An XR Environment for AI Education: Design and First Implementation

Yiyin Gu* Miguel Garcia† Yiqun Zhang‡ Alejandra J. Magana§ Bedrich Benes¶ Voicu Popescu‖

Purdue University, USA

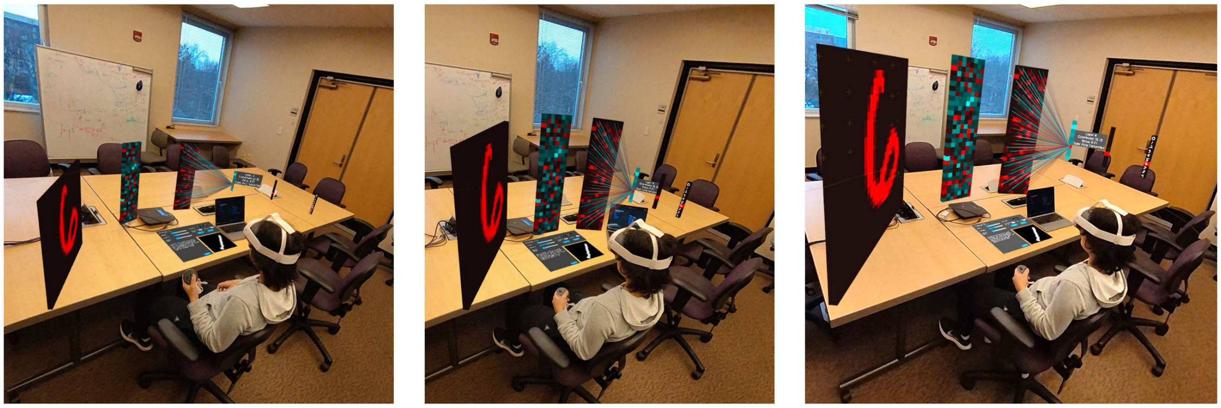


Figure 1: Second-person frame captures of our XR Environment for AI Education. The learner wears an XR headset to see an interactive visualization of a convolutional neural network for hand-written digit recognition. The visualization is integrated into the learner's view of the real world. The visualization is adjustable, here increasing in size from left to right.

Abstract

This work in progress paper presents and motivates the design of a novel extended reality (XR) environment for artificial intelligence (AI) education, and presents its first implementation. The learner is seated at a table and wears an XR headset that allows them to see both the real world and a visualization of a neural network. The visualization is adjustable. The learner can inspect each layer, each neuron, and each connection. The learner can also choose a different input image, or create their own image to feed to the network. The inference is computed on the headset, in real time. The neural network configuration and its weights are loaded from an onnx file, which supports a variety of architectures as well as changing the weights to illustrate the training process.

Keywords: Extended reality, immersive learning, AI education.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Applied computing—Education—Interactive learning environments.

1 Introduction

Artificial Intelligence (AI) is taking our society by storm, finding seemingly every day new and important applications. This wildfire-like growth has left educational institutions scrambling to stand up the AI curricula [18] needed to train the workforce for building AI computational solutions that are not only effective, but also robust and fair [21]. Whereas traditional computer science is difficult to teach and learn, AI is even more so [19]. AI computational solutions rely on deep neural networks with many layers, neurons, and connections, that are trained and then run on vast amounts of data. As such, an AI computational solution acts like a black box, with data paths that cannot be easily traced, inspected, or debugged [6, 17]. Since traditional computer science education tools cannot be readily used in the context of AI education, there is a dire need for tools developed from the ground up to support the teaching and learning of AI [7].

Immersive visualization, such as that afforded by a virtual reality (VR) or augmented reality (AR) headset, has considerable promise in AI education. The natural view selection interface and the depth cues afforded by immersive visualization facilitate the exploration of neural networks which are inherently 3D. Furthermore, immersive visualization removes the boundaries of physical displays, allowing for the application to instantiate virtual displays of a shape and size suitable for the minute exploration of complex neural network architectures. Like for many education applications of immersive visualization, AI education also stands to benefit from the embodied cognition opportunities presented to a learner wearing a VR or AR headset and engaged in motor behavior to assemble, fire, inspect, and modify a neural network.

