“Let’s ask what AI thinks then!”: Using LLMs for Collaborative Problem-Solving in Virtual Environments

Esen Küçüktütüncü*
Institute of Neurosciences
University of Barcelona

Lisa Izzouzi †
University College London
University of Barcelona

Abstract

Integrating Large Language Models (LLMs) in immersive virtual environments presents a significant opportunity for enhancing collaborative problem-solving. As online communication becomes increasingly embedded in both our personal and professional lives, tools like LLMs show significant potential to facilitate real-time interaction and improve group dynamics including problem-solving. Leveraging the capabilities of virtual environments, combined with LLMs, can provide engaging, immersive and interactive experiences that are different from those found in traditional settings. This paper explores how these technologies can effectively be utilized to foster collaborative problem-solving in virtual environments, examining both the benefits and challenges of this integration. Finally, this paper proposes a framework for developing applications with integrated LLMs.

Index Terms: Human-centered computing [Human computer interaction(HCI)]: Interaction Paradigms—Virtual Reality

1 Introduction

While shared virtual reality has gained significant traction in recent years, it is important to note that this technology has been around for over three decades [7]. For individuals not previously acquainted with these technologies, there has been a notable shift from traditional online multiplayer games, chatrooms, and forums to more advanced video-conferencing platforms such as Google Meet and Zoom, a transition that was significantly accelerated by the COVID-19 pandemic. Concurrently, the expansion of online communities like VRChat, Roblox, and Meta Horizon has highlighted the potential of immersive technologies to significantly enhance the sense of connection during interactions. However, despite these advances, it has been observed that communication issues commonly encountered in real-life interactions not only persist but can often become more pronounced within immersive virtual environments [6]. This paradox underscores the need for further research on virtual reality communication tools to effectively address these challenges. There are several factors that contribute to the miscommunications experienced when using immersive technologies. One significant factor is technical limitations, such as issues with latency and resolution [15]. Additionally, there is often a lack of multisensory integration, which can hinder the full immersive experience [18]. Another critical challenge is accurately representing non-verbal cues, which are essential for effective communication [34]. These cues include subtle body movements, eye tracking [1], and facial expressions [53]. The inability to effectively convey these non-verbal signals can lead to misunderstandings and less effective communication within virtual environments.

While these challenges need to be addressed through technical advancements, one promising solution lies in the integration of Large Language Models (LLMs) into the virtual reality paradigm. LLMs, such as GPT-4 and its successors, have demonstrated exceptional abilities in understanding and generating human-like text, thereby facilitating more natural and effective communication in virtual environments [49]. Currently, LLMs like ChatGPT are employed as tools for analysis and research [36], as well as assistants in non-immersive collaborative tasks [3]. When these models are integrated into immersive environments, they can significantly enhance real-time interactions by providing contextually relevant responses and ensuring a smooth flow of communication [51]. This, in turn, helps to reduce misunderstandings and improve overall collaboration within virtual settings [40].

Virtual reality itself has the potential to greatly enhance collaborative learning by actively engaging and motivating learners [21]. It supports distance learning, offers interdisciplinary spaces, and aids in the development of social skills [48]. Integrating LLMs can further enrich this experience by providing interactive and adaptive learning opportunities. These models can customize responses and content to suit individual learners’ needs while facilitating group activities and brainstorming sessions. In these settings, effective communication and mutual understanding among team members are essential, and LLMs can play a crucial role in ensuring these elements are achieved.

Based on this we posit the following research questions:

- What impact do LLMs have on group creativity and innovation in virtual collaborative settings?
- How do different virtual collaboration tools supported by LLMs affect team dynamics and communication?
- How can LLMs help in mitigating the effects of distance in remote collaborations?
- What are the challenges and solutions for maintaining workspace awareness in virtual environments enhanced by LLMs?

2 Methodology

This section goes over the above-mentioned questions while describing the benefits, challenges, design, and development of LLM-integrated virtual environments and methods for evaluating their effectiveness. It includes user studies, experimental setups, and the metrics used to assess collaboration quality and problem-solving efficiency.

