Check for updates

Computer Applications in Engineering Education

WILEY

RESEARCH ARTICLE OPEN ACCESS

Students' Conceptual Explanations of Neural Networks Enabled by Extended Reality Learning: A Multiple Methods Approach

Miguel A. Feijoo-Garcia[1] | Yiqun Zhang[1] | Yiyin Gu[2] | Alejandra J. Magana[1] | Bedrich Benes[2] | Voicu Popescu[2]

[1]School of Applied and Creative Computing, Purdue University, West Lafayette, Indiana, USA | [2]Department of Computer Science, Purdue University, West Lafayette, Indiana, USA

Correspondence: Alejandra J. Magana (admagana@purdue.edu)

Received: 25 January 2025 | **Revised:** 22 May 2025 | **Accepted:** 11 September 2025

Funding: This study was supported in part by the National Science Foundation under award numbers 2412928, 2417510, 2212200, 2219842, 2309564, and 2318657.

Keywords: artificial intelligence | computer science | educational technology | extended reality | immersive learning | multiple methods | qualitative analysis | quantitative analysis | XR learning

ABSTRACT

This study examines the use of extended reality (XR) in helping students with conceptual comprehension of artificial intelligence (AI) concepts, specifically neural networks (NNs) and handwritten digit recognition. Using a multi-methods approach, this study assesses student performance and understanding of such concepts. Student participants ($N = 29$) engaged in an XR environment designed to teach NNs and completed in-lesson assessments consisting of multiple-choice questions and open-ended questions. Quantitative data were analyzed using the *k*-means clustering method to classify performance levels based on the accuracy of the answers. The elbow approach determined the number of clusters, and the average silhouette score showed the cluster quality after clustering. Qualitative data underwent thematic analysis to identify challenges in handwritten digit recognition. Results showed that the accuracy of the students' responses ranged from 17% to 100% and could be classified into three groups, and that factors like handwriting clarity, digit placement, and writing style significantly impacted the accuracy of handwritten digit recognition. The findings suggest the potential of using XR for supporting learning and engagement in studying AI concepts. Future research is encouraged to apply XR across various education levels and explore broader AI concepts. This study contributes to the literature on applying XR in computer science education by providing insights into how XR can enhance conceptual comprehension of complex AI concepts like NNs.

1 | Introduction

Immersive learning is a pedagogical approach that leverages advanced technologies, particularly virtual reality (VR) and augmented reality (AR), to create engaging and interactive learning environments. This method is characterized by its ability to immerse learners in simulated or artificial environments that enhance their understanding and retention of complex concepts. The immersive experience is achieved through high representational fidelity, which fosters a sense of presence and engagement among learners, thereby motivating them to actively participate in the learning process [1, 2].

Abbreviations: MR, mixed reality; STEM, science, technology, engineering, and mathematics; VR, virtual reality; XR, extended reality.

Miguel A. Feijoo-Garcia and Yiqun Zhang contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Computer Applications in Engineering Education* published by Wiley Periodicals LLC.

Specifically, immersive learning refers to learning experiences that are accessed through a head-mounted display in environments with simulated components [3, 4]. Immersive learning initiates with a student-centered design to motivate learners and help them develop the intended skills or learning outcomes [5].

Extended reality (XR) is a technology that enables immersive learning [5, 6]. XR can be considered to encompass mixed reality (MR), AR, and VR technologies [7], mainly combining virtual and real worlds [8]. XR has been advancing at an unprecedented rate [9] as further evidenced by the rapidly increasing number of XR-related publications [10]. Immersive learning with XR has recently become popular and spans different domains, such as medicine and computer science [11]. In a similar manner, XR can positively affect science, technology, engineering, and mathematics (STEM) education [12].

Immersive learning through VR and XR technologies has many educational advantages as it makes learning accessible [13], increases engagement [14], and improves learning outcomes in terms of performance [15] and overall achievement [16]. Moreover, VR and XR can create highly realistic simulations, offering a safe space for learners to practice skills or solve problems at an affordable cost [17]. Therefore, STEM topics can often be taught at different education levels, such as middle school, high school, and undergraduate, using immersive learning [14]. However, VR and XR learning experiences have similar and different advantages for supporting learning. For instance, while VR experiences have the ability to provide learners with a total immersion that promotes creativity and problem-solving by allowing learners to experience scenarios that could not be possible in traditional settings [15], XR overcomes some limitations VR has, such as isolating users from their physical environment and the potential for cybersickness [18], by combining real and virtual elements to create a more balanced and comfortable learning environment.

Unlike VR, XR allows learners to see key parts of their physical surroundings, such as their desk, laptop, peers, or instructor, which reduces feelings of isolation and supports collaboration [14]. By blending real and virtual environments, XR also minimizes the risk of cybersickness [19]. Overall, these features make XR particularly useful for teaching and learning, given its affordability and adaptability for hybrid or group learning, where interaction with both real and virtual spaces is important [17, 19]. XR can create highly realistic simulations, offering a safe space for learners to practice skills or solve problems at an affordable value [17], thus increasing learner motivation and engagement as the novelty of the technology captures their attention throughout the learning process [16].

Educational researchers have documented the learning benefits of using XR technology in STEM education [12] as many STEM concepts are abstract, counterintuitive, and difficult to understand. Researchers have recommended using XR technologies for these topics in STEM, where three-dimensional (3D) visualizations could benefit students by making such concepts more accessible [20]. Artificial intelligence (AI) is one of the abstract concepts in STEM. Learning the underlying AI algorithms and concepts, such as neural networks (NNs), gradient descent, or back propagation, is necessary but presents significant challenges. These topics are difficult to grasp using traditional teaching methods, and there is a gap in finding effective pedagogical approaches to make these complex AI concepts more understandable and engaging. Given the affordances of immersive technologies, we hypothesize that XR can support learners in their understanding of topics related to AI.

This descriptive study investigates how learning of AI concepts, specifically related to NNs, is supported within an XR-based environment. This involves understanding the interplay between the XR technology and the learning activity, such as the sequencing of the content and embedded scaffolding, to help students grasp the AI concepts. For instance, scaffolding may include interactive prompts that guide student interaction with the NN model, feedback (e.g., neuron activation patterns), and integrated conceptual checks that require students to apply their emerging or enhanced knowledge (e.g., in-lesson questions). Particularly, this study evaluates the XR condition for learning NNs and handwritten digit recognition to approach the following research questions:

- *What are learners' conceptual understanding of neural networks while interacting with an XR environment?*
- *How do the learners' conceptual explanations relate to their performance enacted during the immersive learning experience?*

This study used a multi-methods approach to answer these questions, analyzing quantitative data from in-lesson multiple-choice questions (using the *k*-means clustering method) and qualitative data from open-ended voice-to-text responses (using thematic analysis). An important implication of this study is that it informs computer science educators about using XR and its effectiveness for teaching and learning AI-related concepts, addressing important questions in XR studies: (a) for whom, (b) the purposes and conditions, and (c) the potential effectiveness [7].

2 | Related Work

2.1 | Immersive Learning and Frameworks

Immersive technologies offer compelling interactive experiences and have been applied in education [21] to enhance learning participation and outcomes [6]. They transform traditional learning methods by creating environments where learners can interact with realistic simulations or scenarios, leading to deeper engagement [14]. Immersive learning has been defined from educational (e.g., supporting engagement) and technological (e.g., involving simulation aspects) perspectives [21]. Immersive learning frameworks have been developed through technological, pedagogical, and psychological features, such as platform, context, and motivation, respectively [21]. Presence, immersion, cognition, emotion, and motivation can affect immersive learning results [22]. Integrating pedagogical, psychological, and technological aspects in immersive learning emphasizes the need for well-rounded strategies that not only leverage technology but also consider human cognitive and emotional responses [21].

Immersive learning is not only about integrating new technologies into education but also about creating meaningful and

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

engaging experiences for learners [23]. These experiences should allow students to practice, explore, or solve problems in scenarios impossible in traditional learning environments [24]. However, achieving this balance requires careful design of the immersive learning experiences and the system's implementation to avoid making them overly complex, leading to technological distractions or cognitive overload [18]. Instructors can create learner-friendly environments that address learners' emotional engagement, the cognitive load on learning outcomes, embodiment, and social interaction and reflect on how learners feel and think while interacting with technology, ensuring that immersive technologies enhance rather than hinder the educational experience [25, 26].

About two decades ago, a four-dimensional framework was introduced to explore games, simulations, and immersive environments in education, encompassing pedagogical (e.g., learning theories), learner-related (e.g., profile), contextual (e.g., outdoors), and representational (e.g., interactivity) aspects [27]. Its later applications suggested that designing immersive learning experiences that are holistic and interactive is important, making them learner-centric and adaptable to diverse educational needs and technological advancements [28].

Later, the CAMIL framework was developed to understand immersive learning, using immersive VR as a demonstration [4]. This model bridges cognitive science and immersive technology, offering insights into how mental processes interact with technological affordances in learning environments. The emphasis was placed on the interaction between media and methods, highlighting instructional methods that value student presence and agency because they can lead to better learning [4]. The authors discussed how presence and agency impact six cognitive and affective factors: interest, intrinsic motivation, embodiment, cognitive load, self-efficacy, and self-regulation. CAMIL provides a roadmap for educators and developers to optimize immersive learning systems by breaking these factors into actionable components. The connection between these factors and different learning outcomes, including factual, conceptual, and procedural knowledge, along with knowledge transfer, suggests that lower cognitive loads and higher levels of the remaining five factors could improve learning outcomes. Reducing the cognitive load while maintaining other factors ensures that learners can focus on content rather than the mechanics of technology [4].

Another immersive learning evaluation framework was developed specifically in the context of higher education, comprising a game-based approach with five characteristics: goal, results, tasks, scoring criteria, and rating methodology, offering a structured way to measure the effectiveness of immersive learning [29]. This framework was implemented in an educational mobile AR game using interviews to elicit responses, where participants acknowledged the potential and barriers to using immersive technologies in education and generally agreed with the usefulness of the proposed framework in evaluating learning in immersive settings [29]. These findings emphasize the importance of addressing barriers such as accessibility and usability to realize the full potential of immersive technologies. Recommendations for effective practices in game-based learning evaluation in immersive settings include choosing an engaging immersive learning environment [29].

2.2 | Extended Reality

XR has been useful in promoting quality and sustainability in education [11]. XR offers diverse opportunities to enhance learning experiences. XR uses software and hardware to give the audience a comprehensive interactive experience and facilitates interactions [11]. The literature indicates that overall interest in XR in education has grown worldwide [11, 12], and this growth reflects the increasing recognition of XR's potential to revolutionize traditional educational methods by providing innovative tools for instruction and practice. In practice, XR helps with personalized learning and effective interactions and engagement in the learning process [30]. For example, XR scenarios can support learners in learning complex concepts or enhancing hands-on skills in a risk-free environment [30, 31]. XR is also promising in spatial reasoning skill development [32].