Investigating the benefits of immersive visualization in AI education is timely. Like AI, VR and AR are in the midst of a revolution of their own. We now have all-in-one headsets with on-board tracking, rendering, networking, and power, all at consumer-level prices (e.g., Meta's Quest 3, $500). Such headsets provide the user with a completely untethered immersive visualization experience, with a wide field of view, a high resolution, and a comfortable form factor.

We are a team of VR/AR, AI, and educational technology researchers that has set out to harness the potential of VR/AR in AI education. In this work in progress paper we motivate and present the design of our novel immersive visualization system for AI education, and we present its first implementation.

Figure 1 illustrates our current system. The frames show our functioning system and are *not* a concept illustration. The frames were captured by a second person wearing a second XR headset that rendered the visualization seen by the learner. The second-person

*e-mail: gu251@purdue.edu
†e-mail: mfeijoog@purdue.edu
‡e-mail: zhan2897@purdue.edu
§e-mail: admagana@purdue.edu
¶e-mail: bbenes@purdue.edu
‖e-mail:popescu@purdue.edu

frames provide a better illustration of our system than the first-person frames seen by the learner because they also show the learner in relation to the visualization. The learner is seated at a table and wears an XR headset that allows them to see both the real world (i.e., the conference room) and a visualization of a neural network. The visualization is deployed on the table, around the learner. The visualization is parameterizable, which allows tuning the height of the layers, the radius of the cylindrical surface followed by the neural network, and the angle subtended by the neural network.

The learner can inspect each layer, each neuron, and each connection using a virtual laser pointer paradigm to trigger the display of textual information. The learner can also choose a different input image, or create their own image to feed to the network. The inference, i.e., the processing of the input image by the neural network is computed on the headset. The neural network architecture and its weights are loaded from a onnx file, which supports a variety of neural network architectures as well as changing the weights to illustrate the training process. For example, a lesson could load random weights to illustrate the initial, untrained state, it could load weights to illustrate underfitting or overfitting, or it could load weights to illustrate a biased network trained on an incomplete dataset. We also refer the reader to the video accompanying our paper.

2 Prior Work

This work in progress contributes to the field of Extended Reality (XR) in educational settings, particularly focusing on complex abstract concepts such as neural networks. Prior studies, like those conducted by Campos, Hidrogo, and Zavala [9], have successfully demonstrated the utility of Virtual Reality (VR) in enhancing comprehension of visually-oriented concepts, such as three-dimensional vectors in physics. These findings suggest VR's potential in aiding the visualization and understanding of complex ideas.

However, preceding studies have largely concentrated on subject matter that inherently lends itself to visual representation, like vectors and geometric figures [9, 10]. There remains a significant gap in applying VR technology to more abstract subjects like neural networks. A critical observation is that while there is some research on representing neural networks in VR, much of this work has not fully exploited the unique spatial capabilities of VR. Instead, these studies [14] often resort to traditional 2D screen representations, failing to leverage the immersive and three-dimensional potential that VR offers. This underutilization of VR's spatial properties limits the depth and effectiveness of educational experiences in understanding the complex and dynamic nature of neural networks.

In response to these limitations, this work employs the transparent passthrough feature of the Meta Quest 3 [16]. This innovation addresses the common issues of discomfort and motion sickness associated with VR [22], promoting a more accessible and comfortable learning environment [12]. By utilizing the full spatial potential of XR, our approach aims to provide a more intuitive and engaging method for teaching neural network concepts, moving beyond the constraints of traditional 2D visualization.

This work also builds upon existing research in 2D neural network visualization, like Adam W. Harley's study on interactive node-link visualizations of convolutional neural networks [11]. Harley's approach, allowing dynamic input changes, inspires this XR environment. This work aims to extend the techniques into XR, focusing on educational applications to offer learners a richer understanding of neural network data processing, leveraging the unique capabilities of immersive VR and XR environments.

3 System Design

We have designed our immersive visualization system for AI education based on the following considerations.

3.1 Immersive Visualization Technology

VR vs. AR. One consideration is where to place our immersive visualization on the virtual to real continuum.

