NivTA: Towards a Naturally Interactable Edu-Metaverse Teaching Assistant for CAVE

Ye Jia, Zackary P. T. Sin, Xiangzhi Eric Wang, Chen Li, *Member, IEEE,* Peter H. F. Ng, Xiao Huang, Junnan Dong, Yaowei Wang, George Baciu, *Senior Member, IEEE,* Jiannong Cao,*Fellow, IEEE,* and Qing Li, *Fellow, IEEE*

***Abstract*—Edu-metaverse is a specialized metaverse dedicated for interactive education in an immersive environment. Its main purpose is to immerse the learners in a digital environment and conduct learning activities that could mirror reality. Not only does it enable activities that may be difficult to perform in the real world, but it also extends the interaction to personalized and CL. This is a more effective pedagogical approach as it tends to enhance the motivation and engagement of students and it increases their active participation in lessons delivered. To this extend, we propose to realize an interactive virtual teaching assistant called NivTA. To make NivTA easily accessible and engaging by multiple users simultaneously, we also propose to use a CAVE virtual environment (CAVE-VR) as a "metaverse window" into concepts, ideas, topics, and learning activities. The students simply need to step into the CAVE-VR and interact with a life-size teaching assistant that they can engage with naturally, as if they are approaching a real person. Instead of text-based interaction currently developed for large language models (LLM), NivTA is given additional cues regarding the users so it can react more naturally via a specific prompt design. For example, the user can simply point to an educational concept and ask NivTA to explain what it is. To guide NivTA onto the educational concept, the prompt is also designed to feed in an educational KG to provide NivTA with the context of the student's question. The NivTA system is an integration of several components that are discussed in this paper. We further describe how the system is designed and implemented, along with potential applications and future work on interactive collaborative edu-metaverse environments dedicated for teaching and learning.**

***Index Terms*—virtual teaching assistant, LLM agents, cave automatic virtual environment, natural user interface**

I. Introduction

METAVERSE, as an education platform, aims to bring students and educators together into an interactive virtual environment that could potentially unleash a much richer educational content medium due to the highly immersive learning experience. Such a pedagogically focused metaverse is often referred to as an *edu-metaverse* and it is currently an exciting area for e-learning exploration. One interpretation of the edu-metaverse is that it acts like the yellow pages of connected educational concepts and it is visualized as an exploring world in virtual reality (VR) for its immersive benefits [1]. Students and educators alike can enter an immersive environment to discover and discuss different concepts in their discipline.

The driving forces railing the development of engaging education interactions between instructors and students in a metaverse environment stem from the need to expand access and enhance the convenience of learning processes leading to large-scale collaborative learning (CL) processes delivered at lower cost points. This can lead to disruptive transformations in educational experiences and a more streamlined standard content that is both accurate and consistent across all subjects and all fields of study [2].

A metaverse-based future of education will most likely be a hybrid model consisting of online and physical learning spaces that complement each other in ways that are not currently possible in the traditional modes of education and learning. This leads to a primordial need for research in physical spaces that embody virtual or augmented access to fundamental concepts and topics of study, ideas, and exploratory tasks, much like air pilots' training in modern flight simulators. The benefits of a metaverse learning experience are multifaceted and include an immersive learning interaction, risk-free skill applications, the "gamification" of learning tasks, and accelerated or agile learning.

In addition, CL and personalized learning (PL) are emergent pedagogies in the e-learning community. The metaverse is well-poised to fulfill this role. To drive the edu-metaverse toward this vision, we propose a naturally interactable virtual teaching assistant (NivTA) that is informed by a knowledge graph (KG) of concepts and topics. To further encourage a collaborative environment for learning, the virtual tutor is deployed in a CAVE virtual environment (CAVE-VR). The CAVE-VR serves as a "*metaverse window*" through which students can see a life-sized virtual teaching assistant that can naturally interact with them. Although the virtual assistant is empowered by a large language model (LLM), the students can naturally point at an education concept and ask questions about it, and the LLM tutor will better anticipate their points of interest.

In the CAVE-VR environment, it is easier to deploy various exploration modes of interactions for learning from basic traditional 3D visual aids to head-mounted displays (HMD) and augmented reality (AR) interfaces to CAVE-VR environments [3]–[5]. In this environment, multiple agents can co-

This work was supported in full by The Hong Kong Polytechnic University, Strategic Importance Scheme, project code: 1-ZE2N *(Corresponding author: Ye Jia.)*

The authors are with the Hong Kong Polytechnic University, Hong Kong SAR (e-mail: aimmeng-ye.jia@polyu.edu.hk; zackary-p-t.sin@polyu.edu.hk; xiangzhi-eric.wang@polyu.edu.hk; richard-chen.li@polyu.edu.hk; peter.nhf@polyu.edu.hk; xiaohuang@comp.polyu.edu.hk; hanson.dong@connect.polyu.hk; yaowei.wang@polyu.edu.hk; csgeorge@polyu.edu.hk; jiannong.cao@polyu.edu.hk; csqli@comp.polyu.edu.hk).

exist and interact with the immersive virtual content developed based on KGs. This can enhance virtual agents guidance of students through complex concept explorations. NivTA can also be used to further enhance the activity in these VR environments. Briefly, the contributions of this work are: (1) a new pedagogical methodology for a naturally interactable teaching assistant embedded in an immersive CAVE-VR; (2) a new design and implementation of the three-way interaction between students, virtual assistant, and 3D visual concepts; (3) a preliminary user study of the naturally interactable virtual teaching assistant (NivTA).

