CHAPTER 12

AR/VR and Immersive Learning in Mathematics

V. I. Slyusar
Institute of Artificial Intelligence Problems, Ukraine,
E-mail: swadim@ukr.net

Abstract

The composition of agents within a multi-agent educational system can be enhanced by integrating augmented and virtual reality (AR/VR) technologies, alongside the generation of immersive 3D worlds using multimodal large language models (LLMs). Current advancements in LLMs enable the creation of video sequences lasting several minutes from textual descriptions or visual representations of initial and final scenes. Immersing students in such environments, populated by imaginative entities, exposes them to unfamiliar scenarios stemming from limited prior experience, reinforcing the perceived value of continuous learning. In mathematics education, AI-generated virtual environments and characters vividly demonstrate how visualizing mathematical principles, theorems, and concepts can improve comprehension and emotional engagement. VR representations help learners clearly perceive the consequences of mistakes and the importance of gained knowledge, significantly enhancing educational effectiveness. Consequently, integrating 3D technologies into teaching substantially increases student interest and engagement, supporting more effective mastery of mathematical content. A game-based approach offers an engaging alternative to traditional methods by creating virtual mathematical quests, where pupils must solve tasks to unlock new areas within captivating 3D environments. For instance, progressing to the next "level" or activating virtual structures like bridges or stairs requires solving mathematical problems. One illustrative scenario involves students assisting an architect in constructing a dome. Here, pupils must calculate the dome's radius, lengths of arcs. Such interactive quests transform mathematical problems from routine school exercises into exciting adventures, motivating deeper exploration of the subject. Interaction among VR avatar-agents and with external services is managed through unified APIs and protocols such as A2A and MCP. A central coordination agent monitors learning dynamics and identifies functional redundancies among agents, ensuring cohesive and efficient operation of the educational environment.

Keywords: Augmented Reality, Virtual Reality, Traditional approach, Large Language Models (LLMs), Problem solving

INTRODUCTION

Traditional mathematics teaching, as illustrated by the image of a classical school classroom (Fig. 1), is based on direct teacher–student interaction, the use of the blackboard, chalk, and printed instructional materials. This approach emphasizes algorithmic thinking, the step-by-step mastery of formulas and fundamental mathematical structures, such as equations and geometric theorems, for example, the Pythagorean theorem. In a typical scene of frontal teaching (Fig. 1), the teacher explains how to solve a problem while students listen, answer questions, or passively receive information. Although this model is fundamental and well-tested throughout the history of education, it has limitations in visualizing abstract concepts, stimulating students' cognitive activity, and individualizing the learning process.

Modern augmented reality (AR) and virtual reality (VR) technologies offer an alternative that significantly expands the didactic potential of mathematics education. They make it possible to create interactive visual environments in which mathematical objects - such as geometric figures, function graphs, or multidimensional spaces - become dynamic and accessible for real-time manipulation. This contributes to deeper understanding, especially in cases where traditional methods remain too abstract or difficult to grasp.

Within the context of educational transformation, Ch. 2 is devoted to analyzing the opportunities that AR/VR technologies open up for mathematics instruction, particularly in the aspects of immersion, personalization of the learning experience, and strengthening student motivation to explore and learn.

Fig. 1 Traditional approach to education (generated using GPT-4o)

Here, special emphasis is placed on the integration of VR/AR technologies with the capabilities of artificial intelligence [1] to achieve a synergistic effect. In this context, the agent ensemble of the multi-agent educational system described in the previous section can be expanded by integrating augmented and virtual reality (AR/VR) technologies, as well as by generating immersive 3D worlds using multimodal large language models (LLMs). Current advancements in LLMs enable the creation of video sequences lasting several minutes based on textual descriptions or visual representations of initial and final scenes. Immersing students in such environments populated by imaginative entities exposes them to unfamiliar scenarios beyond their prior experience, reinforcing the perceived value of continuous learning.

In mathematics education, AI-generated virtual environments and characters vividly demonstrate how visualizing mathematical principles, theorems, and concepts can improve comprehension and emotional engagement. VR representations help learners clearly perceive the consequences of mistakes and the importance of acquired knowledge, significantly enhancing educational effectiveness [2 – 10]. Consequently, integrating 3D technologies into teaching substantially increases student interest and engagement, supporting more effective mastery of mathematical content.

A gamified approach to learning mathematics opens new opportunities for engaging students, transforming the educational process into a series of missions within immersive virtual spaces focused on mathematical content. In such interactive environments, students perform computational or logical actions to unlock new areas or activate important objects. Instead of the standard presentation of problems, participants engage in contextually justified scenarios where each mathematical solution has functional significance within the game's framework.