2.1 Impact on Creativity and Innovation

One of the main critiques or controversies that surrounded any Transformer Models when they were first introduced was the way they formed sentences, processed information, or ’created’ content from scratch. We believe that its sequential nature does not mean it cannot be creative [4], but here the definition of creativity in itself opens up many discussions [54]. LLMs have already been integrated into non-immersive group ideation tasks, where users speculated that the LLM could play a role in facilitating group brainstorming, not just generating ideas [20, 39]. By diversifying types

*e-mail: esenkucuktutuncu@ub.edu
†e-mail: l.izzouzi@ucl.ac.uk

of inputs and perspectives, LLMs can stimulate creative thinking and help teams overcome mental blocks. Within the context of collaboration in virtual environments, having input from an LLM can significantly enhance creativity and innovation by providing diverse inputs and breaking down cognitive biases. It can generate novel ideas and prompt participants to think outside conventional paradigms [33]. For example, during brainstorming sessions, LLMs can offer suggestions that challenge traditional thinking and stimulate new perspectives, ensuring a dynamic and engaging creative process. This diversity of input is crucial for fostering an innovative atmosphere, as it helps teams explore a broader range of possibilities and solutions [52].

Furthermore, LLMs can analyze ongoing discussions in real-time, providing contextually relevant responses and suggesting novel concepts that keep the creative flow uninterrupted. This continuous input can help teams overcome mental blocks and maintain momentum in their ideation processes. Additionally, LLMs assist in synthesizing ideas by identifying common themes and linking disparate thoughts, which can lead to innovative solutions that might not have been considered otherwise [27]. Apart from integrating the LLMs as agents into the environment, they can also be used as tools for generating the Mixed Reality Experience itself [13].

2.2 Team Dynamics and Communication

While in a shared immersive space, the dynamics of the communication play an equally or even more important role than the content that is being generated [5]. In remote collaboration scenarios, maintaining effective communication and coordination can be challenging, and can hinder collaboration [41]. If the conditions to allow the feeling of presence are not met [44], and elements such as multisensory integration, body ownership or plausibility are not satisfied [42], misunderstandings and miscommunications can occur during virtual meetings or in shared environments [16]. LLms can provide real-time translations and context-aware responses, breaking the language barriers that international teams might have. The work of Olson and Olson highlights the difficulties of remote collaboration and the potential for technology to mitigate these challenges By ensuring that all team members have access to the same information and can communicate effectively [37], LLMs can enhance the cohesiveness and productivity of distributed teams. During collaborative tasks, LLMs can help by mediating conflicts, ensuring balanced participation from all members, diffuse tension, and promotive constructive dialogue by providing input that can be less biased than that of a human moderator.

In their study, Izquierdo-Domenech and al. [25] introduce a VR and LLM architecture and found that this integration led to significant improvements in learning outcomes for the Experimental group compared to the Control group across diverse education contexts. These models can serve as individual virtual assistants for each participant, rather than being presented as embodied avatars. , This allows interactions to be tailored to individual user needs and preferences, making the virtual experience more personalized and engaging. By asking users about their learning outcomes or preferred level of engagement for a problem-solving task, and providing real-time feedback and support when needed, this application can drastically improve engagement for both individuals and the group as a whole.

2.3 Mitigating Distance in Remote Collaboration

Avatars, as digital representations of users, create a sense of presence and identity. The sense of embodiment refers to the sensations of being inside, having, and controlling a body [22]. Without considering LLM, avatars and embodiment play crucial roles in remote collaboration in the context of immersive technologies, as they allow users to develop a stronger connection, have more natural interactions through body language, gestures, and facial expressions, and reduce overall physical barriers in distance [17]. The integration of LLMs into this context further transforms remote real-time collaboration and connectivity among users. Indeed, AI-driven avatars can provide real-time, contextually relevant responses to the environment or the users, ensuring contextual interactions. LLMs-driven avatars can assist in tasks and collaboration, facilitate discussions, and provide practical knowledge that is presented in a more flexible and interactive way [30].

