Virtual Reality and Language Models, a New Frontier in Learning

Juan Izquierdo-Domenech, Jordi Linares-Pellicer, Isabel Ferri-Molla*

Valencian Research Institute for Artificial Intelligence (VRAIN), Universitat Politècnica de València (UPV), València (Spain)

Received 14 September 2023 | Accepted 15 February 2024 | Published 21 February 2024

Abstract

The proposed research introduces an innovative Virtual Reality (VR) and Large Language Model (LLM) architecture to enhance the learning process across diverse educational contexts, ranging from school to industrial settings. Leveraging the capabilities of LLMs and Retrieval-Augmented Generation (RAG), the architecture centers around an immersive VR application. This application empowers students of all backgrounds to interactively engage with their environment by posing questions and receiving informative responses in text format and with visual hints in VR, thereby fostering a dynamic learning experience. LLMs with RAG act as the backbones of this architecture, facilitating the integration of private or domain-specific data into the learning process. By seamlessly connecting various data sources through data connectors, RAG overcomes the challenge of disparate and siloed information repositories, including APIs, PDFs, SQL databases, and more. The data indexes provided by RAG solutions further streamline this process by structuring the ingested data into formats optimized for consumption by LLMs. An empirical study was conducted to evaluate the effectiveness of this VR and LLM architecture. Twenty participants, divided into Experimental and Control groups, were selected to assess the impact on their learning process. The Experimental group utilized the immersive VR application, which allowed interactive engagement with the educational environment, while the Control group followed traditional learning methods. The study revealed significant improvements in learning outcomes for the Experimental group, demonstrating the potential of integrating VR and LLMs in enhancing comprehension and engagement in learning contexts. This study presents an innovative approach that capitalizes on the synergy between LLMs and immersive VR technology, opening avenues for a transformative learning experience that transcends traditional boundaries and empowers learners across a spectrum of educational landscapes.

Keywords

Large Language Models, RetrievalAugmented Generation, Virtual Reality.

DOI: 10.9781/ijimai.2024.02.007

I. Introduction

Technology's rapid expansion, especially in internet-related fields, has revolutionized learning. For today's students, born in this tech-savvy era, accessing information is effortless, but it has raised concerns about their attention and problem-solving abilities. In response, educators are adapting teaching methods, such as Massive Open Online Courses (MOOCs) and the flipped classroom [1]. Virtual reality (VR) has been a research focus within computer science and information technology, especially for its educational applications. Recent advancements have made VR more accessible and immersive, enhancing its potential as a learning tool. VR has been utilized in various educational contexts, from primary and secondary classrooms to professional training programs. It caters to diverse student profiles, including different age groups, learning abilities, and backgrounds. Studies have demonstrated VR's effectiveness across various disciplines like science, history, and medicine, emphasizing its role in providing interactive and realistic learning experiences. Despite its promise, the implementation of VR in education faces challenges such as high development costs and the need for adequate technological infrastructure. Christian et al.'s systematic literature review on VR in superior education distance learning, especially during the COVID-19 pandemic, underscores VR's growing role in higher education. Their review reveals VR's effectiveness in enhancing learning experiences, motivation, and comprehension in fields like engineering and medicine, predominantly among university students, and that technological advancements have made diverse VR applications possible despite equipment issues and budget constraints [2]. Figueiredo et al. explore VR's impact on elementary education, emphasizing its capacity to create captivating learning experiences for young learners. Platforms like Google Expeditions and Nearpod VR have made complex subjects more accessible, promoting student engagement and empathy. The study reflects on the evolution of VR technology, its increasing affordability, and its potential to revolutionize traditional teaching methods despite content development and teacher training challenges [3]. In higher education, particularly in biomedical sciences, Fabris et al. discuss VR's role in enhancing the visual-spatial understanding of complex anatomical structures. The review presents mixed results from various studies regarding VR's effectiveness, highlighting the

* Corresponding author.

E-mail addresses: juaizdom@upv.es (J. Izquierdo-Domenech), jlinares@dsic.upv.es (J. Linares-Pellicer), isfermol@upv.es (I. Ferri-Molla).

importance of interactivity in VR applications for effective learning. It also addresses scalability and cost considerations, pointing to the potential of VR as a valuable tool in education when appropriately integrated into curricula [4].

Numerous investigations have underscored VR's capacity to augment educational outcomes by furnishing learners with genuine and captivating learning settings. Ausburn contends that VR constitutes a potent innovative technology for pedagogy and research, facilitating deeper comprehension and reduced training durations [5]. Correspondingly, the works of Alshammari and Lee et al. scrutinize VR's support for collaborative learning, problem-centric pedagogy, and role-playing scenarios [6],[7]. Some inquiries delve into VR's unique ability to grant access to otherwise unreachable experiences. For example, Asad et al. discern that VR grants students first-hand experiences and amplifies experiential learning [8]. Zakaria et al. elucidate how VR affords simulations of remote and perilous locales [9], while Carruth posits that it permits students to interact with expensive equipment and explore intricate problem domains devoid of risk [10]. Additional investigations explore VR's potential to supplement or even supplant real-world experiences. Oiwake et al. introduce the groundbreaking idea of a "VR Classroom," where students experience the sensation of being in a physical classroom [11]. Similarly, Hunvik et al. have created a VR application tailored for a STEM course. Their research concludes that such an application holds potential as a precursor to conventional learning methods [12]. In a complementary fashion, Smutny et al. review VR applications spanning a wide array of academic disciplines, focusing on curricula including medicine, history, engineering, and music [13]. While VR displays considerable promise in enriching learning experiences, certain constraints persist. Asad et al. underscore the considerable implementation costs [8]. Lopez et al. coincide on the high cost of developing VR experiences, although highlighting that VR is an optimal tool for learning, even in professional contexts [14]. Yet, with these exciting advancements and ongoing inquiries, the future of education seems balanced for a transformative journey into the immersive realms of VR, offering both challenges and opportunities for educators and learners alike.

