Towards AI-Assisted Immersive Learning: Factor Analysis of Learning Effect in K-Cube Edu-Metaverse

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

***Abstract*—This study examines the impact of an AI-powered Virtual Teaching Assistant (NivTA) within a VR-based Edu-Metaverse (K-Cube), highlighting the roles of social presence, trust, and engagement in shaping learning outcomes. Grounded in Social Presence Theory, the Uses and Gratifications framework, and the Cognitive-Affective Theory of Learning with Media (CASTLE), our AI-Assisted Immersive Learning Framework emphasizes both cognitive and affective dimensions.**

In a user study with 21 participants, we collected quantitative and qualitative data on trust, social presence, engagement, workload, and learning performance in a Cave Automatic Virtual Environment (CAVE) setting. Partial Least Squares Structural Equation Modeling revealed that heightened social presence fosters trust, which in turn drives behavioral, cognitive, and affective engagement. Notably, cognitive social presence was directly linked to better knowledge test scores, while confidence in test responses stemmed primarily from all forms of engagement. Overall, these findings underscore the significance of nurturing trust and social presence to enhance learner engagement and outcomes in AI-driven immersive educational environments.

***Index Terms*—Learning environments, Virtual Reality, Data and knowledge visualization, Adaptive and intelligent educational systems, Cave Automatic Virtual Environment .**

I. Introduction

The integration of artificial intelligence (AI) teaching assistants in educational environments has transformed learning experiences, with significant advancements emerging at the intersection of Virtual Reality (VR) technologies and AI. As educational technologies increasingly incorporate immersive VR and knowledge graph (KG) visualization, understanding how these interaction modalities affect trust formation and learning outcomes becomes critical in designing effective educational experiences. The future effectiveness of these systems depends on users' engagement. Ultimately, the goal with generative AI assistance is to gain their full trust.

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: yeaimmeng.jia@connect.polyu.hk; richard-chen.li@polyu.edu.hk; zackary-p-t.sin@polyu.edu.hk; xiangzhi-eric.wang@polyu.edu.hk; jiongning.lian@connect.polyu.hk; peter.nhf@polyu.edu.hk; xiaohuang@comp.polyu.edu.hk;csgeorge@polyu.edu.hk; jiannong.cao@polyu.edu.hk; csqli@comp.polyu.edu.hk).

Recent research has established important foundations in this domain. Studies by Li et al. [1] have demonstrated that multimodal interaction with AI teaching assistants produces superior learning outcomes compared to unimodal approaches. Glikson et al. [2] revealed how different explanation styles in educational AI systems can influence distinct trust dimensions. Li and Kazemitabaar [3] highlighted the benefits of collaborative knowledge visualization in shared VR spaces, showing stronger inter-concept connections compared to individual visualization activities. Shang, Hsieh and Shah [4] made significant contributions with their semantic differential scale for measuring trust in educational AI systems, addressing previous limitations in trust measurement instruments.

The advancements in AI-powered educational technologies have led to the development of NivTA (Natural interaction Virtual Teaching Assistant) [5], an innovative system integrated within the K-Cube VR environment [6], which are detailed in setion IV. This VR platform serves as a "*metaverse window*" into the curriculum knowledge space, enabling students to interact with a life-sized virtual teaching assistant in a naturalistic manner. NivTA distinguishes itself through its multimodal interaction capabilities, incorporating speech, gesture recognition, and text-based communication. Powered by a large language model (LLM), the system facilitates intuitive learning experiences where students can physically point to educational concepts and receive immediate, contextualized responses. This natural interaction paradigm enables the LLM tutor to better anticipate and address students' areas of interest [6]. At its core, NivTA leverages KGexploration to enhance the learning process, creating a dynamic and responsive educational environment within the Cave Automatic Virtual Environment (CAVE) system.

In this study we investigate the critical learning factors that influence user engagement with NivTA and examine how these factors correlate with learning outcomes. Through this investigation, we seek to identify the key elements that contribute to effective AI-assisted learning in immersive environments. However, given the acceptable usability level and the positive feedback from the previous pilot user study, we have not addressed the effects of the NivTa on the learner perspective. According to the prior research on the relevant factors that

2993-396X/25/$31.00 ©2025 IEEE
DOI 10.1109/SWC65939.2025.00278

may influence the learning process, include trust, engagement, learning satisfaction, and the sense of presence. To fill in this gap, in this study, we focus on investigating the main factors that directly affect the learning outcomes.

The main contributions of this study are:

- A new generative AI-assistant Immservise Learning Framework (AI-ILF), that explain the learning process both at the cognitive level and the affective level;
- A user study that shows the potential factors that affect learning in the Edu-Metaverse.

II. Related work

A. Trust research with LLM in Education

The integration of LLMs in educational settings has introduced transformative opportunities while simultaneously challenging traditional trust dynamics between stakeholders.

Key determinants influencing trust in LLM technology include perceived credibility, accuracy, prior AI experience, and the perceived impact on productivity [7]. Cross-cultural studies highlight variations in trust, influenced by factors such as regional disparities, gender, academic level, and socioeconomic background. For students specifically, trust remains a central challenge as they navigate the reliability of AI-generated content [8]. Quantitative research establishes a strong correlation between trust and LLM adoption. Research on designing interactive pedagogical agents for higher education instructors emphasizes features that foster trust, particularly for AI-conservative faculty members [9]. The phenomenon of LLM hallucinations—where models generate incorrect information despite appearing confident—represents a major trust barrier in educational applications [10].