II. Background

With the advent of ChatGPT the aggregation of knowledge has far exceeded in speed, quantity, and quality of knowledge, surpassing predecessors such as Alexa, Siri, and Google Assistant. What values will this bring to higher education? The 3D-KG developed in the spirit of Tim Berners-Lee [6] and formal methods by Wang [7] for the development of the Semantic Web can aid in the extraction of knowledge culminating in a fully immersive education metaverse environment. Additionally, the educational process is significantly fueled by interactions with fellow learners. This aspect becomes especially pertinent when dealing with intricate and equivocal data. Prompt discussions among students foster the consideration of diverse viewpoints and aid in validating the most likely explanations, which is fundamental to CL.

Thus, the edu-metaverse for CL has the potential to become the primary mode of learning where multiple users can interact with each other and enhance their learning activities. Support for joint action requires co-actors to achieve and maintain a high degree of spatial and temporal coordination of individual contributions [8]. Workspace awareness is a critical aspect of collaboration in distributed scenarios, that can significantly affect collaborative performance [9], [10].

The body of research on the educational metaverse, or edu-metaverse, is extensive and varied, encompassing a range of perspectives on its implications and applications in learning environments. Key studies have focused on the transformative potential of the metaverse and the attendant challenges in reshaping educational experiences. Notable contributions include those by Kye et al., Zhang et al., and Stanoevska, who have each examined the metaverse's prospects for education [11]–[13].

III. Pedagogical Motivation

In this section, we will briefly discuss the pedagogy for the edu-metaverse. We address the question of why it necessitates a virtual teaching assistant and new modalities of interaction when operating in a CAVE-VR environment. Foremost, many educational discussions and development are based on *constructivism*.

Constructivism is the synthesis of multiple theories diffused into one form. It combined both behaviorist and cognitive [14]. The key idea of constructivism is that learning is a process of constructing meaning; it is how people sense their experience rather than just receive information from the teacher [15].

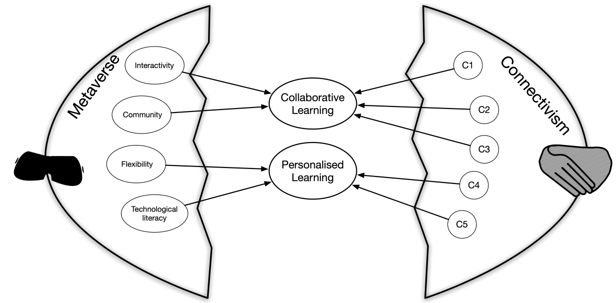

Fig. 1. The proposed edu-Metaverse learning principles with connectivism

Connectionism is a conceptual framework that views learning as a network phenomenon influenced by technology and socialization [16]. Its epistemological framework is based on the concept of distributive knowledge and it is distinct from the epistemological traditions of objectivism, pragmatism, and interpretivism [17]. Connectivism goes beyond the limitations of constructivism. For example, it explains why and how learning happens beyond the common consciousness of the people, that is, learning that can be stored and manipulated by technology and computational entities [18].

We explore new modalities in the edu-metaverse enchanced by constructivist principles as explored in the works of [1], [4]. This leads to further investigation into the integration of connectionism with its pedagogical framework. The edu-metaverse facilitates knowledge sharing, CL, and distance education, leveraging the CAVE-VR environment to create a deeply immersive learning experience distinct from traditional online models. Connectionism, with its focus on networks and technology-driven learning processes, offers a different but coherent view of the learning paradigm within the edu-metaverse.

CL [19] and PL [20] are pedagogies that have been gaining traction in educational methodology research in recent years. They are innately linked with connectionism. In Figure 1, we illustrate how CL and PL can be realized in the edu-metaverse. Ideally, the edu-metaverse would include both CL and PL. A naturally interactable virtual tutor in an immersive CAVE-VR is one way to realize these two methodologies simultaneously. A virtual tutor that is interactively engaging can enhance the attention of students onto the relevant questions related to a specific topic and prepare to answer them with various degrees of detail and depth. In the meanwhile, a CAVE-VR environment also allows multiple students to engage the tutor together, thereby providing an opportunity to learn collaboratively. We further provide details on CL and PL in the following sections.

A. Collaborative Learning

CL can be effectively described as a synchronized effort to jointly build and maintain a mutual understanding of a problem, as discussed in the works of Roschelle [21] and enhanced by the principles of connectivism outlined by Alam [22]. These principles align well with the dynamics of CL.

(C1) Networked Learning: Essential to connectionism, this involves forming networks that include peers, resources, and technologies, facilitating knowledge co-creation.

(C2) Openness and Sharing: Key tenets of connectionism that promote the free exchange of knowledge and resources within the network, enriching collaborative efforts.

(C3) Diversity: The variety of viewpoints within a network enhances the learning process, making it crucial for effective CL.

In the metaverse, CL is enriched by features like Interactivity, which supports dynamic engagement with the environment and virtual objects, and Community, which fosters a socially immersive environment essential for collaborative dialogue and communication. Together, these features make the metaverse an effective platform for CL, utilizing the strengths of constructivism and connectivism to enhance educational experiences in the digital age.