At one end of the spectrum, a *VR system* shows its user only the virtual environment, and none of the user's physical surroundings. One advantage of this option is that the user is completely cut off from the real world, which provides the strongest sense of immersion into the learning environment. Another advantage is that the learning environment can be designed from scratch, without having to accommodate any of the real world constraints. The complete isolation of the user from the real world also has its disadvantages. One is the danger of the user colliding with real world obstacles that do not have a virtual world counterpart and are therefore invisible to the user. Another is that the learner does not see elements of the real world that are important to the learning activity, such as their own body, their own laptop, the instructor, or fellow students. Another challenge of VR is the difficulty of providing haptic feedback to the learner, for the learner to not just see, but also touch the learning environment. Finally, an important challenge of VR is cybersickness, which could preclude the long exposures needed for extensive learning activities.

At the other end of the spectrum, an *AR system* allows its user to see its physical surroundings, limiting the synthetic imagery to annotations overlaid onto the user's view of the real world. An advantage of AR is that the learning activities can involve tangible props, which provide the learner with realistic haptic feedback. Seeing the real world has important advantages: the learner can more easily avoid real world obstacles; the learner can see their own body, their own laptop, their colleagues, and the instructor; finally, the learner is anchored by the real world visualization and is therefore less prone to cybersickness. The reliance on tangible props brings several challenges, such as the need for props, the need to track the props for annotation anchoring, the need for real time depth acquisition for real and virtual depth compositing, and the limiting the learning activity to what can be done with the available props.

We have opted for a compromise position on the virtual to real continuum. The neural network is completely virtual, but the learner can also see their physical surroundings. This hybrid configuration, for which we use the generic term extended reality, is well suited for AI learning because it allows displaying any neural network in large format, which cannot be easily achieved with props, because it alleviates cybersickness, allowing for the extended exposure needed in light of the complexity of the subject matter, and because it does not fully isolate the learner, allowing them to interact with the instructor and fellow learners.

Video pass-through vs. optical pass-through. Once we have decided to employ an XR headset, the next consideration was whether to opt for a video pass-through or an optical pass-through headset.

A *video pass-through headset*, such as Meta's Quest 3, does not allow its user to see the real world directly. Instead, the real world is acquired with two video cameras facing away from the user, one for each eye, and the video feeds are shown to the user on two displays mounted in front of the user's eyes. The advantages of video pass-through include a virtual (i.e., active) field of view equal to the real-world field of view, robustness to ambient lighting, and lower cost. The disadvantages of video pass-through include: a real world field of view limited to the field of view of the camera; a limited dynamic range, i.e., the inability to show both very dark and very bright parts of the real world, in the same frame, without under or over exposure artifacts; limited resolution; the offset between the camera and the user's eyes, which images nearby real world objects at a different location from where the user would see them with their own eyes.

An *optical pass-through headset*, such as Microsoft's HoloLens 2, allows the user to see the real world directly, through a transparent plastic shield. Consequently, the user sees the real world with their natural vision's field of view, dynamic range, and resolution, from the correct viewpoints. A major disadvantage of optical pass-through headsets is a limited active field of view, i.e., the headset can display graphics only over a small region at the center of the display. Other disadvantages include the lack of support for true opacity, i.e., a dark virtual object cannot be overlaid onto a bright part of the real world, and high cost.

Visualizing a neural network requires a large active field of view. Furthermore, high device cost is a severe impediment to mass deployment in education. Consequently, we have opted for an XR headset based on the video pass-through technology.

3.2 Interaction paradigms

Walking vs. standing vs. sitting. Immersive visualization affords natural interactions for view selection. An important consideration is to decide on the range of natural viewpoint translation.

Walking affords the longest range of natural viewpoint translation to the user. Walking has the advantages of allowing the user to examine large virtual spaces naturally, of conveying to the user a sense of presence, and to afford embodied cognition opportunities involving the positioning of the user's entire body with respect to the dataset. A challenge of natural walking is the size discrepancy between the virtual environment that is typically much larger than the physical environment hosting the immersive visualization application. Palliative solutions such as teleportation or redirection can lead to user disorientation and cybersickness. For applications where physical exercise is not a primary goal, walking also brings the challenge of exertion.

Standing, e.g., in a floor area of 1 m × 1 m avoids the limited real world space constraint and improves ergonomics, but the range of viewpoint translations is greatly reduced.

Sitting affords minimal viewpoint translations, reduced to side-to-side head motions, and limits embodied cognition opportunities to the motions the user can make while seated. However, sitting has ideal ergonomics.

It is our belief that immersive visualization can and should be used in education over extended periods of time, the same way non-immersive visualization on a conventional display is used in conventional office work for many hours on end. For this, we have elevated the ergonomics concern above the viewpoint translation range concern and have decided that our user should be seated (Figure 1).