In summary, existing research emphasizes the pivotal role of trust in the adoption and effective use of LLMs in education.

B. LLM-based Virtual Teaching Assistant

Recent advances in LLMs have revolutionized educational technology through virtual teaching assistants. Several notable implementations demonstrate their potential in enhancing learning experiences.

Studies showcase diverse applications: Benedetto et al. [11] integrated IBM Watson with Slack for personalized teaching assistance, while Dong et al. [12] explored LLM integration in AR/VR environments. Hicke et al. [13] developed AI-TA using LLaMA-2, achieving 30% improvement in response quality through search-augmented generation. Further implementations include ChatGPT for programming course assistance [14], WeChat-integrated LLM teaching assistants [15], and LangChain-based systems for large-scale programming courses [16]. Maiti and Goel [17] demonstrated how LLM teaching assistants can support varied cognitive demands and encourage higher-order thinking.

Overall, these studies demonstrate the transformative potential of LLMs in education. Benedetto et al. [11] and Liu and M'hiri [16] emphasize scalability and personalization to meet the diverse needs of students in large courses. Hicke et al. [13] and Anishka et al. [14] emphasize the benefits of feedback for LLMs. Research by Wei et al. [15] and Maiti and Goel [17] suggests that LLM can enhance collaboration and critical thinking, albeit in different ways in different contexts. Together, these studies reveal the great promise of LLM in improving the quality and adaptability of education.

III. AI-Assisted Immersive Learning Framework

With the age of the LLM based AI agent, what is the pedagogy philosophy for the Edu-Metaverse? As we strive to optimize immersive learning experiences for enhanced effectiveness and efficiency, it becomes crucial to establish robust principles governing human-AI interactions in educational contexts. This section presents a comprehensive framework for AI-Assisted Immersive Learning, synthesizing multiple theoretical perspectives, as illisulated in Fig.1. Our theoretical foundation draws upon Social Presence Theory, Uses and Gratifications (U&G) Theory, Cognitive-Affective Theory of Learning with Media (CASTLE), Constructivism and Connectivism learning theories, as previously discussed in our theoretical framework [5], [6]. Through the integration of these theoretical perspectives, we can better understand how AI-enhanced immersive environments facilitate learning processes and shape educational outcomes.

A. Constructivism and Connectivism

Constructivism and Connectivism represent two significant paradigms in educational theory, each reflecting distinct technological and social contexts [18], [19]. Constructivism, pioneered by psychologist Jean Piaget, conceptualizes learning as an intrinsic cognitive process where knowledge is actively constructed through individual mental operations [18]. This theoretical framework emphasizes that learning is fundamentally a learner-centered process.

Connectivism complements traditional constructivist approaches by addressing their limitations in the context of evolving digital technologies [19]. It extends learning processes beyond individual cognition, integrating technological systems as essential to knowledge acquisition [19]. By recognizing that learning can occur within non-human systems and that knowledge is stored, manipulated, and distributed through digital networks, Connectivism highlights the interconnected nature of modern learning environments and the role of AI in constructing and disseminating knowledge.

AI-assisted immersive learning environments, grounded in Constructivist and Connectivist theories, combine VR/XR technologies for active knowledge construction [6] with adaptive AI agents that provide personalized guidance [5]. This integration creates an interactive ecosystem where learners engage directly with virtual environments while receiving individualized support.

B. Social Presence Theory

Social Presence Theory, initially developed for human communication in computer-mediated contexts [20], is now crucial

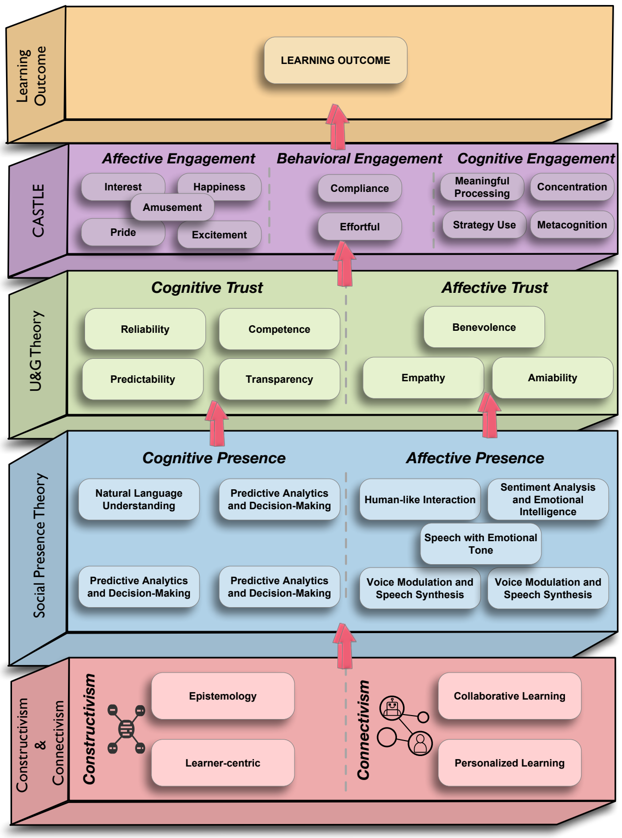

Fig. 1: Proposed AI-assistant Immservise Learning Framework (AI-ILF) incorporate Constructivism, Connectivism, Social Presence Theory, Uses and Gratifications(U&G) theory, and Cognitive-Affective Theory of Learning with Media (CASTLE)

for understanding AI-human interactions in education. It focuses on establishing cognitive and affective presence to foster meaningful learning experiences.