B. Personalized learning

PL in educational environments utilizes technology to cater to individual learning styles and paces, enhancing educational outcomes. Within the framework of connectivism, this approach emphasizes adaptability and a learner-centric focus, crucial for education in the digital age [23]–[25]. In the metaverse, "flexibility" describes the ability of virtual environments to dynamically adapt to diverse educational needs [26], allowing for real-time customization of learning experiences to match individual preferences and requirements. This flexibility supports a variety of interactive and immersive learning activities, enhancing understanding and engagement.

Technological literacy is essential in the metaverse, enabling learners to navigate, interact with, and create within these environments, thus supporting PL by allowing users to adapt educational experiences to their unique needs.

(C4) Self-directed Learning: Encourages learners to direct their own learning paths, setting personal goals and devising strategies to achieve them, a core aspect of connections. It encourages learners to take control of their educational journey, setting personal goals and developing strategies to achieve them, thus personalizing the learning experience to align with individual interests and objectives.

(C5) Lifelong learning: A key aspect of connectivism, underscores the importance of continuous learning to adapt to evolving information and contexts. This ongoing personal adaptation is essential for learners to remain relevant and effective in their educational pursuits, ensuring that education continuously meets individual needs and adapts to the dynamic digital landscape.

IV. Related Educational Systems

The NivTA system is based on two components that could be used as independent systems on their own. Particularly, K-Cube is a crowdsourcing educational knowledge base that structuralizes educational concepts as a KG. PETER, on the other hand, is a prototype for a virtual tutor.

A. K-Cube: A Crowdsourcing Educational Knowledge Graph

The traditional structure of university education, while beneficial, can often be rigid and may not fully accommodate the dynamic nature of individual career progression that many students desire. Moreover, the task of organizing academic knowledge within a discipline poses a challenge due to the difficulty in visualizing the complex interplay of various topics, limitations, and goals for both teachers and learners. To tackle this issue, we previously proposed K-Cube, a system designed to enhance the management and visualization of knowledge in a multi-faceted environment [1], [4]. In particular, K-Cube transforms the content of courses into a structured knowledge network. This network, or KG, encapsulates the fundamental concepts, terms, and subjects of a course by representing them as interconnected nodes.

With a web interface, instructors have the opportunity to refine their course planning each semester by utilizing the most recent KGs tailored to their courses. They are also equipped with the flexibility to personalize these graphs to better suit their teaching objectives. A specialized graph editing tool is made available to educators, enabling them to update and share the revised KGs with their students.

K-Cube automates the creation of structured knowledge from textbooks by developing a tailored ontology for guiding KG construction in VR educational environments. For each course, we start with 'seed entities'—key concepts identified by educators. Using primary sources like textbooks and Wikipedia, we expand the KGs around these entities. A relation extraction model then pulls new triples from relevant text, aligning and filtering them by similarity scores to ensure essential information is captured with minimal redundancy.

In this study, we based on the K-Cube VR to develop the NivTA in Edu-Metaverse.

B. PETER: A LLM Tutor

To develop an LLM tutor, PETER (Proactive Educational Dialogues: Tailored Engagement & Reflection) utilizes Panopto to capture lectures, including slides and speech-to-text transcripts. It employs ConvAI, an LLM tool trained on this content and deployed through Unity WebGL, to craft a tailored educational tool.

Developing a virtual tutor for higher education involves a multi-step process that integrates cutting-edge technologies and platforms to create a personalized and interactive learning experience. First, a total of 21 hours of lectures (3 hours across 7 lecturers) were recorded using Panopto. This platform not only captures the video but also integrates slides and generates transcripts, serving as primary inputs for training the virtual tutor. Then, via a system prompt, the virtual tutor's name and backstory are defined. This narrative foundation enables the AI to contextualize and reference its backstory in relevant conversations, enriching the interaction with students. To strengthen the context of the LLM agent, text-based knowledge, including the transcripts of the course's lectures stored in text files, is inputted to create a knowledge bank. This bank enables the virtual tutor to answer questions related

Fig. 2. The LLM-based tutor, PETER, is trained by the transcript of previously recorded lectures

to the lecture content, providing a rich resource for students. To further improve the question-answering capability of this tutor, critical queries regarding course logistics, such as course and assignment structures, were logged. By recording student questions and analyzing trends, educators can periodically refine the knowledge bank, ensuring the virtual tutor remains a relevant, valuable resource that comprehensively addresses the spectrum of student inquiries.

V. NivTA: Naturally Interactable Virtual Teaching Assistant

The focal point of the proposed work is to develop a system that can support a virtual teaching assistant that can interact naturally and is infused with educational knowledge to share with the students. To make it easily accessible and enable multiple students to use the students together in a collaborative setting, it is proposed that a cave automatically virtual environment is a suitable XR device to offer the interface. It utilizes K-Cube to provide a crowd sourced knowledge base to infuse to the virtual tutor. Further, based on the experience with PETER, we have developed a new tutor suitable for natural interaction. NivTA, however, is a work in progress; as such, we will first share the overall design and architecture that have been laid out. We will then discuss the current progress of our implementation and the consideration that other scholars should take note if they are developing a similar LLM-empowered system. Last, we will discuss the potential applications and the future work we are moving towards.