For example, in the scenario "The Code of the Ancient Labyrinth," students must solve equations with multiple variables to determine the correct code sequence to open a door. In the "Mathematical Expedition to Mars" project, pupils calculate optimal trajectories for a transport rover, relying on geometric constructions, systems of equations, and vectors. In another hypothetical situation - "Restore the Ancient Tower", - students recreate architectural elements using knowledge of proportions, surface areas, and spatial volumes.

These scenarios not only sustain interest but also encourage students to apply their knowledge in non-standard conditions, fostering critical thinking, spatial imagination, and understanding of interdisciplinary connections.

Interaction between VR avatar-agents and external services is managed via unified APIs and protocols [11] such as A2A and MCP [12 - 16]. The central coordinating agent monitors learning dynamics and identifies functional duplication among agents, ensuring coherent and efficient operation of the educational environment.

Let us examine in more detail the functions and operation of a typical AI virtual reality agent.

THE ROLE OF THE VIRTUAL REALITY AGENT (VR AGENT) IN A MULTI-AGENT SYSTEM

A virtual reality agent (VR agent) in a multi-agent system for advanced mathematics education serves as an interactive interface between the user and the three-dimensional virtual environment, which visualizes mathematical objects, operations, and processes. Its main functions include spatial reconstruction of abstract mathematical concepts, adaptive visualization and real-time manipulation of structures, multimodal interaction, and synchronization with other agents (see Fig. 2).

In its operation, the VR agent receives semantic commands from other agents (for example, "visualize the set of eigenvectors of matrix A," "show the limit of the function as $x \to a$," or "construct the level surface for $f(x, y) = c$") and transforms them into three-dimensional scenes with animated elements. At the same time, it automatically adjusts the visualization scale of the scenes, their color palettes, lighting, and viewing perspective to enhance the perception of complex constructions such as multidimensional projections, spatial distortions, topological transitions, or the dynamics of differential equation solutions. When connected to headset devices (HMDs) such as Oculus or HoloLens, the VR agent implements stereoscopic mode, allowing the user to move around the mathematical 3D object, change its parameters using gestures or voice commands, and interact with the environment through feedback sensors.

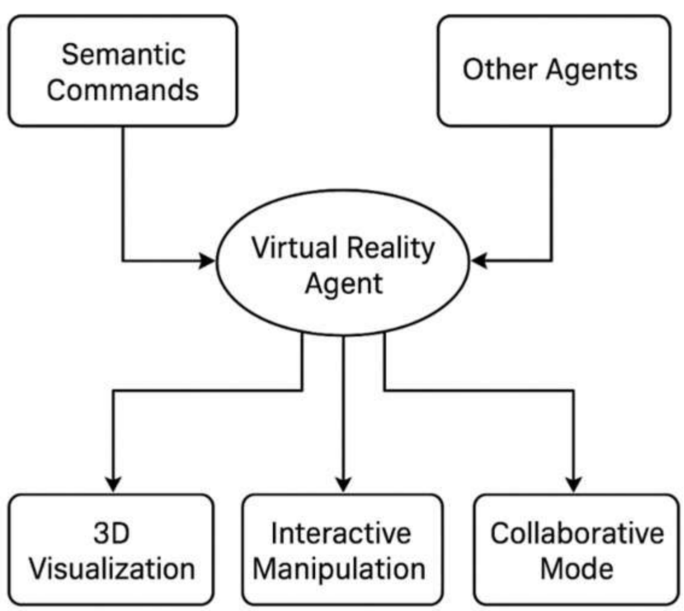

Fig. 2 Diagram illustrating the operation process of the VR agent

An important feature is the VR agent's ability to conduct "live experiments" (Fig. 3), in which the user can change the input parameters of functions or equations and instantly observe changes in graphs, curvature, extremum points, or singularities. In educational scenarios, this agent synchronizes with the mentor agent and the progress agent to dynamically regulate the complexity of visualization and the type of hints depending on the learner's level of preparation. For example, a beginner student will be shown only basic structures (axis cross-sections, fundamental points), while an advanced user will see isolines, tensor fields, or boundary cases.

Fig. 3 "Live" experiments in the virtual environment (generated using GPT-4o).

In addition, the VR agent supports collaborative mode, where several learners are simultaneously connected to the same scene (Fig. 4); in this case, it integrates with the communicator agent, allowing comments, markers, voice discussion, and cooperative manipulation of objects. The VR agent should also be able to save sessions, generate video recordings of virtual demonstrations, or transmit scenes to AR-compatible devices (such as XRAI or Ray-Ban Meta) in a condensed infographic format.