In this evolving educational landscape, the role of Artificial Intelligence (AI) is increasingly significant. As highlighted in "Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4)", AI brings a unique set of opportunities and challenges to the realm of education, with the potential to contribute significantly to achieving Sustainable Development Goal 4 (SDG4) of the UNESCO 2030 Agenda, which emphasizes quality education and lifelong learning opportunities for all. It also emphasizes the ethical considerations and the need for AI to be developed to benefit humanity and respect global norms and standards, making it particularly relevant in the educational context [15]. Empirical investigations conducted in this domain have consistently illuminated the manifold ways AI can be harnessed to ameliorate educational administration, instructional methodologies, and learning outcomes. Notably, AI systems have demonstrated their utility in alleviating the administrative burdens borne by educators. For instance, AI-driven tools have proven instrumental in automating tasks such as assignment grading or personalized teaching, allowing educators to redirect their efforts toward more individualized and engaging endeavors [16]. Furthermore, AI-powered adaptive learning systems have emerged as a pivotal mechanism for tailoring educational curricula and providing content to individual student requisites, thereby supporting student engagement and adapting to specific student needs [17]. The development of virtual classrooms and AI-driven chatbots is concurrently underway, seeking to provide autonomous instruction to students or to function as valuable adjuncts to human educators [18]. Generative AI is a branch of AI focused on creating algorithms and models that produce human-like data or content. These systems use Deep Learning (DL) to learn patterns from large datasets, enabling them to generate contextually relevant and creative outputs, such as text, images, or music. Generative AI has wide-ranging applications, from text generation to creative arts and data synthesis, and is already being applied in education. Leiker et al. highlight that AI-generated synthetic videos can efficiently replace traditional instructional videos, facilitating the cost-effective and time-efficient production of high-quality educational content [19]. Bekeš et al. discovered that AI-generated content is favored by teachers over conventional materials, primarily due to its adaptability and flexibility [20]. However, it is imperative to acknowledge that the advent of AI has prompted discourse regarding its potential to redefine the role of educators. Some studies posit that AI's increasing integration into the educational milieu may gradually transition teachers from traditional lecturers into facilitators as AI assumes instructional responsibilities [21]. Du Boulay argues in favor of enhancing human educators with AI, suggesting that AI can serve as a personalized tutor when necessary, allowing human teachers to concentrate on the broader classroom context [22]. Yang predicts that AI and VR will significantly impact education in the coming years [23]. While AI can detect students' weaknesses and tailor instruction to their needs, VR can foster students' interest and social development.

The potential of technology in education is vast and encompasses a range of innovative tools and methods. While AI plays a crucial role in enhancing educational experiences, it is not the sole driving force. Alongside AI, emerging technologies like VR and generative AI are becoming transformative factors in the educational landscape.

While the application of VR in education has been extensively studied, its combination with Large Language Models (LLMs) represents a novel frontier that holds significant promise for further revolutionizing learning methodologies. The present proposal focuses on integrating VR technology and generative AI to tackle a significant educational challenge: providing rapid and contextually accurate access to information. The system empowers students to access information through Question Answering (QA) mechanisms, offering a unique approach to enhancing comprehension, even in complex laboratory settings. By harnessing the immersive capabilities of VR and the data synthesis abilities of generative AI, this proposal represents an exciting synergy of technological advancements that have the potential to revolutionize education. This research aims to bridge the gap between the immersive experiences provided by VR and the advanced capabilities of LLMs in processing and generating human-like text. The synergy between VR's interactive environments and LLMs' ability to understand and respond to Natural Language (NL) queries presents an unprecedented opportunity to create more engaging, personalized, and effective learning experiences. Our study is positioned at this intersection, exploring how the integration of VR with LLMs can enhance the learning process, particularly in settings where traditional educational methods may fall short, thus filling the gap of research on the combined use of VR and LLMs.

A critical aspect of this system's functionality is using generative AI techniques to generate responses based solely on contextual information, reducing the risk of producing inaccurate or fictitious information. This contextual information can take various forms, such as text documents or .pdf files, with the adaptable library providing access to a wide range of alternative data sources, including databases, spreadsheet files, and even Application Programming Interfaces (APIs). To overcome the challenges associated with physical laboratory access, including scheduling and logistical constraints, VR technology, combined with 360° photos, has been chosen to represent complex environments like laboratories and shopfloors, each containing diverse points of interest. With this approach, the evaluated system empowers users to articulate queries in NL, allowing them to receive responses to their original questions. Moreover, as these answers are derived from contextual knowledge, the application seamlessly guides the users' attention to the relevant elements of interest in the VR environment associated with their questions.

The article is structured into distinct sections, each addressing specific aspects of the research. First, in Section II, the article explores the potential impact of LLMs and generative AI in education. Next, Section III delves into Retrieval-Augmented Generation (RAG) methods and their significance in contextual information retrieval. Subsequently, Section IV explains the system's implementation, including server-side and client-side components. The critical phases of system evaluation are covered in Section V, and a comprehensive examination of limitations is presented in Section VI. Finally, the conclusions are drawn in Section VII.

II. LLM in Education

Large Language Models (LLMs) and generative AI are emerging as transformative educational tools, automating and enhancing various educational processes. While they offer significant advantages in generating high-quality educational content and analyzing student responses, they also present challenges. LLMs can exhibit biases inherited from their training data, leading to ethical concerns. Their lack of deep understanding can result in superficial or inaccurate content, and there is a risk of student overreliance on these models, which may impede the development of critical thinking skills. Additionally, their operation requires considerable computational resources, posing a barrier in some educational settings. Numerous studies have delved into the utilization of LLMs for the generation of high-quality educational content at scale, ranging from programming exercises and code explanations [24] to the creation of comprehensive multimedia course materials [25]. Through techniques like clustering and summarization, LLMs facilitate the rapid and accurate identification of underlying themes and patterns within student responses, surpassing the capabilities of manual analysis alone [26]. Nevertheless, it remains imperative to incorporate human oversight and review mechanisms to ensure the accuracy and reliability of these AI-generated resources before they are made available to students [27]. While the automated generation of educational materials promises to reduce instructors' workload significantly, addressing practical and ethical concerns associated with integrating LLMs into educational settings is essential. A comprehensive analysis of 118 research papers revealed that LLMs have been applied across 53 distinct educational use cases, encompassing tasks such as grading, teaching support, content generation, and recommendation [27]. Although LLMs exhibit the potential to automate and enhance these educational functions, their performance, transparency, privacy implications, commitment to equality, and ethical considerations must be evaluated to ascertain their suitability for educational contexts. Furthermore, LLMs offer a promising avenue for gaining insights into student learning by conducting in-depth analyses of student-generated artifacts, such as essays.

