



# Through the lens of artificial intelligence: A novel study of spherical video-based virtual reality usage in autism and neurotypical participants

Matthew Schmidt<sup>a,\*</sup>, Noah Glaser<sup>b,1</sup>, Heath Palmer<sup>c</sup>, Carla Schmidt<sup>d</sup>, Wanli Xing<sup>e</sup>

<sup>a</sup> Department of Workforce Education and Instructional Technology, Mary Frances Early College of Education, Department of Clinical and Administrative Pharmacy, College of Pharmacy, University of Georgia, River's Crossing, 226, Athens, GA, 30602, Georgia

<sup>b</sup> University of Missouri, United States

<sup>c</sup> University of Cincinnati, United States

<sup>d</sup> Bayada Home Health Care, United States

<sup>e</sup> University of Florida, United States

## ARTICLE INFO

### Keywords:

Computer vision  
Spherical video-based virtual reality  
Artificial Intelligence  
Data mining

## ABSTRACT

The current study explores the use of computer vision and artificial intelligence (AI) methods for analyzing 360-degree spherical video-based virtual reality (SVVR) data. The study aimed to explore the potential of AI, computer vision, and machine learning methods (including entropy analysis, Markov chain analysis, and sequential pattern mining), in extracting salient information from SVVR video data. The research questions focused on differences and distinguishing characteristics of autistic and neurotypical usage characteristics in terms of behavior sequences, object associations, and common patterns, and the extent to which the predictability and variability of findings might distinguish the two participant groups and provide provisional insights into the dynamics of their usage behaviors. Findings from entropy analysis suggest the neurotypical group showed greater homogeneity and predictability, and the autistic group displayed significant heterogeneity and variability in behavior. Results from the Markov Chains analysis revealed distinct engagement patterns, with autistic participants exhibiting a wide range of transition probabilities, suggesting varied SVVR engagement strategies, and with the neurotypical group demonstrating more predictable behaviors. Sequential pattern mining results indicated that the autistic group engaged with a broader spectrum of classes within the SVVR environment, hinting at their attraction to a diverse set of stimuli. This research provides a preliminary foundation for future studies in this area, as well as practical implications for designing effective SVVR learning interventions for autistic individuals.

## 1. Background

The current article presents an exploratory study that investigates the use of computer vision and AI methods for analyzing 360-degree spherical video data, also referred to as spherical, video-based virtual reality, or SVVR (Chien et al., 2020). This study provides valuable insights for researchers who are interested in using AI, computer vision, and machine learning for analyzing immersive learning intervention data collected within autistic and neurotypical groups. The purpose of the study is to demystify the methods and processes used and provide a clear understanding of how these technologies can be applied in the field of immersive learning interventions. This article will be of value to

researchers who are looking to explore the use of AI and computer vision for analyzing 360-degree SVVR data, providing a starting point for further research in this field and shedding light on the potential applications and benefits of these technologies.

Conducting qualitative analysis necessitates substantial allocation of resources and time, which has been well-established in the research community (Smith, 2018; Patton, 2015; Heath et al., 2010). Video analysis, in particular, presents a formidable challenge due to its intricate nature, often demanding specialized technology and labor-intensive manual processes (Atkinson, 2007; Knoblauch et al., 2008). These processes can involve procedures such as: (1) crafting descriptive narratives of actors and activities (Laurier, 2019), (2)

\* Corresponding author.

E-mail address: [matthew.schmidt@uga.edu](mailto:matthew.schmidt@uga.edu) (M. Schmidt).

<sup>1</sup> Authors report equal contributions to the research design, evaluation, and manuscript preparation. Therefore, both the first and second listed authors report equal contribution as co-first author.

frame-by-frame analyses (Pipkin et al., 2016), (3) inductive extraction of themes (Krogh et al., 2015) and audio transcription (Schmidt et al., 2023; Schidt & Glaser, 2021a, 2021b). This becomes even more complicated when analyzing 360-degree spherical videos, due to their unique format and challenges in interpretation (Yaqoob et al., 2020). However, recent advancements in computer vision techniques like object detection and scene recognition, as well as artificial intelligence (AI) analysis methods such as deep learning and neural networks (Hwang et al., 2020) have the potential to revolutionize the way researchers analyze video data (Buch et al., 2011). These technologies can automate many manual processes like data labeling, feature extraction, and pattern recognition and provide more accurate and efficient results owing to their ability to process vast amounts of data swiftly and to learn from iterative processes (Tschang & Almirall, 2021).

### 1.1. Motivation

Interest in the use of 360-degree SVVR is growing, with particular enthusiasm being expressed towards the low cost and perceived ease of development with this medium (Glaser, Newbutt, Palmer, 2022; Glaser, Newbutt, Palmer, et al., 2022). SVVR can provide a range of advantages for autistic learners due to its ability to provide structured, customizable, and controlled simulations that provide a particularly high degree of visual fidelity (Glaser, Newbutt, Palmer, 2022; Glaser, Newbutt, Palmer, et al., 2022). Much like traditional VR interventions, SVVR also can provide predictability, repetition, and visual learning, which have been shown to align well with the needs of autistic learners (Schmidt, Glaser, et al., 2020; Schmidt, Tawfik, et al., 2020). Furthermore, research suggests that SVVR can reduce anxiety among autistic participants by allowing for gradual exposure to unfamiliar situations (Schmidt et al., 2021), aid in development of social competencies through immersive video modeling (Parsons & Cobb, 2016), and offer immediate feedback based on performance (Zhang et al., 2022). As with other immersive technologies, SVVR also can enhance engagement and motivation (Zhang et al., 2022). Taken together, these benefits position SVVR as a particularly promising tool to support skill acquisition and generalization for individuals with autism (Glaser & Schmidt, 2022). However, very little research exists to-date that explores the use of SVVR with autistic groups (Glaser and Schmidt, 2022).

Similar to the research on SVVR with autistic groups, research on the use of AI to analyze data from XR applications developed for autistic groups is nascent and limited, with the vast majority of research focusing on the use of AI to diagnose autism (cf. Anagnostopoulou et al., 2020; Song et al., 2019). Given the challenges of qualitative video analysis and the large investment of time required to do so, automated, AI-based computational approaches to video analysis could present a promising approach for augmenting traditional methodological approaches in this research area (Bohr & Memarzadeh, 2020; Longo, 2020). Hence, the motivation for this research is driven by a convergence of challenges across the domains of qualitative video analysis, recent advancements in computer vision and AI technologies, and the need to provide effective immersive learning interventions for individuals with autism.

To be sure, the resource-intensive and time-consuming nature of video analysis, particularly within the less explored context of 360-degree spherical videos, underscores the demand for more streamlined methodologies. Application of computer vision and AI techniques are promising, as they can automate labor-intensive manual processes and potentially yield more accurate outcomes. However, how to apply these techniques within immersive learning contexts for autism research is an area that remains virtually unexplored, which highlights the overarching research gap that this study aims to address. Situated at the intersection of computer science, learning sciences, and autism research, the approach described here is highly interdisciplinary, which, as some researchers have argued (e.g., Glaser et al., 2021; Schmidt, Glaser, et al., 2020; Schmidt, Tawfik, et al., 2020), is central to overcoming known

constraints of VR development for autistic users.

### 1.2. Objectives and research questions

The purpose of this exploratory analysis is to detail our methodological approach in general and our methods in particular for (1) extracting salient information from SVVR videos using computer vision and (2) applying a range of AI methods to explore behavior sequences and patterns so as to better understand predictability and variability in VR behavior within autistic and neurotypical groups. To this end, we seek to address the following research questions:

RQ1: How does SVVR usage differ between autistic and neurotypical individuals in terms of behavior sequences, object associations, and common patterns?

RQ2: To what extent does the predictability and variability of behavior sequences, object associations, and common patterns in SVVR distinguish autistic participants from neurotypical peers?

While these research questions are focused on differences and distinguishing characteristics of autistic and neurotypical SVVR usage characteristics, the underlying purpose of this study is ultimately to bridge the gap between the utilization of computer vision techniques and AI methods to analyze SVVR data, the potential advantages of such methods in extracting meaningful information, as well as implications for using AI for analyzing immersive learning interventions among autistic and neurotypical groups.

## 2. Literature review

### 2.1. Autism spectrum disorder

Autism is a complex neurodevelopmental condition that impacts individuals' social communication and interaction, as well as the presence of restricted or repetitive behaviors or interests (American Psychiatric Association, 2013). Autism is a lifelong condition that frequently co-occurs with other conditions, such as epilepsy, anxiety, attention deficit hyperactivity disorder, cognitive impairments, and obsessive-compulsive disorder (Brondino et al., 2019). Approximately 2% of people worldwide are affected by autism, with 1 in 36 children in the United States receiving an autism diagnosis (Maenner, 2023). Individuals on the autism spectrum often experience difficulties in educational performance, vocational success, and quality of life, with long-term social inclusion and employment outcomes for autistic people being poor (Brown et al., 2021). Between 50% and 75% of autistic adults are unemployed or underemployed, and vocational and educational activities are critical factors that influence independence and quality of life (Taylor et al., 2014).

Recent advances in technology have brought attention to the potential use of virtual reality (VR) technology in general, and SVVR in particular, as an intervention tool for autism (cf. Glaser & Schmidt, 2022; Schmidt et al., 2019). According to Slater (2018), VR is a computer-generated model of reality that allows users to interact with and obtain information from the model using ordinary human senses, and that is typically associated with three-dimensional environments that can induce psychological sensations of feeling "present" known as telepresence. Research suggests that VR can offer a range of learning benefits such as predictability, structure, customizable task complexity, control, realism, immersion, automation of feedback, assessment, and reinforcement (Bozgeyikli et al., 2018). Given the affordances of the technology, VR can be used to simulate real-life experiences in a controlled and safe environment (Dixon et al., 2020). This can help individuals with autism develop the skills and confidence needed to successfully navigate real-world situations (Parsons & Cobb, 2016).