Cognitive presence ensures meaningful educational interactions. In immersive virtual environments, multimodal learning analytics (e.g., eye-tracking combined with AI dialogue) better capture students' cognitive patterns, improving educational tool design [21]. VR platforms integrating AI tutors (e.g., OpenAI GPT in Unity) enhance personalized language learning through real-time, immersive interactions [22]. However, challenges persist in online classrooms, especially during the COVID-19 pandemic, where cognitive presence is harder to achieve. AI teaching assistants must be designed to foster deeper interactions to address this [23].

Affective presence refers to emotional connections during interactions. Advanced AI voice synthesis systems like METTS enhance emotional expressiveness, improving engagement and motivation [24]. Features like emotional tones and custom voices tailored to specific users or educational contexts deepen the emotional resonance in AI communication [24]. Additionally, social networks often outperform traditional LMS platforms in establishing affective presence through stronger social connections [25]. Cognitive, affective, and teaching presence are interconnected, significantly influencing user engagement and learning outcomes [26]. AI teaching assistants designed with strategies like encouragement, breakout rooms, and personalized feedback can better address the limitations of virtual learning environments, improving their effectiveness in online education [23].

C. U&G Theory

U&G Theory explains how users interact with AI systems based on the benefits they receive. In education, it offers insight into developing trust in AI teaching assistants, encompassing both cognitive and affective dimensions.

Cognitive trust relies on perceptions of reliability, competence, and transparency. Research shows that explainable AI is crucial, as providing transparent explanations enhances user trust without adding cognitive load, overcoming the distrust often associated with "black box" models [27]. Even children develop cognitive trust when they understand how AI systems operate [28]. Affective trust focuses on emotional connections like empathy and benevolence, which can be fostered when children appreciate an AI's practical value and form positive attitudes toward it [28]. Furthermore, combining psychological insights with machine learning can enhance both trust dimensions [29]. While trust and engagement are interconnected [27], excessive trust can lead to over-reliance and reduced critical thinking [30]. Therefore, effective AI teaching assistants must balance trust-building with fostering critical engagement to improve learning outcomes.

D. CASTLE

The CASTLE framework focuses on three key dimensions of engagement—affective, behavioral, and cognitive—essential for effective learning experiences with AI systems [31].

Affective engagement involves emotional responses like interest and happiness that enhance learning outcomes. Research shows that embodied interactions, such as the Shaking-On app for Mandarin learning, improve emotional engagement and learning effectiveness compared to traditional methods [32]. Gamification in training programs also boosts emotional engagement and learning outcomes, suggesting similar strategies could enhance AI teaching assistants [33]. Behavioral engagement refers to active participation and effort in learning activities. Studies show that reducing distractions, such as multitasking during lectures, increases engagement and focus [34]. Gamification further enhances behavioral engagement by encouraging participation and effort, highlighting its potential for improving AI-driven education [33]. Cognitive engagement involves deep processing and analytical thinking. Over-reliance on AI can hinder critical thinking, emphasizing the need for AI tools that support rather than replace cognitive processes [30]. Reducing distractions and fostering focus promotes both cognitive and affective engagement, creating better learning environments [34]. Interactive AI tools also help students develop positive cognitive attitudes and deeper engagement with educational content [28]. The CASTLE framework highlights the need for AI teaching assistants to integrate affective, behavioral, and cognitive engagement strategies. Features like transparency, gamification, emotional expressiveness, and immersive virtual environments can enhance both trust and engagement, leading to more effective learning experiences [22], [24].

IV. K-Cube & NivTA

K-Cube is an Edu-Metaverse platform with the 3D knowledge Graphs (KGs), where individual can naviagting, leanring with the cource knowledge [6]. NivTA represents an innovative implementation of AI-assisted immersive learning within K-Cube Edu-Metaverse. NivTA's novelty lies in its integration of natural interactions (gestures, gaze, spatial awareness) within a CAVE, combined with KG-enhanced prompts, enabling life-sized, human-AI collaborative learning. Unlike text-based virtual assistants, NivTA allows users to point at educational concepts for context-aware explanations, leveraging immersive VR for realistic engagement while grounding responses in structured educational knowledge. The system integrates natural user interfaces, LLMs, and KGs to create an intuitive learning experience. Through multimodal interactions including speech and gesture, students can naturally engage with educational content [5].

The architecture combines OpenAI's GPT-4o for contextual response generation with a virtual avatar interface and sophisticated motion tracking capabilities. Through prompt engineering and KG integration, NivTA delivers personalized learning experiences while facilitating collaborative exploration in the virtual space. This approach exemplifies the practical application of immersive technology in education, enabling both individual engagement and group learning activities within the Edu-Metaverse environment, though future developments aim to enhance its natural interaction capabilities further.