Thus, the virtual reality agent transforms passive perception of mathematical material into visually procedural learning, where abstractions take on physical form and the user becomes an active explorer of the mathematical concept space.

Interaction between the VR agent and the augmented reality agent (AR agent) is implemented as synchronized bidirectional (duplex) communication, aimed at transferring mathematical objects, processes, and scenarios from a fully virtual space into the user's physical environment with superimposed virtual elements. This interaction is particularly relevant for higher mathematics education, where deep visualization needs to be combined with real-world context (for example, classroom instruction or the use of physical materials).

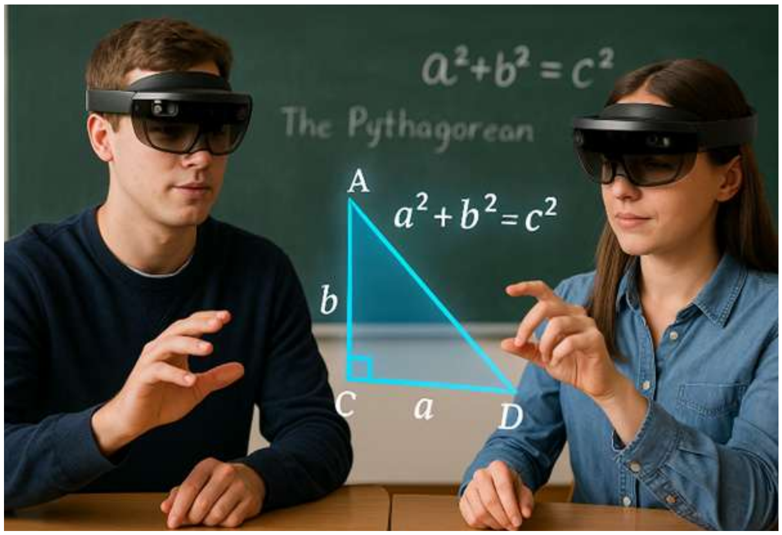

Fig. 4 Collaborative mode in virtual reality (generated using GPT-4o).

At the level of functional exchange (Fig. 5), the VR agent generates three-dimensional objects and scenes based on commands from semantic or subject-specific agents. These objects are stored in an internal representational format (for example, as mesh models, parametric surface descriptors, or voxel-based structures), which can be converted into a lightweight representation suitable for overlay in augmented reality. The transfer of this information is carried out via a shared buffer or protocol, in particular as intermediate visualization packets containing geometric, stylistic, and semantic information (positions, dimensions, markers, etc.).

The AR agent receives these packets and transforms them according to the user's spatial coordinates, determined via the sensors of a headset or mobile device. When using devices such as XRAI Glass or Ray-Ban Meta, the AR agent adapts the content so that it remains "anchored" to the real environment (for example, visualizing a plane as an interactive grid overlaid on a physical table). For mathematical problems, this means the student sees a vector field, singularity, or integral curve directly in the context of their physical surroundings.

The AR agent also supports bidirectional interaction: gestures, voice commands, or gaze are interpreted and transmitted back to the VR agent, which updates the scene. For example, if a student selects a point on a visualized surface with a gesture, the AR agent transmits the coordinates of the selection to the VR agent, which then performs a local operation—calculating a gradient, constructing a normal, or demonstrating the area of a section.

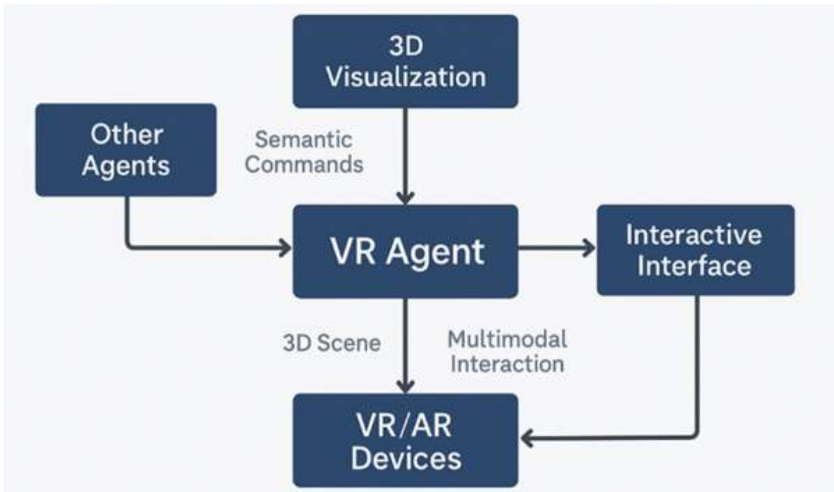

Fig. 5 Interaction between the VR agent and the AR agent

Coordination between the agents is managed by an interface interpreter agent, which ensures consistency of format, accuracy of spatial anchoring, resolution and scene refresh rate synchronization, and adherence to the user's cognitive constraints. Additionally, VR and AR modes can be used in parallel in a collaborative environment: one student works in VR building the scene, while another explores it in AR. This setup enables distributed learning in a multi-agent environment with full system support.