Both the educational and technological sides are crucial for XR in education: for instance, XR classrooms can bring more innovation to education [11]. However, the success of XR implementation also depends on addressing challenges like accessibility, cost, and instructor training [33].

A part of XR, AR provides interaction through an enhanced real-world environment [34]. Literature has shown that AR has been used in distinct subjects and grade levels within education [34]. AR applications in education can be categorized into three themes [34]: (1) hardware-based [years 1995–2009], (2) application-based (years 2010–2019), and (3) device-based (years 2020 and beyond). Although many advantages (e.g., enhanced learner motivation) have been realized over time, there are pending problems like cognitive overload [18]. Additionally, more qualitative AR-in-education studies are needed [35].

The integration of AR in education goes beyond individual learning outcomes to broader pedagogical innovations [36, 37]. The literature indicates that using AR can lead to higher levels of interest and participation in learning activities by allowing students to interact with content at their own pace and revisit challenging concepts [38]. However, educators must ensure AR experiences align with curricular goals and provide meaningful contexts to avoid novelty-driven distractions [36].

VR, on the other hand, digitally represents 3D objects and allows seamless immersion and interaction via head tracking [39]. According to the literature, VR can be used in many fields of education, such as computer science and engineering, and the interest in its usage in education continues to increase [40]. The applications of VR have been studied as a tool, design, and context in education [23]. Common advantages of using VR in education include increased engagement and improved motivation, but issues such as user privacy need to be addressed in research design [23]. VR technology supports deeper understanding through experiential learning, enabling users to interact with digital objects and scenarios dynamically [41]. The increasing interest in VR across various educational fields reflects its ability to engage students more effectively and motivate them to learn [23, 42].

A part of XR, MR combines physical and digital elements [43]. Users can see objects and interact with or manipulate them

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[44]. Compared to AR and VR, MR benefits learners in similar ways, but it may be harder to access due to high computational requirements [44]. However, the use of MR in education and training has become popular and is still growing [43]. The literature indicates that MR in higher education, especially in STEM, helps students learn by combining the real world with digital elements, allowing students to interact with complex systems in an easier way to understand through real-time interaction [45].

2.3 | Teaching Deep Neural Networks

Deep NNs are inspired by human brain functioning. They have become the fundamental element of modern AI and have helped solve long-standing computing problems [46], such as data processing and pattern recognition [47]. An NN model is an algorithm that, mathematically, is a directed graph with distinct properties, including a state variable and weights [48]. The modified National Institute of Standards and Technology (MNIST) data set is popularly used for machine learning or deep learning tasks [49, 50]. MNIST contains 60,000 normalized handwritten digit images for training and 10,000 normalized handwritten digit images for testing [51]. Researchers have used MNIST to learn NNs [52]. Although comprehending the concept of NNs and how they operate is difficult for learners [46], XR provides advanced visualization for learners to explore NNs and understand them better [53]. Hence, the core learning need addressed by this study is based on the challenges students may face in grasping complex and abstract AI concepts.

In education, immersive technologies such as VR and AR can help students comprehend abstract and complex topics [54]. As NNs are important in fields such as AI and data science, learning about them provides students with valuable skills [6]. Immersive technologies make learning more engaging by allowing students to interact with and see NN structures in 3D, helping to connect theory to real-world applications [55]. Literature also suggests immersive technologies improve learning outcomes, increase retention, and develop critical thinking skills [56, 57]. Such technologies allow students to learn at their own pace, making education more inclusive [58, 59]. While immersive technologies have been increasingly explored in various educational domains, the literature has overlooked the conceptual understanding that the students acquire when interacting with complex visualizations aimed at teaching AI-related concepts through XR.

Previous research has investigated the use of XR for supporting the learning of AI [9], and it established the core functionalities, including the video pass-through interface, interactive visualization of NN layers, and controller-based manipulation [9]. The XR system allowed learners to interact with the system to investigate the NN, input images, and subsequently respond to learning questions [9]. While seated, the learner could see the virtual components and physical surroundings, avoiding collision or isolation [9]. Later, a comparative study was conducted against a traditional desktop interface, investigating differences in user experience metrics, such as usability and overall satisfaction [31]. That study emphasized that students were satisfied with the XR system and found it easy to use and learn AI [31].

This manuscript focuses on evaluating how effectively the XR environment [9] (see Figure 1 for system design) helps students gain a conceptual understanding of NNs, which is assessed by looking at their performance and rationales (voice-to-text responses). Recognizing the need for research highlighting learning outcomes, this study addresses the identified gap. The novelty of this study lies primarily in "how" this conceptual understanding is acquired as students interact with an XR environment. Specifically, by applying a multi-methods analysis approach, which combines quantitative performance data with qualitative explanations, we explore how this XR approach impacts conceptual understanding as learners articulate complex AI concepts like NNs and handwritten digit recognition.

3 | Methods

This descriptive study investigates students' conceptual understanding of NN architecture and function in the context of handwritten digit recognition, as students interact with an XR environment designed to teach such concepts.

3.1 | Context and Participants

Conducted in early 2024, this study included 29 participants ($N = 29$) who first provided demographic information and reported their familiarity with NNs and handwritten digital recognition as general background information [31]. Then, the students (i.e., participants) used the educational XR system to learn how NNs work (see Table 1). There were (18, 62.1%) male and (11, 37.9%) female participants (20, 68.9%), undergraduate and (9, 31.1%) graduate students. Most participants were computer science students (18, 62.1%). Moreover, although 17 participants (58.6%) had a basic understanding of NNs, 18 (62.1%)

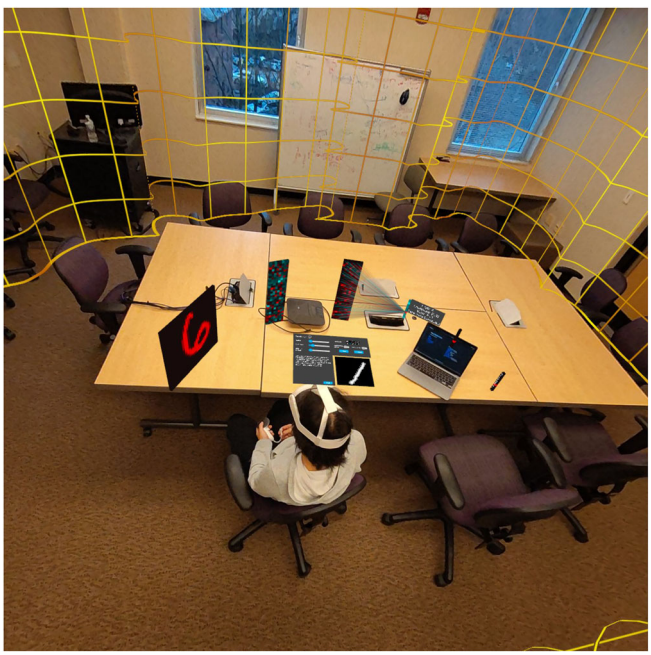


FIGURE 1 | Second person view of the XR system. The learner is seated at a conference table and sees the NN visualization integrated into their physical surroundings.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Participant demographics and background characteristics in the XR group.

Variable	XR ($N = 29$)
Gender	
Male	18 (62.1%)
Female	11 (37.9%)
Other	0 (0%)
VR headset usage	
Never	6 (20.7%)
Once	11 (37.9%)
More than once	12 (41.4%)
XR usage	
Never	15 (51.7%)
Once	8 (27.6%)
More than once	6 (20.7%)
Major	
Computer Science	18 (62.1%)
Data Science	3 (10.3%)
Other	8 (27.6%)
Role	
Undergraduate	20 (68.9%)
Graduate	9 (31.1%)
Familiarity with NNs	
No	12 (41.4%)
Yes	17 (58.6%)
Familiarity with handwritten digit recognition	
No	18 (62.1%)
Yes	11 (37.9%)

Note: Any discrepancies in percentages are due to rounding. Values represent frequencies with percentages in parentheses.

did not have foundational knowledge of handwritten digit recognition. Also, participant familiarity with immersive visualization technology varied, with some students having used VR headsets more than once (12, 41.4%), while a majority had never used XR (15, 51.7%).

3.2 | The XR Learning Environment

The XR learning environment was developed in Unity 3D (version 2022.3.5f1) using Meta's XR All-in-One SDK. The implementation relied on the Barracuda framework [60] to load and run pre-trained NN models, like the MNIST convolutional NN [9]. The system was deployed on Meta's Quest 3 headset (Snapdragon XR2 Gen 2 processor, with 8GB of RAM) [61], supporting a native frame rate of 72 Hz.

The system visualizes an NN trained on the MNIST data set by rendering its input, hidden, and output layers as interactive panels, in 3D, arranged in a cylindrical pattern, on the table, in front of and surrounding the user (see Figure 1). The system loads the weights of a pre-trained NN model and runs it directly on the XR device, on stored or user-generated input. Participants interact with the system using the handheld controllers, through a virtual laser pointer, which can be used to point at and trigger 3D network layers to see details such as neuron activation values (see Figure 2), and to write digits to feed the network (see Figure 3). The user can provide existing or new handwritten digit images as input to the NN, to observe how the input activates nodes across layers and how the digit is ultimately classified [9, 31] (Figure 3).

3.3 | Procedures and Data Collection

Figure 4 illustrates the conceptual design of the study, which had three phases: (1) preparation, (2) learning, and (3) data analysis. The sequence of activities, depicted in Figure 4 and described through steps S1–S7, contains the participant procedural timeline as part of it. The entire session for each participant, encompassing preparation, learning within the XR environment, and initial feedback, was designed to last 30–40 min, with staggered start times implemented to manage the process effectively [31].

Questionnaires were administered using Qualtrics to collect general demographic data and to assess the baseline prior knowledge as part of background information. The training on the headset operation was delivered via Google Slides presented on a tablet, and the research team's guidance during the intervention was provided as needed. Afterward, the core learning experience involved interacting with the educational content through an XR headset equipped with controllers, which allowed manipulation and engagement with the content.