(a) (b)

Fig. 2: (a) K-Cube Edu-Metaverse incorporated the KG and (b) The NivTA is interacting with a participant in the CAVE

V. Methodology

A. Experiment Design & Task

This study employed a within-subjects experimental design. The experiment task is a learning task that instructs the participant to learn the overview of Game Design and Development course within the K-Cube in CAVE system. This course is offered to year three to four undergraduate students as an elective subject carried out by the Department of Computing at Hong Kong Polytechnic University (PolyU).

B. Participant

The study recruited 21 participants (12 females, 9 males) from PolyU. The participants had a mean age of 23.81 years (SD = 4.18). To maintain methodological rigor, we specifically screened participants to ensure they had not previously taken courses related to the experimental task. Consents form were obtained from all participants. Participants were compensated with an HK$100 supermarket voucher. The study was approved by the Institutional Review Board of the PolyU (Application Number: HSEARS20250313009).

C. Measures

1) Trust: Our study employs a validated AI trust measurement instrument [4], comprising 27 semantic differential items that evaluate the cognitive and affective dimensions of trust. Each item is rated on a 5-point Likert scale, with opposing adjective pairs anchoring the endpoints [4]. ***2) Learning Outcomes:*** The knowledge test contains three multiple choice and two single choice questions. Additionally, for each knowledge test, we have a confidence test with a 5-point Likert scale. All the items are from our pervious work [6]. ***3) Engagement:*** The engagement questionnaire were adopted from Classroom Engagement Inventory (CEI) and adapted it to our VR context [35]. This questionnaire includes four sub-dimensions with including eight items of behavioral engagement, five items of affective engagement, eight items of cognitive engagement, and three items of disengagement. The adapted questionnaire assesses engagement using a combination of 5-point and 7-point Likert scales. ***4) Learning Behavioral:*** We used the whole time duration for each participant and also the conversation turns with the NivTA (One user question and one answer from NivTA is considered as one conversation turn) to measure the basic learning behavioral. ***5) Work Load:*** The workload was measured by the NASA Task Load Index (NASA-TLX) which divided the whole workload into six subjective subscales including Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustrations [36]. Each item is on a 10-point Likert scale. ***6) Social Presence:*** The social presence questionnaire to assess the social presence between user and the virtual teaching assistant was implied from Kim et al.'s work [37]. Total 20 items on a 5-point likert scale from 1 (strongly disagree) to 5 (strongly agree) were used. ***7) Qualitative Data:*** A Semi-structured interview was conducted after the experiment, the interview includes questions about: **a)** How effective was the AI teaching assistant in helping you understand and learn the course content? **b)** How did you feel about interacting with the AI using gestures and speech? Which modality worked best for you and why? **c)** Did you feel engaged and connected with the AI during the session?

What contributed to or hindered this experience? **d)** What improvements would you suggest for the AI teaching assistant or the VR learning environment?

D. Procedure

The experiment will follow these steps: **a) Pre-Test** Participants were completed demographic survey to collect their information. **b) Training Session** All participants had train sessions to familiar the VR interaction before the experiment session. **c) Experiment Session** All participants engaged with the K-Cube & NivTA for compleat the learning task, to minimize the time pressure, this session don't have the time limit. **d) Post-test** After experiment, participants were asked to fill in the Knowledge-test, Trust, Social Presence, NASA-TLX, and engagement questionnaire. **e) Semi-Structured Interview** Participants were asked to have the semi-structured interview.

VI. Results

A. Statistical Description Result

The experimental result encompassed multiple dimensions of assessment: Social Presence, Trust, Engagement, Knowledge Acquisition, Learning Behavioral, and Workload. For a detailed overview of the statistical analysis refer to Table I. Figure 3 displays a 3D scatter plot visualizing the normalized relationships between Presence, Trust, and Engagement, with Knowledge Score indicated by color intensity.

TABLE I: The comprehensive statistical results

Main type	Sub Type	Mean	SD
Social Presence	Affective	3.18	1.19
	Cognitive	3.56	1.04
Trust	Affective	2.29	1.16
	Cognitive	2.51	1.02
Engagement	Affective	3.42	1.04
	Behavioral	3.58	1.09
	Cognitive	5.06	1.38
	Disengagement	2.49	0.99
Knowledge Acquisition	Test Score	2.02	1.80
	Test Confidence	3.69	0.81
Learning Behavioral	Time Duration	35.39	16.62
	Conversation Turn	20.62	22.51
Work Load	NASA-TLX	5.92	2.20

B. Path Analysis

Given the study's limited sample size (n=21), Partial Least Squares Structural Equation Modeling (PLS-SEM) was selected as the primary analytical method for path analysis. PLS-SEM is particularly suitable for studies with small sample sizes, as it is less sensitive to sample size limitations compared to covariance-based SEM and does not require normal data distribution assumptions.