It should be noted that, ideally, the VR and AR agents should function as complementary modules: the former creates and manages the complete virtual learning space, while the latter adapts these objects to the real world, enabling the learner to study without losing connection to physical reality, with maximum interactivity, precision, and immersion in mathematical logic.

MATHEMATICAL VR QUESTS FOR PRIMARY SCHOOL PUPILS

Returning to the topic of mathematical VR quests for primary school pupils, let us consider an example implementation of one possible scenario, conventionally titled "The Dungeon of Numbers." This quest combines gamified learning with virtual immersion in a fairy-tale world, where students solve a sequence of logic-mathematics problems to achieve the goal: to find the keys to a magical gate.

The scenario assumes that the student, in virtual reality, finds themselves in a fantastic dungeon divided into several chambers, each representing a distinct mathematical topic: addition, subtraction, geometric shapes, number comparison, object counting, and multiplication within 10.

In the first chamber -"The Cave of Addition", - glowing number tiles light up on the floor (see Fig. 6). A voice guide, played by the mentor agent, gives the task: "Find the tile with the answer to the sum 5 + 3." The student must physically move and step onto the correct tile. If the answer is correct, the tile lights up green, a key appears, and the door to the next room opens.

Fig. 6 Virtual "Cave of Addition" (generated using GPT-4o).

In the second chamber - the "Geometry Labyrinth", - the pupil sees three-dimensional shapes floating in the air. The task: "Find all the cubes and bring them to the basket." The child manipulates objects in VR (using controllers), picking up and moving the shapes. At the same time, the visualization agent pronounces the names of the shapes and emphasizes the number of faces. If the pupil mistakenly grabs a different shape, a hint appears: "This is not a cube—it has fewer faces."

In the "Subtraction Treasury" chamber, the task is implemented as a game with chests (see Fig. 7), each labeled with a math example. There are also boards with numerical answers in the room. The pupil must carry each board with the correct answer to the appropriate chest. Upon successful completion, a new magical artifact appears, activating a portal to the final challenge.

The final stage of the quest is the "Number Tower." This level involves ascending a staircase where each step is labeled with a number. The pupil must move along the stairs by adding or subtracting a specific number (for example, "move three steps forward, then two steps back"), thus modeling mental arithmetic. An incorrect step requires returning to the beginning of the section, but with a hint from the mentor agent.

Throughout the described quest, the VR agent controls the dynamics of scenes and transitions between them; adapts the complexity of problems depending on the child's level of success; provides multimodal hints (by voice, gesture, light effects); tracks progress and sends it to the analytics agent for feedback to the teacher.

The described and similar VR-quest options make it possible to create an engaging, intuitive, and active mathematical environment, where the child not only completes exercises but experiences learning as a meaningful and coherent adventure.

Fig. 7 The "Subtraction Treasury" chamber (generated using GPT-4o).

The process of generating images for VR visual scenes can be built on the same principle as autoregressive text generation by language models, predicting the next element in a sequence based on the previous ones. However, instead of text tokens, a sequence of visual tokens is used - discrete image fragments produced by a tokenizer (e.g., VQ-VAE, DALL-E tokenizer, or a similar visual codec).

For VR scene generation, an autoregressive multimodal LLM such as transframer [17], or modern models like GPT-4o, constructs the VR visual scene as a sequence of tokens encoding 3D space elements (colors, shapes, depth, textures) in a certain order - pixel by pixel or patch by patch. The model sequentially predicts each next token based on those already generated, taking into account both the global context (e.g., the scene is "Subtraction Treasury" or "rocky dungeon with torches on the walls and chests with numbers") and the spatial relationships among objects.

The transformer architecture [18] underlying such models preserves spatial dependencies in a way analogous to how semantic and grammatical connections are maintained in texts. This includes information about perspective, lighting, and spatial distance, ensuring conformity to the physical logic of the scene.