The structured instructional design began with the preparation phase, where participants provided their consent and completed a demographic questionnaire. This phase corresponds to *S1* (Consent and Demographics), including sub-activities *S1.1* and *S1.2*. After that, the basic knowledge questionnaire (not a pre-test) was provided to gather information about the participants' understanding of NN concepts using closed-ended questions on topics such as nodes, layers, and hardware components. This corresponds to *S2* (Baseline Questionnaire). The participants answered five questions randomly selected from the following list:

- What is a neural network?
- What is a node in a neural network?
- What is a layer in a neural network?
- Arrange the following in the correct sequence of processing in a neural network:
  - Hidden Layer
  - Output Layer
  - Input Layer
- (True/False) Unsupervised learning involves training models on labeled data with explicit guidance.
- (True/False) The architecture configuration of a neural network can affect its energy consumption, with larger and more complex networks generally consuming more energy.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

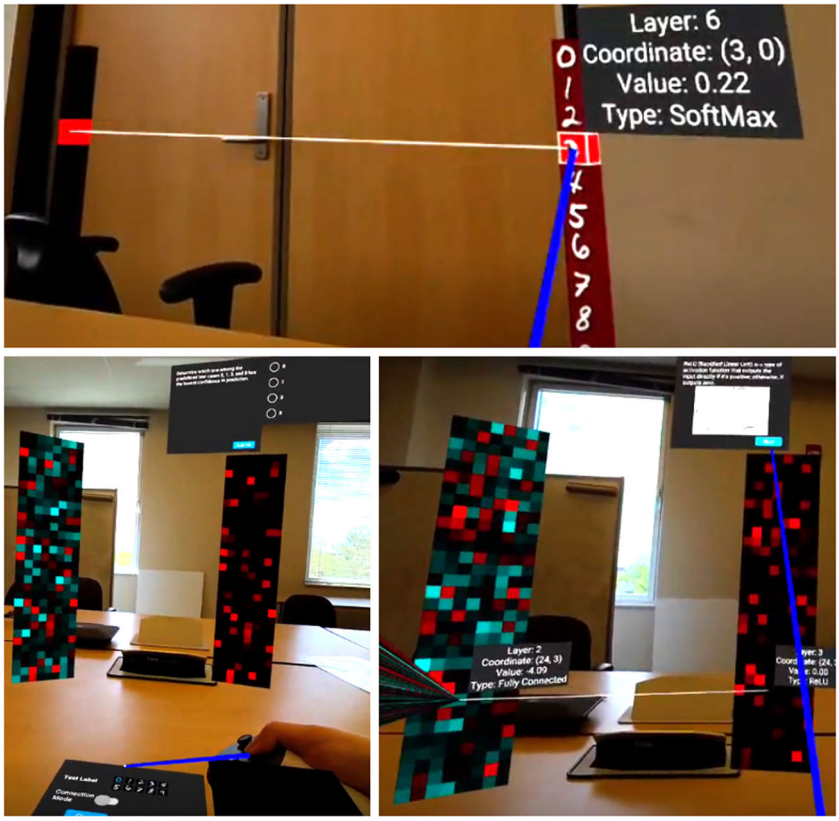


FIGURE 2 | Frames captured by the XR headset, showing the learner's view of the AI concept lesson delivered by the XR system.

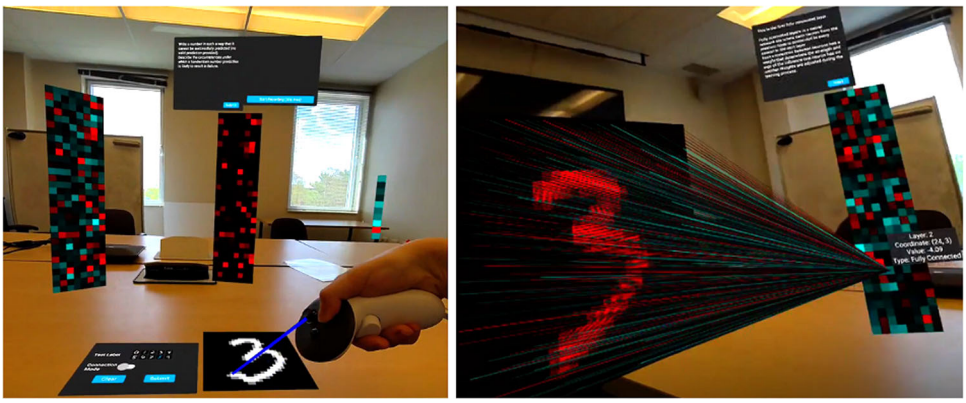


FIGURE 3 | Additional XR headset frames, showing the learner's view of the AI concepts delivered by the XR system.

- (True/False) In supervised learning, the model is trained on labeled data with correct outputs provided during training.
- How are GPU, CPU and TPU important for neural networks or CNNs?
- What is a CPU?
- What is a GPU?
- What is a TPU?
- Which of the following best describes the MNIST database?
- (True/False) The architectural design of a neural network, including the arrangement of layers, nodes, and connections, has no impact on its energy consumption.
- (True/False) A Convolutional Neural Network (CNN) is a specialized type of neural network designed for processing grid-like data, such as images and videos, by using a series of convolutional layers.

Additionally, participants were given slides to explain how to use the VR headset before interacting with the XR environment (*S3*, VR Headset Training, including *S3.1*, Slides-Based Training).

In the learning phase with the XR educational environment (*S4–S6*, Learning Phase), participants engaged with the lesson. Within this phase, the participants interacted with the designed XR immersive environment to learn about NNs and handwritten digit recognition. The headset was set in video passthrough mode,

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

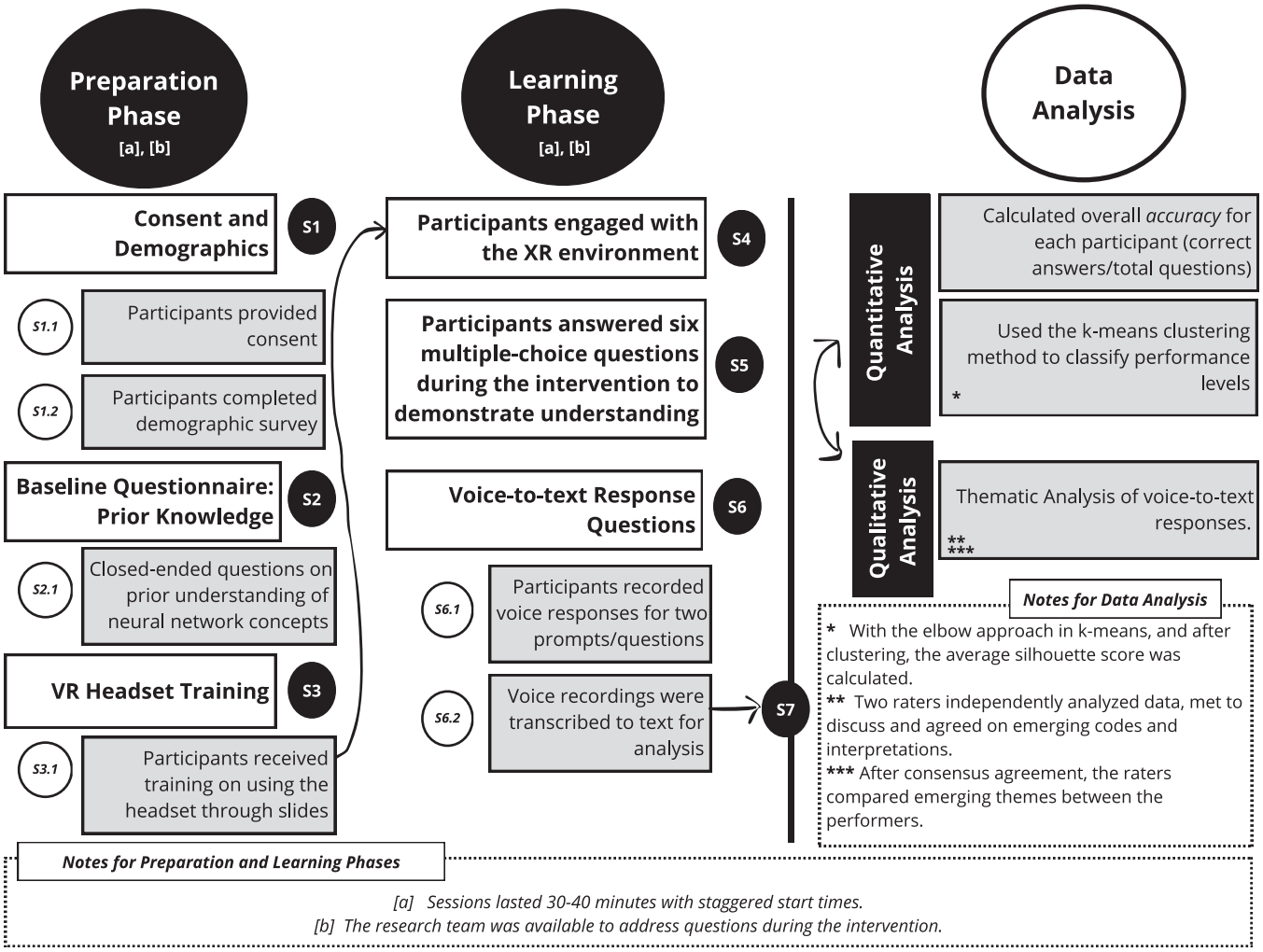


FIGURE 4 | Conceptual design of the intervention introducing AI concepts (NNs and handwritten digit recognition) delivered by the XR system, along with the data analysis procedures.

allowing learners to see virtual 3D elements integrated into their physical surroundings (e.g., the NN panels appear to stand on the real-world table). The participants actively engaged with the visual representation of the NN, exploring its structure, observing how input images (handwritten digits) were processed through layers, and examining the influence of network components. The learning was reinforced through integrated assessments, where participants answered multiple-choice questions about NN concepts and provided voice-recorded explanations in response to prompts requiring them to manipulate inputs and predict NN behavior, thereby demonstrating their conceptual understanding in real-time.

As mentioned, in the learning phase of the lesson, participants answered six multiple-choice questions to demonstrate their understanding (*S5*, Multiple-Choice Questions), which were the following:

- How many input neurons are there in this network for MNIST images?
- What do weights in a neural network represent?
- What does ReLU do if the input is negative?
- What is ReLU used to introduce?
- What is the main purpose of the softmax function in a neural network?
- Determine which one among the predefined test cases 0, 1, 3, and 8 has the lowest confidence in prediction.

In addition, the learning phase included open-ended voice responses, recorded, and transcribed into text for further analysis (*S6.1*, Voice Recording, and *S6.2*, Voice-to-Text Transcription). The prompts for the open-ended questions were:

- *Write a number that looks like a "3" and position it toward the edge of the writing pad so that it might not be easily recognized as a "3." Explain how the placement of the digit near the side of the panel could influence the prediction result.*
- *Write a number in such a way that it cannot be successfully predicted (with no valid prediction provided). Describe the circumstances in which a handwritten number prediction is likely to result in failure.*

The research team was available during this phase to answer questions and provide support (*S4–S6*). This assistance ensured an effective learning experience.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

After the learning phase, participants completed a post-questionnaire to evaluate their likelihood of recommending the system and provide feedback. Each session lasted approximately 30–40 min in total (*S4*, Session Duration). Staggered start times minimized distractions caused by vocal responses, and the research team's presence helped participants stay focused and address any issues that arose.