In alignment with our proposed AI-ILF framework, we examined the interrelationships among three primary factors influencing learning outcomes. To achieve a more nuanced understanding, we disaggregated the learning outcomes into

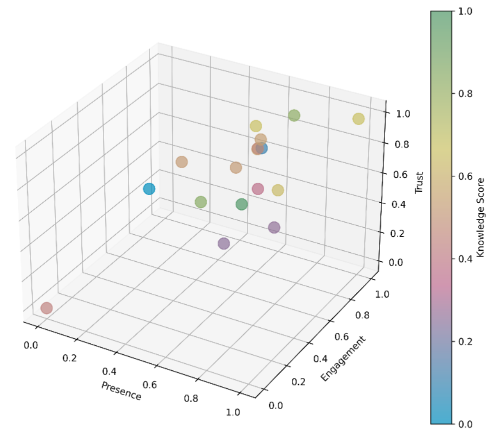

Fig. 3: 3D scatter plot visualizing the normalized relationships between Presence, Trust, and Engagement with Knowledge Score indicated by color intensity

two distinct components. Knowledge Test Performance and Response Confidence. The analysis framework focused on specific domains within each major factor. For the Social Presence and Trust, we focus on both cognitive and affective domains. Regarding the engagement, our analysis is focused on the cognitive, affective, and the Behavioral.

As shown in Fig 4, all identified paths in the model (11 total) were statistically significant ($p < 0.05$), indicating robust relationships among the variables despite the small sample size. The model demonstrated weak but acceptable fit with an overall R^2 of 0.279, which just exceeds the threshold for weak explanatory power (>0.25) according to established PLS-SEM criteria. Individual R^2 values for endogenous variables ranged from 0.206 to 0.370, indicating that the model explains a modest but meaningful portion of variance in the outcome variables.

The model revealed several important relationships. Regarding the Trust domain, the Cognitive Trust and Affective Trust can be significantly predicted by the Cognitive Social Presence (Path = p3, $\beta = 0.5994$, $p = 0.0045$) and Affective Social Presence(Path = p5, $\beta = 0.6285$, $p = 0.0034$), respectively. For the Engagement domain, the Cognitive Engagement was significantly predicted by the Cognitive Social Presence(Path = p2, $\beta = 0.5572$, $p = 0.0325$). The Affective Engagement were significantly predicted by the Affective Trust (Path = p6, $\beta = 0.8434$, $p = 0.0015$) and Affective Social Presence (Path = p4, $\beta = 0.6094$, $p = 0.0387$). The Behavioral Engagement were significantly predicted by the both Cognitive Trust (Path = p7, $\beta = 0.4997$, $p = 0.0375$) and Affective Trust (Path = p8, $\beta = 0.6402$, $p = 0.0040$). For the Learning Outcome, the Knowledge Test Score was significantly predicted by the Cognitive Social Presence(Path = p1, $\beta = 0.4188$, $p = 0.0268$). The Knowledge Test Confidence were significantly predicted by the all Engagement subdomains including Cognitive Engagement(Path = p9, $\beta = 0.6319$, $p = 0.0005$) , Affective

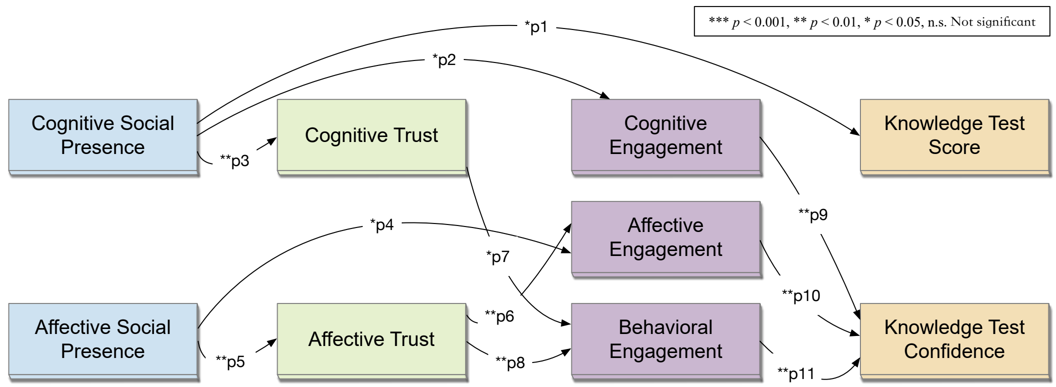

Fig. 4: Path analysis results

Engagement(Path = p10, $\beta = 0.4518$, $p = 0.0080$) , and Behavioral Engagement(Path = p11, $\beta = 0.5904$, $p = 0.0038$).

VII. Discussion

A. Social Presence is Important

Social Presence emerges as a crucial component for AI avatars in immersive educational environments. Our findings demonstrate that Social Presence facilitates higher trust levels, which subsequently enhances engagement and ultimately leads to improved confidence in knowledge assessment. Notably, our analysis revealed a direct effect of Cognitive Social Presence on Knowledge test scores, underscoring its fundamental importance in the learning process.

This relationship can be better understood through an illustrative case from our qualitative data. One participant's experience provides valuable insight into the dynamics of human-AI interaction. As the participant reported: *"When I asked he question, he give me the wrong answer, then I decided to not use with it"*. This statement reveals two significant phenomena: The participant's disengagement was triggered by the AI's failure to meet expected performance standards. The linguistic shift from treating NivTA as a social entity to referring to it as something to "use" suggests a degradation from social actor to mere tool. In the contrary, another participant stated that *" I feel connected to the Virtual TA (NivTA), because he was just stand there, (he) is very helpful, I trust him"*.