When creating scenes, conditional information is considered—viewer position, expected perspective (field of view), spatial coordinates of objects, focus zones, entry and exit points. These parameters are input to the model as conditions, so that the probability $p(x_t|x_{<t}, c)$ of generating the next scene element x_t, such as an object description or its properties, is determined not only as a sequence of all previous elements $x_{<t}$, but also by additional context c, which can include navigation parameters, spatial masks, or depth maps. Thus, $p(x_t|x_{<t}, c)$ formalizes the dependence of each new scene element's generation on the generation history (what has already been generated) and on conditions that define the geometric and semantic structure of the environment.

Scene generation occurs in layers: first, the model generates the rough geometry and general layout of objects, then adds surface details, textures, lighting, and shadows. Thanks to this approach, the scene can adapt to interactive VR viewing: the user moves freely, and the system dynamically completes the necessary zones around the field of view in real time (foveated rendering), continuing autoregressive scene refinement on the GPU or edge modules.

Thus, the VR scene is formed as a sequential reconstruction of a conditionally continuous space, where each new token adds detail in much the same way that words form an idea in a sentence. Scenes are not static models, but are dynamically created according to the task, the user's position, and the learning content, implementing principles characteristic of autoregressive text generation models.

JOINT EMBEDDING OF TEXT DESCRIPTIONS, 3D SCENES, AND AUDIO NARRATION

In the generation of VR environments, the mechanism of joint embedding of text descriptions, 3D scenes, and audio accompaniment plays a crucial role. The result of such embedding is a multidimensional vector space in which heterogeneous modalities (text, three-dimensional graphics, and audio) are encoded in a shared latent representation that preserves semantic correspondence and logical relationships between objects, events, and descriptions. To achieve such alignment, multimodal neural architectures are used, trained so that identical content, regardless of the form of presentation (verbal, visual, or auditory), is transformed into neighboring points in the latent space.

The text component - for example, the description “a chest stands in the dungeon, and the muffled sound of coins is heard”, - is encoded by a transformer-based text encoder into a vector that contains information about objects, spatial relationships, and actions. The 3D scene—a set of mesh objects with coordinates, textures, lighting, and spatial connections - is encoded by graph neural networks or 3D convolutional encoders into a vector of the same dimension. In turn, audio accompaniment, such as the sound of coins clinking or a voice explanation, is encoded by audio transformers or spectrogram-based models into the corresponding embedding.

At the level of joint training (multimodal contrastive learning or alignment loss), these embeddings are optimized so that the embedding of a scene, its verbal description, and its accompanying audio are maximally close in the shared space if they belong to the same context, and maximally distant for different contexts. This makes it possible to perform text-based searches and instantly retrieve the relevant 3D scene and audio file, or conversely, to generate text descriptions or voice prompts based on a visual scene or audio accompaniment.

In educational VR scenarios, joint embedding enables coherent integration: changing a text task automatically restructures the visualization and audio, while interacting with an object in the scene triggers corresponding explanation fragments or sound effects. This creates a unified multimodal educational environment in which the content of different modalities is harmonized in terms of meaning and context.

CAPABILITIES OF AR/VR AND IMMERSIVE LEARNING IN THE STUDY OF IUT

It should be noted that the use of AR/VR technologies is appropriate not only in secondary school, but also for the study of advanced mathematics. In particular, these technologies have exceptionally powerful potential to assist in understanding and applying extremely complex mathematical theories, such as Inter-universal Teichmüller Theory (IUT) [19 - 22]. Thanks to their immersive properties, AR/VR allows the user to literally “step into” a new mathematical universe, revealing theoretical aspects that might otherwise remain unclear.

First and foremost, AR/VR can be used to spatially visualize Frobenioids and θ-functors of IUT. For example, the multi-level structure of factorizations in VR space can be represented as floating nodes in a three-dimensional graph, where θ-functors are tunnels between parallel universes that can be physically traversed or jumped across with a gesture. This representation allows objects to retain their "morphology" while changing arithmetic context.

Instead of an abstract concept of a log-structure, one can show how a "halo" around an object modulates its properties. The user can interact with these halos: remove, transform, or combine them. Such operations help to understand how manipulating log-structures "deforms" the proof space.

Volumetric structures can also reproduce compositions of functors, and commutative diagrams of categories can be rendered as holograms, which are difficult to imagine in 2D. In this case, commutativity is visible through motion, as a figure "transforms" when moving an object along different paths. VR makes it possible to see this from any angle, thereby realizing visual recognition of commutativity and "Explorable Explanation" style learning. Each virtual IUT block - Frobenioids, θ-functors, log-schemes, ABC-graphs, - can be explored, rotated, and subjected to morphisms in real time, like in an interactive mathematics museum.