Outside looking in vs. inside looking out. Another aspect of the visualization design is whether the learner should focus on a workspace in front of them, i.e., an outside-looking-in design, or whether the learner should be surrounded by the workspace from all direction, i.e., an inside-looking-out design. For a seated learner, the outside-looking-in design limits the size of the workspace. On the other hand, the inside-looking-out design requires the ability to pan the view direction, possibly in all directions. A swivel chair allows a seated learner to look in all directions easily, and the real world provides the landmarks to assist the learner in remembering where each layer is located.

We have opted to allow the application to choose how big of an angle in front of the user is subtended by the display volume. A 360° angle provides a true inside-looking-out visualization design (Figure 2).

Small vs. large display volume. Regardless of whether the visualization is in front of the learner or whether it is surrounding the learner, the 3D display volume is an orthogonal and important parameter. A smaller display volume places the visualization closer to the learner, resulting in stronger depth cues. On the flip side, a small display volume just doesn't have sufficient space to display a complex neural network eloquently (e.g., the many lines between two fully connected layers). A large display volume pushes the visualization farther from the user dulling the depth cues, but is also can provide an adequately low visualization density for complex networks.

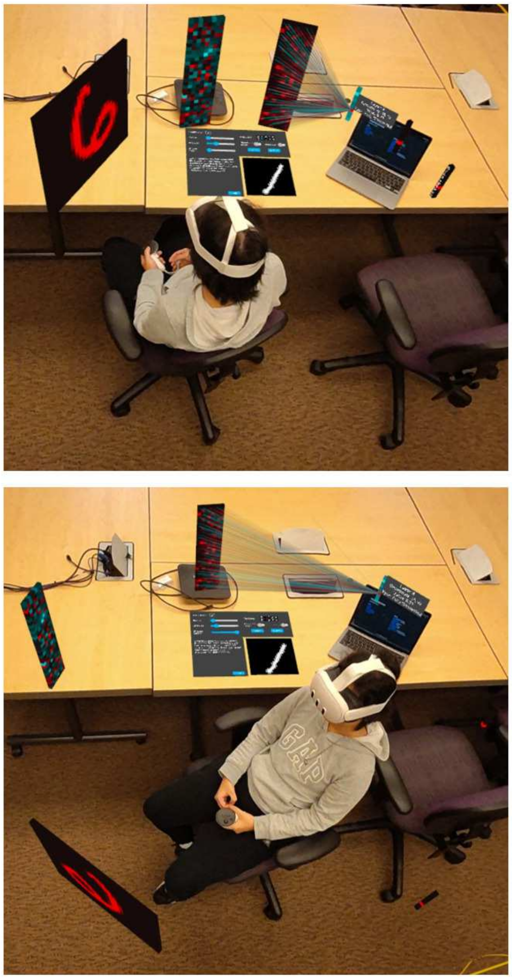


Figure 2: Neural network visualization deployed in front of the learner (top), and all around the learner (bottom).

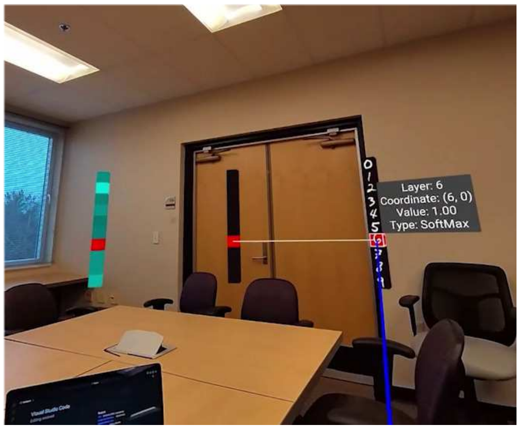
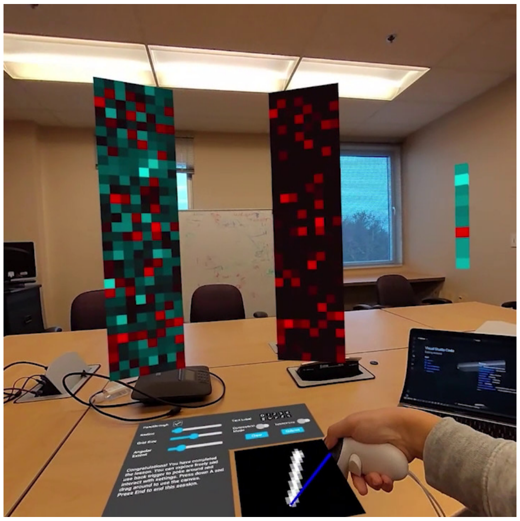