Through these two opposite altitudes we can find that social presence it matter to the Trust. This observation aligns with Social Presence theory, suggesting that the perceived humanness of AI behavior directly correlates with trust formation. When AI behavior closely mimics human interaction patterns, participants are more likely to engage in social rather than purely instrumental interactions. This human-like social presence appears to be a foundational element in establishing trust and maintaining meaningful educational interactions, rather than reducing the AI to a mere utilitarian tool.

B. Trust Boosts Engagement

Our analysis reveals that trust serves as a fundamental catalyst for enhanced engagement in AI-mediated learning environments. The relationship between trust and engagement manifests through multiple pathways, as evidenced by both our quantitative findings and participants' qualitative feedback.

Trust facilitates engagement through three primary mechanisms: **Information Reliability:** When participants develop trust in the AI's competence, they engage more readily with the system. As one participant noted, *"When I asked the question, it gave me the answers very quickly and very reliable. So, I don't need to go through the learning materials."* **Emotional Support:** Trust creates a psychological safety net that encourages continued interaction, particularly during challenging learning moments. This is illustrated by the participant comment: *"When I felt helpless, I could ask for the assistant's help."* **Interactive Motivation:** Trust in the system's capabilities promotes sustained engagement and exploratory behavior. As evidenced by one participant's experience: *"I personally feel I connect with the AI. Sometimes I just want to bring up other questions."*

Our results suggests that when learners establish trust in both the cognitive capabilities and affective responsiveness of the NivTA, they demonstrate increased willingness to engage deeply with the learning process. This trust-engagement dynamic creates a virtuous cycle that enhances the overall learning experience and potentially leads to better learning outcomes. This finding can be supported by the U&G Theory, which posits that users actively seek out and engage with media that satisfies their specific needs and expectations.

In the context of AI-enhanced learning, when learners' initial interactions with the NivTA fulfill their academic needs (cognitive gratification) and provide positive social experiences (affective gratification), they develop trust in the system. This trust, in turn, motivates them to engage more deeply with the NivTA, seeking both instrumental support for learning tasks and social-emotional satisfaction from the interaction. Thus our proposed AI-ILF incorporated the U&G theory is able to explain why trust serves as a crucial mediator between initial interaction and sustained engagement - it validates the user's expectation that continued interaction will yield both practical and psychological benefits, thereby maintaining the motivation to engage actively with the NivTA.

C. Learning Outcome

Our analysis of learning outcomes encompassed two distinct measures: Knowledge Test Performance and Knowledge Test Confidence. The results reveal a nuanced relationship between these outcomes and various factors in the AI-enhanced learning environment.

1) Knowledge Test Performance: The analysis revealed that Knowledge Test Performance was directly influenced by Cognitive Social Presence. This relationship can be understood through participants' qualitative feedback, particularly regarding the AI's role in content organization and understanding. As one participant noted, *"The AI can help me to determine which part to start first, and then it can give me the overview of each of the topics, and also help me to make a summary for each topic."* This suggests that when learners perceive strong cognitive presence from the AI teaching assistant, they are better able to grasp and retain the learning content.

2) Knowledge Test Confidence: Knowledge Test Confidence demonstrated significant relationships with all three dimensions of engagement - cognitive, behavioral, and affective. These relationships were supported by participant experiences, particularly in cases where positive engagement fostered learning confidence. However, limitations were noted when AI responses were inconsistent or overly simplistic, as illustrated by one participant's experience: *"... the answer (from NivTA) is that there is no gradient involved. But a few slides later, there is actually gradient updating."*

The findings suggest that while cognitive social presence directly influences knowledge acquisition, confidence in that knowledge is more closely tied to how learners engage with the AI teaching assistant across multiple dimensions. This highlights the importance of designing AI educational systems that not only deliver content effectively but also foster meaningful engagement through cognitive, behavioral, and affective channels. The relationship between engagement and confidence appears to be particularly crucial, as it affects learners' willingness to apply their knowledge and participate actively in the learning process.

D. Interesting Finds

1) User Preferences for Interaction Modes: As our system provided participants with two ways to interact with the NivTA, one is through gesture (point to a node and ask a question), another way is to directly talk with NivTA. Initially, we thought gesture would be more effective for participants due to its precision and ability to directly link questions to specific visual elements. However, the feedback revealed mixed preferences. Some participants preferred speech for its simplicity and directness: *"I think speech directly is better for me because I can get the answer directly and to understand the learning."* For these participants, speech seems to allow for a more conversational and natural interaction, probably reducing the cognitive load using VR. On the other hand, others favored gestures, citing issues with speech recognition or difficulties pronouncing technical terms: *"Sometimes he might listen to my words, and sometimes I might not know how to pronounce the professional terms."* Gestures also provided a more visual and context-specific form of interaction, enabling users to focus on the material rather than formulating precise verbal questions. These findings suggest that neither modality is universally superior, and user preferences depend on individual needs and technical limitations. A hybrid interaction model that integrates both modalities could provide greater flexibility.