In augmented reality environments (for example, using HoloLens or a smartphone), one can see "visual additions" over abstract formulas, such as θ-arrows, log-layers, or multi-level function maps. This is useful for classroom teaching or remote collaboration.

An additional VR simulation of "universe transitions" would allow a virtual journey in which the user moves from theater (universe) G_1 to G_2 and then to G_3, experiencing changes in structure, color, and orientation, to intuitively understand the change of context in IUT.

In each IUT world, a multilingual AI agent can serve as an AR/VR guide, interactively explaining IUT specifics, commenting on what is happening with log-structures, conducting dialogue, asking and answering questions. Multi-agent assistants can help the learner construct their own θ-morphisms or simulate ABC proofs.

Possible technical platforms for implementing this concept include:

- Unity 3D + C# or Unreal Engine – for building full-fledged VR scenes.
- Combination of WebXR + Three.js – for browser-based VR/AR viewing.
- Blender + Python – for preliminary modeling of structures.
- OpenAI + Whisper + Text-to-Speech – for voice guides in space.
- Jupyter + pyvista / ipyvolume – for embedded 3D research in scientific notebooks.

In summary, IUT is a mathematical theory that requires not only logical understanding, but also intuitive switching between mathematical contexts, which is almost impossible in standard 2D. This is why AR/VR can become the most effective environment for its study, visualization, explanation, and even creative modeling.

As an example, let us consider the concept of designing a simplified AR/VR scene for learning IUT. For this, we use a basic immersive module that visualizes three Frobenioid universes with θ-transitions between them, which the user can explore in the VR/AR space with commentary from an AI assistant.

The key component of the scene is sphere-objects that symbolize IUT theaters or universes (see Fig. 8), grouped into three zones: G_1 (X coordinate = 0); G_2 (X coordinate = +3, θ_1-universe); G_3 (X coordinate = +6, θ_2-universe).

Fig. 8 The set of theaters (universes)

Another element is arrows (morphisms). They are divided into internal arrows (×2, ×3, ×5) within each world, shown as colored semi-transparent lines (blue, red, green), as well as θ-transitions between the worlds—displayed as light, curved arrows labeled θ_1 and θ_2.

The explanation system in the space operates in such a way that when hovering over a node, a label with the factorization formula appears, and when clicking on a θ-arrow, a voice explanation of what a θ-functor is is played.

An interactive agent (in the form of a sphere or avatar) comments on objects and responds to queries, for example:

- "Explain how θ_1 works at this node."
- "What is the difference between multiplication by 2 and the transition to G_2?"

The described scenario can be implemented in two versions:

- WebXR (Three.js + A-Frame) – works in the browser, including VR headsets or a phone in AR mode;
- Unity 3D (for Oculus, SteamVR, Hololens) – a professional scene with support for gestures, voice, and multimodal control.

If special equipment (such as VR headsets, HoloLens, etc.) is not available, it is best to start with WebXR in the browser. This option requires nothing but a browser (Chrome, Firefox), works even on a regular laptop or phone, and supports simple 3D navigation with a mouse and keyboard. At the same time, it is possible to display all the necessary elements (nodes, arrows, labels, interactive explanations) and gradually expand the visualization scene into a fully-fledged virtual environment.

Using Perplexity to Evaluate the Effectiveness of LLMs in Generating VR Spaces for Mathematical Quests

Various metrics can be used to assess the quality of virtual world generation [23]. Applying perplexity as a metric for evaluating the effectiveness of LLMs [24] that generate VR spaces for mathematical quests is a highly appropriate idea. This metric allows one to formalize the concepts of coherence, predictability, and semantic alignment between the prompt and the generated space. Given the specifics of such generation tasks, this idea can be developed in three interrelated directions.

1. Perplexity as a Metric of Text-to-Scene Semantic Control (Text-to-Scene Alignment).

   In most cases, each VR quest begins with a textual prompt or scenario ("build a room with geometric shapes where the task is to find the area of a circle"). If an LLM transforms this description into a scene where the required shapes are absent or only squares and rectangles are present, one can assess how "non-perplexed" the model's output is. In this case, perplexity at the generation level can be formalized as:

   $$PPL_{\text{text}\to\text{scene}} = exp\left(-\frac{1}{N}\sum_{i=1}^{N}\log P\left(s_i \middle| x_1,\ldots,x_i\right)\right)$$

   where s_i are tokens of the scene description or parametric descriptors in a scene language (such as VRML, glTF), and x_i are text tokens of the prompt.

   This metric reflects how confidently the model constructs the scene based on the scenario's semantics.