Figure 3: Frame captured from the learner's XR headset. The learner uses the virtual laser pointer (blue) to probe the firing value of neuron "6" of the SoftMax layer. The value is 1.0, indicating the correct classification of the input image (see Figure 1).



Congratulations! You have completed the lesson. You can explore freely and use back trigger to poke around and interact with settings. Press down A and drag around to use the canvas. Press End to end this session.

Figure 4: Frame captured from the learner's XR headset. The learner uses the virtual laser pointer to draw a digit to be classified by the neural network.

The eloquence of an XR visualization also depends on whether the real world scene has adequate empty space to host the visualization. A large visualization requires sufficient real world empty space in front of the user. Consider the case of a 0.5 m desk placed against a wall. If the visualization exceeds the available space of 0.5 m, the visualization will appear to be deeper than the wall, resulting in a depth perception artifact.

We have opted to allow the application to control the size of the display volume through a layer panel height and a display radius parameter (Figure 1 from left, to middle, to right).

Hands vs. controller. State of the art XR headsets support hand tracking, which opens the door to freeing the learner from burdensome handheld controllers. One challenge is that the user has to learn pointing and grabbing gestures. Another challenge is that the ergonomics of the interface deteriorate when the visualization is large and the learner has to reach far and move their hands over large distances.

We have opted for a conventional virtual laser pointer paradigm. The laser beam is always on, and the trigger is the only controller button used. The simple and ergonomic interface is sufficient to select elements of the neural network and to request additional information (Figure 3). The interface also allows drawing (”spraying”) on a tablet to create new input images (Figure 4).

Passive haptics vs. active haptics. Providing haptic feedback to the user of an immersive visualization application can be done with one of two fundamental approaches.

Passive haptics relies on aligning real world objects with the virtual entity with which the user makes contact. The advantage is the realism of the touching sensation, and the disadvantage is that opportunities for passive haptic feedback are sparse, since it is unlikely that a physical object is aligned with each virtual object with which the user would like to interact. Increasing the number of passive haptic opportunities can be done by either modifying the physical world using an encountered type haptic device (ETHD), or by modifying the virtual world, through redirection. A possible ETHD implementation is a robotic arm that places a ”carried” object in alignment with the virtual object the user is about to interact. A possible redirection technique takes advantage of moments of user inattention, such as saccades, to shift unbeknownst to the user the virtual object into alignment with a nearby physical object.

Active haptic devices are worn by the user and apply pressure to the user's body in sync with the user's contact with the virtual object. The advantage is that the user takes the haptic device with them, so no alignment is needed. One disadvantage is encumbrance, e.g., a user feels a haptic glove at all times and not just when touching virtual objects. Another disadvantage is that the haptic feedback can lack realism, e.g., the user pressing their finger against a virtual wall will feel a pinch from the haptic glove, and not just a pressure, and the user's finger can still move right through the virtual wall. Finally, active haptic devices increase the equipment and logistical cost of the educational intervention.

As described above, we have opted for a virtual laser pointer interaction paradigm, which makes haptic feedback less imperative than hand interaction. Our system does calibrate the physical desk, which could be used to provide passive haptic feedback, for example when editing a new input image. The handheld controller provides *active* haptic feedback, which can be used to guide user interaction, for example to indicate when the laser beam hits a layer, or to stabilize the selection of an individual neuron.

Text input. Text input is notoriously challenging in the context of immersive visualization. One option is an ”air keyboard” that the learner punches with the virtual laser beam. The advantage is suitability to the immersive visualization interface, and the disadvantage is the slow rate of the text input. Two recent developments hold the promise of providing effective solutions to the problem of text input in the context of immersive visualization. One is the advent and proliferation of large language models (LLMs) that improve speech to text robustness, taking into account accents and domain specific vocabulary. Speech to text is now performing well enough to become the preferred text entry interface not only when the keyboard is missing, but also when the keyboard is small or not easily accessible, such as that of phones. The second development is the quality of the video pass-through mode of XR headsets which allow users to see physical keyboards well enough to type proficiently.