2.4 Considerations and Challenges

Despite both technologies having very distinct roles, there are many areas where immersive technologies can benefit from the use of AI and LLMs. First, VR and AI can be used in education to create immersive learning experiences allowing students to engage in hands-on activities usually not possible in traditional educative methods. AI can also analyze a student's performance and learning style, while providing real-time feedback, adapting then the educational content to a specific individual's needs. Second, VR and AI can be used to create new tools for healthcare professionals helping them to interact better with the patient and the medical data, potentially revolutionizing medical diagnostics by sporting anomalies that could escape the human eye. Finally, in the context of problem-solving and conflict resolutions, VR and AI can create immersive interactive, data-driven experiences where users can experience role-playing scenarios, crisis management or other trainings, and virtual workspaces. In those, AI mediators can also be used to guide conflict resolution by assisting users. The AI's capacity to adapt makes it a potential tool to use for effective and engaging problem-solving and conflict resolution in a professional or intimate environment. These example areas presented above, are just examples of the potential use cases that could benefit from immersive technologies-technologies. But a non-exhaustive list of other potential use cases, already being used in the wild includes military training, collaborative tools, gaming, enhanced entertainment, education, healthcare, and training. The power of the combination of these two technologies lies in the possibility of generating personalized content, of recognizing and tracking in real-time objects with computer vision techniques. This powerful pairing of technologies transforms how we learn, work, and interact with digital environments and with each other.

Despite the benefits listed above, several challenges need to be addressed when talking about the integration of LLMs into virtual environments. One of the key elements is ensuring a seamless integration of the models into the application. One of the primary issues is the complexity of the underlying infrastructure required to support both VR and AI simultaneously. This requires significant computational power and robust network infrastructure to handle the real-time processing of both the graphics and the AI models. Studies have highlighted the need for optimized software architectures that allow this integration without compromising performance, especially if the aim is to deploy to standalone headsets. Given it is a continuous stream of data, managing the flow of it and ensuring low latency becomes indispensable. Of course, this also depends on the current limitations of the network speeds. Furthermore, making the application cross-platform poses an additional layer of complexity.

Improving the accuracy and responsiveness of LLMs in real-time collaborative tasks is crucial for their effectiveness. One of the primary challenges is ensuring that LLMs provide accurate and contextually relevant responses in a timely manner. Delays or inaccuracies can disrupt the flow of communication and reduce the effectiveness of collaboration. It is important to optimize LLM algorithms and utilize advanced machine-learning techniques to enhance response accuracy and reduce latency. Techniques such as fine-tuning LLMs on domain-specific datasets and employing real-time feedback loops can significantly improve performance. More-