2. Perplexity as an Indicator of Visual-Semantic Consistency (Scene Consistency).

   Since a VR space is three-dimensional and contains spatial relationships, the scene can be quantized as tokens (via voxel coders, DALL·E-like visual encoders, or Latent Diffusion Tokenizers). The LLM should then recover these tokens in a sequence with low perplexity. If the scene is illogical (e.g., objects intersect, lighting contradicts light source positions, or the task is placed in an inaccessible location), perplexity increases:

$$PPL_{\text{scene}} = exp\left(-\frac{1}{M}\sum_{j=1}^{M}\log P\left(z_i | z_{<j}\right)\right),$$

where z_j are the visual scene tokens.

3. Combined Perplexity as a Measure of Multimodal Alignment (Multimodal PPL Fusion).

   Since virtual reality is a complex environment with circulating text, visual, and audio information, a weighted score can be introduced that accounts for data multimodality:

   $PPL_{\text{total}} = \alpha \cdot PPL_{\text{text}\rightarrow\text{scene}} + \beta \cdot PPL_{\text{scene}} + \gamma \cdot PPL_{\text{audio}}$,

   where the weights α, β, γ can be optimized through user feedback (e.g., scene quality rating) or determined based on the scenario (for example, in the "Subtraction Treasury" audio may be more important than geometry).

These approaches make it possible to compare different LLMs for their suitability in generating VR spaces; to identify weak points (e.g., "the model generates objects well but poorly places them in space"); and to automatically reject poorly aligned generations or send them for re-refinement.

CONCLUSION

Summarizing the materials discussed, it can be concluded that the integration of AR/VR technologies and modern LLMs fundamentally transforms the paradigm of mathematics education, expanding both the didactic and cognitive tools of teaching. The implementation of virtual and augmented environments provides unique opportunities for the visual representation of abstract concepts, immersion of students in deeply personalized and contextualized scenarios, as well as the creation of situations that stimulate critical thinking and the development of spatial imagination. Thanks to the multilevel interaction of agents in multi-agent VR/AR educational systems, the student moves from passive perception of material to active participation in the construction of mathematical knowledge, while the educational process becomes dynamic, interactive, and motivationally enriched. It is especially important that such technologies open the way to studying complex mathematical theories, in particular IUT, through immersive simulations and multimodal explanations, which cannot be realized in the traditional 2D format.

The proposed approaches to VR scene generation and their multi-component control make it possible to formalize and automate the creation of educational virtual environments with a high degree of scenario correspondence, semantic consistency, and modality interaction. The quality assessment of such environments using modifications of a perplexity metric provides a new level of objectivity and controllability in applying LLMs for educational purposes. Thus, the synergy of AR/VR and agentic architectures not only increases the effectiveness of mastering mathematical knowledge but also creates the conditions for the formation of a new type of educational experience - interactive, adaptive, and deeply immersive in content.

BIBLIOGRAPHY

Bhardwaj, D., Beniwal, A., Chaudhari, S., Kalyan, A., Rajpurohit, T., Narasimhan, K. R., Deshpande, A., & Murahari, V. (2025). *Agent context protocols enhance collective inference*. arXiv. https://arxiv.org/abs/2505.14569

Gao, H., Huai, H., Yildiz-Degirmenci, S., Bannert, M., & Kasneci, E. (2024). DataliVR: Transformation of data literacy education through virtual reality with ChatGPT-powered enhancements. *arXiv preprint arXiv:2409.08738*. https://arxiv.org/abs/2409.08738

I. Habler, K. Huang, V. S. Narajala, and P. Kulkarni, "Building A Secure Agentic AI Application Leveraging A2A Protocol," arXiv, Apr. 2025. [Online]. Available: https://arxiv.org/abs/2504.16902

Izquierdo-Domenech, J., Linares-Pellicer, J., & Ferri-Molla, I. (2024). *Virtual reality and language models, a new frontier in learning*. International Journal of Interactive Multimedia and Artificial Intelligence, 8(5), 46–54. https://doi.org/10.9781/ijimai.2024.02.007.

Jelinek, F., Mercer, R. L., Bahl, L. R., Baker, J. K. (1977). Perplexity – a measure of the difficulty of speech recognition tasks. *J. Acoust. Soc. Am*. 62, S63. DOI: 10.1121/1.2016299.