The final stage, data analysis (*S7*), combined quantitative and qualitative methods to evaluate the responses. The quantitative analysis involved calculating the overall accuracy and classifying the performance levels using the *k*-means clustering method. The elbow method in *k*-means was used to determine the cluster number, and the mean silhouette score displayed the quality after clustering. The qualitative analysis involved two raters independently analyzing the transcribed voice responses. They discussed their findings, agreed on codes and interpretations, and compared themes for different groups of performers to extract meaningful insights. This systematic approach ensured reliable and thorough conclusions about participants' learning.

This study used a multi-methods approach to provide a comprehensive picture of participants' conceptual understanding (*S7*). Quantitative analysis was about measurable metrics, such as response accuracy and clustering results, to identify performance categories. Qualitative analysis explored contextual and subjective elements of participants' voice responses, offering explanations and deeper interpretations. These methods jointly bridged numerical data with thematic understanding, ensuring a balanced evaluation and interpretation of participants' overall learning through the intervention (*S7*). This structured conceptual design created a consistent and supportive learning environment for all participants.

3.4 | Data Analysis Methods

The data collected includes quantitative data, that is, the responses to multiple-choice questions, and qualitative data, that is, the students' verbal explanations of their understanding of the AI-related concepts, as they were interacting with the scene.

On the *quantitative* side, we used descriptive methods and calculated the overall learner performance for each student by dividing the number of questions correctly answered by the total number of questions. The results were accuracies, given as proportions truncated to two decimal places, or described as two-digit percentages. The quantitative data points for this study were from the same source as an earlier publication [31], but here, we focused on students' conceptual understanding in the XR group and used the *k*-means clustering method [62, 63] to classify students into different performance groups based on overall accuracy. The overall accuracy of each participant in the data was used as originally saved (i.e., proportion data), meaning no normalization or sorting by the overall accuracy was done for clustering. The elbow graph was produced in RStudio (Version 2025.05.0+496), adapting available codes [64]. The quantitative data were analyzed using IBM SPSS Statistics (Versions 29 and 30), with the final number of clusters determined from the elbow method [62]. After clustering, the cluster quality was determined by the average silhouette score [65, 66], adapting available codes [67], using the "cluster" package [68] in RStudio (Version 2025.05.0+496).

On the *qualitative* side, we chose thematic analysis techniques to perform a qualitative analysis based on the participants' voice-to-text responses during the intervention. Thematic analysis systematically organizes and interprets qualitative data, revealing emerging categories to help draw meaningful conclusions and insights [69]. Two raters (graduate student researchers) jointly participated in the qualitative thematic analysis to examine the data. After independently analyzing the available data, they met to discuss the emerging codes and their interpretations. Through these meetings, the raters discussed and worked toward reaching an agreement, using a consensus approach to ensure alignment in their interpretations and conclusions. This process allowed for a reliable and consistent understanding of the data.

To increase the trustworthiness of the analysis, the researchers performed inter-rater reliability by examining the consistency and agreement between the hand-coding codification and categorization procedures. Both researchers maintained the same coding protocol and utilized a consensus coding approach, where they independently coded the data and later convened to review the codes and determine the final implementation of the coding scheme across the entire data set. The two researchers worked together by holding weekly meetings to review the codification process and initial findings, intending to ensure consistency and reliability in the coding process. During these discussions, the researchers sought clarity in understanding the coding process and the overall criteria to ensure consistency in this step. Additionally, both researchers could recognize and rectify differences in coding interpretations found in the initial coding method over a limited data set of observations. Moreover, through engaging in these regular conversations, the researchers carefully assessed and reflected on their own biases and how they might have influenced the coding process and their self-awareness in enhancing the overall reliability of the study.

The inter-rater reliability analysis revealed a Cohen's Kappa value $\kappa = 0.84$, indicating a high agreement between the two raters. This result showed consistency during the independent codification process and judgments during the analysis. This high reliability confirmed that the coding protocol followed by both raters was stable, produced consistency in results, and was based on clear and accurate criteria.

Finally, to jointly understand the relationships between the two research questions, we related the accuracy of the students' multiple-choice responses to their voice-to-text responses in our discussion. Recall that accuracy was operationalized into performance outcomes. Learners were classified based on their understanding of the multiple-choice questions. For example, the high-performing group showed more understanding, while the low-performing group did not. In the discussion later, the researchers will show the main points in qualitative themes across the learner groups, focusing on the varying understanding or reasoning. By discussing these differences, we could see what factors contribute to performance, providing insights into how the intervention worked for different learners.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.5 | Ethical Considerations

The Institutional Review Board (IRB) of Purdue University reviewed this study, under the study number IRB-2024-57 [31]. Valuing participant consent, the research team developed a multi-page consent form to provide the participants with insights regarding different aspects of this study, such as the participants' rights and confidentiality. Since this study occurred in person, each potential participant was given a hard copy of the consent form when they arrived. Only individuals who signed and dated the consent form participated in the study. Withdrawals were allowed if participants decided to leave during the study.

4 | Results

This section is organized into two major subsections, following our research questions. Section 4.1 investigates students' achieved knowledge during the XR lesson and Section 4.2 focuses on how students understood the concept of NNs while they interacted with the XR environment.

4.1 | Accuracy as Levels of Performance

The 29 student participants ($N = 29$) who used XR to learn completed five random questions out of a total of 14 to demonstrate their basic understanding of NNs and handwritten digit recognition. Those general background questions (not a pre-test) were asked before the lesson, together with the demographics questionnaire. In general, the learners had some understanding of the concepts, and the percentage of correctly answered questions on basic knowledge ranged from 40% to 100%, with an average of 78.6%.

As mentioned earlier, we used in-lesson multiple-choice questions to test the students' understanding of the material presented. Table 2 shows that most students ($n = 24$, 82.8%) correctly answered at least four multiple-choice questions out of six, suggesting an overall clear understanding of the concepts taught regarding NNs and handwritten digit recognition.

Specifically, 10 students (34.5%) provided four correct answers, 12 students (41.4%) provided five correct answers, and two students (6.9%) answered all questions correctly. Only a few students ($n = 5$, 17.2%) answered fewer than than four questions correctly. The correct answer to Question 4 was chosen by 83% of students ($n = 24$). Questions 2, 3, and 5 were answered correctly by most students ($n = 27$, 93% for each). However, many students struggled with Questions 1 and 6, with fewer than half of the students ($n = 12$, 41%) answering Question 1 correctly, and even fewer students ($n = 5$, 17%) answering Question 6 correctly.

Figure 5 illustrates a box plot of the distribution regarding the overall accuracy of the in-lesson multiple-choice questions among the students. The box plot provides a visual summary of the data, displaying the distribution. Based on the empirical rule, although three cases (below 50% accuracy, at the bottom of the data) were not as close to the remaining data, they were not considered outliers because their Z-scores were within ± 3. Thus, all 29 cases were included in the clustering process. Figure 6 shows that the elbow point was identified as three clusters [62]. The cluster number informed the grouping of learners by level of performance based on the overall accuracy of the students' answers. Each cluster (described next) represented a subset of students with similar performance on the in-lesson multiple-choice questions.

TABLE 2 | Overall performance of students on in-lesson multiple-choice questions.

Category	Number	Percentage
Total students	**29**	
At least 4 correct answers	24	82.8
Correct answers breakdown		
4 correct answers	10	34.5
5 correct answers	12	41.4
6 correct answers	2	6.9
Fewer than 4 correct answers	5	17.2
Correct answer to Question 1	12	41.0
Correct answers to Questions 2, 3, and 5 (each)	27	93.0
Correct answer to Question 4	24	83.0
Correct answer to Question 6	5	17.0

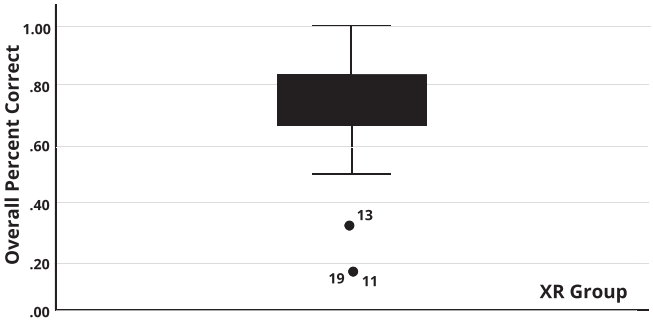


FIGURE 5 | Boxplot for the overall accuracy of the in-lesson multiple-choice questions.

The first cluster contained students who had 83% overall accuracy ($n = 12$) and 100% overall accuracy ($n = 2$), which was considered the group of *high performers* (HP). The second cluster included students who had 17% overall accuracy ($n = 2$) and 33% overall accuracy ($n = 1$), which was considered the group of *low performers* (LP). The third cluster involved students who had 50% overall accuracy ($n = 2$) and 67% overall accuracy ($n = 10$), which was considered the group of *moderate performers* (MP). Among the three clusters, the first HP group was the largest in size, indicating that about half of the students performed well (see Table 3 for the final cluster information).

In Table 3, the cluster centers (rounded to two decimal places) were based on the overall accuracy as proportion data (e.g., 0.33). These categories were identified through the k-means clustering method. The number of cases displays how many

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

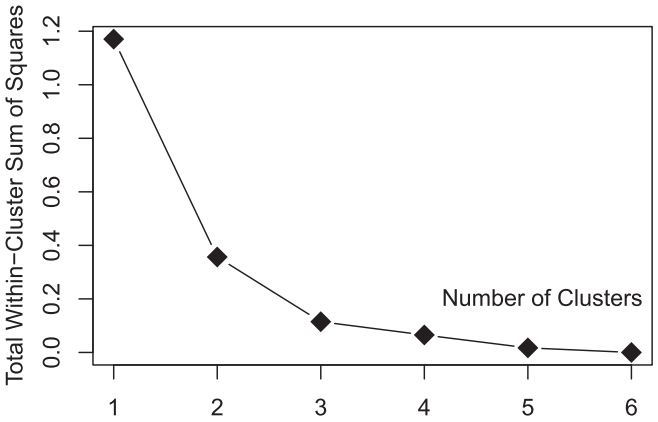


FIGURE 6 | Elbow method for *k*-means clustering to determine the number of learner groups.

TABLE 3 | Summary of final cluster information based on the overall accuracy.