A. Natural Interface Design

The design of NivTA is driven by the aim to realize a virtual teaching assistant that can be naturally interacted with as if it is like a real person. In order to achieve this, it is important to design the natural user interface [27] based on how people interact with one another. Here, we will discuss some of the particular interactions that can be mimicked. It is noted, however, although it will be ideal to replicate human interactive function, it is a challenging endeavor that will require extensive research. As an intermediate solution that can reason as a proof of concept that can also drive preliminary investigation on NivTA, we can simply select some key human interaction that is relatively easy to implement while having a noticeable impact on immersion.

1) Speech: Of course, speech itself is an important mode of communication, and even more so in edu-metaverse as it is usually the case that an immersive XR environment has difficulties offering a keyboard for text-based communication. Thus, oral speech will be the main method of communicating, particularly, when relatively complex semantics need to be conveyed. In our case, the student can directly ask questions for NivTA to answer (Fig. 3*a*). However, it is should be noted that communication can be a mixture of signals from different forms as well [28].

2) Gesture: Prior to language, gesture is important for communication [29], and gesture is still a very important tool for us to convey messages. The most obvious gesture in our context is that the user can point at a point of interest to infer what the user is referring to. For example, the user can simply point at a node in the KG and ask "What is this?", and NivTA will be able to know that the user is asking the question about the node (Fig. 3*b*).

3) Gaze: When people are communicating, we usually look into the eyes. In addition, gaze can also serve as a mode to signal some information to other people [30]. As such, it is proposed that NivTA can be designed to occasionally look at the users to make the conversation between NivTA and the users more authentic (Fig. 3*c*).

4) Spatial: Space is another factor to consider during human interactions [31]. It is proposed that NivTA can also take into consideration the spatial location of the user to determine what interaction to perform. One such example is greeting. When we approach another person, it is a courtesy to greet him/her when s/he is approaching. As such, NivTA can also initiate a conversation when the user is coming into the CAVE (Fig. 3*d*).

B. K-Cube CAVE: Immersive Knowledge Exploration

In our previous work [1], we investigated on K-Cube VR, an edu-metaverse based on K-Cube. It is a primarily VR application that is accessible with VR HMD. Here, as we aim to place NivTA in a CAVE system, We have transferred the edu-metaverse application for CAVE deployment, which we separately refer to as K-Cube CAVE since the user control scheme is different. Briefly, within K-Cube CAVE, the user will face a visualized educational KG, providing an interactive and immersive interface to explore and learn educational concepts.

In the immersive CAVE, students will encounter a visual array of interconnected nodes. Courses are symbolized by cube-shaped nodes, while the associated keywords of a course are depicted by star-shaped nodes. The lines connecting these nodes signify the hierarchical relationship between them, with one being a subtopic of another. Additionally, a circular boundary delineates the domain of each course.

In addition to rotating the viewing angle, users can move within this virtual space using node-teleportation which allows users to directly transport themselves in front of a selected

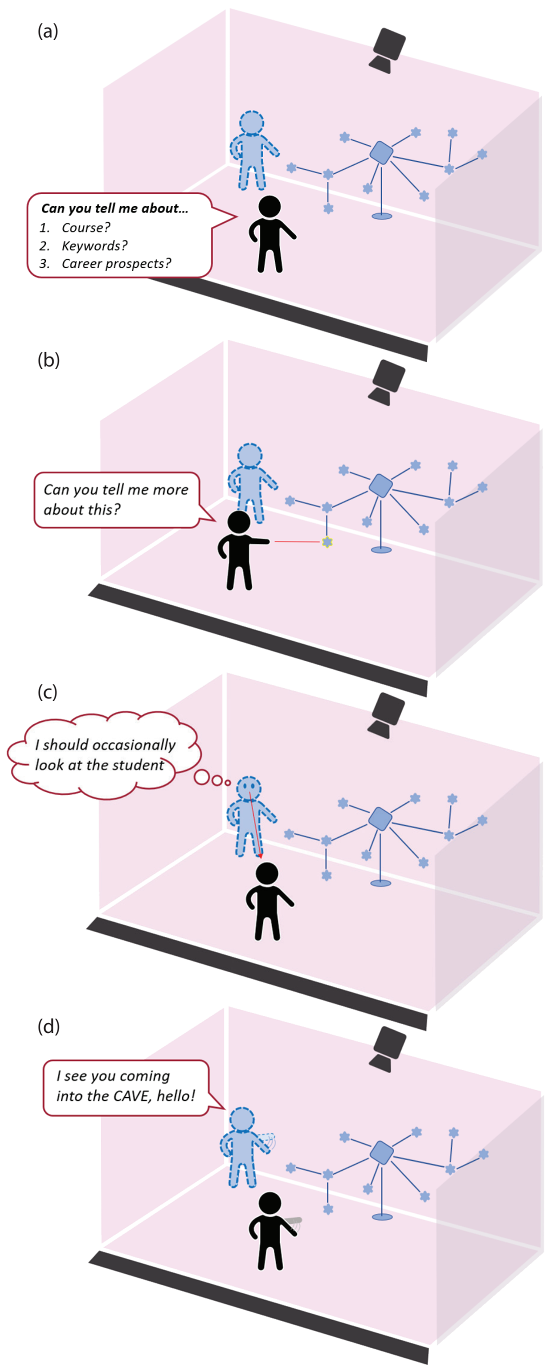

Fig. 3. The goal of NivTA is to use natural interactions such as (a) speech, (b) gesture, (c) gaze and (d) spatial and a mixture of communication modes to engage with the users.

node by pointing at it and activating the controller's trigger. This method also reorganizes the graph layout, bringing the child nodes into the user's field of view, which facilitates the exploration of related subtopics.