III. RAG for Contextual Information Retrieval

RAG methods have recently gained significant interest since they allow to combine neural generation models (i.e., parametric memory) with contextual information (i.e., non-parametric memory), as depicted in Fig. 1. RAG is an approach in Natural Language Processing (NLP) that combines the power of language models with information retrieval, enabling the generation of more informed and contextually relevant responses by dynamically fetching and integrating external knowledge sources during the generation process. Numerous articles have explored RAG models for open-domain question answering and found that they can achieve state-of-the-art performance. Lewis et al. introduced a general RAG recipe, showing RAG models outperform parametric seq2seq models and task-specific architectures on knowledge-intensive NLP tasks like open-domain QA [28]. Ranjit et al. built on this work, proposing a RAG model for radiology report generation that achieved the best metrics [29]. While early RAG work focused on retrieving text, recent papers have expanded to multimodal knowledge. Yu discussed obstacles to single-source retrieval and provided solutions for RAG over heterogeneous knowledge [30]. Chen et al. introduced the first multimodal RAG, accessing images and text to answer questions [31]. Zhao et al. surveyed RAG methods across modalities, reviewing image, code, table, graph, and audio retrieval for generation [32]. Some work has aimed to improve RAG domain adaptation. Siriwardhana et al. proposed an end-to-end trained RAG variant with an auxiliary loss for reconstructing sentences from retrieved knowledge [33]. They showed significant gains in adapting RAG to COVID-19, news, and conversation domains. Finally, Mao et al. presented an alternative approach: Generation-Augmented Retrieval (GAR) [34]. GAR uses generation to expand queries before retrieving relevant passages. On open-domain QA, GAR with sparse retrieval matched or outperformed dense retrieval methods, achieving state-of-the-art extractive QA performance.

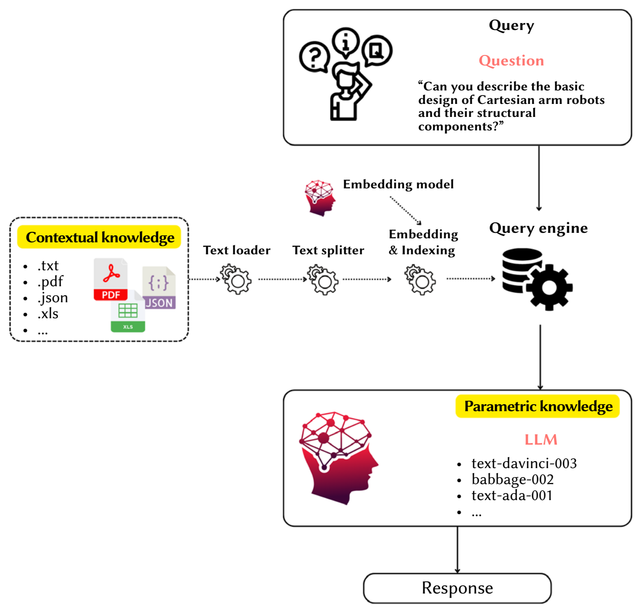

Fig. 1. RAG complements parametric knowledge with contextual knowledge.

The RAG methodology is a powerful tool for tailoring NLP and generation to specific domains, such as education. It offers a unique advantage over techniques such as few-shot learning by constraining generated responses to verified information, effectively reducing the risk of "hallucination" or generating incorrect or irrelevant answers. This feature ensures that students receive accurate and consistent information, enhancing the overall learning experience.

Few-shot learning, while valuable, has limitations, particularly concerning the maximum token size for prompts. This constraint can lead to less accurate or incomplete answers. RAG addresses this issue by using semantic similarity to pull the most relevant information from a given context, thereby creating a more precise final prompt. This ensures the generated answer is accurate and relevant to the student's query.

However, it is essential to consider the limitations of fine-tuning, especially in QA environments. While fine-tuning offers granular control over model behavior, it often requires substantial training data and computational resources. More critically, fine-tuning can be less reliable in generating precise answers to specific questions, as it does not inherently constrain the model's responses to verified information. This makes it less suitable for applications where the accuracy of each individual answer is paramount, such as educational settings where incorrect information could have enduring impacts. RAG's efficiency and focus on accuracy offer a more reliable alternative in these contexts.

IV. System Implementation

This section is dedicated to explaining the specific implementation that was carried out for the evaluated system. The architecture presented here is based on a client-server model, with the server responsible for processing LLMs and contextual data to respond to user queries and the client serving as a VR interface for users to explore educational environments and pose NL questions.

A. Server-Side Implementation

The server-side implementation is responsible for tasks and actions related to the processing of LLMs and, crucially, the use of RAG for extracting specific contextual information. In this implementation, the LlamaIndex library has been employed due to its ability to provide an interface enabling developers to work with various LLMs, such as gpt-3.5 or text-davinci-003 [35]. Furthermore, LlamaIndex facilitates the execution of RAG, which means that it is possible to enhance the parametric knowledge of the LLM with contextual information. Fig. 1 details how contextual information is accessed to enable RAG.

In the analyzed context, the contextual information is based on text documents in formats such as .pdf and .txt, which describe various elements present in a classroom (e.g., an Angular arm robot or a Cartesian arm robot). However, the flexibility of LlamaIndex allows access to a broad spectrum of alternative data sources, encompassing databases, spreadsheet files, and even Application Programming Interfaces (APIs). Besides, a .json file is utilized to define the relationships between the elements in a classroom and the VR scenes located on the client side, as shown in the appendix in Listing 1. Further details can be found in subsection IV.B.

To facilitate RAG's utilization of contextual knowledge exclusively, LlamaIndex incorporates the concept of an Index. Illustrated in Fig. 1, the indexing stage assumes responsibility for allowing rapid access to relevant context for a user query. These generated indexes streamline the retrieval process, automating vector embedding calculations. While the `VectorStoreIndex` is a prevalent index type, the system's preference in this instance is the `KeywordTableIndex`. This choice aligns with the system's approach, wherein each node (i.e., each textual chunk produced during the text splitting task) additionally factors in specific keywords. During a query operation, the node selection containing the relevant text chunk is determined based on keywords extracted from the query, enhancing answer reliability.