Metric	Cluster 1 Value	Cluster 2 Value	Cluster 3 Value
Cluster center[a]	0.86	0.22	0.64
Number of cases[b]	14	3	12

[a]This is based on the overall accuracy, which is measured as a proportion.
[b]Number of cases represents the count of instances in each cluster.

students fall into each of the three performance categories, reflecting the distribution of all students. After clustering, the metric to determine the cluster quality was the mean silhouette score, which in this case was 0.79 (rounded to two decimal places), showing good quality [66]. This result suggests that the student clusters were well-separated [66], indicating that the three student clusters were different. In other words, the three student clusters (i.e., the clusters of HP, LP, and MP) could be considered to represent the three distinct groups of levels of student performance (i.e., overall accuracy).

4.2 | Conceptual Explanations

The findings from the qualitative analysis were diverse challenges related to handwriting recognition in machine learning systems, focusing mainly on factors involving the placement of the digit, the clarity of handwriting, the style of handwriting, and the probability that one number resembles another. There could also be limitations in the training data or in the algorithm itself that impact the prediction results, as suggested by a few students. These elements could influence how well the system can predict the correct number. The analysis revealed that multiple variables can interfere with accurate number recognition or prediction, pointing to the complex interplay between data inputs and algorithmic interpretation. This complexity highlighted the need for machine learning models to be more adaptable and robust in handling diverse inputs and conditions to improve accuracy and reliability.

Table 4 depicts the resulting codification of the rationales of the respondents for their responses to the first prompt asking them to write a number that looks like a "3" and position it toward the edge of the writing pad so that it might not be easily recognized as a "3," and explain how the placement of the digit near the side of the panel could influence the prediction result. Moreover, it also shows the resulting codification of the rationales of the respondents for their responses to the second prompt, asking them to write a number in such a way that it cannot be successfully predicted (with no valid prediction provided), and then describe the circumstances in which a handwritten number prediction is likely to fail. Table 4 presents for each emerging theme of the six, two columns for the codification of the rationales of each prompt, representing the rationales of the first prompt (R_1) with light gray and the rationales of the second prompt (R_2) with black.

One of the most prominent issues discussed was the placement of the digit near the edge of the panel. Several responses suggested that placing a number at the edge would make it harder for the algorithm to recognize it. For example, a student claimed:

> *[W]hen the number is at the center of the image, the algorithm is able to recognize the pattern and match it with '3'. But when you place it on the side of the panel, it kind of recognizes only the bottom half.*

This means that if a number is drawn too close to the edge, the system may not fully capture all the pixels necessary to understand the shape of the digit. As a result, the number could be misidentified or not recognized at all. In addition, another student mentioned:

> *[I]f the model is not trained to recognize numbers on the side, it may not be able to [...] because the weights are designed to recognize numbers in the center.*

This suggests that the model would be more effective when the numbers were placed in the center because of how it was trained. Digit placement could disrupt the model's ability to use its trained weights effectively, highlighting a limitation in the system's adaptability to variable digit positions. This implied that incorporating more diverse training data with digits in various positions could enhance the robustness of the model. A student elaborated, saying:

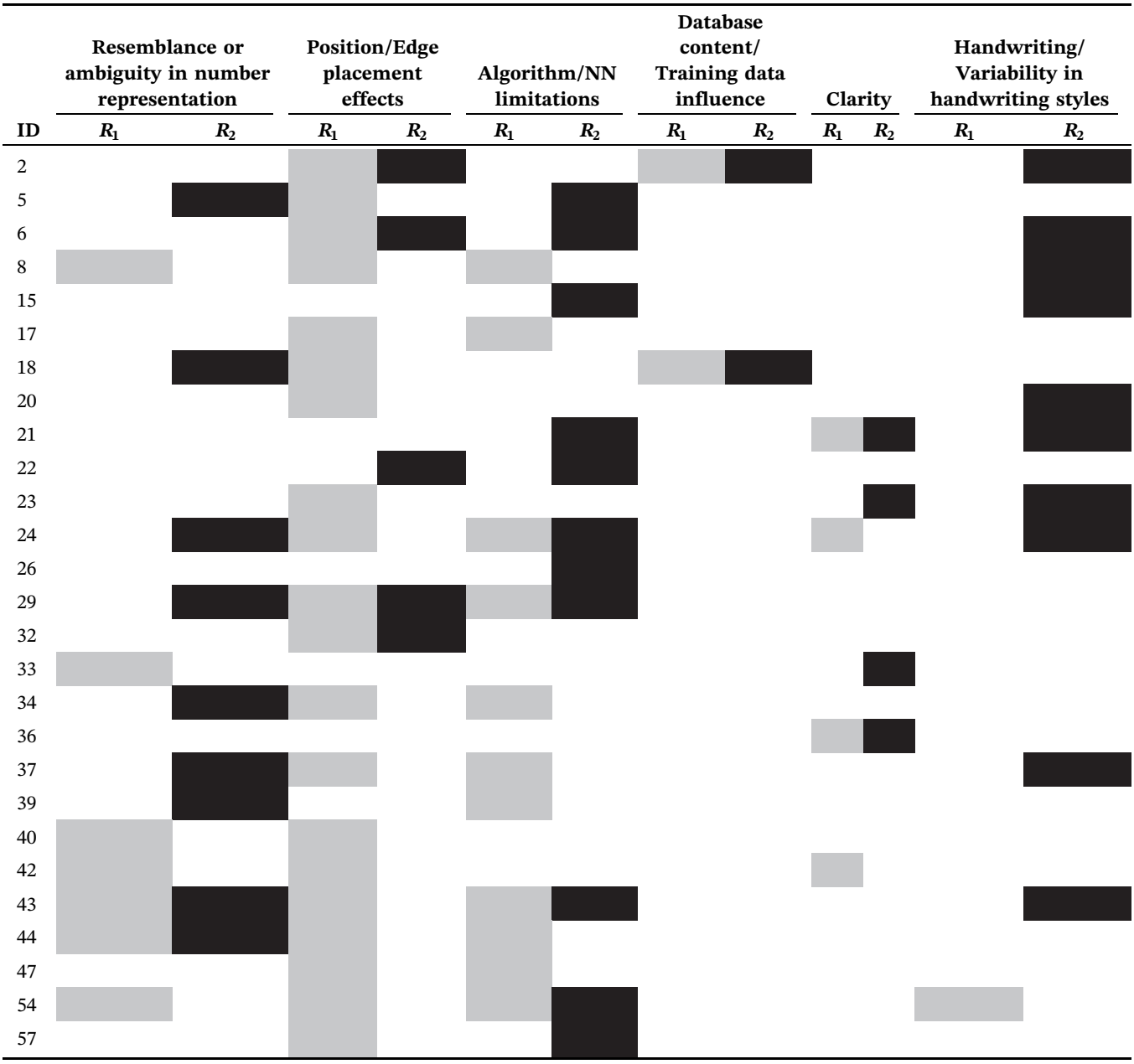
> *[W]hen I try to draw at the edge of the box instead of the middle, it seems that the models could not successfully predict what numbers I'm writing.*

The NN has a higher prediction power when the digit is centered and scaled. Therefore, the NN can make a wrong prediction if a user writes a digit off-center and close to the edge. This can be alleviated by data amplification during the training, to familiarize the NN with a variety of digit placement, or by normalizing the digit images, to center the digit.

The second emergent code referred to the resemblance or ambiguous similarity between different numbers. Many respondents mentioned that numbers can look very similar, making it difficult for the algorithm to distinguish them. For instance, a student said:

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | Codification process of the rationales for prediction accuracy and failure in handwritten digit recognition.



Note: R_1 refers to the codification of the rationales (indicated in light grey shading) for the question: *How does placing a "3" near the edge affect prediction?*
R_2 refers to the codification of the rationales (indicated in black shading) for the question: *When does handwritten number prediction fail?*

> *[I]f we write 3 like this [...] it would be difficult for the model to recognize it because it also seems to resemble 8.*

In this case, the number "3" might look like "8" if drawn in a certain way, causing confusion. This was a problem when the system tried to match the handwritten number with one of the examples in its database, especially if the number was written in a non-standard way that resembled another digit. This problem relates to the expressivity of the training data set. If similar examples were absent, the network could not generalize and detect them correctly. Similar problems occurred when "the number looks like multiple numbers in the test label," as another student pointed out, which could lead to a failure in the prediction. This issue underscored the importance of training data diversity and the need for algorithms capable of discerning resemblance in digit appearance to reduce misclassification. Furthermore, another student noted:

> *I wrote number '3' close to the edge. The lower part of number '3' [...] changed the format of the number [...] and I think this will become more difficult to recognize.*

This quote illustrates how minor changes can lead to significant confusion for the model, reinforcing the idea that even small variations in how numbers are written, whether due to position or the writer's style, can greatly impact algorithm performance. Hence, machine learning systems should incorporate greater flexibility to handle a wider range of written input, including

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

numbers that may resemble each other in writing. The ability to correctly distinguish similar digits is directly related to the size and completeness of the training set. The network cannot generalize the input for which it has not seen similar examples in training.

In addition, students frequently stated the importance of handwriting styles. Different students had different ways of writing numbers, and these variations might affect how well the model recognizes the digits. For example, a student claimed that "there are different ways that each person writes the numbers," and this variation could lead to misclassification if the system were not trained to recognize all the possible ways a number could be written. As one student said, "if someone's handwriting is really bad, it is hard to recognize," highlighting how differences in handwriting can confuse the system. This example emphasized how handwriting style variability poses a challenge for machine learning models, which must be generalized across different writing forms. In short, handwriting quality could directly impact recognition.

Another important factor that emerged from the data was the clarity of the number. Several students suggested that, in addition to handwriting styles, it would be more difficult for the model to predict accurately when the data is unclear. One person said that "if it's too messy or written in an unconventional way, it may result in failure." This indicated that slanted, distorted, or unclear handwriting could confuse the model and cause it to make incorrect predictions. Similarly, when participants wrote numbers near the edge of the panel, it often made the digits unclear or incomplete, leading to mistakes in prediction. In fact, one student described:

> *[T]he placement of the number near the panel side could make it so that the program cannot determine the number correctly and misses some of the pixels.*

These findings highlighted that handwriting quality and placement are crucial factors in successful digit recognition. Hence, to improve accuracy, machine learning models must be able to handle variability in handwriting styles, digit clarity, and positioning. Just as humans cannot understand some handwritten digits, the NN also has its limits. However, the NN reports its confidence in the answer, which is visualized. Future iterations of the lesson can clarify that the NN will not always provide a correct answer.

Moreover, as discussed under the themes "Resemblance or Ambiguity" and "Algorithm/NN Limitations" described in this same subsection, the content and size of the training data set are important. If the training data do not include enough examples of various handwriting styles, edge placements, or ambiguous digits that challenge specific aspects of the NN's feature extraction or classification layers, the network's ability to generalize and classify new, unseen digits correctly will be limited.