Grasping the key terms and principal ideas of a course is crucial for students to form a general understanding. To aid in this, multimedia resources are linked to each keyword. These include lecture slides and related Wikipedia entries. When a user targets a keyword node, the corresponding lecture slide appears, and they can navigate through the slides using the thumbstick. Accessing the Wikipedia page is as simple as pressing a designated button. Users can also pin nodes to keep the content visible and move it to a convenient location for persistent viewing.

The introduction of K-Cube CAVE serves two purposes. First, it provides context to the user to interact with NivTA. Instead of making enquiry based on the learner's knowledge, s/he can simply find topics within K-Cube CAVE to discuss with NIVAT. Second, the KG of K-Cube CAVE also provides context for NivTA to discuss the educational concepts. We will further discuss how prompts are driving NivTA to answer coherently based on the KG.

C. Implementation

The current implementation of NivTA focuses on investigating how to integrate the LLM model, virtual avatar, and K-Cube CAVE-VR in such a way that enables a more natural interaction between the user and the virtual tutor. Fig. 4 provides a system overview of NivTA. For the LLM model, we opted to use OpenAI's GPT 3.5, and naturally, the LLM model is used to generate answers given user query. Additional information on the query such as where the user is pointing at is provided by K-Cube CAVE. The virtual avatar, on the other hand, is designed to act authentically. So, it is designed to occasionally look at the user when answering his/her question, and play relevant animations. Further, it can greet the user when s/he is approaching the CAVE.

To interface with the user, the CAVE-VR has a motion capture device to track the head (via eyeglasses) and hands (via controller) of the user. The tracking of hands allows the user to point at an object to ask a question. The tracking of the user's head informs the virtual avatar on where to look. A speech-to-text model is also used to convert the user's speech into text. The text is then combined with additional information (such as where the user is pointing) and sent to the LLM model in JSON format. When the LLM replies, a text-to-speech model is used to act as the voice of NivTA. The Fig.5 demonstrates how user can interact with the NivTA.

D. Prompt Engineering

The interaction between LLM tools and students is significantly shaped by the initial prompts set for the LLMs. Various styles of prompts can also influence the quality of responses generated by LLMs. As indicated by [32], prompts that include question-and-answer (Q&A) examples tend to be particularly effective.

In developing our edu-metaverse system, we strategically crafted four Q&A examples based on the most likely question scenarios anticipated from users. These scenarios include: 1) general inquiries about the KG where no specific node is mentioned, 2) questions directly related to the node the user is interacting with, 3) queries that are unrelated to the learning content or the KG, and 4) questions concerning the identity of our virtual assistant, NivTA. These assumptions guided

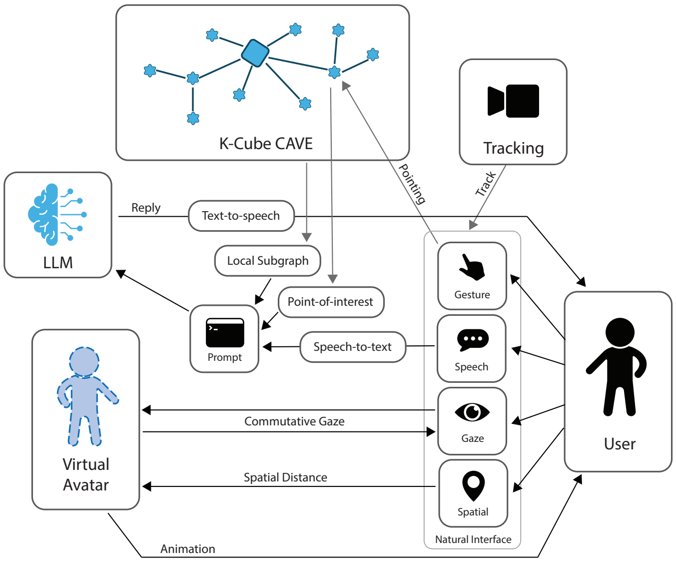

Fig. 4. Nivta integrates LLM agent, tracking system, virtual avatar and K-Cube CAVE-VR.

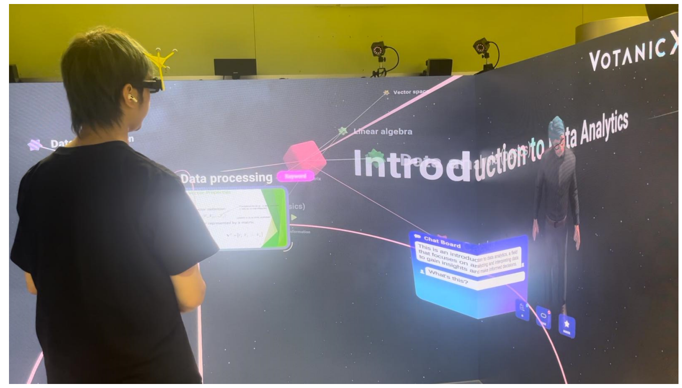

Fig. 5. A user (left) currently engaging with, NivTA, the naturally interactable virtual teaching assistant (right). He can engage with the virtual tutor with natural interaction such as combining speech with gesture. For example, pointing at a concept and asking NivTA what the concept is about with "What's this?".

us in creating tailored Q&A examples that help steer the responses of the LLM, ensuring that they are relevant and contextually appropriate for the user's educational experience in the edu-metaverse. Furthermore, to make NivTA even more exclusive to our edu-metaverse, we also create our customized knowledge base which includes all the nodes' relationships of KG in edu-metaverse for the LLM.