The requests that the server handles are summarized in the following routes:

1. `GET /summaries` This route returns a list of summaries organized by scene, briefly explaining the elements present in each of them, as illustrated in Fig. 2.
2. `POST /query` When requesting this route, based on the user's query, a response is returned based solely on the available contextual information. In addition to the response, a unique identifier of the queried element, the name of the element of interest, and the VR scene in which it is located are provided.

When a user submits a question through the immersive application (refer to Section IV.B), the query is transmitted to the POST /query endpoint. Here, a two-step process is employed to enhance the system's contextual understanding and ensure meaningful responses.

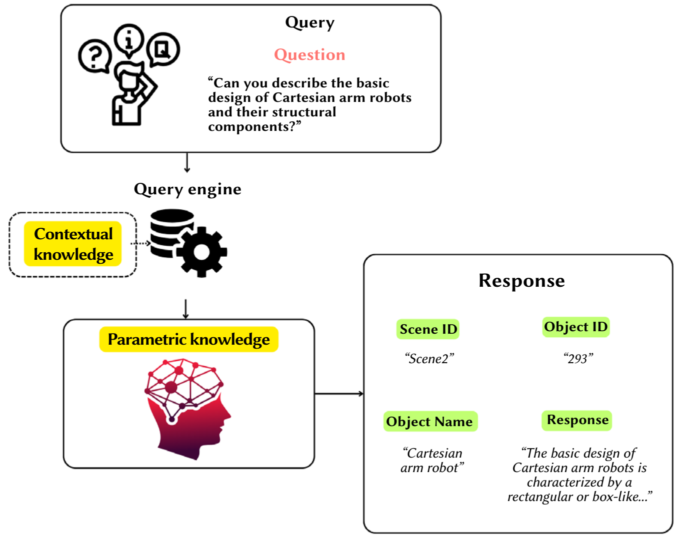

Fig. 2. Example of a GET /query request diagram.

First, we use prompting to integrate the user's question, leading to a format like this:

```
Answer the question using ONLY THE CONTEXT, and if
you're not TOTALLY sure of the answer, say
'Sorry, I don't know'. Q: {question} A:
```

This approach compels the system to rely on contextual knowledge to answer the question effectively. To pinpoint the specific index or chunk of text where the answer resides, the SubQuestionQueryEngine, available in LlamaIndex, is used.

Once a valid response (i.e., an answer different from 'Sorry, I don't know') is obtained, the system follows up with another prompt:

```
Based on the question "{original_question}" and its
answer "{query_answer}", please return an answer in the
format
"scene_id:_,object_id:_,object_name:_" If you don't
know the answer, respond
"scene_id:N/A,object_id:N/A,object_name:N/A" A:
```

The success of this query to the LLM dramatically depends on the .json file that establishes relationships between scenes and elements of interest. In this regard, the results achieved have been quite promising.

Eventually, in response to the user's query, they receive the answer and identifiers for the scenes and objects in question, which they can utilize in the client-side application.

Although RAG techniques, as used in the project, significantly improve LLM performance in front of users' questions, RAG cannot improve the current well-known limitations in reasoning questions or multi-hop questions on documents that are still part of current LLM solutions [36].

The underlying technology in the server development is based on Python 3.10.11 as the primary programming language, with FastAPI and Uvicorn for implementing the REST server. To access the content of contextual information in PDF format, the PyPDF library is utilized.

B. Client-Side Implementation

Unity was chosen as the development engine for the client-side component due to its outstanding capabilities in creating cross-platform applications. Specifically, it has enabled the efficient and effective development of VR applications. In order to optimize the cost associated with VR application development, the decision was made to employ an accessible yet entirely valid technique for exploring a specific environment, namely, an educational laboratory. This technique uses 360º photographs to circumvent the complexities associated with 3D modeling and physical laboratory access, including scheduling and logistical constraints. The VR device used to deploy the evaluated system was the Meta Quest 2; nevertheless, Unity's cross-platform architecture enables executing the same application to similar devices, such as the Pico VR or the HTC Vive.

The VR application comprises several scenes, each hosting various points of interest. In the example of the scene depicted in Fig. 3, the most prominent element is a Cartesian robotic arm; however, it is essential to note that the system is adaptable and can accommodate different points of interest in the same scene, and spread among different scenes. All this information must be explicitly detailed in the .json file described in Subsection IV.A (Listing 1 in the appendix).

Fig. 3. VR scene with a Cartesian arm robot in the middle.

Users can pose questions in NL using speech recognition during the virtual environment exploration. These questions are transmitted to the server through the GET /query request. Suppose the sought-after information is part of the contextual knowledge (described in Section III). In that case, the provided response will include the requested information and metadata related to the point of interest and the user's current scene. Consequently, students are not obliged to be in the VR scene containing the point of interest they inquire about; they can ask about any point of interest encompassed by the contextual knowledge. Fig. 4 visually represents the user's post-response perspective. Notably, alongside the textual answer, a visual cue is strategically employed to direct the user's attention towards the specific element relevant to the initial question.

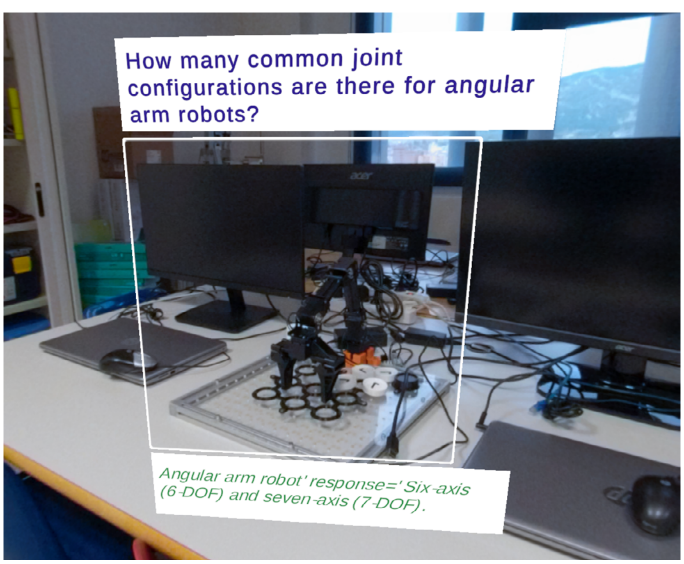

Fig. 4. In VR, user questions trigger dual feedback - textual responses and visual cues, so students get textual answer and VR interaction.