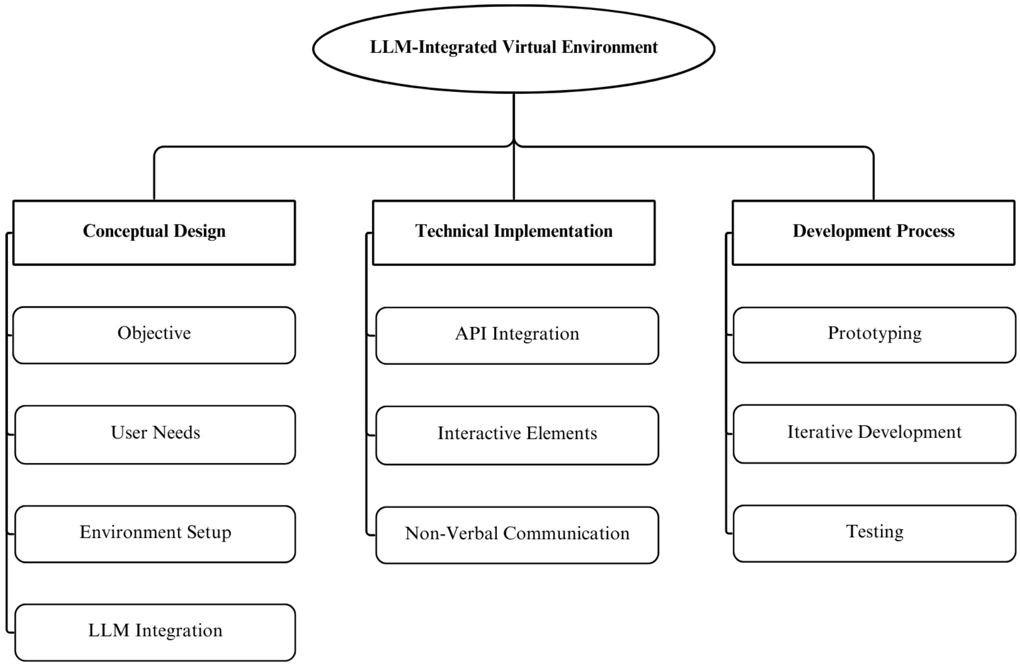

Figure 1: A flowchart for creating an LLM-integrated virtual environment

over, the responsiveness of LLMs can be affected by the computational load and network conditions. Ensuring high availability and reliability of the underlying infrastructure is critical to maintaining the system's responsiveness. Studies by Izquierdo-Domenech et al. suggest that incorporating edge computing and distributed processing techniques can help mitigate these issues by offloading processing tasks closer to the data source, thereby reducing latency and improving responsiveness [26]. Continuous monitoring and optimization of the system's performance are essential to ensure that LLMs can meet the demands of real-time collaborative environments.

The use of LLMs raises several ethical issues related to privacy and data security, which in the case of VR with multi-modal data input, or the level of realism it offers [43], should be carefully managed. One of the major concerns is the potential misuse of personal data [10]. Pre-trained models already require or have processed vast amounts of data to function effectively, but in order for the model to be adaptive to the environment it is in, it requires even more data to ensure cohesive input which could be exploited if not properly secured. In order to prevent this, stringent data protection measures and robust encryption protocols are essential to prevent unauthorized access. Moreover, depending on the location of use, compliance with regulations such as the General Data Protection Regulation (GDPR) is crucial to maintain user trust and avoid legal repercussions [23].

As the data that the model is based on, be it the dataset or the previous users, potential bias or misinformation in LLM outputs is also a concern [11]. The risk of perpetuating and even amplifying the existing biases in interactions can cause major issues, especially in the context of education, where unbiased and fair communication is critical. Addressing these ethical concerns requires continuous efforts in algorithm transparency, bias detection, and the development of fair AI practices.

VR is not a new technology, but it still is unknown to many demographics [2, 29]. New users may need time to adapt to new technologies, as it has been with previous examples. Effective training and support are indispensable and if not done correctly, can hinder its adoption and effectiveness. Prior to introducing any application, users should be familiarized with the functionalities of LLMs, and interaction dynamics of the environment through hands-on experience, and should be offered continuous support to address any issues that may arise. In addition to training, the design choices for the virtual environment and the interface must be user-friendly and intuitive. Complex interfaces can deter users and reduce the overall effectiveness of the system. Simplifying the user interface and providing clear instructions can significantly enhance user satisfaction and ease the adaptation process [38]. Incorporating feedback mechanisms can also allow users to report issues and suggest improvements, which can help developers refine the system to better meet user needs.

Here we have divided the process into three main phases: Phase 1 is the conceptual design, Phase 2 is the technical implementation and finally Phase 3 is the development process.

2.5 Design and Development of LLM-Integrated Virtual Environments

Conceptual Design

The process of conceptual design begins with defining clear objectives and user-centric goals. This objective can range from educational experience to professional training or mental health/therapy applications, but it should be clearly defined as each application has distinct requirements. Understanding the user and their needs, and from there specifying the desired outcomes is key when setting the direction of the development process.

Another key thing in this phase is selecting the appropriate game engine and hardware for creating an immersive and interactive experience. Game development tools like Unity and Unreal Engine are popular choices due to their advanced graphics capabilities, scalability, and compatibility, whereas Godot and WebXR have proven to be almost equally competent in those realms [12]. Have in mind that these choices will affect the architecture and integration when it comes to planning how the LLM will interact with users within the virtual environment through different modalities of interaction (e.g., voice, text, virtual avatar) and the contexts in which the LLM will provide assistance or feedback [19].