Building upon the above prompt structure, we further explore the impact of different client-to-LLM input formats on the naturalness of the LLM's responses. Currently, we offer two user input formats. The first format sends only the name of the node currently selected by the user. The second format is more comprehensive, including not only the name of the selected node but also the names of its child nodes and the course to which the node belongs. We believe these variations in input could affect how naturally the LLM responds.

In scenarios where the LLM receives only the current node's name, the responses can sometimes seem less natural. In contrast, a real teaching assistant would understand not just the queried node but also its relationship to adjacent nodes. While we have partially mitigated this limitation through a customized knowledge base, a new challenge emerges as the KG expands: increased input complexity can slow down the LLM's response time, as detailed in Yuan et al. (2024) [33]. As the KG grows more complex, this decrease in response speed becomes more pronounced, rendering our first input format merely a temporary solution.

To better support the scalability and functionality of our educational metaverse platform, we propose adopting the latter user input structure. This approach, by incorporating all relevant information directly into the user input, obviates the need for the LLM to rely extensively on external knowledge bases for context. This comprehensive approach not only enhances the LLM's efficiency but also improves the naturalness of its responses, as it aligns more closely with how a human assistant would process such queries.

To evaluate the effectiveness of these two input formats, we have designed a pilot user study to assess the qualitative user experience. This study aims to gather direct feedback from users interacting with both input formats, providing insights into their preferences and the perceived naturalness of the LLM's responses.

VI. Pilot User Study

To evaluate the user experience of our current proposed NivTA, we conducted our user study with two input forms. The objective of this user study is to provide us the feedback from the user as well as the system's usability. After the user experience. A focus group discussion was conducted to collect qualitative feedback. The task of the user study is to explore the KG with the help of NivTA.

1) Comparative User Input To LLM: As mentioned above (refer to V-D), this user study has two forms of under-input. Besides that, all settings are the same. LLM will receive two types of user input, one contains the current user-selected node. Another contains the user's current pointing node's name its children nodes and the course that it belongs to.

2) Participants: Four participants with expertise in VR and E-learning were invited to attend this user study. They were randomly divided to use two forms of user input to explore the KG and interact with the KG. Node name-only group with one male and one female. Another group with one male and one female as well. We asked them to try to use this KG and under the assistance of NivTA helped to grasp the overview of the course.

3) Equipment: The experiment was conducted in our laboratory. The equipment used to deliver NivTA was a CAVE immersive VR system, an X-Box wireless controller, and a mic attached to the ceiling in a camera.

4) Procedure: The procedure of the experiment was: *(1)* introduce how to use the CAVE Immersive VR system; *(2)* explain to the participant how to interact with NivTA; *(3)* let participant familiarizes with the VR navigation and chat with NivTA. Then reset the application when they were ready; *(4)* explain the goal of the experiment; *(5)* allow the participant to interact with NivTA; *(6)* ask them to fill in the useability questionnaire and *(7)* conduct the focus group discussion.

5) Measures: We are interested in understanding the user experience and how the prompts affect the learning experience. To this end, we utilized the System Usability Scale (SUS) [34]. Additionally, after participants completed their sessions, we conducted a focus group to gather more in-depth feedback.

6) Focus Group: Semi-structured focus groups were conducted to collect the participants' personal feelings about the content and presentation. The interview questions were asked about their Initial Impression, usability, NivTA's Understanding and Accuracy, Response Time, Navigation Help, Communication Style, Engagement, and Learning Experience.

VII. Results and Discussion

In this section, we present the result of our pilot user study. Generally, despite its preliminary implementation, there is some positive feedback from the expert users.

1) NivTA is good at summarizing: NivTA is acknowledged for its proficiency in summarizing content within the edu-Metaverse. However, opinions vary slightly in detail. Student(01) comments, *"it helps a lot, although it misses some content in detail"*, suggesting that while NivTA is generally effective, it sometimes overlooks finer points. On the other hand, Student(02) observes, *"it is weak in presenting generalized structures and concepts but excels in detailed aspects"*, highlighting strength in detailed analysis despite weaknesses in broader conceptual frameworks. This divergence in views is intriguing and may stem from differences in how they interact with NivTA. Specifically, the student who perceives NivTA as better at providing an overview typically in its inputs containing the names of the child nodes of the currently selected node along with the source name. In contrast, the second student inputs only the name of the currently selected node. This variation in user input likely contributes to their contrasting opinions, though both students agree that NivTA is effective at summarizing. This indicates the prompt structure can potentially largely affect the learning experience in edu-metaverse.