V. System Evaluation

In this study, 20 participants divided into two distinct groups were selected to evaluate the effectiveness of our architecture in enhancing the learning process. This sample size was determined based on the available resources, the innovative nature of the technology involved, and the need for in-depth interaction with each participant. While a larger sample could provide more generalizable results, this exploratory study's specific constraints and focus guided this decision. The first group, referred to as the 'Experimental group,' had access to the immersive VR application, which empowered participants to interact with their educational environment, pose questions, and receive informative responses in text and as visual cues in VR. The second group, termed the 'Control group,' followed a more traditional learning approach devoid of VR technology, where participants relied on conventional methods to access educational content and resources, such as PDF files. By juxtaposing these two groups, this study aims to discern the transformative potential of our VR and LLM architecture compared to established learning practices, thus providing a comprehensive evaluation of its impact across diverse educational contexts.

In this study, Robotics was selected as the primary subject for system evaluation due to its relevance in modern education and the potential to benefit from VR and LLM technologies. The practical nature of robotics, involving theoretical knowledge and hands-on skills, makes it an ideal candidate for the presented educational architecture. The application was pivotal in bridging the gap between theory and practice. It enabled students to understand complex robotics concepts and later apply them to manipulating and controlling the robot arms. Integrating immersive VR experiences with enriched theoretical insights by LLMs illustrates the system's capability to offer a comprehensive learning experience, particularly in subjects where practical skills are as crucial as theoretical knowledge. In class, students have access to a variety of robotic arms, including angular and cartesian types, which they were required to manipulate after completing the necessary learning modules.

To ensure the accuracy and trustworthiness of the presented system within the limitations of RAG systems, as explained in section III, a set consisting of 10 questions per type of robot was defined. These questions were designed to cover an overall spectrum of topics, including definitions, historical backgrounds, design features, and applications. Using prompting strategies to reduce hallucinations and given that all the answers to the questions could be found within the provided contextual knowledge, the system presented a high accuracy rate in delivering correct responses. This testing protocol ensured a comprehensive evaluation of the system's capacity to handle varied inquiries and affirmed its effectiveness in providing precise and relevant information.

The participants in this study had some prior familiarity with the subject matter as they had been students in a course that involved working with robots; however, they had not previously interacted with the specific robots featured in the VR setting. The system evaluation took place during four 2-hour sessions outside of regular class hours, during which participants received training on how to use the robots (both groups), familiarized themselves with the system (Experimental group), and completed the tests (both groups). It is important to remark that the primary purpose of this experiment was to learn how to operate the robots and to compare the Control group with the Experimental group.

Before engaging with the VR and LLM application or the traditional learning method, participants completed pre-tests to establish their baseline knowledge and skills in the subject matter. While it was acknowledged that some participants might have had limited prior exposure to the subject, the pre-tests captured their initial understanding. Subsequently, post-tests were administered after

TABLE I. Descriptive and Statistical Contrasts

	Score			Intra-subject Effects Tests	
	Pre-test	**Post-test**		**Score**	**Score*Group**
	Mean (Sd)	*Mean (Sd)*		**F(d.f.); *p*-value (η^2)**	**F(d.f.); *p*-value (η^2)**
Group			*Mean difference (p-value)*	$F(1;18) = 239.63$; $p < 0.001$ (0.93)	$F(1;18) = 8.53$; $p < 0.009$ (0.322)
Control	1.50 (1.08)	5.80 (1.23)	-4.3 (<0.001)		
Experimental	1.30 (0.95)	7.60 (0.97)	-6.3 (<0.001)		
Mean difference (p-value)	0.2 (0.455)	-1.8 (0.002)			

d.f.: degrees of freedom. η^2: partial eta-squared (effect size)

participants had interacted with their respective learning methods. The post-tests allowed to measure the extent of learning gains and the overall impact of this educational approach. By comparing pre-test and post-test results, it was possible to evaluate the system's effectiveness in fostering learning and comprehension, even among those with little prior knowledge.

To ensure that participants in the Experimental group engage effectively with the application, participants were encouraged to explore the application at their own pace while highlighting the significance of thorough knowledge acquisition. They were informed about the availability of a diverse range of learning resources within the application. They were guided on how navigating and asking questions was performed and the contents they needed to review for the subsequent assessment. A standardized VR experience across all participants in the Experimental group was ensured. Each participant used the same VR hardware and software configurations to minimize variability in the quality of the VR experience. Additionally, the technological background of each participant was assessed through a pre-study questionnaire. This assessment helped understand the participants' familiarity and comfort with VR and other digital technologies, which could influence their interaction with the VR environment.

In Table I, the result of the two-way repeated measures ANOVA test is displayed to determine whether the methodology influences the scores obtained in the test. The result shows a statistically significant difference between the pre-test and post-test, regardless of the group. However, the interaction between the group and the score was significant, indicating that the test scores depend on the group. Thus, the students in the Experimental group significantly increased their scores in the post-test compared to the pre-test, as did the Control group, although to a lesser extent. In the beginning, no differences were observed between the groups. In contrast, at the end of the study (i.e., post-test), the scores of the students in the Experimental group were significantly higher than those of the Control group. Fig. 5 displays the evolution of the scores of the groups.

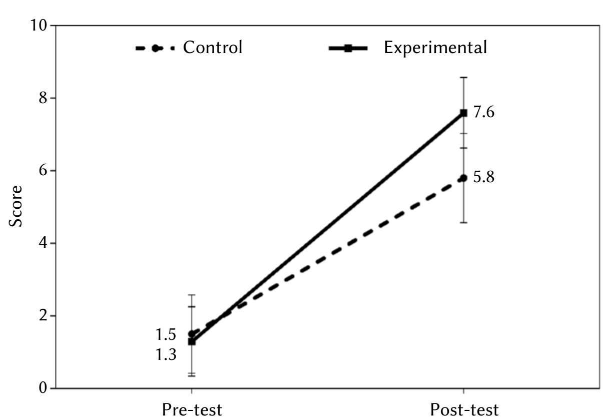

Fig. 5. Score evolution per group.