2) Learning Style is Different: We observed a variations in how participants engaged with NivTA, influenced primarily by their learning style preferences and prior knowledge. This heterogeneity in learning approaches emerged as a crucial factor affecting the overall learning experience. Some participants demonstrated a clear preference for traditional reading-based learning, as evidenced by one participant's statement: *"I think for me it's better to read the information and remember one by one by myself rather than ask AI and get the answer from them."* This suggests that despite the availability of interactive AI support, some learners maintain a stronger affinity for self-directed reading and processing of information. That could be the reason why the participants in the Fig. 3 achieve a such low social presence. The perceived utility of NivTA's responses appeared to correlate strongly with participants' prior knowledge levels. Beginners found the AI's assistance more valuable, particularly for basic concept clarification: *When I ask very simple questions, it gives me very effective answers."*

This observed variation in user satisfaction underscores a critical implication for the future development of AI-enhanced learning environments: the necessity for adaptive and personalized learning systems. The current implementation of AI teaching assistants, while effective for novice learners, reveals the need for more sophisticated algorithms that can dynamically calibrate their responses to match individual learning styles and knowledge levels.

VIII. Conclusion

This study investigated AI-assisted immersive learning within the K-Cube Edu-Metaverse, validating our AI-ILF framework. Our findings reveal a clear causal chain: social presence is the foundational element that builds learner trust; this trust, in turn, is the critical catalyst for cognitive, affective, and behavioral engagement; and finally, learning outcomes are shaped by distinct pathways of cognitive presence (impacting knowledge) and engagement (impacting confidence).

The implications for broader educational contexts are direct. To be effective, AI tutors on any platform must be designed to: *1)* cultivate a strong sense of social presence to initiate the learning relationship, *2)* deliberately build trust through reliability and transparency to unlock deeper engagement, and *3)* adapt to individual learning styles and knowledge levels to maintain that engagement. By focusing on these core human-AI interaction principles, designers can create more effective and trustworthy AI-assisted learning experiences, moving beyond simple information delivery to foster genuine knowledge and confidence in learners.

REFERENCES

[1] C. Li, X. Weng, Y. Li, and T. Zhang, "Multimodal learning engagement assessment system: An innovative approach to optimizing learning engagement," *International Journal of Human–Computer Interaction*, pp. 1–17, 2024.

[2] E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Academy of management annals*, vol. 14, no. 2, pp. 627–660, 2020.

[3] J. Li and M. Kazemitabaar, "Collaborative knowledge visualization in shared vr spaces: Strengthening inter-concept connections," *Educational Technology Research and Development*, 2023.

[4] R. Shang, G. Hsieh, and C. Shah, "Trusting your ai agent emotionally and cognitively: Development and validation of a semantic differential scale for ai trust," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, vol. 7, pp. 1343–1356, 2024.

[5] Y. Jia, Z. P. Sin, X. E. Wang, C. Li, P. H. Ng, X. Huang, J. Dong, Y. Wang, G. Baciu, J. Cao, *et al.*, "Nivta: Towards a naturally interactable edu-metaverse teaching assistant for cave," in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, pp. 57–64, IEEE, 2024.

[6] 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.

[7] M. Pawłowska-Nowak, "Identification of trust determinants in llm technology using the dematel method," *European Research Studies Journal*, vol. 27, no. S2, pp. 694–711, 2024.

[8] R. Li, M. Li, and W. Qiao, "Engineering students' use of large language model tools: An empirical study based on a survey of students from 12 universities," *Education Sciences*, vol. 15, no. 3, p. 280, 2025.

[9] S. Chen, R. Metoyer, K. Le, A. Acunin, I. Molnar, A. Ambrose, J. Lang, N. Chawla, and R. Metoyer, "Bridging the ai adoption gap: Designing an interactive pedagogical agent for higher education instructors," *arXiv preprint arXiv:2503.05039*, 2025.

[10] F. Leiser, S. Eckhardt, M. Knaeble, A. Maedche, G. Schwabe, and A. Sunyaev, "From chatgpt to factgpt: A participatory design study to mitigate the effects of large language model hallucinations on users," in *Proceedings of Mensch und Computer 2023*, pp. 81–90, 2023.

[11] L. Benedetto *et al.*, "A virtual teaching assistant for personalized learning," *Proceedings of Politecnico di Milano*, 2023.

[12] X. L. Dong *et al.*, "Towards next-generation intelligent assistants leveraging llm techniques," *Meta Reality Labs & Columbia University*, 2023.

[13] Y. Hicke *et al.*, "Ai-ta: Towards an intelligent question-answer teaching assistant using open-source llms," *Cornell University*, 2023.

[14] Anishka *et al.*, "Can chatgpt play the role of a teaching assistant in an introductory programming course?," in *Proceedings of ITiCSE 2024*, pp. 1–8, 2024.

[15] R. Wei *et al.*, "Improving collaborative learning performance based on llm virtual assistant," in *Proceedings of the 13th International Conference on Educational and Information Technology*, pp. 1–6, 2024.

[16] M. Liu and F. M'hiri, "Beyond traditional teaching: Large language models as simulated teaching assistants in computer science," in *Proceedings of SIGCSE 2024*, pp. 744–750, 2024.

