Integrating Virtual Reality and Artificial Intelligence for Mass Casualty Incident Training

Peng Xia

Department of Engineering, Computer and Mathematical Science

Auckland University of Technology

This dissertation is submitted for the degree of

Doctor of Philosophy

September 2025

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Peng Xia

September 2025

Acknowledgements

I would like to express my heartfelt gratitude to my primary supervisor, Dr. Ji Ruan, and my secondary supervisor, Prof. Dave Parry, for their steadfast support and guidance throughout this journey. Despite challenges such as the COVID-19 pandemic, this experience has been both rewarding and joyful. Your invaluable mentorship and unwavering encouragement have been instrumental in helping me achieve this significant milestone in my life. I would also like to extend my sincere thanks to Dr. Stephen Aiello for his full support of my research. Your recommendations for the VR learning tool and your efforts in arranging participants for the experiment were crucial to the success of this work. Without your assistance, this achievement would not have been possible.

I would like to extend my heartfelt gratitude to my family. To my parents, thank you for your unwavering support and encouragement. A special thanks to my aunt, Ms. Dan (Julia) Huang, for helping me settle in New Zealand and being a steadfast pillar of support throughout my life here. I am also deeply grateful to my grandmother. Although she is now struggling with Alzheimer's and lost in her own world, unable to remember me, the wisdom and love you imparted to me in my childhood continue to guide and inspire me. Lastly, I want to thank my grandfather. I hope you are watching from above, feeling proud and happy as the first overseas PhD in our family is completed, a milestone that owes much to your love and guidance.

I am deeply grateful to my good friends who have stood by me throughout this journey. My sincere thanks to my former and current colleague, Dr. Wenwang Pang, and his wife, Mrs. Zeng Yuan, for their support in both academic and professional matters. I would like to thank all current and former members of the INKFAN band, as well as our guest performers, including Yuting Gao, Shuailin Zhao, Shihong Zheng, Jing Gan, Yuran Zou, Zhijun Guo, Jianan Yang, and Derek Li.

I would also like to thank Liu Yang for standing by me during the challenging times of the COVID-19 pandemic, especially when New Zealand closed its borders and I was stuck in China. To Yue Wu, thank you for your support during my PhD journey, and to Ms. Miaoling Wei, your unwavering support means so much to me—thank you.

Finally, I want to take a moment to thank myself. Only I truly understand how challenging this journey has been. The experiences and stories from 20 years ago have shaped the life I am living today. I hope the years ahead bring happiness, a healthy body, the freedom to make your own decisions, and all the resources needed to pursue your personal interests.

You have come so far—may the journey ahead be just as fulfilling.

Abstract

This thesis explores the integration of Virtual Reality (VR) and Artificial Intelligence (AI) for enhancing pre-hospital triage training in the context of Mass Casualty Incidents (MCIs). Addressing the challenges posed by traditional training methods, which often involve significant costs and logistical complexities due to the infrequent nature of MCIs, this research introduces an innovative VR learning tool designed to realistically simulate emergency scenarios. The primary objectives of this study were to develop the VR learning tool using a Design Science Research methodology and to detail the advanced data analysis methodologies employed in addressing the research questions (RQ1–RQ3.3).

By leveraging VR and AI technologies, the tool aimed to provide emergency healthcare professionals with a dynamic and immersive training environment, enabling them to practice and refine their triage skills without the constraints of traditional physical simulations. A significant portion of the research was dedicated to addressing the limitations of current VR interaction methods by prototyping more interactive and lifelike user experiences through advanced VR controllers.

This enhancement allowed for deeper and more comprehensive data collection, including audio metrics, thereby facilitating a nuanced understanding of trainees' performance and engagement within the simulated environment.

The development process of the VR learning tool, characterized by the integration of AI techniques and statistical methods, reflected the study's commitment to both innovation and effective assessment of training outcomes. The experimental phase outlined the preparation,

execution, and ethical considerations of implementing the VR training, providing insights into the tool's potential to quantitatively and qualitatively evaluate emergency response skills.