A Likert-scale questionnaire was also administered to evaluate user satisfaction, chosen for its versatility and effectiveness in capturing nuanced user sentiments. The Likert-scale questionnaire provides a structured, user-friendly format that encourages participants to express their opinions across various dimensions, accommodating diverse user backgrounds and preferences. This inclusivity makes it a valuable tool for assessing user satisfaction in the context of our transformative educational technology. The questionnaire results can be found in the appendix, specifically in Table III, along with the questionnaire questions in Table II. In terms of satisfaction, the median score for students in the Control group was 1.5 (IQR = 1-2), while students in the Experimental group scored 5 points (IQR = 4-5). The Mann-Whitney U test for independent samples revealed that the difference in satisfaction between students in the Experimental group was significantly higher than that of students in the Control group ($U = 0$, $p < 0.001$).

TABLE II. User Experience Questionnaire

Index	Question
Q1	I found the learning experience engaging and immersive.
Q2	The learning materials/methods provided me with valuable information and learning opportunities.
Q3 (Experimental)	Using the VR application improved my understanding of the subject.
Q3 (Control)	The traditional learning materials/methods improved my understanding of the subject matter.
Q4 (Experimental)	I felt more confident in applying the knowledge gained through the system.
Q4 (Control)	I felt more confident in applying the knowledge gained through the traditional learning methods.
Q5	Overall, I am satisfied with my learning experience.

VI. Discussion and Limitations

In exploring the integration of AI in educational contexts, it is crucial to consider both the transformative potential and the challenges posed by these technologies. As García et al. highlight in their study, the emergence of tools like ChatGPT has significantly influenced teaching and learning processes, raising important questions about AI's biases, ethical considerations, and social implications in education. Their work underscores the need for a nuanced understanding of AI's role in education, balancing its benefits with a critical awareness of its limitations and potential risks [37].

The presented architecture, which integrates VR technology with LLMs for educational purposes, has shown promising results in enhancing the learning process. The significant improvement is evident from the post-test scores of the Experimental group compared to the

TABLE III. Likert-Scale Questionnaire

Participant	Group	Q1	Q2	Q3	Q4	Q5
1	Experimental	5	4	5	4	5
2	Experimental	5	3	5	5	4
3	Experimental	5	5	5	5	5
4	Experimental	4	5	5	5	5
5	Experimental	5	5	5	4	5
6	Experimental	3	4	4	4	4
7	Experimental	4	4	5	4	4
8	Experimental	5	3	5	5	5
9	Experimental	5	5	4	4	5
10	Experimental	4	5	5	4	5
11	Control	1	3	4	2	2
12	Control	1	2	3	4	1
13	Control	2	3	4	3	3
14	Control	1	3	4	2	1
15	Control	3	4	4	3	2
16	Control	2	3	3	3	1
17	Control	3	2	3	3	2
18	Control	1	4	2	2	2
19	Control	1	4	4	2	1
20	Control	2	2	4	3	1

Control group. The substantial increase in post-test scores among the Experimental group suggests that an immersive learning experience, coupled with the assistance of LLMs, can effectively foster knowledge acquisition and comprehension, even among participants with limited prior knowledge of the subject matter; however, several noteworthy limitations must be acknowledged to provide a more comprehensive understanding of the system's potential and its impact on education.

- The effectiveness of this approach heavily depends on the quality and comprehensiveness of the contextual knowledge provided to the LLM. Incomplete or inaccurate contextual knowledge may lead to suboptimal responses to user queries. Ensuring the accuracy and relevance of the information fed into the system is critical for its success.
- The success of this system also depends on the availability of appropriate 360º photographs and the accurate mapping of points of interest within VR scenes.
- The study design involved two distinct groups: the Experimental group and the Control group. While the Experimental group experienced the immersive VR and LLM-based learning environment, the Control group followed traditional learning methods. This design might introduce biases related to individual learning preferences and engagement levels. Some participants in the Control group may have needed more motivation due to the absence of the novel VR experience, potentially affecting their post-test performance. The novelty of the VR and LLM integration in the Experimental group might have influenced the motivation and engagement levels, which could affect learning outcomes. Future iterations of the research will aim to equalize engagement potential between the groups. Therefore, it is essential to consider the potential impact of participant motivation and engagement as a limitation when interpreting the study's results.
- The limited number of participants could influence the generalizability of our findings. However, it is noteworthy that similar exploratory studies in VR and LLM integration have also operated with small sample sizes. Future studies with larger samples are necessary to confirm these initial observations and to understand the broader implications of VR and LLMs in educational settings.
- Users unfamiliar with VR technology may face a learning curve when using the system, necessitating adequate training and guidance to ensure a smooth educational experience. Although measures were taken to standardize the VR experience and assess the participants' technological backgrounds, the potential impact of technological issues must be considered. Variations in individual comfort levels and familiarity with VR technology may have influenced the results. Although no significant technological issues were reported during the study, future research should further explore the role of technological familiarity in the effectiveness of VR-based educational interventions.
- While 360º photographs offer a cost-effective means of environment representation, they might not capture all aspects of a physical laboratory, potentially limiting users' tactile and spatial experiences.
- Response times for user queries may fluctuate depending on query complexity and contextual data volume, leading to occasional delays in receiving responses.
- Scalability becomes a crucial consideration in terms of time allocation, the availability of physical resources such as VR headsets, and the number of students, pointing out that as long as these essential materials and resources are accessible, the proposed system can effectively accommodate a growing number of participants linearly without experiencing a substantial decline in performance. This attribute holds significant importance as it ensures that educational institutions can seamlessly expand the adoption of immersive VR and LLM technologies to reach a broader student audience, enhancing the scalability and widespread applicability of innovative educational methodologies.