[17] P. Maiti and A. K. Goel, "How do students interact with an llm-powered virtual teaching assistant in different educational settings?," in *Proceedings of the 17th International Conference on Educational Data Mining (EDM) Workshop*, 2024.

[18] A. Sardar, "Constructivism in education: A comprehensive review of theory and research," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, 2023.

[19] Y. Wang, C. Liu, Y. Zhao, W. Huang, B. You, and J. Lin, "The new theory of learning in the era of educational information 2.0—connected constructivism," in *Intelligent Computing and Block Chain: First BenchCouncil International Federated Conferences, FICC 2020, Qingdao, China, October 30–November 3, 2020, Revised Selected Papers 1*, pp. 417–428, Springer, 2021.

[20] R. E. Rice, "Media Appropriateness.: Using Social Presence Theory to Compare Traditional and New Organizational Media," vol. 19, no. 4, pp. 451–484.

[21] S. Yanjie, T. Lei, D. Hao, and F. Jiachen, "Developing a multimodal learning analytics approach to examine students' cognitive presence and metacognition in a metaverse environment," in *International Conference on Computers in Education*, 2024.

[22] A. TG, G. Srinivasa, *et al.*, "Leveraging virtual reality and ai tutoring for language learning: A case study of a virtual campus environment with openai gpt integration with unity 3d," *arXiv preprint arXiv:2411.12619*, 2024.

[23] T.-m. Wut and J. Xu, "Person-to-person interactions in online classroom settings under the impact of covid-19: a social presence theory perspective," *Asia Pacific Education Review*, vol. 22, no. 3, pp. 371–383, 2021.

[24] S. S. Kolekar, D. J. Richter, M. I. Bappi, and K. Kim, "Advancing ai voice synthesis: Integrating emotional expression in multi-speaker voice generation," in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 193–198, IEEE, 2024.

[25] P. Kaewsaiha, "Lms vs social network: Content analysis with the social presence theory approach," in *International Conference" Actual economy: local solutions for global challenges"*, pp. 238–247, 2019.

[26] A. F. Ghazali, A. K. Othman, Y. Sokman, N. A. Zainuddin, A. Suhaimi, N. A. Mokhtar, and R. M. Yusoff, "Investigating social cognitive theory in online distance and learning for decision support: The case for community of inquiry," *International Journal of Asian Social Science*, vol. 11, no. 11, pp. 522–538, 2021.

[27] D. Wang, C. Bian, and G. Chen, "Using explainable ai to unravel classroom dialogue analysis: Effects of explanations on teachers' trust, technology acceptance and cognitive load," *British Journal of Educational Technology*, vol. 55, no. 6, pp. 2530–2556, 2024.

[28] F. Martin, S. Vahedian Movahed, J. Dimino, A. Farrell, E. Irankhah, S. Ghosh, G. Jain, V. Mahipal, P. Rayavaram, I. Temitayo Sanusi, *et al.*, "Perception, trust, attitudes, and models: Introducing children to ai and machine learning with five software exhibits," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, pp. 1740–1741, 2024.

[29] S. A. Alimour, J. H. A. Attiyeh, and M. Y. Balawi, "Designing novel trust models: Merging psychological insights with machine learning to enhance ai user trust," in *2024 Global Congress on Emerging Technologies (GCET)*, pp. 16–18, IEEE, 2024.

[30] C. Zhai, S. Wibowo, and L. D. Li, "The effects of over-reliance on ai dialogue systems on students' cognitive abilities: a systematic review," *Smart Learning Environments*, vol. 11, no. 1, p. 28, 2024.

[31] S. Schneider, M. Beege, S. Nebel, L. Schnaubert, and G. D. Rey, "The cognitive-affective-social theory of learning in digital environments (castle)," *Educational Psychology Review*, vol. 34, no. 1, pp. 1–38, 2022.

[32] J.-C. Hong, K.-H. Tai, and T.-Y. Tai, "Enhancing mandarin learning through classroom response apps: A comparison of cognitive and affective factors in touch and shaking responses," *Journal of Computer Assisted Learning*, vol. 40, no. 6, pp. 3360–3368, 2024.

[33] S. Liu, G. Ma, P. Tewogbola, X. Gu, P. Gao, B. Dong, D. He, W. Lai, and Y. Wu, "Game principle: enhancing learner engagement with gamification to improve learning outcomes," *Journal of Workplace Learning*, vol. 35, no. 5, pp. 450–462, 2023.

[34] N. Michinov and J. Hutain, "Displaying the teacher's slideshow on students' devices prevents multitasking and promotes engagement during lectures," *Active Learning in Higher Education*, vol. 25, no. 3, pp. 551–567, 2024.

[35] Z. Wang, C. Bergin, and D. A. Bergin, "Measuring engagement in fourth to twelfth grade classrooms: the classroom engagement inventory.," *School Psychology Quarterly*, vol. 29, no. 4, p. 517, 2014.

[36] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908, Sage publications Sage CA: Los Angeles, CA, 2006.

[37] J. Kim, "Developing an instrument to measure social presence in distance higher education," *British Journal of Educational Technology*, vol. 42, no. 5, pp. 763–777, 2011.