2) NivTA leads to a high Engagement: All students concur that NivTA facilitates their engagement within the virtual environment. Student(03) remarks, *"I think it's fine when I need it, and if it answers something more interesting, I'll use it more"*, indicating a desire for more captivating responses to increasing usage. Meanwhile, Student(02) notes, *"I think its interactivity is not particularly strong; perhaps adding some clear prompts to different nodes could help"*. Besides them, all students suggest that enhancing NivTA with more intuitive hints could significantly improve their interaction experience. Despite these suggestions for improvement, they unanimously find NivTA to be quite intriguing.

3) NivTA has a notable natural interaction ability: The students particularly emphasized that the interaction with the virtual agent feels very natural and realistic. Student(01) commented, *"The interaction is relatively strong, even standing there will have the action of shaking the head, and it feels like there is a real person standing next to you for questions, not just a virtual object."* Meanwhile, Student(03) remarked, *"It's a brand new thing. It's really practical when you run into problems."* Hence, our proposed NivTA is recognized for its notable natural interaction ability. However, it is worth noting that sometimes its responses are perceived as too formal, as suggested by Student(01).

4) Results of SUS: The average SUS score of NivTA is 60, which indicates that our proposed prototype offers a fair level of usability. Although this has not reached the 'Good' level as indicated by [35], this score seems to generally align with other prototypical systems.

VIII. Potential Application: Lecture Reenactment

Other than acting as a virtual tutor for introducing and discussing educational concepts, NivTA can also be used as a system to reenact lectures. Many lecture halls in higher education institutions already have cameras installed. Thus, it is possible to capture the pose and speech of the tutor during a lecture. To this end, we propose that NivTA can be fed with a recording of the lecture and reenact past lectures, replicating both pose and speech; this will provide a more immersive experience compared to videos of the recorded lecture. Further, the students can also ask questions about the reenacted lecture, and NivTA, again, should be able to answer them based on the context provided by the recorded lecture and the K-Cube KG (refer to Fig. 6). It is also noted that unlike typical lectures where students may be reluctant to ask questions due to concern about interrupting the flow of the lesson, students should have no such concern when interacting with a virtual tutor.

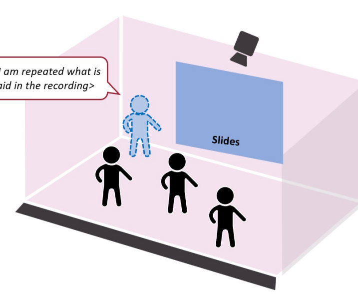

Fig. 6. NivTA can be used as a system for lecture reenactment, providing an immersive revisit of previous lectures while enabling students to interrupt with questions.

IX. Future Work and Conclusion

Despite some early encouraging results, improvements in user tracking and reactive content generated by our virtual assistant can be further explored. Our current implementation of NivTA relies on a tracking controller and tracked eyeglasses to infer the general location of the user's hands and head. This configuration, however, limits some of the modalities in user interactions. In the future, we plan to utilize RGBD tracking cameras such as Azure Kinect to better track the pose of the user in a CAVE-VR environment. This will enable more natural interaction such as the ability to wave at NivTA to greet the virtual assistant and use the user's actual hand for pointing at objects of interest within K-Cube CAVE-VR. However, it is noted that this will require synchronization between the motion capture cameras and the RGBD cameras as otherwise they will interfere with each other.

To conclude, we have proposed a new design for CAVE-VR educational environments, called NivTA, a naturally interactable virtual teaching assistant. This virtual tutor can interact naturally as if it is a real person. For example, the student can simply point at an object and ask the virtual tutor what it is. Based on our previous experience on educational KGs and LLM-based tutor, we have developed a preliminary prototype that showcases the potential of this edu-metaverse agent. NivTA can be used as an easily accessible virtual tutor for explaining educational concepts to the user. We can also configure it for lecture reenactment in the future. Expert users in e-learning and VR believe that NivTA can become a life-like agent to communicate with. This shows the potential for NivTA to act as a substitute teaching assistant. In the future, we will implement a better pose capture system and generally improve the natural interaction in NivTA to further improve realism and collaboration in the edu-metaverse.

References

[1] Z. P. Sin, Y. Jia, A. C. Wu, I. D. Zhao, R. C. Li, P. H. Ng, X. Huang, G. Baciu, J. Cao, and Q. Li, "Towards an edu-metaverse of knowledge: Immersive exploration of university courses," *IEEE Transactions on Learning Technologies*, 2023.

[2] P. H. Ng, P. Q. Chen, Z. P. Sin, Y. Jia, R. C. Li, G. Baciu, J. Cao, and Q. Li, "From classroom to metaverse: A study on gamified constructivist teaching in higher education," in *International Conference on Web-Based Learning*. Springer, 2023, pp. 92–106.

[3] Q. Li, G. Baciu, J. Cao, X. Huang, R. C. Li, P. H. Ng, J. Dong, Q. Zhang, Z. P. Sin, and Y. Wang, "Kcube: A knowledge graph university curriculum framework for student advising and career planning," in *International Conference on Blended Learning*. Springer, 2022, pp. 358–369.

[4] Z. P. Sin, I. D. Zhao, A. C. Wu, R. C. Li, P. H. Ng, X. Huang, G. Baciu, J. Cao, and Q. Li, "Towards a metaverse of knowledge: a constructivist proposition on immersive learning and visual literacy," in *International Conference on Web-Based Learning*. Springer, 2022, pp. 214–225.