### 2.2. 360-Degree spherical video-based virtual reality

Spherical, video-based VR is a type of VR technology that uses a 360-

degree spherical view to create immersive, interactive virtual environments. It creates the illusion of being in a virtual environment. In SVVR, users view the virtual environment through a headset or other device, and can view the environment with 3 degrees of freedom simply by moving their head (Glaser, Newbutt, Palmer, 2022; Glaser, Newbutt, Palmer, et al., 2022). This creates a sense of immersion, but less so than other immersive technologies (i.e., head-mounted display-based VR such as Meta Quest 2), as users are still aware that they are viewing a video and are not actually present in the simulated environment (Snelson & Hsu, 2020). This is in contrast to fully immersive VR that uses computer-generated graphics and other technology to create a completely artificial environment that users can interact with and explore. In this type of VR, users wear a headset or other device that tracks their movements and provides visual, auditory, and sometimes tactile feedback to create a sense of being fully immersed in the virtual environment (Glaser, Newbutt, Palmer, 2022; Glaser, Newbutt, Palmer, et al., 2022). Fully immersive VR allows users to move freely and interact with the virtual environment in a way that is not possible with spherical video-based virtual reality. This distinction between SVVR and fully immersive VR and their user interaction and sensory engagement makes SVVR objectively less immersive than fully immersive VR (Miller & Bugnariu, 2016; Slater et al., 2009).

Spherical video-based virtual reality has the potential to be a valuable tool for learning, instruction, and assessment (Schidt & Glaser, 2021a, 2021b). By providing a realistic and immersive environment, spherical video-based virtual reality can help engage and motivate learners, and can provide new opportunities for exploration and experimentation (Dixon et al., 2020). In terms of instruction, spherical video-based virtual reality can be used to create interactive and engaging learning experiences (Pirker & Dengel, 2021). For example, students could use spherical video-based virtual reality to explore and learn about different historical periods (Yildirim et al., 2018) or geographic locations (Prisille & Ellerbrake, 2020), or to engage in hands-on simulations of complex scientific (Nersesian et al., 2019) or mathematical concepts (Ferdig & Kosko, 2020). SVVR can also be used to create authentic and realistic scenarios for evaluating student performance (Yoganathan et al., 2018). For example, students could be assessed on their ability to navigate and interact with a virtual environment, or to solve complex problems in a virtual setting (Pirker & Dengel, 2021), which can help provide a more complete and accurate assessment of student learning and skills. In one example, researchers deployed a SVVR training to provide nurse training around blood transfusion safety, with results suggesting that use of SVVR improved students' learning achievement in blood transfusion safety and that the technology had potential for developing other professional capabilities (Huang et al., 2021).

### 2.3. Artificial intelligence and data mining in education

Research concerning the use of artificial intelligence (AI) technologies in education has been ongoing for over 30 years (O'shea & Self, 1986). Since these endeavors first began, researchers have explored the potential of AI technologies to facilitate the teaching and learning process with the goal of furthering inferences, judgments, and/or predictions in educational settings (Holmes et al., 2016). In recent years, advances in computational systems and processing power has led to a renewed interest in using AI, and there is a growing number of studies being published that broadly explore the potential of this technology (Hwang et al., 2020). More specifically, AI has been widely applied to education technologies and learning environments including intelligent tutoring systems, teaching robots, learning analytics dashboards, adaptive learning systems, etc. (Hwang et al., 2020).

Data mining serves as the foundation of AI. Its use in education consists of applying data mining techniques and tools to analyze educational data in order to identify patterns and trends that can be used to improve teaching and learning (Koedinger & Alevan, 2007). This can

involve using data mining algorithms to analyze large datasets of student performance data, such as test scores, grades, and attendance records, in order to identify factors that are associated with academic success (Ramaphosa et al., 2018). This information can then be used by educators to tailor their instruction and support to the needs of individual students, and to identify areas where additional resources or interventions may be needed. Data mining in education can also be used to identify trends and patterns in student behavior and engagement, such as the effectiveness of different teaching methods or the impact of various factors on student motivation and retention (Aldowah et al., 2019). By analyzing this data, educators can better understand the factors that influence student learning and achievement, and can use this information to make more informed decisions about how to improve teaching and learning in their classrooms (Bienkowski et al., 2012).

Data mining methods have been increasingly utilized to identify previously unknown patterns in raw data, thereby providing valuable insights across various domains. One such method, clustering, has been applied to training software named VideoTote, with the aim of enhancing job training for autistic adults (Burke et al., 2013). This software, which was implemented on a tablet, employs video modeling and video prompting techniques to facilitate the learning process of individuals in the context of packing and shipping operations in a manufacturing and shipping warehouse. During the usage of the VideoTote software, user events were meticulously recorded and the device utilization patterns were analyzed. The clustering method was then applied to the collected data, which revealed that different participants exhibited distinctive patterns of utilizing the software.

### 2.4. Computer vision in education

Computer vision (CV) is a field of study within computer science and artificial intelligence that focuses on enabling computers to interpret and understand visual information from the world, in the same way that human vision does (Stockman & Shapiro, 2001; Venables, 2019). It fits into the broader framework of artificial intelligence as it aims to develop algorithms and models that allow machines to gain high-level understanding from digital images and videos (Forsyth & Ponce, 2002). The purpose of using CV is to build systems that can perform tasks that typically require human-level perception and cognition, such as object recognition, image classification, scene understanding, etc. (Fan et al., 2022). Perhaps one of the most well-known applications of CV is self-driving cars, which utilize CV algorithms to navigate roads and detect obstacles (Dikmen & Burns, 2016). In the field of education, research studies have used CV methods for analyzing students' behavior and engagement in the classroom, developing educational technologies that can provide personalized feedback to students, etc. (Hwang et al., 2020).

Computer vision technology has received significant attention for its potential to provide supports for individuals with autism (de Belen et al., 2020). One common application is the use of CV algorithms to detect and analyze facial expressions, body language, and other nonverbal cues to help individuals with autism better understand and interpret social cues and interactions (Ali et al., 2022; Benssassi et al., 2018; Hashemi et al., 2012). This can help individuals with autism better navigate social situations, improve communication skills, and also has shown promise in the area of autism diagnosis (Hashemi et al., 2012). In addition, CV technology has been used to develop assistive technologies for individuals with autism, such as apps that use facial recognition to help individuals with autism identify and remember the names of people they meet (Drimalla et al., 2021). Other applications of CV technology for autistic individuals include using visual cues to help with scheduling and organization, and using computer vision to track and analyze behavior patterns to identify potential triggers for challenging behaviors (Marinoiu et al., 2018).





Fig. 1. Screenshots from the Virtuoso-SVVR app, illustrating a participant (1) selecting a training scenario and (2) an associated activity, then (3) viewing an associated SVVR video that models (4) how to use the university shuttle system.

### 3. Methods

#### 3.1. Proposed Approach

We created and evaluated a novel VR intervention for autistic adults called Virtuoso, which was designed to support autistic individuals in their transportation needs. Virtuoso is a virtual reality intervention that was designed for and in collaboration with a day program called *Impact Innovation* that provides support and services for over 20 autistic adults in a year-round program with an emphasis on promoting vocational and social opportunities within the community. Given the inherent dangers associated with the use of public transportation, the intervention was specifically designed to improve autistic users' ability to navigate transportation systems safely and effectively. One component of Virtuoso was a SVVR experience designed to model the behaviors of using public transportation for autistic adults. The target goals of the SVVR intervention were situated around providing a formalized shuttle training routine as transportation is often cited as being one of the greatest barriers to community integration for individuals with disabilities (Allen & Mor, 1997; Carmien et al., 2005). Virtuoso uses a stage-wise technique that progresses from simple to complex across a spectrum of low-tech to high-tech tools (Schidt & Glaser, 2021a, 2021b). The stage-wise approach includes (1) skill introduction, (2) 360-degree video modeling of the skill, (3) rehearsal of the skill within a fully immersive VR scenario, and (4) real-world practice of the skill. The work presented in this manuscript focuses entirely on the second stage of our approach. In this stage, users were presented within a series of 360-degree videos that broke the task of catching a shuttle bus into four discrete subskills, including: (1) check the daily schedule, (2) walk to the shuttle, (3) wait for the shuttle, and (4) get on the shuttle. Virtuoso-SVVR was deployed in an Oculus Rift head-mounted display.

#### 3.2. Participants

The current study took place at a large Midwestern university in the United States. Research participants from the autistic group were recruited from an autistic adult day program. Neurotypical participants

were also recruited from the day program, as they were student staff members. Purposeful sampling was used to identify autistic participants who would represent a broad range of comorbidities across the autism spectrum. Six autistic participants were recruited, all of whom were male, had a medical diagnosis of autism, and were between 22 and 35 ( $M = 26.6$ ) years old. Autism diagnosis included the Peabody Picture Vocabulary Test ( $M = 134$ ;  $\min = 57$ ;  $\max = 196$ ;  $SD = 58$ ), Social Responsiveness Scale ( $M = 71$ ;  $\min = 65$ ;  $\max = 82$ ;  $SD = 6$ ), and the Behavior Rating Inventory of Executive Function ( $M = 79$ ;  $\min = 50$ ;  $\max = 94$ ;  $SD = 17$ ). In addition, six neurotypical participants were recruited. Convenience sampling was used to recruit neurotypical participants, all of whom were staff members in the adult day program, were evenly divided between male and female, and were between 21 and 26 ( $M = 22.6$ ) years old. All participants viewed the SVVR videos in the principal investigator's office, with autistic participants having a trained staff member present. This research and all procedures were approved by the IRB.

#### 3.3. Procedures

This study took part within a structured usage test of the Virtuoso intervention, with all data being collected during June of 2019. Two separate groups of participants took part in the study: a group of autistic adults ( $n = 6$ ) and a group of neurotypical peers ( $n = 6$ ). Participants from each group took part in one data collection session that lasted around 1 h. During this time, participants viewed SVVR videos presented using an Oculus Rift. Four videos were presented. The videos followed the four tasks that were the focus of the transportation intervention described in the Proposed Approach section above (see Fig. 1 for an example of the system interface and a learning scenario).