Jeong, C. (2025). *A study on the MCP × A2A framework for enhancing interoperability of LLM-based autonomous agents*. arXiv. https://arxiv.org/abs/2506.01804

Liu, J., Yu, K., Chen, K., Li, K., Qian, Y., Guo, X., Song, H., & Li, Y. (2025). *ACPs: Agent collaboration protocols for the Internet of Agents*. arXiv. https://arxiv.org/abs/2505.13523

Madden, J., Pandita, S., Schuldt, J. P., Kim, B., Won, A. S., & Holmes, N. G. (2019). Ready student one: Exploring the predictors of student learning in virtual reality. *arXiv preprint arXiv:1910.10939*. https://arxiv.org/abs/1910.10939.

Martineau, K. (2025, May 28). *The simplest protocol for AI agents to work together*. IBM Research. https://research.ibm.com/blog/agent-communication-protocol-ai

Motejlek, J., & Alpay, E. (2021). Taxonomy of virtual and augmented reality applications in education. *arXiv preprint arXiv:2112.04619*. https://arxiv.org/abs/2112.04619

Nash, C., Carreira, J., Walker, J., Barr, I., Jaegle, A., Malinowski, M., & Battaglia, P. (2022). *Transframer: Arbitrary frame prediction with generative models*. arXiv. https://doi.org/10.48550/arXiv.2203.09494.

Nasir, M. U., James, S., & Togelius, J. (2024). *Word2World: Generating stories and worlds through large language models*. arXiv. https://doi.org/10.48550/arXiv.2405.06686

Novita, S., Setyowibowo, H., Wijayanti, P. A. K., Erwina, W., Yudiana, W., Dermawan Purba, F., et al. (2025). Exploring the impact of virtual reality-based mathematics learning on students' motivation: Protocol for a systematic review and meta-analysis. *PLOS ONE*, *20*(4), e0319664. https://doi.org/10.1371/journal.pone.0319664.

Shinichi Mochizuki, Inter-universal Teichmüller Theory I: Construction of Hodge Theaters. Publ. Res. Inst. Math. Sci. 57 (2021), no. 1/2, pp. 3–207 DOI 10.4171/PRIMS/57-1-1.

Shinichi Mochizuki, Inter-universal Teichmüller Theory II: Hodge–Arakelov-Theoretic Evaluation. Publ. Res. Inst. Math. Sci. 57 (2021), no. 1/2, pp. 209–401. DOI 10.4171/PRIMS/57-1-2.

Shinichi Mochizuki, Inter-universal Teichmüller Theory III: Canonical Splittings of the Log-Theta-Lattice. Publ. Res. Inst. Math. Sci. 57 (2021), no. 1/2, pp. 403–626. DOI 10.4171/PRIMS/57-1-3.

Shinichi Mochizuki, Inter-universal Teichmüller Theory IV: Log-Volume Computations and Set-Theoretic Foundations. Publ. Res. Inst. Math. Sci. 57 (2021), no. 1/2, pp. 627–723. DOI 10.4171/PRIMS/57-1-4.

Simonetti, M., Perri, D., Amato, N., & Gervasi, O. (2021). Teaching math with the help of virtual reality. *arXiv preprint arXiv:2111.01973*. https://arxiv.org/abs/2111.01973.

V.I. Slyusar, Y.P. Kondratenko, A.I. Shevchenko, T.V. Yeroshenko, 'Some Aspects of Artificial Intelligence Development Strategy for Mobile Technologies', Journal of Mobile Multimedia, Vol. 20_3, pp. 525–554, 2024. DOI: 10.13052/jmm1550-4646.2031.

V. I. Slyusar, 'Applications of Large Language Models in the Military Sphere', *Artificial Intelligence: Achievements and Recent Developments*, pp. 53 – 81, 2025.

V. Slyusar, 'Distributed Multi-agent Systems Based on the Mixture of Experts Architecture in the Context of 6G Wireless Technologies'. In: Dovgyi, S., et al. (eds) Applied Innovations in Information and Communication Technology. ICAIIT 2024. Lecture Notes in Networks and Systems, vol. 1338. Springer, 2025, pp. 81-110. DOI: 10.1007/978-3-031-89296-7_6.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention Is All You Need. *Neural Information Processing Systems*, 30, pp. 5998-6008.

Vincent-Lancrin, S., & van der Vlies, R. (2020). *Trustworthy artificial intelligence (AI) in education: Promises and challenges* (OECD Education Working Papers No. 218). OECD Publishing. https://doi.org/10.1787/a6c90fa9-en.

Y. P. Kondratenko, V. I. Slyusar, M. B. Solesvik, N. Y. Kondratenko, and Z. Gomolka, 'Interrelation and inter-influence of artificial intelligence and higher education systems', *Research Tendencies and Prospect Domains for AI Development and Implementation*, pp. 31–58, 2024.