Methods for Evaluating Effectiveness and Techniques for Data Collection and Analysis

Experimental Design	Data Collection	Metrics for Evaluation	Advanced Data Analysis Techniques
Randomized Controlled Trials	Surveys and Questionnaires	Communication Quality	Predictive Analytics
Longitudinal Studies	Automated Logging	Creativity and Innovation	Multivariate Analysis
Within-Subject Designs	Behavioral Analytics	Problem-Solving Efficiency	Natural Language Processing
		Team Dynamics and Collaboration Quality	

Figure 2: A table for methods for evaluating effectiveness and techniques for data collection and analysis

Technical Implementation

For the technical implementation, opting for a modular software architecture allows easy integration of LLMs and various other components, such as the OpenAI API for LLM integration or WebXR for web-based VR applications [9]. In this framework, after the conceptual design of the interactive elements, the integration of voice recognition and synthesis technologies such as Whisper, AssemblyAI, Google Cloud Speech-to-Text, or Eleven Labs, Coqui TTS, Bark, Tortoise TTS allows users to communicate effectively with the LLM and enable natural language communication. Additionally, gesture recognition can enhance non-verbal communication, making interactions among users and with LLM more intuitive and engaging. Here, it is worth emphasizing the importance of non-verbal communication, given that avatars that can convey gestures, facial expressions, and eye movements add a great depth to the interaction. Operationalizing these modalities can greatly enhance the presence and realism of the virtual environment.

Development Process

Third and the final phase starts with the prototyping. This step should focus on core functionalities and interactions, allowing initial testing and feedback. Starting the prototyping process early helps with identifying potential issues and setbacks, giving way for improvement and iterative development. The idea behind the iterative development approach is based on continuous refinement of the system, with the help of user feedback and performance metrics. This process ensures that the final applications meets the goals of the developers and the expectations of the users.

While this framework can seem broad, it aims to be a general guide in order to address and be adaptable to various scenarios. After implementing any application that somehow follows this flow, we can use these metrics for evaluation:

2.6 Methods for Evaluating Effectiveness and Techniques for Data Collection and Analysis

Experimental Design

Although experimental design in itself forms part of the design and development of the virtual environment, it can also serve as a method for evaluating the effectiveness of the LLM integration. For this Randomized Controlled Trials(RCTs) are a robust method where participants are randomly assigned to control and experimental groups, and relevant task scenarios can allow assessing the impact of the system on joint problem-solving and engagement.

Another alternative is longitudinal studies, which is something that is still lacking in this field due to its novel and rapidly changing nature. Regular follow-up assessments can serve to track changes in user behavior, performance, and satisfaction and provide great insights into the sustained impact (or lack thereof) of the application.

Finally, within-subject designs are experiments where the same participants experience both control and the experimental conditions at different times and it allows control for individual differences. For obvious reasons, counterbalancing the order of conditions mitigates order effects, ensuring that the observed differences are solely due to the interaction caused by the LLM integration.

Data Collection

The type of data that is being collected might be the collimating part that defines the results you'd reach and the range of it. Surveys and questionnaires System Usability Scale (SUS) [31], NASA Task Load Index (NASA-TLX) [24], Presence and Co-Presence Questionnaire [47], measure user experience, perceived effectiveness and the feeling of being with others in the virtual environment. Semi-structured interviews or sentiment analysis via essay inputs can provide in-depth qualitative insights. Logging the user actions, timestamps, talking patterns, and interaction sequences can give insight and highlight the patterns [14], and quantify the interaction with the user-LLM. Physiological sensors (e.g including heart rate, galvanic skin response (GSR), eye-tracking) can be used for individual assessment.

Metrics for Evaluation

Depending on the format of the data, the quality of problem-solving can be assessed: by looking into response time among users, the accuracy and the relevance, also the users reaction or approach to the

LLM-generated actions and responses. Various Natural Language Processing (NLP) tools can analyze the logged data coherence and sentiment [28, 32]. For measuring the abstract concepts such as creativity and innovation, the Torrance Tests of Creative Thinking (TTCT) [45] can be used to both ensure the baseline for the participants and have a pre-post test for the possible effects LLM had on an individual scale. The physiological measures mentioned above can be used to assess team dynamics, collaboration quality, engagement, and immersion [35]. Social network analysis applied to the turn-taking data can visualize and quantify metrics such as centrality, density and modularity [50]. Finally measuring the task completion time (depending on the nature of the task), can provide the error rates and the quality of the solutions.