#### 3.4. Data collection

Participants viewed the corpus of Virtuoso SVVR videos using an Oculus Rift head-mounted display (HMD). The video from the HMD was mirrored to a connected laptop. Participants' view of the SVVR videos was captured using Open Broadcaster Software (OBS), which recorded

the full field-of-view presented to the participant. All recordings were captured at the same resolution. Each participant viewed four SVVR videos, resulting in four recordings—one for each of the videos watched. In total, 48 videos were recorded. Recorded video duration ranged from 47 s to 2 min, 30 s, depending on the length of the SVVR video that participants viewed.

### 3.5. Ethical considerations

All research was approved by the lead author's university institutional review board. Before their involvement, all participants (including both autistic and neurotypical) were informed about the nature of the study, its purpose, procedures, potential risks, and benefits. Informed consent was obtained from all participants, ensuring their voluntary participation and understanding of the study's objectives. Particularly for the autistic participants, who might require additional support in comprehending the study's details, trained staff members were present during the sessions to assist and address any concerns. All participants' identities were protected through the use of pseudonyms. All collected data was securely stored using encrypted digital storage methods to ensure the confidentiality and integrity of participant information. Recruitment was conducted equitably from both autistic and neurotypical participant groups, with a view to achieve balanced representation for comprehensive analysis.

### 3.6. Automated data extraction from SVVR videos using computer vision methods

A Python script was developed to detect objects in the video recordings (Appendix A). The prerequisite technologies that power the script were TensorFlow (<https://www.tensorflow.org/>) and Image AI 2.0.3 (<http://www.imageai.org/>). TensorFlow is an open-source software library for machine learning and artificial intelligence. It is developed by Google and is used for a wide range of applications, including natural language processing, image recognition, and predictive modeling. TensorFlow is designed to be flexible and scalable, allowing developers to build and deploy machine learning models easily and efficiently. ImageAI is an open-source library for Python that allows developers to easily integrate state-of-the-art computer vision technologies into their applications. ImageAI is built on top of TensorFlow and provides a simple and powerful approach to training custom object detection models using the YOLOv3 algorithm (<https://viso.ai>). We used the Anaconda distribution of Python (<https://www.anaconda.com/>) along with the Spyder (<https://www.spyder-ide.org/>) integrated development environment (IDE) to simplify management of the various software tools described here. Anaconda is a no-cost distribution of the Python and R programming languages that includes a variety of packages and tools for data science, scientific computing, and machine learning. Spyder is a no-cost IDE for scientific computing and data science, written in Python. It is included in the Anaconda distribution.

Our Python script first segmented our SVVR usage test videos into individual image stills at 30 frames per second and provided a detected object classifier and probability based on the pre-trained YOLOv3 neural network weights. The frame number, object class, confidence value, and bounding box coordinates were exported to one CSV file per video. Each set of four videos required, on average, 12 h to process using a high-end, custom-built development machine. Total processing time was approximately 144 h. Exported CSV files were reviewed, and the classifications with a confidence score of 80 and above were reviewed for accuracy (Bian et al., 2021) and organized into a separate worksheet for further analysis.

### 3.7. Data mining & analysis

Following data extraction, we conducted three analyses using machine learning: (1) entropy, (2) Markov chains, and (3) sequential

pattern mining, because these methods offer unique insights into the patterns and behaviors exhibited in the video data. Entropy helps gauge the unpredictability of behavior sequences (Shannon, 1948), Markov chains provide probabilistic transitions between observed behaviors (Han & Kamber, 2006), and sequential pattern mining identifies recurring behavior patterns (Baker & Inventado, 2014). Together, these analyses provide a comprehensive understanding of the data, allowing for more nuanced interpretations and conclusions. These analyses were conducted using a custom script in RStudio (<https://www.rstudio.com/>).

#### 3.7.1. Entropy

Entropy, or information density, measures how much content an event carries, or the uncertainty in a series of numbers (Shannon, 1948). The value of entropy normally lies between 0 and 1 when there are two possible categories. When there are multiple categories, the value may exceed one. The higher the value is, the more information the unit carries, and the higher level of uncertainty it is. Entropy analysis has been used to effectively identify malware in computer science research (Lyda & Hamrock, 2007), as well as to measure the complexity of learning behaviors presented in solving ill-structured problems (Tawfik et al., 2018). We conducted entropy analysis to understand the uncertainty on behavior sequences between the autistic and neurotypical groups of study participants.

#### 3.7.2. Markov Chains

Markov chains are stochastic models describing a sequence of events where the probability of a state depends on the previous state (Han & Kamber, 2006). Markov chains use network structures to condense the data. They present a global representation of the data, wherein each state is an item and each item is taken into consideration and can be used to understand students' performance in conducting an experiment, for example, whether 3D virtual environment simulations improve learner achievement in handling lab equipment during physical science experiments (Paxinou et al., 2021). We implemented this method to determine the sequence of focal objects in the entire SVVR experience.

#### 3.7.3. Sequential pattern mining

Sequential pattern mining (SPM) is a technique used in data mining to identify patterns in a sequence of data. It involves finding occurrences of patterns in data sets that can be ordered in some way, such as time series data or transaction data. SPM is one of the commonly used learning analytics methods to discover the frequent patterns from input sequences (Baker & Inventado, 2014). It can be used to understand students' learning behaviors that are associated with more successful and less successful collaborative learning groups in online learning or computer-supported learning environments (Zheng et al., 2019). It takes a number of sequences as input, and calculates statistically relevant patterns. Its output is a number of patterns and their support values. The support value which lies between 0 and 1 indicates the possibility of the pattern occurring in the given data set. The higher the support value is, the more frequently the pattern occurs. We utilized this method to discover meaningful patterns and compare the differences between the autistic and neurotypical groups of participants.

## 4. Results

Three machine learning analyses were performed, including: (1) entropy, (2) Markov chains, and (3) sequential pattern mining. Findings from these analyses and comparisons between the autistic (ASD) and neurotypical (NT) groups of participants are presented in the following sections.

### 4.1. Entropy results

To understand the uncertainty on behavior sequences between the

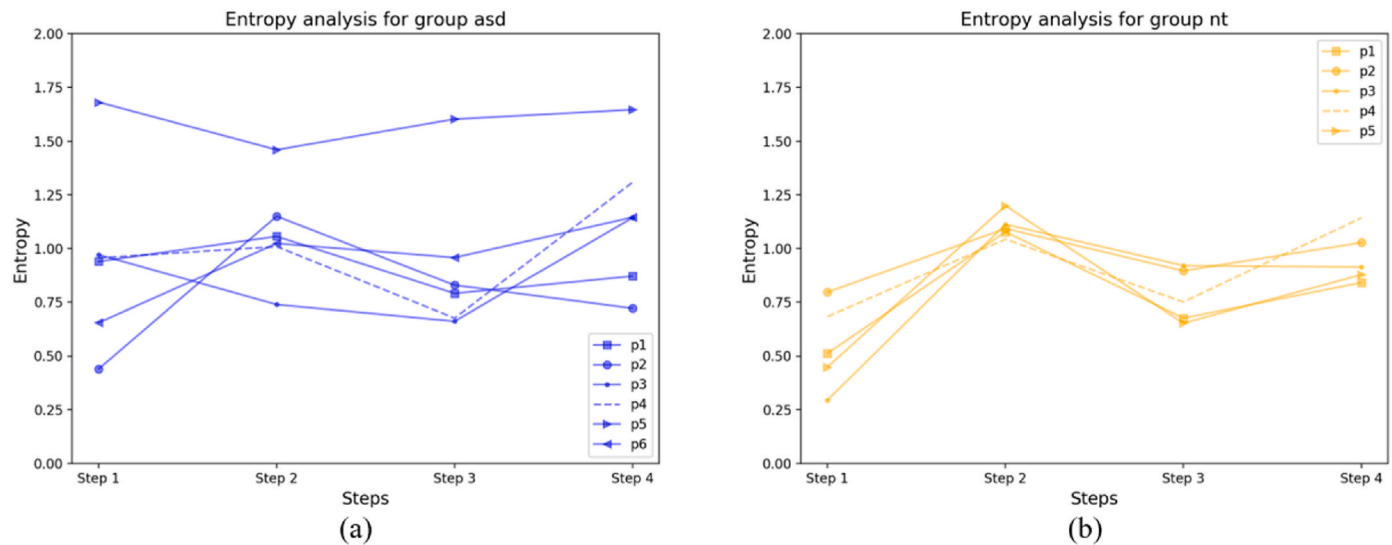


Fig. 2. Entropy analysis for two groups of participants: (a) group ASD; (b) group NT.

ASD and NT groups of participants, we calculated the entropy value for each participant at each step, as shown in Fig. 2. In general, participants in group NT shared similar patterns on the entropy values, while the entropy values were highly variable in group ASD. To investigate this further, we performed a one-way ANOVA test on the entropy value between group ASD ( $N = 4 \times 6$ ) and group NT ( $N = 4 \times 5$ ) ( $F(1, 42) = 3.61$ ,  $p = 0.06$ ). The F-value of 3.61 suggests that there is some difference between the means of the ASD and NT group, with the p-value of 0.06 approaching significance.

#### 4.2. Abbreviated Markov Chains results

Output from our Markov Chains analysis resulted in 48 separate matrices, with four matrices for each participant—one for each step of the instructional sequence. The data represents a set of probabilistic transition matrices that show the probability of transitioning from the current state to the next state for a given set of identified objects from the SVVR videos. Each matrix has a set of objects in both the current state and the next state, and the probabilities are represented as decimal values ranging from 0 to 1.