These limitations underscore the need for ongoing refinement, quality assurance, user support, and research efforts to optimize the system's educational utility and ensure its effectiveness in diverse educational settings. Future research should also explore strategies to mitigate biases in study designs and improve the overall user experience within this innovative educational framework.

VII. Conclusions

Integrating VR technology and generative AI provided by LLMs within the educational landscape represents a transformative approach to learning. This research has explored the potential of combining VR and generative AI to address the challenge of providing rapid and contextually accurate access to information. Through developing an immersive VR application and using RAG, this proposal has demonstrated the ability to enhance comprehension and learning outcomes. Indeed, as highlighted by García et al., the ongoing evolution and refinement of generative AI technologies, including those used in our VR application, are rapidly shaping the future of education, promising transformative changes and novel approaches in teaching and learning methodologies [38].

The findings from the system evaluation suggest a clear advantage for the Experimental group, which had access to the immersive VR and LLM-based learning environment, over the Control group that followed traditional learning methods. The significant improvement in post-test scores among the Experimental group highlights the effectiveness of this innovative approach in fostering knowledge acquisition and comprehension, even when the participants had limited prior knowledge of the subject matter. Moreover, the user satisfaction scores from the Experimental group were significantly higher, underlining the appeal and user-friendliness of this novel educational system.

Nevertheless, it is essential to acknowledge the limitations identified in this research. The system's success is contingent on the accuracy and comprehensiveness of the contextual knowledge provided to the AI models. Incomplete or inaccurate information may lead to suboptimal responses. Besides, the availability of appropriate 360° photographs and accurate establishment of points of interest within VR scenes are critical for the system's effectiveness.

While this study provides insightful initial findings on the integration of VR and generative AI in education, it is important to note that the use of RAG was limited to handling text-based data without multimodal integration from the VR environment. This focus is due to the scope of the current research. However, exploring integrating multimodal data, such as visual inputs from VR, into RAG is a promising direction for future work. The small sample size in this study limits generalizability. Hence, future research with larger, more diverse participant groups needs to be done. Such studies would validate and expand upon the findings and explore the full potential of multimodal VR and LLM systems in educational contexts. As technology advances, addressing these limitations, education is poised for transformation through immersive VR, presenting both challenges and opportunities. Continued research and development are essential for realizing this transformative potential in global education.

Appendix

Listing 1: .json file linking scenes with elements

```
{
   "scenes": [
      {
         "scene_id": "Scene1",
         "objects": [
            {
               "object_id": 776,
               "object_name": "Angular arm robot"
            },
            // Other objects in Scene 1
         ]
      },
      {
         "scene_id": "Scene2",
         "objects": [
            {
               "object_id": 293,
               "object_name": "Cartesian arm robot"
            }
         ]
      },
      // Other scenes , with their objects
   ]
}
```

References

[1] A. Cordero, C. J. Lluch, E. Sanabria Codesal, J. Torregrosa, "Towards a Better Learning Models Through OCWs and MOOCs," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 4, pp. 26–30, 2015, doi: 10.9781/ijimai.2015.345.

[2] B. Christian, C. Salvador, G. Christian, "Virtual Reality (VR) in Superior Education Distance Learning: A Systematic Literature Review," *JOIV: International Journal on Informatics Visualization*, vol. 5, no. 3, pp. 264–270, 2021.

[3] M. Figueiredo, R. Mafalda, A. Kamensky, "Virtual reality as an educational tool for elementary school," in *Smart Innovation, Systems and Technologies*, vol. 198 SIST, 2021, pp. 261–267, Springer. doi: 10.1007/978-3- 030-55374-6_26.

[4] C. P. Fabris, J. A. Rathner, A. Y. Fong, C. P. Sevigny, "Virtual Reality in Higher Education," *International Journal of Innovation in Science and Mathematics Education*, vol. 27, no. 8, pp. 69–80, 2019.

[5] L. J. Ausburn, F. B. Ausburn, "Desktop virtual reality: A powerful new technology for teaching and research in industrial teacher education," *Journal of Industrial Teacher Education*, vol. 41, no. 4, pp. 1–16, 2004.

[6] S. H. Alshammari, "The Role of Virtual Reality in Enhancing Students' Learning," *International Journal of Educational Technology and Learning*, vol. 7, no. 1, pp. 1– 6, 2019, doi: 10.20448/2003.71.1.6.

[7] E. Ai-Lim Lee, K. Wai Wong, "A Review of Using Virtual Reality for Learning," *Transactions on edutainment I*, pp. 231–241, 2008.

[8] M. M. Asad, A. Naz, P. Churi, M. M. Tahanzadeh, "Virtual Reality as Pedagogical Tool to Enhance Experiential Learning: A Systematic Literature Review," *Education Research International*, vol. 2021, pp. 1–17, 2021, doi: 10.1155/2021/7061623.

[9] G. Zakaria, S. Wilkie, "Applications for virtual reality experiences in tertiary education," *ASCILITE Publications*, pp. 186–193, 2020.

[10] D. W. Carruth, "Virtual reality for education and workforce training," in *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 2017, pp. 1–6.

[11] K. Oiwake, K. Komiya, H. Akasaki, T. Nakajima, "VR classroom: enhancing learning experience with virtual class rooms," in *2018 Eleventh International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, 2018, pp. 1–6.

[12] S. R. B. Hunvik, F. Lindseth, "Making Use of Virtual Reality for Artificial Intelligence Education," in *Bridges and Mediation in Higher Distance Education: Second International Workshop, HELMeTO 2020, Bari, BA, Italy, September 17–18, 2020, Revised Selected Papers 2*, vol. 1344 of *Communications in Computer and Information Science*, Cham, 2021, pp. 56–70, Springer International Publishing. doi: 10.1007/978-3-030-67435-9.

[13] P. Smutny, M. Babiuch, P. Foltynek, "A review of the virtual reality applications in education and training," in *2019 20th International Carpathian Control Conference (ICCC)*, 2019, pp. 1–4.

[14] M. A. Lopez, S. Terron, J. M. Lombardo, R. Gonzalez- Crespo, "Towards a solution to create, test and publish mixed reality experiences for occupational safety and health learning: Training-MR," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 2, pp. 212–223, 2021, doi: 10.9781/ijimai.2021.07.003.