Advanced Data Analysis Techniques

The final step of the methods for evaluation can be the more statistical or machine learning-based techniques for predictive analytics. Feature engineering can be used to extract relevant features from interaction logs and telemetry data [46]. In order to identify patterns and relationships in the data without any priors, Principal Component Analysis (PCA) can be implemented alongside with cluster and regression analysis.PCA reduces the dimensionality of the data while retaining most of the variance, thus simplifying the complexity and revealing the underlying structures [8]. By combining multiple regression techniques, it is possible to quantify the impact of various features on indicators such as engagement among users, problem-solving task outcomes, and efficiency. These models can also be used to predict future user behaviors and can allow proactive adjustments if they are fed back to the LLM.

3 Future Directions and Conclusion

The integration of LLMs into immersive virtual environments represents a promising approach to enhancing collaborative problem-solving. As online communication becomes increasingly integral to both personal and professional spheres, the potential of LLMs to facilitate real-time interaction and improve group dynamics is profound. This paper has explored the multifaceted benefits and challenges of this integration, offering a framework for developing LLM-enhanced applications in virtual environments.

LLMs can significantly enhance real-time interactions in virtual settings by providing contextually relevant responses and smoothing the flow of communication, thus reducing misunderstandings and improving overall collaboration. The ability of LLMs to generate diverse and creative ideas stimulates group innovation and problem-solving, making them invaluable in brainstorming and ideation sessions. Additionally, LLMs can mediate conversations, ensuring balanced participation and managing conflicts, which enhances team dynamics and communication.

Despite these benefits, several challenges must be addressed to fully realize the potential of LLMs in virtual environments. Technical integration requires significant computational power and robust network infrastructure to handle real-time processing. Moreover, ethical considerations related to privacy and data security must be managed carefully to maintain user trust. Ensuring user adaptation through effective training and intuitive interface design is also crucial for the successful adoption of these technologies. Furthermore, considering the concept of distributed cognition, future research should explore how interactions between LLMs and users will evolve, especially compared to traditional collaborative interactions. Understanding these dynamics could reveal new ways of using LLMs for more effective teamwork and cognitive enhancement.

Future research should focus on enhancing non-verbal communication capabilities and developing personalized user experiences while ensuring ethical and responsible AI use. Future directions could include exploring cross-disciplinary applications to advance scalability and accessibility. Large multimodal models offer promising possibilities for these advancements. Gamification techniques and longitudinal studies are needed to assess the sustained impact of LLM integration on collaborative learning and problem-solving. Moreover, considering the intent and level of participation expected from LLM agents, it is important to explore how delayed productivity might actually foster creativity. This could influence how we aim to optimize collaboration outcomes, potentially embracing certain inefficiencies to enhance the creative process. By addressing these areas, we can innovate and create more effective, ethical, and inclusive AI systems.

In conclusion, combining LLMs and immersive virtual environments holds immense promise for revolutionizing collaborative interactions. By addressing the technical, ethical, and user adaptation challenges, we can harness the full potential of these technologies to create engaging, effective, and innovative virtual collaboration tools.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the project GuestXR, with grant agreement No. 101017884. EK is also partially funded by a grant from Facebook.

References

[1] I. B. Adhanom, P. MacNeilage, and E. Folmer. Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality*, 27(2):1481–1505, 2023.

[2] C. Ball, K.-T. Huang, and J. Francis. Virtual reality adoption during the covid-19 pandemic: A uses and gratifications perspective. *Telematics and Informatics*, 65:101728, 2021.

[3] A. Bastola, H. Wang, J. Hembree, P. Yadav, N. J. Mcneese, and A. Razi. Llm-based smart reply (lsr): Enhancing collaborative performance with chatgpt-mediated smart reply system. *ArXiv*, abs/2306.11980, 2023.

[4] A. Bellemare-Pepin, F. Lespinasse, P. Thölke, Y. Harel, K. Mathewson, J. A. Olson, Y. Bengio, and K. Jerbi. Divergent creativity in humans and large language models. *arXiv preprint arXiv:2405.13012*, 2024.

[5] F. Biocca and M. R. Levy. Virtual reality as a communication system. *Communication in the age of virtual reality*, pp. 15–31, 1995.

[6] F. Biocca and M. R. Levy. *Communication in the age of virtual reality*. Routledge, 2013.

[7] C. Blanchard, S. Burgess, Y. Harvill, J. Lanier, A. Lasko, M. Oberman, and M. Teitel. Reality built for two: a virtual reality tool. In *Proceedings of the 1990 symposium on Interactive 3D graphics*, pp. 35–36, 1990.