Given the breadth of these results, a full presentation and analysis of these matrices is beyond the scope of the current article. Below, we provide representative sample data in the form of two matrices (see Table 1). This is followed by an analysis of these specific data, which illustrates our overall analysis approach. We then provide a general overview of findings from the full analysis of the 48 matrices.

In the above example, ASD P1 and ASD P2 matrices both have the object ‘cell-phone’ in the current and next states. However, the probabilities of transitioning from the current state to the next state for this object differ between the two matrices, suggesting that object viewing characteristics may have changed over time or in different contexts. ASD P3 differs from ASD P1 and ASD P2 in that this participant has several objects in the current state with no transitions to the next state, potentially indicating a decreased likelihood of associating these objects with other objects in the viewing process. ASD P4 only has three objects, which are all related to household items. This matrix may represent a more focused or specific viewing task compared to the previous three participants. ASD P5 has a much larger set of objects compared to the other matrices, which may indicate a more complex or varied viewing task. The matrix also has more objects with non-zero probabilities. ASD P6 shares several objects with the previous participants, but there are notable differences in the probabilities of transitioning between objects. For example, the probability of transitioning from ‘chair’ to ‘person’ is

much higher in the sixth matrix compared to the first two matrices, which may suggest a shift in the way that ASD P6 associates these objects with each other over time or in different contexts.

Shifting focus from this specific example to a more general analysis of the full 48 matrices, we found that each of the matrices had a unique set of objects, and there were no consistent or shared relationships between the objects across the matrices. Although some level of uncertainty or randomness was found across all matrices regarding transitions from the current state to the next state for a given object, there was far more variability in the ASD participants’ Markov Chains than in the NT participants’, and the probabilities predicting which objects might come next in the chain were more predictable for NT participants than for ASD participants. For instance, the given probabilities for NT P1 in Step 1 suggested that if the current state is a person, there was a high chance (88.4%) of staying in the same state and a low chance (7.2%) chance of transitioning to a different state. This trend extended to all other NT participants, suggesting that the object ‘person’ had a very high probability of following an inanimate object such as a cell-phone or a screen. In contrast to this, for ASD P1 in Step 1, the probability of transitioning from an object to a person varies. While for ASD P1, P3, and P4 there was a high probability of transitioning to a person state from an object state (96.8%, 95.8%, 70.3% respectively), the probability was lower for P6 (45.6%). For ASD P2 and P5, there was a much lower probability of transitioning to a person state from an object state (35% and 39.2%, respectively).

#### 4.3. Sequential pattern mining results

An essential goal of SPM is to extract meaningful sequences from large datasets. The current research studied the sequences of classes in the same frame. If a frame had two or more classes, then a sequence was generated based on the temporal order of these classes. Table 2 and Table 3 present the number of extracted sequences from each participant from each group at each step. In general, for each participant in both the ASD and NT groups, the number of extracted sequences varied significantly across different steps. However, within each step, the number of extracted sequences shows less variability, which could potentially be attributed to high implementation fidelity.

There were 27 classes in the sequences of group NT, while the number for group ASD was 46. Participants in the ASD group registered more classes than the NT group. While the ASD group shared nearly all classes found in the NT group data, the class ‘baseball bat’ was not found in the ASD group data. The total frequency for each class is shown

**Table 1**

Markov Chain results from Participants 1 and 2 for Step 1 of Task Analysis.

ASD P1 Step 1						
			Next State			
			cell-phone	chair	person	refrigerator
Current State	cell-phone	0	0.574468085	0	0.425531915	0
	chair	0	0	0.96797153	0.028469751	0.003558719
	person	0.002673797	0.053475936	0.018716578	0.906417112	0.018716578
	refrigerator	0	0	0.111111111	0.777777778	0.111111111
ASD P2 Step 1						
		Next States				
Current State	book	book	cell-phone	chair	person	refrigerator
		0	0	0	1	0
	cell-phone	0	0.857142857	0	0.142857143	0
	chair	0	0	0.282051282	0.641025641	0.076923077
	person	0.005008347	0.001669449	0.038397329	0.941569282	0.013355593
	refrigerator	0	0	0.25	0.35	0.4
ASD P3 Step 1						
		Next State				
Current State	book	book	cell-phone	chair	person	refrigerator
		0	0	0	1	0
	cell-phone	0	0.656976744	0	0.343023256	0
	chair	0	0	0	0.958333333	0.041666667
	person	0.035335689	0.104240283	0.040636042	0.78975265	0.030035336
	refrigerator	0	0	0.016393443	0.278688525	0.704918033
ASD P4 Step 1						
		Next State				
Current State	chair	chair	person	refrigerator		
		0.827338129	0.172661871	0		
	person	0.065826331	0.703081232	0.231092437		
	refrigerator	0	0.763888889	0.236111111		
ASD P5 Step 1						
		Next State				
Current State	backpack	backpack	bench	book	bottle	cell-phone
		0	0	0	0	0
	bench	0	0	0	0	0
	book	0	0.125	0	0	0
	bottle	0	0	0	0	0
	cell-phone	0	0	0	0	0.114285714
	chair	0	0	0.008663366	0	0.055693069
	dining-table	0	0	0	0.023391813	0
	handbag	0	0	0.027777778	0	0
	knife	0	0	0	0	0
	laptop	0	0	0.010204082	0	0
	oven	0	0	0	0	0
	person	0.000834725	0.010016694	0.039232053	0	0.088480801
	refrigerator	0	0	0	0	0.012269939
	sink	0	0	0	0	0
	sports-ball	0	0	0	0	0
	suitcase	0	0	0	0	0
tv	0	0	0	0	0	
ASD P6 Step 1						
		Next State				
Current State	book	book	chair	laptop	person	refrigerator
		0	0	0	1	0
	chair	0	0.527868852	0.003278689	0.455737705	0.013114754
	laptop	0	0	0	1	0
	person	0.006178288	0.110326567	0	0.854368932	0.029126214
	refrigerator	0	0.473684211	0	0.5	0.026315789

in Fig. 3. A class was included only if it was presented in extracted sequences. In other words, the frequency of each class did not equal the total frequency of each class in log files. Class “car” and “person” had extremely high frequency due to the nature of the subject matter (i.e., public transportation). This introduced difficulties in visualizing differences between other classes. The difference in the frequency of class “car” between the two groups was too small to be visible in the chart

illustrated in Fig. 4, resulting in two columns of similar height. We therefore removed these two classes and re-plotted the figure, as shown in Fig. 3. For classes shared by the NT and ASD groups, we did not observe significant differences in frequencies between the two groups (Table 4). However, frequencies for a some classes (e.g., baseball bat ( $N = 6$ ), dog ( $N = 5$ )) were too low to be visible in Fig. 3.

Based on extracted sequences (see Table 5), we applied sequential



**Table 2**

Number of sequences extracted for group ASD (total: 23,690).

	P1	P2	P3	P4	P5	P6
Step 1	36	33	117	222	846	192
Step 2	2433	2326	1198	1326	2793	2237
Step 3	817	818	774	818	1107	636
Step 4	637	968	856	687	945	868

**Table 3**

Number of sequences extracted for group NT (total: 19,855).

	P1	P2	P3	P4	P5
Step 1	114	57	62	142	115
Step 2	2031	2339	2727	2401	2532
Step 3	862	561	673	683	625
Step 4	767	819	793	875	677

pattern mining to identify the common sequences for both groups. A support value of 0.5 was used to determine whether a sequence was common. Specifically, there were 36 sequences in Step 1 for a participant, then any sequence with a frequency over 18 was considered common and would be documented. Table 5 shows all the common sequences for two groups. In general, group ND had fewer sequences than group ASD. Even for the identified sequences, the sequences are no longer than those of group ASD.

## 5. Discussion

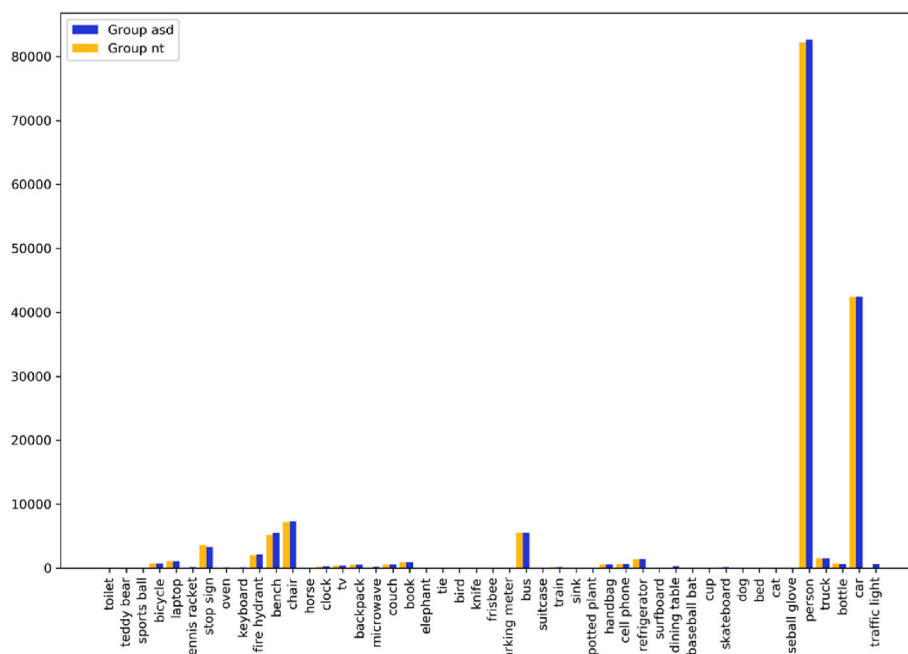
We conducted a usage study to explore research questions centering around (RQ1) how SVVR usage differs between autistic and neurotypical individuals in terms of behavior sequences, object associations, and common patterns, and (RQ2) the extent to which the predictability and variability of behavior sequences, object associations, and common patterns in SVVR might distinguish autistic participants from neurotypical peers. The data set we used for the analysis was drawn from six NT and six ASD participants' videos created while they were using a SVVR learning environment. Three machine learning analyses were performed, including: (1) entropy, (2) Markov chains, and (3) sequential pattern mining. In the following sections, we discuss our findings

relative to our research questions.