[5] J. C. Chan, Y. Wang, Q. Li, G. Baciu, J. Cao, X. Huang, R. C. Li, and P. H. Ng, "Intelligent instructional design via interactive knowledge graph editing," in *International Conference on Web-Based Learning*. Springer, 2022, pp. 41–52.

[6] J. Heldler, T. Berners-Lee, E. Miller, and R. Masuoka, "Integrating applications on the semantic web," *Journal of the Institute of Electrical Engineers of Japan*, vol. 122, no. 10, pp. 676–680, 2002.

[7] Y. Wang, "A formal theory of ai trustworthiness for evaluating autonomous ai systems," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2022, pp. 137–142.

[8] N. Sebanz and G. Knoblich, "Progress in joint-action research," *Current Directions in Psychological Science*, vol. 30, no. 2, pp. 138–143, 2021.

[9] F. Yu, P. Zhang, X. Ding, T. Lu, and N. Gu, "Exploring how workspace awareness cues affect distributed meeting outcome," *International Journal of Human–Computer Interaction*, vol. 39, no. 8, pp. 1606–1625, 2023.

[10] A. Prouzeau, A. Bezerianos, and O. Chapuis, "Awareness techniques to aid transitions between personal and shared workspaces in multi-display environments," in *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, 2018, pp. 291–304.

[11] B. Kye, N. Han, E. Kim, Y. Park, and S. Jo, "Educational applications of metaverse: possibilities and limitations," *Journal of educational evaluation for health professions*, vol. 18, 2021.

[12] X. Zhang and Y. Chen, "The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics," *Frontiers in Psychology*, vol. 13, p. 1016300, 2022.

[13] K. Stanoevska-Slabeva, "Opportunities and challenges of metaverse for education: a literature review," *Edulearn22 Proceedings*, pp. 10 401–10 410, 2022.

[14] R. J. Amineh and H. D. Asl, "Review of constructivism and social constructivism," *Journal of social sciences, literature and languages*, vol. 1, no. 1, pp. 9–16, 2015.

[15] M. Larochelle, N. Bednarz, and J. W. Garrison, *Constructivism and education*. Cambridge University Press, 1998.

[16] G. Siemens, "Connectivism: Learning theory or pastime of the self-amused," 2006.

[17] S. Downes, "Learning networks and connective knowledge," in *Collective intelligence and E-Learning 2.0: Implications of web-based communities and networking*. IGI global, 2010, pp. 1–26.

[18] G. Siemens, "Elearnspace. connectivism: A learning theory for the digital age," *Elearnspace. org*, pp. 14–16, 2004.

[19] M. Laal and S. M. Ghodsi, "Benefits of collaborative learning," *Procedia-social and behavioral sciences*, vol. 31, pp. 486–490, 2012.

[20] A. Shemshack and J. M. Spector, "A systematic literature review of personalized learning terms," *Smart Learning Environments*, vol. 7, no. 1, p. 33, 2020.

[21] J. Roschelle and S. D. Teasley, "The construction of shared knowledge in collaborative problem solving," in *Computer supported collaborative learning*. Springer, 1995, pp. 69–97.

[22] M. A. Alam, "Connectivism learning theory and connectivist approach in teaching and learning: a review of literature," *Bhartiyam International Journal Of Education & Research*, vol. 12, no. 2, 2023.

[23] M. Montebello, "Personalized learning environments," in *2021 International Symposium on Educational Technology (ISET)*. IEEE, 2021, pp. 134–138.

[24] R. D. Caytiles and H.-j. Kim, "e-friendly personalized learning," *International Journal of Internet, Broadcasting and Communication*, vol. 4, no. 2, pp. 12–16, 2012.

[25] H. Smidt, M. Thornton, and K. Abhari, "The future of social learning: A novel approach to connectivism," 2017.

[26] N. Chuk, "Flexibility, flâneurie, and affinity in the metaverse," *Baltic Screen Media Review*, vol. 10, no. 2, pp. 235–251.

[27] J. Jain, A. Lund, and D. Wixon, "The future of natural user interfaces," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 2011, pp. 211–214.

[28] P. Bernardis and M. Gentilucci, "Speech and gesture share the same communication system," *Neuropsychologia*, vol. 44, no. 2, pp. 178–190, 2006.

[29] A. Meguerditchian, H. Cochet, and J. Vauclair, "From gesture to language," *Primate communication and human language*, pp. 91–120, 2011.

[30] M. Garau, M. Slater, S. Bee, and M. A. Sasse, "The impact of eye gaze on communication using humanoid avatars," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2001, pp. 309–316.

[31] M. Patterson, "Spatial factors in social interactions," *Human Relations*, vol. 21, no. 4, pp. 351–361, 1968.

[32] S. Arora, A. Narayan, M. F. Chen, L. Orr, N. Guha, K. Bhatia, I. Chami, and C. Re, "Ask me anything: A simple strategy for prompting language models," in *The Eleventh International Conference on Learning Representations*, 2022.

[33] Z. Yuan, Y. Shang, Y. Zhou, Z. Dong, C. Xue, B. Wu, Z. Li, Q. Gu, Y. J. Lee, Y. Yan *et al.*, "Llm inference unveiled: Survey and roofline model insights," *arXiv preprint arXiv:2402.16363*, 2024.

[34] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[35] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.