[8] R. Bro and A. K. Smilde. Principal component analysis. *Analytical methods*, 6(9):2812–2831, 2014.

[9] R. Capilla and M. Martínez. Software architectures for designing virtual reality applications. *Software Architecture*, p. 135–147, 2004. doi: 10.1007/978-3-540-24769-2_10

[10] M. Carter and B. Egliston. What are the risks of virtual reality data? learning analytics, algorithmic bias and a fantasy of perfect data. *New media & society*, 25(3):485–504, 2023.

[11] C. Chen and K. Shu. Combating misinformation in the age of llms: Opportunities and challenges. *arXiv preprint arXiv:2311.05656*, 2023.

[12] E. Coronado, S. Itadera, and I. G. Ramirez-Alpizar. Integrating virtual, mixed, and augmented reality to human–robot interaction applications using game engines: A brief review of accessible software tools and frameworks. *Applied Sciences*, 13(3):1292, 2023.

[13] F. De La Torre, C. M. Fang, H. Huang, A. Banburski-Fahey, J. Amores Fernandez, and J. Lanier. Llmr: Real-time prompting of interactive worlds using large language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642579

[14] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283, 1972.

[15] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler. Toward low-latency and ultra-reliable virtual reality. *IEEE network*, 32(2):78–84, 2018.

[16] N. Fereydooni and B. N. Walker. Virtual reality as a remote workspace platform: Opportunities and challenges. 2020.

[17] G. Freeman and D. Maloney. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on human-computer interaction*, 4(CSCW3):1–27, 2021.

[18] A. Gallace, M. K. Ngo, J. Sulaitis, and C. Spence. Multisensory presence in virtual reality: possibilities & limitations. In *Multiple sensorial media advances and applications: New developments in MulSeMedia*, pp. 1–38. IGI Global, 2012.

[19] R. Gallotta, G. Todd, M. Zammit, S. Earle, A. Liapis, J. Togelius, and G. N. Yannakakis. Large language models and games: A survey and roadmap. *arXiv preprint arXiv:2402.18659*, 2024.

[20] G. E. Gonzalez, D. A. S. Moran, S. Houde, J. He, S. I. Ross, M. J. Muller, S. Kunde, and J. D. Weisz. Collaborative canvas: A tool for exploring llm use in group ideation tasks. In *IUI Workshops*, 2024.

[21] J.-Q. Guan, S.-F. Ying, M.-L. Zhang, and G.-J. Hwang. From experience to empathy: An empathetic vr-based learning approach to improving efl learners' empathy and writing performance. *Computers & Education*, p. 105120, 2024.

[22] M. Guy, J.-M. Normand, C. Jeunet-Kelway, and G. Moreau. The sense of embodiment in virtual reality and its assessment methods. *Frontiers in Virtual Reality*, 4:1141683, 2023.

[23] L. Halawi and A. Makwana. The gdpr and uk gdpr and its impact on us academic institutions. *Issues in Information Systems*, 24(2), 2023.

[24] 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.

[25] J. Izquierdo-Domenech, J. Linares-Pellicer, and I. Ferri-Molla. Virtual reality and language models, a new frontier in learning. *Int. J. Interact. Multim. Artif. Intell.*, 8:46, 2024.

[26] J. Izquierdo-Domenech, J. Linares-Pellicer, and I. Ferri-Molla. Virtual reality and language models, a new frontier in learning. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5):46, 2024. doi: 10.9781/ijimai.2024.02.007

[27] K. S. John, G. A. Roy, and P. Bindhya. Llm based 3d avatar assistant. In *2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST)*, pp. 1–5. IEEE, 2024.

[28] B. R. Kaur, K. Navneet, K. Ravneet, and K. Gurpreet. Opinion mining and sentiment analysis. In *3rd international conference on computing for sustainable global development (INDIACom). IEEE, New Delhi, India*, pp. 452–455, 2016.

[29] J. W. Kelly, L. A. Cherep, A. F. Lim, T. Doty, and S. B. Gilbert. Who are virtual reality headset owners? a survey and comparison of headset owners and non-owners. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 687–694. IEEE, 2021.