Besides the themes, the responses revealed significant connections between the emerging themes, highlighting the key components of the explanations. Table 5 illustrates the connection between these emerging codes with the resemblance or ambiguity in the number representation. Also, Table 6 depicts the connection between some main themes, such as digit placement and variability in handwritten style, with algorithm/NN confusion/limitations.

TABLE 5 | Connection between emerging themes with "Resemblance or Ambiguity in the Number Representation."

Resemblance or ambiguity in number representation	Position/ Edge placement effects	Handwriting/ Variability in handwriting styles
X	X	X
X	X	
X	X	
X	X	
X	X	
X	X	
X	X	
X	X	
X		X
X		X
X		X
X		X
X		
X		
X		
X		
X		

Note: The shaded cells show the themes that have the strongest connection with the theme "*Resemblance or Ambiguity in the Number Representation*." These cells are shaded to highlight the most relevant connections found in the analysis.

The data showed how digit placement, variability in handwritten style, and ambiguity or resemblance between similar numbers were closely linked in the digit recognition or prediction process (see Table 5). Notably, Table 5 shows that out of the 17 instances where "Resemblance or Ambiguity in Number Representation" was identified, students also linked the confusion to where the number was placed in eight cases (47.1%). About one-third of the time (29.4%), they thought handwriting style also played a role, with five cases. These results show that ambiguity in numbers is often influenced by multiple factors. Many students reported that placing a digit near the edge often leads to a misinterpretation by the model. For example, a student noted:

> *[T]he number 3 near the left side could cause it to look like something else, like 8, because the left side of the digit is cut off.*

The ambiguity or resemblance between digits compounded this issue. In other words, when poorly placed or written, the number "3" can resemble an "8," making it harder for the model to differentiate between the two. This overlap in shape, caused by digit placement and digit similarity, could create a challenging scenario for the model, making accurate predictions difficult.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 | Connection between emerging themes with "Algorithm/NN Limitations."

Resemblance or ambiguity in number representation	Position/Edge placement effects	Algorithm NN limitations	Handwriting/Variability in handwriting styles
X		X	
	X	X	X
X	X	X	
		X	X
	X	X	
		X	X
X	X	X	
	X	X	
X		X	X
		X	
	X	X	
X	X	X	
	X	X	
	X	X	
		X	
X	X	X	
X		X	X
X	X	X	
	X	X	
X	X	X	X
		X	X
X		X	X

Note: The shaded cells show the themes that are most strongly connected with the category "*Algorithm/NN Limitations.*" These cells are shaded to highlight the most relevant connections found in the analysis.

Furthermore, the relationship between variability in handwriting styles and the algorithmic limitations of the NN was also evident. Table 6 shows that 10 out of the 22 students (45.5%) who mentioned "Algorithm/NN Limitations" also referred to "Resemblance or Ambiguity in Number Representation." Similarly, 13 students (59.1%) also mentioned "Position/Edge Placement Effects," and 8 students (36.4%) brought up "Handwriting/Variability in Handwriting Styles." These results suggest that students might have attributed the NN's limitations to issues related to digit placement, followed by challenges involving ambiguous number forms and handwriting variation. When students noted a number being too close to the edge, resembling another number, or written unclearly, they likely concluded that the NN was not equipped or strong enough to handle such common real-world variations.

Unclear or messy handwriting was often likely to cause incorrect predictions, as one student claimed that "if it's too messy, the system can't determine what the number is." This denoted the limitations of the model in handling handwriting variability. The algorithm may work best with clear and standard handwriting. Still, real-world conditions could often involve a broader range of writing styles and clarity, resulting in discrepancies between reality and prediction.

As brought up earlier, in real-world scenarios, when digits were placed on the edge of the writing panel (i.e., position/edge placement effects), the system struggled to recognize them. As one student noted, "the left side of the digit is cut off, making it look like something else." This incomplete data confused the algorithm, preventing it from analyzing the full shape of the digit (i.e., hindering complete feature extraction). Variability in handwriting style further increased confusion for the NN. For instance, another student reflected on how the model failed to identify numbers when handwriting deviated from its training data, emphasizing the need for more variety in training cases to achieve better adaptability. Specifically, this student noted that:

> *[D]ifferent people write numbers in different ways, and if the model is not trained for those variations, it cannot make an accurate prediction.*

Moreover, as indicated before, the reliance of the model on centered digits limits its ability to handle placement effects, resulting in misidentifications. Hence, improving the ability of the model to process unclear handwriting at different positions would reduce these limitations and increase prediction accuracy in more diverse settings. Larger and more diverse training datasets and NN architectures with a larger number of trainable parameters may be needed to improve accuracy and robustness.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

However, larger NNs can memorize the training data set, which leads to overfitting.

5 | Discussion and Implications

Recall that based on the clustering of the results from the in-lesson multiple-choice questions, the group of HP included the learners who had accuracy from 83% to 100%, the group of MP had the learners who showed 50% to 67% accuracy, and the group of LP had 17% to 33% accuracy.

The LP group demonstrated the lowest average level of basic understanding in the baseline questions. This suggests that students with less prior knowledge struggled more with complex topics like digit recognition, which require understanding algorithms and data representation [70, 71]. In other words, students who lacked a strong foundation in the NN architecture or its characteristics might have found it more challenging to grasp how algorithms work in the particular task of handwritten digit recognition.

The MP group generally showed a wider range of prior knowledge compared to the high-performing group, although some MP had a similar level of prior understanding to some HP. In the lesson, both groups recognized similar challenges in handwritten digit recognition or prediction, such as clarity issues and algorithmic limitations. This highlights the importance of clear and consistent support for all students. Since the XR system provided the same guidance to everyone, students with similar prior knowledge appeared to benefit equally overall, although the support was not tailored to individual needs.

The HP group demonstrated varying levels of prior understanding, despite a good average level of basic understanding: some understood only a few concepts, such as GPU and CPU, while others understood all of them. This observation suggests that learning strategies that help grasp key ideas are important and can assist learners in achieving a certain level of success, even with varying knowledge of the topic. Achieving accurate results may have depended on effective learning strategies and engagement with the lesson based on their prior understanding [71, 72]. Research suggests that good instruction can help learners succeed, even with limited background knowledge [72, 73], emphasizing the importance of well-designed teaching strategies.

The HP usually provided more detailed responses to the in-lesson open-ended questions. They discussed multiple factors, such as the placement effect and ambiguity in digit representation, which affected handwritten digit recognition. Hence, they had a deeper cognitive engagement with the material. Overall, these students demonstrated good conceptual understanding in recognizing various challenges in digit recognition systems, suggesting they could analyze complex issues more effectively. Their explanations suggest a stronger ability to think about the "why" and "how" behind the system's behavior. For instance, one high-performing student attempted to explain the *internal mechanism* within the NN regarding misclassification:

> *If a number is written in such a way that it's not something that's been taught in the database then it could possibly find a different pattern that either results in a different number or no possible match for what was drawn.*

In addition, another high-performing student even designed a test case and predicted the outcome based on specific details, showing a deep understanding of how a specific visual change could directly lead to a specific error, indicating an ability to analyze feature-level details and predict system vulnerabilities. This student noted:

> *I have written a variant of 2 which might resemble a little bit like 3 as the bottom line is like more inclined. So in that circumstance, the recognition of 2 would be difficult as it might confuse with 3 and it might actually predict the other class.*

In contrast to the HP, the MP and LP still grasped some key aspects that impacted handwritten digit recognition, such as digit placement and algorithmic limitations, but their answers were usually less detailed. For instance, non-high-performing students were able to identify algorithmic limitations, "I assume test label 6 is harder to recognize because it's not standard; I guess," or "Since the number does not take the whole writing pad it might cause to the wrong results." However, those students generally did not elaborate on how a particular challenge might interact with other challenges. Cases like these suggest that those students simply described what they saw or could identify a relevant concept but did not apply it thoroughly to the situation, with some MP briefly claiming: "It looks cut off," "the placement of 3 near the panel side makes it look like M instead of 3," or "it's unable to read it." This suggests that their ability to explain the material was limited by their incomplete understanding of how NNs can be used for handwritten digit recognition and what might influence the outcome [74]. That is, some MP and LP struggled to explain more complex foundational details due to gaps in their knowledge, although they may have understood the concepts based on the outcomes observed during the interaction.

Overall, XR tools can be effective for supporting the learning of NNs and handwritten digit recognition, even with learners at different levels of understanding. Research emphasizes that XR technologies can improve learning by engaging students with different levels of knowledge, helping them interact with complex concepts like NNs [53, 75, 76]. This suggests that XR technology can make learning more engaging and interactive, even when the topics are difficult. Nevertheless, it is important to interpret the "effectiveness" with caution, as the outcomes from the learning experience are a result not only of the XR technology itself but also of how the intervention was designed, how the content was sequenced, and how embedded supports helped students engage with the material. The qualitative findings also match this, as learners were able to engage with the XR tool despite challenges like handwriting style and digit placement. However, not all learners are familiar with XR tools, which can affect their learning outcomes. Their success with XR depends on how easily they adapt to the tool and overcome

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

difficulties navigating it [76]. There is also a need for user-friendly designs and support systems to maximize the educational benefits of immersive technologies.

XR shows a promise in STEM education, especially for abstract topics. The decreasing cost of an XR headset may facilitate more educational implementations. For example, in 2023, Meta's Quest 3 headset was released at around $500 with premium passthrough performance [77]. Later in 2024, Meta released the Quest 3S headset, at a cost of only around $300–400, which is clearly on the lower end of price but still is a quite capable version of the Quest 3 [77]. While Meta admits that it is currently selling the headsets at a loss, it is clear that, together with Apple, these are two trillion-dollar companies firmly invested in XR technology, betting on its widespread adoption across different fields.

6 | Conclusions, Limitations, and Future Work

This study used a multi-methods approach to examine how students learned NNs and handwritten digit recognition (i.e., AI concepts) using XR. The results were based on quantitative and qualitative data collected during an XR lesson. Quantitatively, students achieved 17%–100% accuracy on all multiple-choice questions about AI concepts. Three different performance clusters (i.e., high, moderate, and low) were formed based on the accuracy using the k-means clustering method with the elbow method; the cluster quality was good, given an average silhouette score of 0.79. Many students ($n = 26$) demonstrated moderate and high performance regarding answering questions about AI concepts while engaging in the lesson, suggesting that XR is relatively effective in learning AI concepts. Qualitatively, students identified challenges in handwritten digit recognition, such as unclear handwriting and the placement of digits in recognition tasks. The HP generally provided detailed explanations, reflecting a deeper understanding of the material. In contrast, the MP and LP usually offered less detailed responses.