In our context of a learner seated at a table, involving a physical keyboard is practical–the learner can use a laptop or a standalone wireless keyboard placed on the table to enter text.

3.3 Software and Hardware System Concerns

Neural network description file format It is important that the XR environment import neural network architectures developed using AI software development suites such as PyTorch [20]. For this the XR environment has to parse popular neural network description formats. We have opted for the onnx [8] file format that most AI tools export and import.

Computational load partitioning. The XR headset has significant CPU and GPU computation capabilities, but the XR headset remains a ”thin” client that has to be shielded from excessive computational and rendering load. Capping the computational load is particularly important in immersive visualization as frame fluctuations have been shown to induce cybersickness [23]. We distinguish between the computational load needed to run a trained network on a given input and the that needed to train the network.

For small neural networks, such as the two-layer convolutional neural network running on the the MNIST hand-written digit recognition dataset [13] used as an example throughout this paper, the XR headset can run the inference in real time. This allows the learner to create novel input images and to investigate the network behavior on the new input. When the neural network is too complex to be run on the headset, a client/server architecture should be used to offload the computation to a nearby server, e.g., a laptop, leveraging the headset's wireless communication capabilities.

For the foreseeable future, XR headsets will not and should not be able to train networks. One reason is that it would take too long and would interfere with the learning activity. Another reason is that such intense computation would prematurely deplete the battery of the headset, requiring that it be plugged in, again interfering with the learning activity. Shielding the headset from the computational load of training networks will require pre-training the networks offline and saving training states (i.e., weight values) as needed for the learning activity, e.g., to illustrate under-fitting, over-fitting, or incorrect training leading to bias.

4 Implementation

We started development in September 2023 using Meta's Quest Pro [3], for its color pass-through mode as opposed to the black and white pass-through mode of the Quest 2 [1]. In December 2023 we switched to Meta's newly released Quest 3 [2], which improved on the pass-through mode of the Quest Pro while at the same time reducing its price tag three fold. We developed our XR environment in Unity 3D [5], release 2022.3.5f1. We also use Meta's XR All-in-one SDK [15]. The onnx [8] files are loaded leveraging the Barracuda framework [4]. The application does not exceed the rendering capabilities of the Quest 3 XR headset so it runs at the native frame rate of 72Hz.

5 Conclusions, Limitations, and Future Work

We have described the design and first implementation of an XR environment for AI education. The system relies on extended reality technology, which allows the learner to see the neural network in 3D, and also the surrounding real world environment; this alleviates cybersickness concerns and it allows the learner to see important elements of the real world such as their own body, desk, laptop, colleagues, and instructor. The learner is seated, which alleviates exertion and collision concerns. The visualization is reconfigurable in terms of panoramic height, radius, and learner wraparound angle.

One limitation stems from the inadequate handling of occlusions. The neural network is fundamentally a 3D structure and the user cannot gain adequate line of sight to all layers and to all the connections between them. After testing the initial implementation we have added a parameter to the visualization that defines the orientation of the layers, from tangential to the cylindrical display, i.e., perpendicular to the user's view direction (Figure 5, left), to radial, i.e., parallel to the user's view direction (Figure 5, right). The radial configuration better reveals the connections to the detriment of the layers, whereas the tangential configuration better reveals the layers, to the detriment of the connections.

The visualization has a number of parameters (radius, height, angle, orientation) which are presently set manually. The settings can be saved and reused in future sessions. However, future work should examine setting these parameters automatically, based on the neural network to be displayed, on the size of the available physical space, as well as on user preference.

In the current implementation, our application employs a pre-trained model loaded onto the device, keeping the neural network's weights static during use. Future iterations could include multiple neural network states to simulate the dynamic evolution of weights during training. Implementing live training sessions would require a backend system with GPUs to support semi-real-time training, thereby enhancing the application's educational value.

Regarding future user studies, a variety of methodologies are available. Parallel Learning Groups involve dividing participants into two groups, one using the XR interface and the other a desktop version. Care will be taken to ensure demographic similarity and baseline knowledge parity between groups for a direct comparison of learning outcomes. Alternatively, a Cross-Over Design could be employed, where participants are exposed to both interfaces but with different content or modules. This method allows for an assessment of user preference and engagement levels while controlling for learning material variability.