### 5.1. Discussion of entropy results

In the context of the two research questions, the conducted entropy analysis aimed to shed light on the behavioral sequences' uncertainty between the ASD and NT groups of study participants. Findings suggest that entropy within the NT group of participants was largely similar. While some variation was found in specific entropy values between the four steps of the instruction, overall entropy patterns across steps remained mostly uniform. Contrariwise, findings suggest that entropy within the ASD group of participants was largely dissimilar. Not only was substantial variation found in specific entropy values between the four steps of the instruction, overall entropy patterns across the steps are remarkably different. Comparatively speaking, the NT group entropy data suggest greater homogeneity and therefore lower uncertainty relative to the ASD group, which shows remarkable heterogeneity and therefore higher uncertainty. This implies that behavior is more variable and less predictable among the ASD group than the NT group relative to objects that appear in the user's field-of-view while using SVVR. The findings reveal intriguing insights into these research questions. Among the NT group participants, the analysis of entropy indicates a considerable level of similarity in their behavior sequences. Although some slight variations in entropy values are observed across the four instructional steps, the overall entropy patterns remain relatively consistent. Conversely, the ASD group participants exhibit significantly dissimilar entropy patterns. Notably, substantial variations are identified in specific entropy values across the instructional steps, contributing to distinctly different entropy patterns. Comparing the two groups, the data from the NT group suggests higher homogeneity and lower uncertainty. In contrast, the ASD group displays remarkable heterogeneity and higher uncertainty in their behavior sequences within the SVVR context.

Regarding the behavior dynamics between the two groups, the NT group's greater homogeneity and lower uncertainty point towards a more predictable and less variable usage behavior when interacting with objects in their field of view. Conversely, the remarkable heterogeneity and higher uncertainty observed within the ASD group highlight the heightened variability and less predictable behavior among these

**Fig. 3.** Class distribution for both groups (all classes).



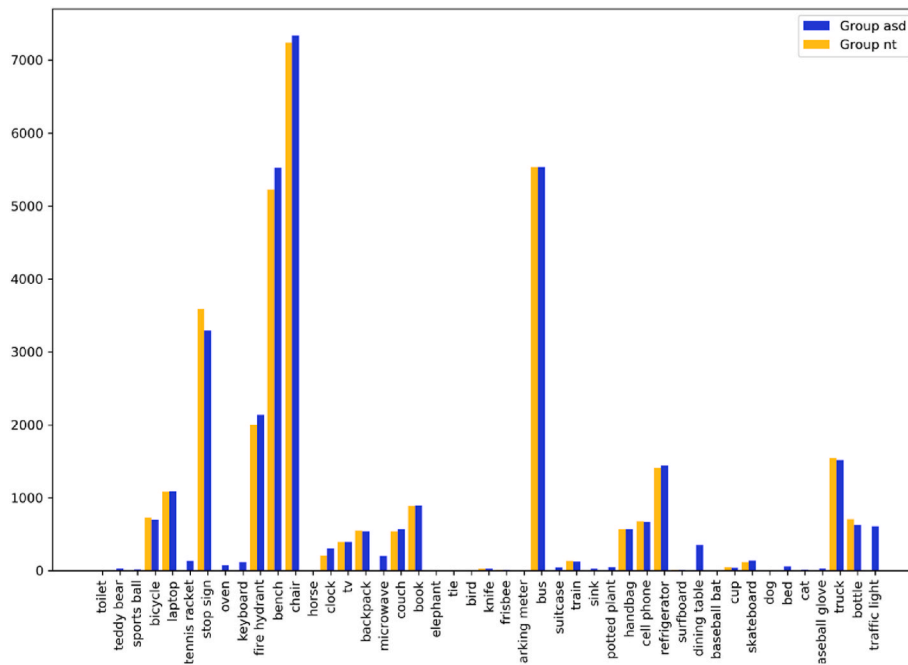


Fig. 4. Class distribution for both groups after removing “car” and “person” classes.

Table 4

One-way ANOVA test on the support between group ASD and NT.

	Step 1	Step 2	Step 3	Step 4
F (1, N) =	0.18	0.21	0.37	0.24
p-value =	0.68	0.65	0.55	0.63

Table 5

Common sequences for two groups (support = 0.5).

	ASD Group	NT Group
Step 1	person → chair person → cell phone refrigerator → person	person → refrigerator
Step 2	person → car car → car person → car → car	person → car
Step 3	person → person → person	person → person → person
Step 4	bus → person person → person → person	bus → person person → person

individuals in relation to the objects presented in the SVVR environment. This provides valuable insights into the distinctive ways these two groups engage with SVVR technology and the potential cognitive and perceptual factors at play. For example, known differences associated with autism such as sensory processing differences, attentional mechanisms, executive functioning, object perception, and individualized special interests could have contributed to the distinct patterns of behavior sequences, object associations, and usage patterns observed in the study (American Psychiatric Association, 2013; Robertson & Simmons, 2015). The heightened variability and unpredictability within the ASD group could reflect the influence of these factors on usage behavior, which aligns with Parsons’ (2016) call for better understanding of “which technologies work for whom, in which contexts, with what kinds of support, and for what kinds of tasks or objectives?” (p. 153).

## 5.2. Discussion of Markov Chains results

Addressing RQ1, our exploration of Markov Chains yielded 48

distinct matrices, each capturing the probabilistic transitions between states across four instructional steps. These matrices, representing SVVR video objects’ transition probabilities, were key to comparing behavior sequences, object associations, and patterns between autistic and neurotypical participants. In alignment with RQ2, our findings reveal noteworthy distinctions in the predictability and variability of these matrices, differentiating the two participant groups and providing provisional insights into the dynamics of their usage behaviors.

Examining the matrices’ content through the lens of RQ1, instances such as ASD P1 and ASD P2 matrices, both featuring the object ‘cell-phone’ in current and next states, highlight contextual shifts in autistic adults’ perceptions over time. Meanwhile, ASD P3’s matrix, with objects present in the current state yet lacking transitions, suggests a reduced likelihood of object associations during viewing. ASD P4’s matrix, with limited objects related to household items, suggests a focused task, differing from the more complex ASD P5 matrix. These observations hint at variations in SVVR engagement strategies across ASD participants, not entirely unlike the heterogeneity observed in the entropy results discussed previously.

Generalizing our analysis, we found the 48 matrices to be unique, indicating a lack of consistent relationships between objects across them. Addressing RQ2, this variability highlights the distinctness between autistic and neurotypical participants’ Markov Chains. While both groups exhibited some uncertainty in transitions, the ASD participants’ matrices showed heightened variability in comparison with the NT group. Moreover, probabilities predicting next objects in the chain were more stable for the NT group, implying more predictable behavior patterns, thereby distinguishing the two groups’ SVVR usage patterns.

Further distinguishing ASD participants from the NT group, as considered in RQ2, we propose that the variability observed in autistic participants’ Markov Chains might stem from cognitive diversity, potentially related to increased sensitivities to sensory stimuli and differences in perceptual processing. This aligns with literature on autism-related sensory processing differences (Robertson and Simmons, 2015), which could contribute to the wide range of observed transition probabilities and object associations.

The Markov Chains analysis presented here contributes insights into SVVR usage disparities between autistic and neurotypical participants. As with results of entropy analysis, these findings further suggest that

autistic individuals may process SVVR environments differently, thereby prompting further investigation into the relationship between sensory processing and SVVR usage. From a design perspective, these findings also suggest that designers could benefit from simplifying scenes and including only essential objects (cf. Schmidt et al., 2023) to ease analysis and enhance engagement for neurodiverse learners, while catering to the specific object associations autistic individuals are prone to make (Schwarz et al., 2018; Siris et al., 2021, pp. 4156–4166).

### 5.3. Discussion of sequential pattern mining results

Findings from sequential pattern mining are discussed here primarily from the perspective of RQ2. Specifically, the distinctiveness of autistic individuals' engagement with SVVR content was evident in the broader class spectrum observed in the data. Specifically, the ASD group data contained a strikingly wider range of registered classes (46) in comparison to the NT group (27), which could suggest that autistic participants are drawn to a more diverse set of stimuli. However, the mechanisms driving this behavior are unclear. On the one hand, heightened engagement with varying stimuli could signify an enriched sensory experience in the virtual environment, which has ramifications for motivation and engagement. On the other hand, this could suggest distraction or dysregulation, which could have deleterious effects on potential learning benefits (Keenan et al., 2023). From a design perspective, this could suggest that removing extraneous objects or objects that are not pertinent to the learning objectives potentially could enhance the visual salience of objects that are intended to be focused on (Mayer, 2005). By extension, the shared sequences and the variance in sequence lengths we identified suggest distinct engagement patterns, which could further suggest a need for more individualized learning pathways based on users' profiles and usage patterns (Romero & Ventura, 2007).

Interestingly, our analysis unveiled a lack of significant differences in shared class frequencies between the ASD and NT groups. This observation hints at a commonality in viewing preferences, despite the distinctive sequence patterns. The shared classes indicate that certain aspects of SVVR content capture the attention of both autistic and neurotypical participants, suggesting that focus areas of interest in SVVR videos may have naturalistic properties that invite focus and attention, not unlike real-world counterparts. This common ground suggests a potential direction for the design of generalizable content, which requires sufficient perceptual salience for a range of cognitive profiles in order for generalization to occur (cf. Schmidt et al., 2023). Understanding the convergence in shared class frequencies has implications for the creation of SVVR content that resonates across diverse audiences. Recognizing the alignment in focal points for both autistic and NT groups could guide content developers to emphasize key elements that capture attention universally.