The researchers of this study acknowledge several limitations. First, the sample size may not be large enough to generalize the findings to domains beyond AI education. The sample diversity was limited by recruiting from only a few STEM departments in higher education, which may also impact the generalizability of the findings. Further, since this intervention was short without repeated measures, learning gains over time were not reflected. Last but not least, this study did not employ a pretest–posttest design or a true experimental design. The prior knowledge questionnaire, which was not a pre-test, was used to elicit general background information and was not designed for quantitative analysis (e.g., categorizing learner groups for the intervention). In addition, the within-lesson assessment was not considered a posttest.

There are several avenues for future work. One would be studying the long-term effect of attained knowledge. Researchers can speculate that the impact of the visualization on the students can have a long-standing effect, as opposed to learning from textual descriptions or traditional presentations. A longitudinal design and analysis could also be deployed. For example, one could investigate whether XR improves the retention of NN concepts over 6 months. Similarly, more measures could be taken beyond conceptual learning to also include measures of engagement. Future studies may also involve developing longer lessons about AI, incorporating other AI concepts beyond NNs or handwritten digit recognition.

Also, this study may be replicated in other educational contexts and with different populations. Other future studies may carry out the intervention at different education levels, such as high school, to further examine the overall effectiveness of XR in learning AI concepts across different educational levels. In general, XR could also be blended with traditional educational methods to comprehensively support learning needs.

Another avenue for future work could involve practical implications related to added functionality to the system. First, researchers could allow low-level control over the different parameters and settings of the experiments. The users could, for example, modify the resolution of the individual layers, the connectivity, the activation functions, and so forth. Second, the users could work with the training data set and see how its content affects the training and the network's accuracy. However, such experiments should be carefully designed as the number of options could be overwhelming.

Despite its limitations, findings from this study suggest that XR is a useful tool for learning AI concepts (i.e., NNs and handwritten digit recognition). In particular, XR experiences can be more effective when technology is thoughtfully integrated with instructional strategies such as clear sequencing of content and appropriate scaffolding to guide the learning process. While the results of this study suggest that XR is a versatile tool for learning AI and may be suitable for learners with varying levels of basic knowledge, caution should be taken when developing and implementing nontraditional learning methods, like XR, to better engage students in learning: for example, designing user-friendly interfaces and providing training materials are essential. In this context, the effectiveness of XR should be seen and interpreted considering the overall learning design, where the positive outcomes likely result from the combination of XR technology and the structured experience (i.e., the way the content is organized and supported, including the pedagogical sequencing and scaffolding provided).

Author Contributions

The individual contributions are as follows: Conceptualization: M.A.F.-G., Y.Z., Y.G., B.B., A.J.M., and V.P. Methodology: M.A.F.-G., Y.Z., Y.G., B.B., A.J.M., and V.P. Validation: Y.Z. and M.A.F.-G. Formal analysis: Y.Z. and M.A.F.-G. Investigation: A.J.M., B.B., and V.P. Resources: A.J.M., B.B., and V.P. Data curation: M.A.F.-G., Y.Z., and Y.G. Writing – original draft preparation: M.A.F.-G., Y.Z., B.B., A.J.M., and V.P. Writing – review and editing: M.A.F.-G., Y.Z., Y.G., B.B., A.J.M., and V.P. Supervision: A.J.M. Project administration: V.P. Funding acquisition: A.J.M., B.B., and V.P. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors thank all participants for their voluntary involvement and engagement in the study. Any opinions, findings, and conclusions or

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This study was supported in part by the National Science Foundation under award numbers 2412928, 2417510, 2212200, 2219842, 2309564, and 2318657.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson, "Immersive Virtual Reality as a Pedagogical Tool in Education: A Systematic Literature Review of Quantitative Learning Outcomes and Experimental Design," *Journal of Computers in Education* 8, no. 1 (2021): 1–32.

2. T. Mühling, I. Späth, J. Backhaus, et al., "Virtual Reality in Medical Emergencies Training: Benefits, Perceived Stress, and Learning Success," *Multimedia Systems* 29, no. 4 (2023): 2239–2252.

3. A. Dengel, "What Is Immersive Learning?," In *2022 Eighth International Conference of the Immersive Learning Research Network (iLRN)*, 2022, 1–5.

4. G. Makransky and G. B. Petersen, "The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in immersive Virtual Reality," *Educational Psychology Review* 33 (2021): 937–958.

5. S. Mystakidis and V. Lympouridis, "Immersive Learning," *Encyclopedia* 3, no. 2 (2023): 396–405, https://doi.org/10.3390/encyclopedia3020026.

6. G. Baxter and T. Hainey, "Using Immersive Technologies to Enhance the Student Learning Experience," *Interactive Technology and Smart Education* 21, no. 3 (2024): 403–425.

7. A. Çöltekin, I. Lochhead, M. Madden, et al., "Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions," *ISPRS International Journal of Geo-Information* 9, no. 7 (2020): 439, https://doi.org/10.3390/ijgi9070439.

8. K. W. Kosko, R. E. Ferdig, and L. Roche, "Conceptualizing a Shared Definition and Future Directions for Extended Reality (XR) in Teacher Education," *Journal of Technology and Teacher Education* 29, no. 3 (2021): 257–277.

9. Y. Gu, M. A. Feijoo-Garcia, Y. Zhang, A. J. Magana, B. Benes, and V. Popescu, "An xr Environment for ai Education: Design and First Implementation," in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (IEEE, 2024), 157–162.

10. M. Vasarainen, S. Paavola, and L. Vetoshkina, "A Systematic Literature Review on Extended Reality: Virtual, Augmented and Mixed Reality in Working Life," *International Journal of Virtual Reality* 21, no. 2 (2021): 1–28.

11. X. Guo, Y. Guo, and Y. Liu, "The Development of Extended Reality in Education: Inspiration From the Research Literature," *Sustainability* 13, no. 24 (2021): 13776, https://doi.org/10.3390/su132413776.

12. Y. Zhang, M. A. Feijoo-Garcia, Y. Gu, V. Popescu, B. Benes, and A. J. Magana, "Virtual and Augmented Reality in Science, Technology, Engineering, and Mathematics (STEM) Education: An Umbrella Review," *Information* 15, no. 9 (2024): 515, https://doi.org/10.3390/info15090515.

13. G. Zwoliński, D. Kamińska, A. Laska-Leśniewicz, et al., "Extended Reality in Education and Training: Case Studies in Management Education," *Electronics* 11, no. 3 (2022): 336, https://doi.org/10.3390/electronics11030336.

14. M. A. Kuhail, A. ElSayary, S. Farooq, and A. Alghamdi, "Exploring Immersive Learning Experiences: A Survey," *Informatics* 9, no. 4 (2022): 75, https://doi.org/10.3390/informatics9040075.

15. S. C. Chang and G. J. Hwang, "Impacts of an Augmented Reality-Based Flipped Learning Guiding Approach on Students' Scientific Project Performance and Perceptions," *Computers and Education* 125 (2018): 226–239, https://doi.org/10.1016/j.compedu.2018.06.007.

16. R. Liu, L. Wang, J. Lei, Q. Wang, and Y. Ren, "Effects of an Immersive Virtual Reality-Based Classroom on Students' Learning Performance in Science Lessons," *British Journal of Educational Technology* 51, no. 6 (2020): 2034–2049.

17. A. Logeswaran, C. Munsch, Y. J. Chong, N. Ralph, and J. McCrossnan, "The Role of Extended Reality Technology in Healthcare Education: Towards a Learner-Centred Approach," *Future Healthcare Journal* 8, no. 1 (2021): e79–e84.

18. M. Akçayır and G. Akçayır, "Advantages and Challenges Associated With Augmented Reality for Education: A Systematic Review of the Literature," *Educational Research Review* 20 (2017): 1–11, https://doi.org/10.1016/j.edurev.2016.11.002.

19. R.-d.G. Lázaro and J. M. Duart, "You Can Handle, You Can Teach It: Systematic Review on the Use of Extended Reality and Artificial Intelligence Technologies for Online Higher Education," *Sustainability* 15, no. 4 (2023): 3507, https://doi.org/10.3390/su15043507.

20. P. Acevedo, A. J. Magana, B. Benes, and C. Mousas, "A Systematic Review of Immersive Virtual Reality in Stem Education: Advantages and Disadvantages on Learning and User Experience," *IEEE Access* 12 (2024): 189359–189386.

21. F. A. Fernandes, C. S. C. Rodrigues, E. N. Teixeira, and C. M. L. Werner, "Immersive Learning Frameworks: A Systematic Literature Review," *IEEE Transactions on Learning Technologies* 16, no. 5 (2023): 736–747, https://doi.org/10.1109/TLT.2023.3242553.

22. A. Dengel and J. Mägdefrau, "Immersive Learning Explored: Subjective and Objective Factors Influencing Learning Outcomes in Immersive Educational Virtual Environments." *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 2018, 608–615.

23. M. A. Rojas-Sánchez, P. R. Palos-Sánchez, and J. A. Folgado-Fernández, "Systematic Literature Review and Bibliometric Analysis on Virtual Reality and Education," *Education and Information Technologies* 28 (2023): 155–192.

24. C. Y. Chang, H. C. Kuo, and Z. Du, "The Role of Digital Literacy in Augmented, Virtual, and Mixed Reality in Popular Science Education: A Review Study and an Educational Framework Development," *Virtual Reality* 27, no. 3 (2023): 2461–2479.

25. C. L. Huang, Y. F. Luo, S. C. Yang, C. M. Lu, and A. S. Chen, "Influence of Students' Learning Style, Sense of Presence, and Cognitive Load on Learning Outcomes in an Immersive Virtual Reality Learning Environment," *Journal of Educational Computing Research* 58, no. 3 (2020): 596–615.

26. N. Wenk, J. Penalver-Andres, K. A. Buetler, T. Nef, R. M. Müri, and L. Marchal-Crespo, "Effect of Immersive Visualization Technologies on Cognitive Load, Motivation, Usability, and Embodiment," *Virtual Reality* 27, no. 1 (2023): 307–331.

27. S. de Freitas and M. Oliver, "How Can Exploratory Learning With Games and Simulations Within the Curriculum Be Most Effectively Evaluated?," *Computers and Education* 46, no. 3 (2006): 249–264, https://doi.org/10.1016/j.compedu.2005.11.007.

28. S. De Freitas, G. Rebolledo-Mendez, F. Liarokapis, G. Magoulas, and A. Poulovassilis, "Learning as Immersive Experiences: Using the Four-Dimensional Framework for Designing and Evaluating Immersive

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Learning Experiences in a Virtual World," *British Journal of Educational Technology* 41 (2010): 69–85.

29. C. Udeozor, J. L. Dominguez Alfaro, and J. Glassey, "Assessment Framework for Immersive Learning: Application and Evaluation," in *Immersive Learning Research Network* (Springer, 2024), 195–208.