In conclusion, the environment is fully functional and ready to be deployed in educational activities. While there is scope for future enhancements and research, its current state presents a robust platform for engaging and innovative learning experiences in the field of AI education.

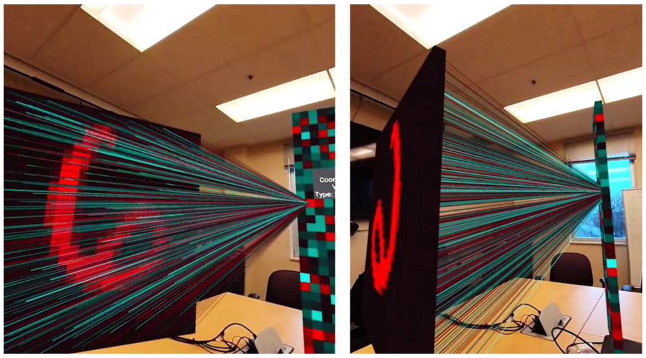


Figure 5: Layers oriented tangentially (left) and radially (right) with respect to the cylindrical volumetric display axis.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Nos. 2212200, 2219842, 2309564, and 2318657.

References

[1] Meta quest 2: meta.com/quest.

[2] Meta quest 3: www.meta.com/quest/quest-3.

[3] Meta quest pro: meta.com/quest/quest-pro.

[4] Unity barracuda: https://github.com/unity-technologies/barracuda-release.

[5] Unity real-time development platform: unity.com.

[6] A. Adadi and M. Berrada. Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160, 2018.

[7] AI4K12. The artificial intelligence (ai) for k-12 initiative (ai4k12), 2022.

[8] J. Bai, F. Lu, K. Zhang, et al. Onnx: Open neural network exchange. `https://github.com/onnx/onnx`, 2019.

[9] E. Campos, I. Hidrogo, and G. Zavala. Impact of virtual reality use on the teaching and learning of vectors. *Frontiers in Education*, 7, 2022. doi: 10.3389/feduc.2022.965640

[10] Q. Chen, L. Zhang, B. Dong, and Y. Zhou. Interactive cues on geometry learning in a virtual reality environment for k-12 education. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 524–526, 2023. doi: 10.1109/VRW58643.2023.00114

[11] A. W. Harley. An interactive node-link visualization of convolutional neural networks. In *ISVC*, pp. 867–877, 2015.

[12] M. Kaufeld, M. Mundt, S. Forst, and H. Hecht. Optical see-through augmented reality can induce severe motion sickness. *Displays*, 74:102283, 2022. doi: 10.1016/j.displa.2022.102283

[13] Y. LeCun, C. Cortes, and C. Burges. The mnist database of handwritten digits, 2022.

[14] N. Meissler, A. Wohlan, N. Hochgeschwender, and A. Schreiber. Using visualization of convolutional neural networks in virtual reality for machine learning newcomers. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 152–1526, 2019. doi: 10.1109/AIVR46125.2019.00031

[15] Meta. Meta xr all-in-one sdk: https://developer.oculus.com/downloads/package/meta-xr-sdk-all-in-one-upm/.

[16] Meta. Getting started with quest pro: Full-color passthrough, 2023. Accessed: 2023-12-31.

[17] NASEM. Data science for undergraduates. opportunities and options, 2018.

[18] NCES. The artificial intelligence (ai) for k-12 initiative (ai4k12) aaai and csta., 2013.

[19] D. T. K. Ng and S. K. W. Chu. Motivating students to learn ai through social networking sites: A case study in hong kong. *Online Learning*, 25(1):195–208, 2021.

[20] S. Paszke, Adam abd Gross, S. Chintala, and G. Chanan. Pytorch, 2016.

[21] C. Pazzanese. Ethical concerns mount as ai takes bigger decision-making role in more industries. *The Harvard Gazette, October*, 26:2020, 2020.

[22] D. Saredakis, A. Szpak, B. Birckhead, H. A. D. Keage, A. Rizzo, and T. Loetscher. Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, 14, 2020. doi: 10.3389/fnhum.2020.00096

[23] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cyber-sickness: Impact, causes, and measures. a review. *Frontiers in Virtual Reality*, 1:582204, 2020.