### 5.4. Implications

The majority of research investigating autistic gaze patterns has been performed using eye tracking technologies (Boraston & Blakemore, 2007; Guillon et al., 2014; Papagiannopoulou et al., 2014). The current study provides a novel method for gaining insights into how autistic users interact with and view SVVR videos and provides a comparison with neurotypical participants. The insights gained from the three AI

methods employed in this study are summarized below:

1. Entropy analysis suggests that there is greater uncertainty in the ASD group's data than in the NT group's;
2. Markov chain analysis suggests that object sequences can be predicted with greater probability with the NT group than with the ASD group; and
3. SPM analysis suggests that there is variation between the NT group and the ASD group in terms of meaningful sequences of detected objects, but also that the overall distribution of objects is not statistically different.

Individually, these findings provide some insights regarding usage behavior; however, taken together, the data provide mutually supportive and parsimonious representation of how groups viewed the SVVR videos, as well as differences between the groups, as per the research questions. That is, entropy analysis provides a general understanding that the ASD group's viewing behavior is more variable than the NT group's. Markov chain results complement this finding in that they provide further details regarding this variability. Finally, SPM results extend the Markov chain and entropy findings by showing that not only does overall behavior and object prediction vary between groups, but also sequences of objects detected. As a whole, these findings provide evidence that autistic users view 360 videos differently than NT participants. This finding is corroborated by many existing studies (e.g., Grossman et al., 2019). However, what the current study adds to the literature is the manner in which these insights were gleaned. Extending this is the finding that, when compared, the overall count of identified objects was not significantly different between groups, suggesting that although differences were found in how the groups viewed the videos, the overall content (and perhaps underlying instructional message) of the videos was being attended to in similar ways by both groups. This finding has implications for the design of SVVR systems for autistic users, as discussed above.

In addition to these findings, by comparing the behavior of participants in our SVVR learning environment to the behaviors predicted by a Markov chain model, it is possible that researchers can evaluate the effectiveness of SVVR learning systems. For example, if the model predicts that participants should spend a certain amount of time on a particular task or activity (for example, checking the daily schedule or walking to the shuttle), researchers can compare this to the actual behavior of users to see if the SVVR system is effective at guiding users to engage with the content as designed. Markov chain models can also be used to optimize the design of SVVR learning systems. By analyzing data on user behavior and comparing it to the predictions of the Markov chain model, researchers can identify areas where the SVVR system is not performing as well as expected, and make changes to the design to improve its effectiveness (Talton et al., 2012, pp. 63–74).

### 5.5. Limitations

Findings from this study should be considered in light of the limitations described in the following sections. We begin with specific limitations related to the study, followed by general limitations related to our methodological approach.

### 5.5.1. Specific limitations with the current study

First, the dataset used for this research was relatively small, consisting of a limited number of participants from both ASD and NT groups. A larger and more diverse dataset could provide a more comprehensive understanding of SVVR usage patterns and potentially uncover subtler distinctions between the two groups. Second, the current study focused on SVVR usage within a specific learning context. As such, the findings may not be directly applicable to other SVVR scenarios or contexts. Different SVVR experiences, such as gaming or recreational use, might yield different patterns of engagement and behavior that were not explored here. Third, the sequence extraction process was limited in that the order of classes within the same time frame was used to generate sequences, potentially overlooking meaningful interactions between objects within a different frame. A more sophisticated approach that considers intra-frame interactions could provide a more nuanced perspective on SVVR usage patterns. Fourth, the application of sequential pattern mining inherently involves the selection of parameters, such as the support value used to define common sequences. Varying these parameters might lead to different patterns being identified, and the chosen threshold could influence the interpretation of results. Fifth, the SPM method itself has certain limitations. It operates on the assumption of independence between different time steps, which might not hold true in all SVVR scenarios. Additionally, SPM is sensitive to the order and frequency of sequences and might not capture more complex relationships or subtle variations in behavior. Sixth, the study mainly focused on identification of patterns and did not deeply explore underlying cognitive processes that may have driven observed behaviors. Future research should incorporate additional data collection techniques, such as cognitive interviews (Schmidt, Glaser, et al., 2020; Schmidt, Tawfik, et al., 2020), user surveys (Lu et al., 2022), or electroencephalography devices (Davidesco et al., 2023), to provide deeper insights into participants' motivations and perceptions during SVVR usage. Finally, because the current study was intentionally exploratory, questions pertaining to its validity and reliability remain unanswered. While our research design and methodologies were thoughtfully chosen to ensure rigor and objectivity, the nature of exploratory research inherently invites further investigation and replication to establish the generalizability and robustness of the findings. Future research with larger and more diverse datasets, as well as a broader range of SVVR contexts, could contribute to a more comprehensive understanding of SVVR behavior dynamics across different populations and scenarios.

### 5.5.2. General limitations with methodological approach

While CV algorithms can be highly accurate, they falter with understanding and interpreting visual data as compared to the human brain (Zhuang et al., 2017). This means that there are still many tasks that are difficult for computers to perform when it comes to analyzing visual data. CV algorithms are only able to recognize and understand the objects and scenes that they have been trained on (Leo et al., 2018; O'Mahony et al., 2020, pp. 128–144). If an algorithm has not been trained to recognize a particular object or scene, it will not be able to understand it and may interpret it as something else. For example, in dataset ASD P3, the model had high accuracy for the “train” and “bus” classes, but low accuracy for the “fire-hydrant” and “backpack” classes. It is likely that researchers will need to spend substantial effort in improving their training data to improve predictions for these classes (e.

g., Narejo et al., 2021). Using generic object detection models (for example, we used the general YOLO algorithm for object detection) is likely to be insufficient for specialized studies. Ultimately, real world applications need greater certainty in predicting the next state of objects (Zhao & Li, 2020). For better results, researchers need to train CV algorithms using the specific objects included in their data set (Ouyang & Wang, 2019).

Secondly, computers can only process visual data within certain parameters as CV algorithms are highly sensitive to the quality and resolution of the visual data that they are processing. In the case of low-resolution or poor quality videos, the algorithm may not be able to accurately interpret the data (Paneru & Jeelani, 2021). In addition, CV algorithms are often highly complex and require substantial computational power to run. This can serve as a barrier to implementation when working in resource-constrained contexts or using lower-powered devices, such as smartphones or tablets. As the field of VR interventions for autistic adults continues to evolve, future studies will need more robust participant samples, enhanced control measures, and longitudinal assessments.

## 6. Conclusion

In this study, we aimed to explore the potential of CV and AI methods in analyzing data extracted from SVVR videos. Our findings suggest that these AI methods, including entropy analysis, Markov chain analysis, and sequential pattern mining, can provide valuable insights into the behavior patterns of autistic and neurotypical users in SVVR-based video data. Our study highlights the potential of AI methods in analyzing SVVR-based video data, which can provide insights into the behavior patterns of autistic and neurotypical users in a VR learning environment. These findings provide a preliminary foundation for future studies in this area, as well as practical implications for designing effective SVVR-based educational materials.

### Statements on open data and ethics

The participants in this study were safeguarded through the concealment of their personal information. They participated voluntarily and were informed of their right to withdraw from the experiment at any point. The dataset utilized in this research can be made available upon request by contacting the corresponding author.

### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 in order to improve language and readability, with caution. After using this tool/service, the author(s) reviewed and edited the content as needed and takes full responsibility for the content of the publication.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A

### Python Script for Machine Learning-based Object Detection

```
# import relevant libraries
from imageai.Detection import ObjectDetection
import os
from os import listdir
from os.path import isfile, join
import mimetypes
import cv2
import os
import pandas as pd

# prepare YOLO algorithm
execution_path = os.getcwd()
detector = ObjectDetection()
detector.setModelTypeAsYOLOv3()
detector.setModelPath(os.path.join(execution_path, "yolo.h5"))
detector.loadModel()

# point the script to SVR videos & prepare videos
videos_path = "./p5" #This should be changed to your videos folder
path
fnames = [f for f in listdir(videos_path) if isfile(join(videos_path,
f))]
videos_names = []
for vname in fnames :
    try:
        if mimetypes.guess_type(join(videos_path,
vname))[0].startswith('video') :
            videos_names.append(vname + "__FOLDER__")
    except :
        pass

print(videos_names)

# convert video to still images
def video2Frames(vname):
    vidcap = cv2.VideoCapture(os.path.join(videos_path,
vname.replace("__FOLDER__", "")))
    video_path = os.path.join(".", vname)
    os.mkdir(video_path)
    success,image = vidcap.read()
    count = 0
    while success:
        cv2.imwrite(os.path.join(video_path, "%d.jpg" % count),
image) # save frame as JPEG file
        success,image = vidcap.read()
```



```

        count += 1

for vname in videos_names :
    video2Frames(vname)

# Output to csv
out_dict = {
    "video_name" : [],
    "frame" : [],
    "class" : [],
    "probability" : [],
    "bounding_box" : [],
    "center" : []
}

for vname in videos_names :
    print("-"*8, vname.replace("__FOLDER__", ""), "-"*8)
    images = [f for f in listdir(os.path.join(".", vname)) if
isfile(os.path.join(".", vname, f))]
    vpath = os.path.join(".", vname)
    count = 1
    for image in images :
        img_path = os.path.join(vpath, image)
        print("-"*2, "Detection, frame = ", count, "of", len(images) )
        detections =
detector.detectObjectsFromImage(input_image=img_path,
                                output_image_path="./dummy.jpg",

minimum_percentage_probability=30)
        count += 1
        for detection in detections :
            out_dict["video_name"].append(vname.replace("__FOLDER__",
""))
            out_dict["frame"].append(image.replace(".jpg", ""))
            out_dict["class"].append(detection["name"])

        out_dict["probability"].append(detection["percentage_probability"])
        out_dict["bounding_box"].append(detection["box_points"])
        out_dict["center"].append([(detection["box_points"][0] +
detection["box_points"][2])/2.0, (detection["box_points"][1] +
detection["box_points"][3])/2.0])

df = pd.DataFrame.from_dict(out_dict)
df.to_csv("data.csv") # writes to a csv file called 'data.csv'

```

. (continued).

