the level of understanding of each student, the educational agent can adjust the content's difficulty, provide additional exercises, or simplify complex explanations. This can enhance learning effectiveness by providing an individualized

experience tailored to each student's needs, focusing only on what the student doesn't understand.
On the other hand, there are challenges such as the availability and quality of educational datasets used to train the models. Having a representative and diverse dataset that covers the entire subject matter and additional resources is crucial. Additionally, ethical and privacy concerns must be addressed when handling and using sensitive student educational data.\

6 Conclusion

In this study, we have explored the development of an Intelligent Educational Agent (IEA) for content adaptation based on large language models. Our research aimed to investigate the role of intelligent educational support agents in enhancing the online learning experience for students, the utilization of large language models like GPT-3 for personalized education support, and the challenges of implementing these models in resource-constrained educational settings.
Our technical work in the development of the IEA involved utilizing open-source technologies within a layered architecture, enabling us to focus on each of the functionalities presented by the educational assistant. We have successfully demonstrated that content adaptation based on large language models offers substantial opportunities to personalize the learning experience of students. These models exhibit a deep understanding of context and semantics, allowing precise content adaptation based on student questions and teacher-provided resources.
However, the journey is not without its challenges. Issues such as the availability and quality of educational datasets, interpretation and explanation of the agent's reasoning, data transparency, and privacy concerns, as well as the essential computational infrastructure, must be addressed effectively. These challenges represent a significant investment cost, both in terms of resources and expertise, to transition from research to production.
Despite these hurdles, the development of an Intelligent Educational Agent for content adaptation based on large language models is not only feasible but holds great promise for the improvement of education. It can provide individualized learning experiences, generate valuable educational resources for educators, and enhance the overall quality and effectiveness of education.
In closing, our work underscores the importance of fostering collaboration between education experts, software developers, researchers, and other stakeholders. This interdisciplinary approach enables us to harness the full potential of technology to transform and enrich the educational process. By providing students with tools and resources rooted in research and educational methodologies, we can empower them to reach their full potential.



References

1. Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.
2. Kraus, M., Bingler, J. A., Leippold, M., Schimanski, T., Senni, C. C., Stammbach, D., ... & Webersinke, N. (2023). Enhancing Large Language Models with Climate Resources. arXiv preprint arXiv:2304.00116.
3. Xu, W., Ouyang, F. A systematic review of AI role in the educational system based on a proposed conceptual framework. Educ Inf Technol 27, 4195–4223 (2022). https://doi.org/10.1007/s10639-021-10774-y
4. Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. L., ... & Hajishirzi, H. (2021). Generated knowledge prompting for commonsense reasoning. arXiv preprint arXiv:2110.08387.
5. Richardson, R., & North, M. (2020). Transition and migration to online learning environment. International Management Review, 16(2), 5-105.
6. Kavitha, M. G., & Raj, D. L. (2017). Educational data mining and learning analytics-educational assistance for teaching and learning. arXiv preprint arXiv:1706.03327.
7. Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., & Ahamed, J. (2023, May). Artificial Intelligence in Education: An Argument of Chat-GPT Use in Education. In 2023 9th International Conference on Information Technology Trends (ITT) (pp. 151-156). IEEE.
8. Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., ... & Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. Asian Journal of Distance Education, 18(1).
9. Biswas, S. (2023). Role of Chat GPT in Education. Available at SSRN 4369981.
10. Lesage, J., Brennan, R., Eaton, S. E., Moya, B., McDermott, B., Wiens, J., & Herrero, K. (2023). Exploring natural language processing in mechanical engineering education: Implications for academic integrity. International Journal of Mechanical Engineering Education, 03064190231166665.
11. Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large language models challenge the future of higher education. Nature Machine Intelligence, 5(4), 333-334.
12. Bouzenada, S. N. E., Boissier, O., & Zarour, N. E. (2018). An agent-based approach for personalised and adaptive learning. International journal of technology enhanced learning, 10(3), 184-201.
13. Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Lucas, P., Sauzéon, H., & Oudeyer, P. Y. (2023). Gpt-3-driven pedagogical agents to train children's curious question-asking skills. International Journal of Artificial Intelligence in Education, 1-36.
14. Kim, S., Yun, S., Lee, H., Gubri, M., Yoon, S., & Oh, S. J. (2023). Propile: Probing privacy leakage in large language models. arXiv preprint arXiv:2307.01881.
15. https://python.langchain.com/docs/get_started/introduction
16. https://qdrant.tech/
17. Li, Y., Tan, Z., & Liu, Y. (2023). Privacy-preserving prompt tuning for large language model services. arXiv preprint arXiv:2305.06212.