30. V. Kuleto, P. Mi, M. Stanescu, et al., "Extended Reality in Higher Education, a Responsible Innovation Approach for Generation Y and Generation Z," *Sustainability* 13, no. 21 (2021): 11814, https://doi.org/10.3390/su132111814.

31. M. A. Feijoo-Garcia, Y. Zhang, Y. Gu, A. J. Magana, B. Benes, and V. Popescu, "Exploring Extended Reality (XR) in Teaching AI: A Comparative Study of XR and Desktop Environments," in *The 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 1: GRAPP, HUCAPP and IVAPP*, (2025), 472–482.

32. E. Baumgartner, R. E. Ferdig, and E. Gandolfi, "Exploring the Impact of Extended Reality (XR) on Spatial Reasoning of Elementary Students," *TechTrends* 66 (2022): 825–836.

33. F. Škola, A. Karanasiou, M. Triantafillou, H. Zacharatos, and F. Liarokapis, ""Perceptions and Challenges of Implementing XR Technologies in Education: A Survey-Based Study," in *Smart Mobile Communication & Artificial Intelligence*, eds. M. E. Auer and T. Tsiatsos (Springer Nature Switzerland, 2024), 297–306.

34. J. Garzón, "An Overview of Twenty-Five Years of Augmented Reality in Education," *Multimodal Technologies and Interaction* 5, no. 7 (2021): 37, https://doi.org/10.3390/mti5070037.

35. H. Altinpulluk, "Determining the Trends of Using Augmented Reality in Education Between 2006-2016," *Education and Information Technologies* 24 (2019): 1089–1114.

36. D. Kamińska, G. Zwoliński, A. Laska-Leśniewicz, et al., "Augmented Reality: Current and New Trends in Education," *Electronics* 12, no. 16 (2023): 3531.

37. M. Li, Y. T. Chen, C. Q. Huang, G. J. Hwang, and M. Cukurova, "From Motivational Experience to Creative Writing: A Motivational AR-Based Learning Approach to Promoting Chinese Writing Performance and Positive Writing Behaviours," *Computers & Education* 202 (2023): 104844.

38. C. Avila-Garzon, J. Bacca-Acosta, J. Duarte, J. Betancourt, others, "Augmented Reality in Education: An Overview of Twenty-Five Years of Research," *Contemporary Educational Technology* 13, no. 3 (2021): ep302.

39. S. Kavanagh, A. Luxton-Reilly, B. Wuensche, and B. Plimmer, "A Systematic Review of Virtual Reality in Education," *Themes in Science and Technology Education* 10, no. 2 (2017): 85–119.

40. J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda," *Computers and Education* 147 (2020): 103778, https://doi.org/10.1016/j.compedu.2019.103778.

41. A. Marougkas, C. Troussas, A. Krouska, and C. Sgouropoulou, "Virtual Reality in Education: A Review of Learning Theories, Approaches and Methodologies for the Last Decade," *Electronics* 12, no. 13 (2023): 2832.

42. Y. T. Chen, M. Li, M. Cukurova, and M. S. Y. Jong, "Incorporation of Peer-Feedback Into the Pedagogical Use of Spherical Video-Based Virtual Reality in Writing Education," *British Journal of Educational Technology* 55, no. 2 (2024): 519–540.

43. Y. M. Tang, K. M. Au, H. C. Lau, G. T. Ho, and C. H. Wu, "Evaluating the Effectiveness of Learning Design With Mixed Reality (MR) in Higher Education," *Virtual Reality* 24 (2020): 797–807.

44. M. J. Maas and J. M. Hughes, "Virtual, Augmented and Mixed Reality in K–12 Education: A Review of the Literature," *Technology, Pedagogy and Education* 29, no. 2 (2020): 231–249, https://doi.org/10.1080/1475939X.2020.1737210.

45. A. Banjar, X. Xu, M. Z. Iqbal, and A. Campbell, "A Systematic Review of the Experimental Studies on the Effectiveness of Mixed Reality in Higher Education Between 2017 and 2021," *Computers & Education: X Reality* 3 (2023): 100034.

46. R. Kiraly, S. Kiraly, and M. Palotai, "Investigating the Usability of a New Framework for Creating, Working and Teaching Artificial Neural Networks Using Augmented Reality (AR) and Virtual Reality (VR) Tools," *Education and Information Technologies* 29 (2024): 13085–13104.

47. C. M. Bishop, "Neural Networks and Their Applications," *Review of Scientific Instruments* 65, no. 6 (1994): 1803–1832.

48. B. Müller, J. Reinhardt, and M. T. Strickland, *Neural Networks: an Introduction* (Springer Science and Business Media, 2012).

49. A. Baldominos, Y. Saez, and P. Isasi, "A Survey of Handwritten Character Recognition With Mnist and EMNIST," *Applied Sciences* 9, no. 15 (2019): 3169, https://doi.org/10.3390/app9153169.

50. L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," *IEEE Signal Processing Magazine* 29, no. 6 (2012): 141–142, https://doi.org/10.1109/MSP.2012.2211477.

51. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE* 86, no. 11 (1998): 2278–2324, https://doi.org/10.1109/5.726791.

52. Z. Wang, S. Wu, C. Liu, S. Wu, and K. Xiao, "The Regression of MNIST Dataset Based on Convolutional Neural Network," in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)*, eds. A. E. Hassanien, A. T. Azar, T. Gaber, R. Bhatnagar, and M. F. Tolba (Springer International Publishing, 2020), 59–68.

53. D. Reiners, M. R. Davahli, W. Karwowski, and C. Cruz-Neira, "The Combination of Artificial Intelligence and Extended Reality: A Systematic Review," *Frontiers in Virtual Reality* 2 (2021): 721933.

54. Z. Turan and S. C. Karabey, "The Use of Immersive Technologies in Distance Education: A Systematic Review," *Education and Information Technologies* 28, no. 12 (2023): 16041–16064.

55. A. Suh and J. Prophet, "The State of Immersive Technology Research: A Literature Analysis," *Computers in Human Behavior* 86 (2018): 77–90.

56. T. Tene, J. A. Marcatoma Tixi, M. d. L. Palacios Robalino, M. J. Mendoza Salazar, C. Vacacela Gomez, and S. Bellucci, " Integrating Immersive Technologies With STEM Education: A Systematic Review," in Frontiers in Education. 9. (Frontiers Media SA, 2024), 1410163.

57. M. Won, D. A. K. Ungu, H. Matovu, et al., "Diverse Approaches to Learning With Immersive Virtual Reality Identified From a Systematic Review," *Computers & Education* 195 (2023): 104701.

58. D. Buragohain, S. Chaudhary, G. Punpeng, A. Sharma, N. Am-in, and L. Wuttisittikulkij, "Analyzing the Impact and Prospects of Metaverse in Learning Environments Through Systematic and Case Study Research," *IEEE Access* 11 (2023): 141261–141276.

59. R. Lindgren, M. Tscholl, S. Wang, and E. Johnson, "Enhancing Learning and Engagement Through Embodied Interaction Within a Mixed Reality Simulation," *Computers & Education* 95 (2016): 174–187.

60. D. Rivera, "Visualizing Machine Learning in 3D," in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, 2022, 1–2.

61. M. Aros, C. L. Tyger, and B. S. Chaparro, "Unraveling the Meta Quest 3: An Out-of-Box Experience of the Future of Mixed Reality Headsets," in *International Conference on Human-Computer Interaction* (Springer, 2024), 3–8.

62. T. M. Kodinariya and P. R. Makwana, "Review on Determining Number of Cluster in k-Means Clustering," *International Journal of*

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Advance Research in Computer Science and Management Studies 1, no. 6 (2013): 90–95.

63. Z. Zhang, Q. Feng, J. Huang, Y. Guo, J. Xu, and J. Wang, "A Local Search Algorithm for k-Means With Outliers," *Neurocomputing* 450 (2021): 230–241, https://doi.org/10.1016/j.neucom.2021.04.028.

64. S. Anand, "Finding Optimal Number of Clusters," 2017, https://www.r-bloggers.com//02/finding-optimal-number-of-clusters/.

65. P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics* 20 (1987): 53–65.

66. K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," *2020 IEEE Seventh International Conference on Data Science and Advanced Analytics (DSAA)* (IEEE, 2020), 747–748.

67. RDocumentation, "silhouette: Compute or Extract Silhouette Information From Clustering," accessed May 14, 2025, https://www.rdocumentation.org/packages/cluster/versions/2.1.8.1/topics/silhouette.

68. M. Maechler, P. Rousseeuw, A. Struyf, et al., "Cluster: 'Finding Groups in Data': Cluster Analysis Extended Rousseeuw et al.," accessed May 14, 2025, https://doi.org/10.32614/CRAN.package.cluster.

69. V. Braun and V. Clarke, "Reflecting on Reflexive Thematic Analysis," *Qualitative Research in Sport, Exercise and Health* 11, no. 4 (2019): 589–597, https://doi.org/10.1080/2159676x.2019.1628806.

70. D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, Big, Simple Neural Nets for Handwritten Digit Recognition," *Neural Computation* 22, no. 12 (2010): 3207–3220.

71. I. Tuba, U. Tuba, and M. Veinović, "Classification Methods for Handwritten Digit Recognition: A Survey," *Vojnotehnicki Glasnik* 71 (2023): 113–135, https://doi.org/10.5937/vojtehg71-36914.

72. V. Mane, R. Sapate, S. Raut, R. Sonji, and A. Khairnar, "Handwritten Digit Recognition," *International Journal for Research in Applied Science and Engineering Technology* 12 (2024): 4878–4882, https://doi.org/10.22214/ijraset.2024.62557.

73. M. A. Nur, M. Abebe, and R. Sharma, "Handwritten Geez Digit Recognition Using Deep Learning," *Applied Computational Intelligence and Soft Computing* 2022 (2022): 1–12, https://doi.org/10.1155/2022/8515810.

74. D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional Neural Network Committees for Handwritten Character Classification," in *2011 International Conference on Document Analysis and Recognition* (IEEE, 2011), 1135–1139.

75. A. N. Ghanbaripour, N. Talebian, D. Miller, et al., "A Systematic Review of the Impact of Emerging Technologies on Student Learning, Engagement, and Employability in Built Environment Education," *Buildings* 14, no. 9 (2024): 2769.

76. Z. N. Khlaif, A. Mousa, and M. Sanmugam, "Immersive Extended Reality (XR) Technology in Engineering Education: Opportunities and Challenges," *Technology, Knowledge and Learning* 29, no. 2 (2024): 803–826.

77. Meta, "Latest Meta Quest News," accessed May 8, 2025, https://about.fb.com/news/category/technologies/oculus/.

10990542, 2025, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cae.70084, Wiley Online Library on [25/10/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License