## References

- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–49. <https://doi.org/10.1016/j.tele.2019.01.007>
- Ali, A., Negin, F. F., Bremond, F. F., & Thümmel, S. (2022). Video-based behavior understanding of children for objective diagnosis of autism. In *VISAPP 2022-17th International Conference on Computer Vision Theory and Applications*.
- Allen, S. M., & Mor, V. (1997). The prevalence and consequences of unmet need. Contrasts between older and younger adults with disability. *Medical Care*, 35(11), 1132–1148.
- American Psychiatric Association, D. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5). DC: American psychiatric association Washington.
- Anagnostopoulou, P., Alexandropoulou, V., Lorentzou, G., Lykothanasi, A., Ntaountaki, P., & Drigas, A. (2020). Artificial intelligence in autism assessment. *International Journal of Emerging Technologies in Learning*, 15(6), 95–107. <https://doi.org/10.3991/ijet.v15i06.11231>
- Atkinson, P. (2007). *Ethnography: Principles in practice*. Routledge.
- Baker, R. S., & Inventado, P. S. (2014). In J. A. Larusson, & B. White (Eds.), *Learning analytics: From research to practice* Educational data mining and learning analytics (pp. 61–75). Springer. [https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4).
- de Belen, R. A. J., Bednarz, T., Sowmya, A., & Del Favero, D. (2020). Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Translational Psychiatry*, 10(1). <https://doi.org/10.1038/s41398-020-01015-w>. Article 1.
- Benssassi, E. M., Gomez, J.-C., Boyd, L. E., Hayes, G. R., & Ye, J. (2018). Wearable assistive technologies for autism: Opportunities and challenges. *IEEE Pervasive Computing*, 17(2), 11–21.
- Bian, J. W., Zhan, H., Wang, N., Li, Z., Zhang, L., Shen, C., ... Reid, I. (2021). Unsupervised scale-consistent depth learning from video. *International Journal of Computer Vision*, 129(9), 2548–2564.
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. In *Office of Educational Technology*. US Department of Education.
- Bohr, A., & Memarzadeh, K. (2020). In A. Bohr, & K. Memarzadeh (Eds.), *Artificial intelligence in healthcare, The rise of artificial intelligence in healthcare applications* (pp. 25–60). Academic Press. <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>.
- Boraston, Z., & Blakemore, S. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology*, 581(3), 893–898.
- Bozgeyikli, L., Raj, A., Katkooi, S., & Alqasemi, R. (2018). A survey on virtual reality for individuals with autism spectrum disorder: Design considerations. *IEEE Transactions on Learning Technologies*, 11(2), 133–151. <https://doi.org/10.1109/TLT.2017.2739747>

- Brondino, N., Fusar-Poli, L., Miceli, E., Di Stefano, M., Damiani, S., Rocchetti, M., & Politi, P. (2019). Prevalence of medical comorbidities in adults with autism spectrum disorder. *Journal of General Internal Medicine*, 34(10), 1992–1994. <https://doi.org/10.1007/s11606-019-05071-x>
- Brown, H. M., Stahmer, A. C., Dwyer, P., & Rivera, S. (2021). Changing the story: How diagnosticians can support a neurodiversity perspective from the start. *Autism*, 25(5), 1171–1174. <https://doi.org/10.1177/13623613211001012>
- Buch, N., Velastin, S. A., & Orwell, J. (2011). A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3), 920–939.
- Burke, R. V., Allen, K. D., Howard, M. R., Downey, D., Matz, M. G., & Bowen, S. L. (2013). Tablet-based video modeling and prompting in the workplace for individuals with autism. *Journal of Vocational Rehabilitation*, 38(1), 1–14. <https://doi.org/10.3233/JVR-120616>
- Carmien, S., Dawe, M., Fischer, G., Gorman, A., Kintsch, A., & Sullivan, J. F., JR. (2005). Socio-technical environments supporting people with cognitive disabilities using public transportation. *ACM Transactions on Computer-Human Interaction*, 12(2), 233–262. <https://doi.org/10.1145/1067860.1067865>
- Chien, S. Y., Hwang, G. J., & Jong, M. S. Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions. *Computers & Education*, 146, Article 103751. <https://doi.org/10.1016/j.compedu.2019.103751>
- Davidesco, I., Glaser, N., Stevenson, I. H., & Dagan, O. (2023). Detecting fluctuations in student engagement and retention during video lectures using electroencephalography. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13330>
- Dikmen, M., & Burns, C. M. (2016). Autonomous driving in the real world: Experiences with tesla autopilot and summon. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 225–228).
- Dixon, D. R., Miyake, C. J., Nohely, K., Novack, M. N., & Granpeesheh, D. (2020). Evaluation of an immersive virtual reality safety training used to teach pedestrian skills to children with autism spectrum disorder. *Behavior Analysis in Practice*, 13, 631–640. <https://doi.org/10.1007/s40617-019-00401-1>
- Drimala, H., Baskow, I., Behnia, B., Roepke, S., & Dziobek, I. (2021). Imitation and recognition of facial emotions in autism: A computer vision approach. *Molecular Autism*, 12, 1–15.
- Fan, J., Zheng, P., & Li, S. (2022). Vision-based holistic scene understanding towards proactive human-robot collaboration. *Robotics and Computer-Integrated Manufacturing*, 75, Article 102304.
- Ferdig, R. E., & Kosko, K. W. (2020). Implementing 360 video to increase immersion, perceptual capacity, and teacher noticing. *TechTrends*, 64(6), 849–859. <https://doi.org/10.1007/s11528-020-00522-3>
- Forsyth, D. A., & Ponce, J. (2002). *Computer vision: A modern approach*. Prentice Hall Professional Technical Reference.
- Glaser, N., Newbutt, N., & Palmer, H. (2022a). Not perfect but good enough: A primer for creating spherical video-based virtual reality for autistic users. *Journal of Enabling Technologies*, 16(2), 115–123.
- Glaser, N., Newbutt, N., Palmer, H., & Schmidt, M. (2022b). Video-based virtual reality technology for autistic users: An emerging technology report. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-022-09594-x>
- Glaser, N., & Schmidt, M. (2022). Systematic literature review of virtual reality intervention design patterns for individuals with autism spectrum disorders. *International Journal of Human-Computer Interaction*, 38(8), 753–788. <https://doi.org/10.1080/10447318.2021.1970433>
- Glaser, N., Schmidt, M., Schmidt, C., Palmer, H., & Beck, D. (2021). In B. Hokansen, M. Exeter, A. Grinczewicz, M. Schmidt, & A. Tawfik (Eds.), *Intersections across disciplines: Interdisciplinarity and learning: The centrality of interdisciplinarity for overcoming design and development constraints of a multi-user virtual reality intervention for adults with autism: A design case*. New York, NY: Springer.
- Grossman, R. B., Mertens, J., & Zane, E. (2019). Perceptions of self and other: Social judgments and gaze patterns to videos of adolescents with and without autism spectrum disorder. *Autism*, 23(4), 846–857. <https://doi.org/10.1177/1362361318788071>
- Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, 42, 279–297. <https://doi.org/10.1016/j.neubiorev.2014.03.013>
- Han, J., & Kamber, M. (2006). Classification and prediction. *Data Mining: Concepts and Techniques*, 2006, 347–350.
- Hashemi, J., Spina, T. V., Tepper, M., Esler, A., Morellas, V., Papanikolopoulos, N., & Sapiro, G. (2012). A computer vision approach for the assessment of autism-related behavioral markers. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1–7).
- Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage Publications.
- Holmes, W., Luckin, R., & UCL Knowledge Lab. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson : UCL Knowledge Lab.
- Huang, H., Hwang, G.-J., & Chang, S.-C. (2021). Facilitating decision making in authentic contexts: An SVVR-based experiential flipped learning approach for professional training. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2021.2000435>
- Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers & Education: Artificial Intelligence*, 1, Article 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
- Keenan, E. G., Gurba, A. N., Mahaffey, B., Kappenberg, C. F., & Lerner, M. D. (2023). Leveling up dialectical behavior therapy for autistic individuals with emotion dysregulation: Clinical and personal insights. *Autism in Adulthood*. <https://doi.org/10.1089/aut.2022.0011>
- Knoblauch, H., Baer, A., Laurier, E., Petschke, S., & Schnettler, B. (2008, September). Visual analysis. New developments in the interpretative analysis of video and photography. In *Forum qualitative sozialforschung/Forum Qualitative Social Research*, 9(3).
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. <https://doi.org/10.1007/s10648-007-9049-0>
- Krogh, K., Bearman, M., & Nestel, D. (2015). Expert practice of video-assisted debriefing: An Australian qualitative study. *Clinical Simulation in Nursing*, 11(3), 180–187. <https://doi.org/10.1016/j.ecns.2015.01.003>
- Laurier, E. (2019). The panel show: Further experiments with graphic transcripts and vignettes. *Social Interaction. Video-Based Studies of Human Sociality*, 2(1). <https://doi.org/10.7146/si.v2i1.113968>
- Leo, M., Furnari, A., Medioni, G. G., Trivedi, M., & Farinella, G. M. (2018). Deep learning for assistive computer vision. In *Proceedings of the European Conference on Computer Vision (ECCV) workshops*.
- Longo, L. (2020). In A. Costa, L. Reis, & A. Moreira (Eds.), *Computer supported qualitative research. WCQR 2019. Advances in intelligent systems and computing: Vol. 1068. Empowering qualitative research methods in education with artificial intelligence*. (pp. 1–21). Cham: Springer. [https://doi.org/10.1007/978-3-030-31787-4\\_1](https://doi.org/10.1007/978-3-030-31787-4_1)
- Lu, J., Schmidt, M., Lee, M., & Huang, R. (2022). Methodological properties of usability studies on educational and learning technologies: A systematic review. *Educational Technology Research & Development*, 70, 1951–1992. <https://doi.org/10.1007/s11423-022-10152-6>
- Lyda, R., & Hamrock, J. (2007). Using entropy analysis to find encrypted and packed malware. *IEEE Security Privacy*, 5(2), 40–45. <https://doi.org/10.1109/MSP.2007.48>
- Maenner, M. J. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 Years—autism and developmental disabilities monitoring network, 11 sites, United States, 2020. *MMWR. Surveillance Summaries*, 72. <https://doi.org/10.15585/mmwr.mm7202a1>
- Marinoiu, E., Zafir, M., Olaru, V., & Sminchescu, C. (2018). 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2158–2167.
- Mayer, R. E. (2005). In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning: Principles of multimedia learning based on social cues: Personalization, voice, and image principles* (pp. 201–212). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.014>
- Miller, H. L., & Bugnariu, N. L. (2016). Level of immersion in virtual environments impacts the ability to assess and teach social skills in autism spectrum disorder. *Cyberpsychology, Behavior, and Social Networking*, 19(4), 246–256. <https://doi.org/10.1089/cyber.2014.0682>
- Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering*, 2021, 1–9. <https://doi.org/10.1155/2021/9975700>
- Nersesian, E., Spryszynski, A., & Lee, M. J. (2019). Integration of virtual reality in secondary STEM education. In *2019 IEEE integrated STEM education conference (ISEC)*, Princeton, NJ, USA, 2019 (pp. 83–90). <https://doi.org/10.1109/ISEC.2019.8882070>
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2020). *Deep learning vs. Traditional computer vision*.
- O'shea, T., & Self, J. (1986). *Learning and teaching with computers: The artificial intelligence revolution*. Prentice Hall Professional Technical Reference.
- Ouyang, L., & Wang, H. (2019). Vehicle target detection in complex scenes based on YOLOv3 algorithm. In *IOP Conference Series: Materials Science and Engineering*, 569. IOP Publishing, Article 052018.
- Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*, 132, Article 103940. <https://doi.org/10.1016/j.autcon.2021.103940>
- Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience*, 9(6), 610–632. <https://doi.org/10.1080/174770919.2014.934966>
- Parsons, S., & Cobb, S. (2016). State-of-the-art of virtual reality technologies for children on the autism spectrum. In *Technology and students with special educational needs* (pp. 77–88). Routledge.
- Patton, M. Q. (2015). *Qualitative research and evaluation methods*. Sage Publications.
- Paxinou, E., Kalles, D., Panagiotakopoulos, C. T., & Verykios, V. S. (2021). Analyzing sequence data with Markov chain models in scientific experiments. *SN Computer Science*, 2, 1–14. <https://doi.org/10.1007/s42979-021-00768-5>
- Pipkin, A., Kotecki, K., Hetzel, S., & Heiderscheid, B. (2016). Reliability of a qualitative video analysis for running. *Journal of Orthopaedic & Sports Physical Therapy*, 46(7), 556–561. <https://www.jospt.org/doi/10.2519/jospt.2016.6280>
- Pirker, J., & Dengel, A. (2021). The potential of 360 virtual reality videos and real VR for education—a literature review. *IEEE computer graphics and applications*, 41(4), 76–89. <https://doi.org/10.1109/mcg.2021.3067999>
- Prisille, C., & Ellerbrake, M. (2020). In D. Edler, C. Jenal, & O. Kühne (Eds.), *Modern Approaches to the Visualization of landscapes. RaumFragen: Stadt – region – landschaft/Virtual reality (VR) and geography education: Potentials of 360° 'experiences' in secondary schools*. Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-30956-5\\_18](https://doi.org/10.1007/978-3-658-30956-5_18)