[15] J. M. Flores-Vivar, F. J. García-Peñalvo, "Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4)," *Comunicar: Revista cientifica de comunicacion y educacion*, vol. 31, no. 74, pp. 37–47, 2023, doi: 10.3916/C74-2023-03.

[16] L. Chen, P. Chen, Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264– 75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[17] M. Fahimirad, S. S. Kotamjani, "A Review on Application of Artificial Intelligence in Teaching and Learning in Educational Contexts," *International Journal of Learning and Development*, vol. 8, p. 106, 12 2018, doi: 10.5296/ijld.v8i4.14057.

[18] J. Huang, S. Saleh, Y. Liu, "A review on artificial intelligence in education," *Academic Journal of Interdisciplinary Studies*, vol. 10, pp. 206–217, 5 2021, doi: 10.36941/AJIS-2021-0077.

[19] D. Leiker, A. R. Gyllen, I. Eldesouky, M. Cukurova, "Generative AI for learning: Investigating the potential of synthetic learning videos," *arXiv preprint arXiv:2304.03784*, 2023.

[20] E. R. Bekeš, V. Galzina, "Utilizing smart digital technology and artificial intelligence in education for transforming the way content is delivered," in *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2022, pp. 573–578, IEEE.

[21] K. Mondal, "A Synergy of Artificial Intelligence and Education in the 21 st Century Classrooms," in *2019 International Conference on Digitization (ICD)*, 2019, pp. 68–70.

[22] B. Du Boulay, "Artificial intelligence as an effective classroom assistant," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 76–81, 2016.

[23] H. Yang, "Explore How Artificial Intelligence and VR Technology

will Change the Development of Future Education," *Journal of Physics: Conference Series*, vol. 1744, no. 4, 2021, doi: 10.1088/1742-6596/1744/4/042146.

[24] S. Sarsa, P. Denny, A. Hellas, J. Leinonen, "Automatic Generation of Programming Exercises and Code Explanations Using Large Language Models," in *ICER 2022 - Proceedings of the 2022 ACM Conference on International Computing Education Research*, vol. 1, 8 2022, pp. 27–43, Association for Computing Machinery, Inc. doi: 10.1145/3501385.3543957.

[25] D. Leiker, S. Finnigan, A. Ricker Gyllen, M. Cukurova, "Prototyping the use of Large Language Models (LLMs) for adult learning content creation at scale," *arXiv preprint arXiv:2306.01815*, 2023.

[26] A. Katz, U. Shakir, B. Chambers, "The Utility of Large Language Models and Generative AI for Education Research," *arXiv preprint arXiv:2305.18125*, 5 2023.

[27] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin, D. Gašević, "Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review," *British Journal of Educational Technology*, 3 2023, doi: 10.1111/bjet.13370.

[28] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[29] M. Ranjit, G. Ganapathy, R. Manuel, T. Ganu, "Retrieval Augmented Chest X-Ray Report Generation using OpenAI GPT models," *arXiv preprint arXiv:2305.03660*, 5 2023.

[30] W. Yu, "Retrieval-augmented generation across heterogeneous knowledge," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, 2022, pp. 52–58.

[31] W. Chen, H. Hu, X. Chen, P. Verga, W. W. Cohen, "MuRAG: Multimodal Retrieval-Augmented Generator for Open Question Answering over Images and Text," *arXiv preprint arXiv:2210.02928*, 10 2022.

[32] R. Zhao, H. Chen, W. Wang, F. Jiao, X. L. Do, C. Qin, B. Ding, X. Guo, M. Li, X. Li, S. Joty, "Retrieving Multimodal Information for Augmented Generation: A Survey," *arXiv preprint arXiv:2303.10868*, 3 2023.

[33] S. Siriwardhana, R. Weerasekera, E. Wen, T. Kaluarachchi, R. Rajib, S. Nanayakkara, "Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering," *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1–17, 2023, doi: 10.1162/tacl.

[34] Y. Mao, P. He, X. Liu, Y. Shen, J. Gao, J. Han, W. Chen, "Generation-Augmented Retrieval for Open-domain Question Answering," *arXiv preprint arXiv:2009.08553*, 9 2020.

[35] J. Liu, "LlamaIndex," 11 2022. [Online]. Available: https://github.com/jerryjliu/llama_index.

[36] S. Es, J. James, L. Espinosa-Anke, S. Schockaert, "RAGAS: Automated Evaluation of Retrieval Augmented Generation," *arXiv preprint arXiv:2309.15217*, 9 2023.

[37] F. J. García Peñalvo, F. Llorens-Largo, J. Vidal, "The new reality of education in the face of advances in generative artificial intelligence," *RIED-Revista Iberoamericana de Educación a Distancia*, vol. 27, pp. 9– 39, 7 2023, doi: 10.5944/ried.27.1.37716.

[38] F. García-Peñalvo, A. Vázquez-Ingelmo,"What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 4, pp. 7–16, 2023, doi: 10.9781/ijimai.2023.07.006.

Juan Izquierdo-Domenech

Juan Jesús Izquierdo Doménech is an Adjunct Professor of Computer Science in Universitat Politècnica de València. He received his Bachelor's degree in Computer Science Engineering from Universitat Politècnica de València (UPV, Spain) and holds a Master's degree in Multimedia Applications from Universitat Oberta de Catalunya (UOC, Spain). He is currently performing his Ph.D. studies in UPV in the field of Human-Computer Interaction, Mixed Reality, and Artificial Intelligence.

Jordi Linares-Pellicer

Jordi Linares Pellicer is an Associate Professor at Universitat Politècnica de València (UPV, Spain), where he leads the VertexLit research group at the Valencian Research Institute for Artificial Intelligence (VRAIN). He received his Ph.D. in Computer Science from UPV and holds a Master's degree in Artificial Intelligence from Universidad Internacional de La Rioja (UNIR, Spain).

Isabel Ferri-Molla

Isabel Ferri Mollá is currently pursuing her Master's Degree in Artificial Intelligence, Pattern Recognition, and Digital Imaging at Universitat Politècnica de València (UPV, Spain). She received her Bachelor's Degree in Computer Science Engineering from UPV in 2022. Her research interests include areas of artificial intelligence, augmented reality, and human-computer interaction.