[30] C. Krauss, L. Bassbouss, M. Upravitelev, T.-S. An, D. Altun, L. Reray, E. Balitzki, T. El Tamimi, and M. Karagülle. Opportunities and challenges in developing educational ai-assistants for the metaverse. In *International Conference on Human-Computer Interaction*, pp. 219–238. Springer, 2024.

[31] J. R. Lewis. The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018.

[32] B. Liu. *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[33] Y. Liu, S. Chen, H. Cheng, M. Yu, X. Ran, A. Mo, Y. Tang, and Y. Huang. How ai processing delays foster creativity: Exploring research question co-creation with an llm-based agent. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–25, 2024.

[34] D. Maloney, G. Freeman, and D. Y. Wohn. " talking without a voice" understanding non-verbal communication in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020.

[35] D. Mønster, D. D. Håkonsson, J. K. Eskildsen, and S. Wallot. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & behavior*, 156:24–34, 2016.

[36] M. Namvarpour and A. Razi. Apprentices to research assistants: Advancing research with large language models. *ArXiv*, abs/2404.06404, 2024.

[37] G. Olson and J. Olson. Mitigating the effects of distance on collaborative intellectual work. *Economics of Innovation and New Technology*, 12(1):27–42, 2003.

[38] G. M. Richard E. Mayer and J. Parong. The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human–Computer Interaction*, 39(11):2229–2238, 2023. doi: 10.1080/10447318.2022.2108563

[39] O. Shaer, A. Cooper, O. Mokryn, A. L. Kun, and H. B. Shoshan. Ai-augmented brainwriting: Investigating the use of llms in group ideation. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024.

[40] A. Shoa, R. Oliva, M. Slater, and D. Friedman. Sushi with einstein: Enhancing hybrid live events with llm-based virtual humans. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, pp. 1–6, 2023.

[41] M. Slater. Real people meeting virtually real people—a review of some experiments in shared virtual environments. *BT Technology Journal*, 17(1):120–127, 1999.

[42] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in virtual reality*, 3:914392, 2022.

[43] M. Slater, C. Gonzalez-Liencres, P. Haggard, C. Vinkers, R. Gregory-Clarke, S. Jelley, Z. Watson, G. Breen, R. Schwarz, W. Steptoe, et al. The ethics of realism in virtual and augmented reality. *Frontiers in Virtual Reality*, 1:512449, 2020.

[44] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.

[45] E. P. Torrance. Torrance tests of creative thinking. *Educational and psychological measurement*, 1966.

[46] C. R. Turner, A. Fuggetta, L. Lavazza, and A. L. Wolf. A conceptual basis for feature engineering. *Journal of Systems and Software*, 49(1):3–15, 1999.

[47] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence*, 9(5):497–503, 2000.

[48] N. van der Meer, V. van der Werf, W.-P. Brinkman, and M. Specht. Virtual reality and collaborative learning: a systematic literature review. *Frontiers in Virtual Reality*, 4, 2023. doi: 10.3389/frvir.2023.1159905

[49] H. Wan, J. Zhang, A. A. Suria, B. Yao, D. Wang, Y. Coady, and M. Prpa. Building llm-based ai agents in social virtual reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2024.

[50] S. Wasserman and K. Faust. Social network analysis: Methods and applications. 1994.

[51] S. Xu, Y. Wei, P. Zheng, J. Zhang, and C. Yu. Llm enabled generative collaborative design in a mixed reality environment. *Journal of Manufacturing Systems*, 74:703–715, 2024.

[52] S. Zha, Y. Qiao, Q. Hu, Z. Li, J. Gong, and Y. Xu. Designing child-centric ai learning environments: Insights from llm-enhanced creative project-based learning. *arXiv preprint arXiv:2403.16159*, 2024.

[53] Z. Zhang, J. M. Fort, and L. Giménez Mateu. Facial expression recognition in virtual reality environments: challenges and opportunities. *Frontiers in psychology*, 14:1280136, 2023.

[54] Y. Zhao, R. Zhang, W. Li, D. Huang, J. Guo, S. Peng, Y. Hao, Y. Wen, X. Hu, Z. Du, et al. Assessing and understanding creativity in large language models. *arXiv preprint arXiv:2401.12491*, 2024.