- Ramaphosa, K. I. M., Zuva, T., & Kwuimi, R. (2018). *Educational data mining to improve learner performance in Gauteng primary schools* (pp. 1–6).
- Robertson, A. E., & Simmons, D. R. (2015). The sensory experiences of adults with autism spectrum disorder: A qualitative analysis. *Perception*, 44(5), 569–586. <https://doi.org/10.1068/p7833>
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146. <https://doi.org/10.1016/j.eswa.2006.04.005>
- Schmidt, M., & Glaser, N. (2021a). Investigating the usability and learner experience of a virtual reality adaptive skills intervention for adults with autism spectrum disorder. *Educational Technology Research & Development*. <https://doi.org/10.1007/s11423-021-10005-8>
- Schmidt, M., & Glaser, N. (2021b). Piloting an adaptive skills virtual reality intervention for adults with autism: Findings from user-centered formative design and evaluation. *Journal of Enabling Technologies*, 15(3), 137–158. <https://doi.org/10.1108/JET-09-2020-0037>
- Schmidt, M., Glaser, N., Schmidt, C., & Palmer, H. (2020a). In B. Hokanson, G. Clinton, A. A. Tawfik, A. Grincewicz, & M. Schmidt (Eds.), *Educational technology beyond content: A new focus for learning* Promoting acquisition and generalization of embodied skills in a 3D collaborative virtual learning environment for individuals severely impacted by autism. New York, NY: Springer.
- Schmidt, M., Lee, M., Francois, M., Lu, J., Huang, R., Cheng, L., & Weng, Y. (2023). Learning experience design of project PHoENIX: addressing the lack of autistic representation in extended reality design and development. *Journal of Formative Design in Learning*, 2023(7), 27–45. <https://doi.org/10.1007/s41686-023-00077-5>
- Schmidt, M., Newbutt, N., Schmidt, C., & Glaser, N. (2021). A process-model for minimizing adverse effects when using head mounted display-based virtual reality for individuals with autism. *Frontiers in Virtual Reality*, 2, Article 611740. <https://doi.org/10.3389/frvir.2021.611740>
- Schmidt, M., Tawfik, A. A., Jahnke, I., & Earnshaw, Y. (2020b). In M. Schmidt, A. A. Tawfik, I. Jahnke, & Y. Earnshaw (Eds.), *Learner and user experience research: An introduction for the field of learning design & technology* Methods of user centered design and evaluation for learning designers. EdTech Books. Retrieved from [https://www.edtechbooks.org/ux/ucd\\_methods\\_for\\_lx](https://www.edtechbooks.org/ux/ucd_methods_for_lx).
- Schwarz, M., Milan, A., Periyasamy, A. S., & Behnke, S. (2018). RGB-D object detection and semantic segmentation for autonomous manipulation in clutter. *The International Journal of Robotics Research*, 37(4–5), 437–451. <https://doi.org/10.1177/0278364917713117>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Siris, A., Jiao, J., Tam, G. K., Xie, X., & Lau, R. W. (2021). *Scene context-aware salient object detection*.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431–433. <https://doi.org/10.1111/bjop.12305>
- Slater, M., Lotto, B., Arnold, M. M., & Sánchez-Vives, M. V. (2009). *How we experience immersive virtual environments: The concept of presence and its measurement*. <http://diposit.ub.edu/dspace/handle/2445/49643>.
- Smith, B. (2018). Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative research in sport, exercise and health*, 10(1), 137–149.
- Snelson, C., & Hsu, Y. C. (2020). Educational 360-degree videos in virtual reality: A scoping review of the emerging research. *TechTrends*, 64, 404–412. <https://doi.org/10.1007/s11528-019-00474-3>, 2020.
- Song, D. Y., Kim, S. Y., Bong, G., Kim, J. M., & Yoo, H. J. (2019). The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: A literature review. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, 30(4), 145. <https://doi.org/10.5765/jkacap.190027>
- Stockman, G., & Shapiro, L. G. (2001). *Computer vision*. Prentice Hall PTR.
- Talton, J., Yang, L., Kumar, R., Lim, M., Goodman, N., & Mëch, R. (2012). *Learning design patterns with bayesian grammar induction*.
- Tawfik, A. A., Law, V., Ge, X., Xing, W., & Kim, K. (2018). The effect of sustained vs. Faded scaffolding on students' argumentation in ill-structured problem solving. *Computers in Human Behavior*, 87, 436–449. <https://doi.org/10.1016/j.chb.2018.01.035>
- Taylor, J. L., Smith, L. E., & Mailick, M. R. (2014). Engagement in vocational activities promotes behavioral development for adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44, 1447–1460.
- Tschang, F. T., & Almirall, E. (2021). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, 35(4), 642–659. <https://doi.org/10.5465/amp.2019.0062>
- Venables, M. (2019). An overview of computer vision. Towards Data Science. Retrieved from <https://towardsdatascience.com/an-overview-of-computer-vision-1f75c2ab1b66>.
- Yaqoob, A., Bi, T., & Muntean, G.-M. (2020). A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials*, 22(4), 2801–2838.
- Yildirim, G., Elban, M., & Yildirim, S. (2018). Analysis of use of virtual reality technologies in history education: A case study. *Asian Journal of Education and Training*, 4(2), 62–69. <https://doi.org/10.20448/journal.522.2018.42.62.69>
- Yoganathan, S., Finch, D. A., Parkin, E., & Pollard, J. (2018). 360 virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *International Journal of Surgery*, 54, 24–27. <https://doi.org/10.1016/j.ijssu.2018.04.002>
- Zhang, M., Ding, H., Naumceska, M., & Zhang, Y. (2022). Virtual reality technology as an educational and intervention tool for children with autism spectrum disorder: Current perspectives and future directions. *Behavioral Sciences*, 12(5), 138. <https://doi.org/10.3390/bs12050138>
- Zhao, L., & Li, S. (2020). Object detection algorithm based on improved YOLOv3. *Electronics*, 9(3), 537. <https://doi.org/10.3390/electronics9030537>
- Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education*, 136, 34–48. <https://doi.org/10.1016/j.compedu.2019.03.005>
- Zhuang, Y., Wu, F., Chen, C., & Pan, Y. (2017). Challenges and opportunities: From big data to knowledge in AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, 18, 3–14.