The findings from this study indicated that VR technology can be a supplementary tool in emergency healthcare training, particularly in scenarios involving mass casualty incidents. The analysis of training sessions with 10 participants showed variability in performance during simulated car crash and earthquake scenarios: some participants were quick but less accurate, while others were slower yet more precise, reflecting diversity in emergency response approaches. Survey results showed that participants, predominantly aged 18–24 with varying levels of experience, found the VR training highly immersive and engaging, although some reported physical discomfort, highlighting the need for ergonomic improvements. Additionally, AI-driven analysis of speech data demonstrated improved consistency and accuracy in participants' communication over time, emphasizing the importance of clear communication in emergencies.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The growing demands in healthcare, highlighted by the United Nations' projection of a significant population increase by 2025 [148], with a substantial portion being individuals aged 65 or older, underscore the urgency of emergency healthcare. This demographic shift, coupled with the challenges posed by global health crises like the COVID-19 pandemic, calls for innovative training and preparedness approaches in healthcare. Virtual Reality (VR) and Artificial Intelligence (AI) technologies present themselves as effective, cost-efficient solutions, particularly in addressing the complexities of Mass Casualty Incidents (MCIs).

MCIs, often resulting from natural or human-made disasters, require emergency personnel to make swift, critical decisions under extreme stress [103]. The infrequent nature of such events complicates the effectiveness of traditional training methods, which are generally expensive and logistically cumbersome. VR technology can provide simulations with effectiveness similar to live simulations but at a lower cost for participant paramedic training [107]. Currently, some researchers are focusing on leveraging the capabilities of VR to enhance emergency healthcare training in different scenarios, such as car accidents (e.g., [86, 12]) and natural disasters (e.g., [51, 72]). Moreover, as the development of hardware and software progresses, AI is beginning to assume a more critical role in emergency healthcare. The primary focus of most research combining AI and healthcare revolves around computerassisted triage and optimizing resources in the Emergency Department (ED). For instance, research efforts like those in the works of [17, 28, 30] are directed towards employing AI to enhance resource distribution and reduce wait times in the ED. For example, [28] specifically aims to optimize staff configurations to decrease the duration of casualties' stays in the ED. Conversely, some studies concentrate on applying AI for automatic casualty triaging or to assist in making triage decisions (e.g., [84, 42, 56, 55]). Remaining research pertains to predictions within the ED, including casualty conditions [125] and waiting times [79].

This project aims to bridge these gaps by developing methodologies and applications that better incorporate AI and VR into MCI training, focusing on objective evaluation techniques and extending the application of AI to pre-hospital triage training. Despite advances in the field, notable gaps remain in the integration of VR, AI, and MCI training. Firstly, the current approaches to VR MCI training use questionnaires or interviews (e.g., [95, 90, 71, 49]), and they are subjective and lack the use of more objective sensor data for evaluating participant performance. The predominant reliance on questionnaires for gathering user feedback and assessing participant outcomes introduces subjective biases, as individuals may interpret and engage with VR differently. Secondly, the majority of research on AI within emergency healthcare is concentrated on in-hospital triage, overlooking pre-hospital scenarios and MCI training specifically (e.g., [48, 159, 144]). Lastly, the engagement of AI in current research is focused on vital sign detection and emergency healthcare resource optimization. More advanced AI techniques such as speech and image recognition are not yet involved in pre-hospital MCI training research.

This project introduces a VR learning tool that integrates AI, tailored specifically for casualty triage simulation in MCIs, offering an immersive and realistic experience. This tool stands as a practical alternative to traditional training methods, overcoming many of the financial and logistical barriers they present.

The project's main objectives are to develop the tool's innovative design and functionalities and to establish robust data collection and analysis methodologies. The VR learning tool brings emergency responders into a virtual environment that simulates real-life MCI scenarios, providing a risk-free platform for training and evaluation [6, 37, 72, 15]. The interactive VR component allows responders to develop their skills in casualty assessment, triage, and critical decision-making within a controlled environment, consistent with previous findings that VR training can enhance clinical decision-making under pressure [96, 152, 1]. AI and Machine Learning (ML) methods are integral to evaluating performance, analyzing diverse data types collected during training, such as timespan, audio speech, and decision-making processes [31, 144, 33]. This approach delivers detailed insights into performance, identifying strengths and areas needing improvement in responders' triage and casualty management strategies, building on earlier calls for more objective and data-driven evaluation of emergency training outcomes [78]. The VR learning tool's effectiveness in evaluating the triage performance of paramedic students during MCIs was tested through extensive training sessions in experiments, reflecting similar approaches adopted in other VR-based disaster preparedness studies [1, 96]. These training sessions are designed to ensure the VR learning tool's applicability across a range of MCI scenarios, emphasizing its potential as a valuable training resource in emergency preparedness and response [152, 15].

To guide and structure the research, the following key research questions are proposed:

1. **What are the essential components required to build an effective VR learning tool?**

   This question focuses on identifying the necessary development and system components for the VR learning tool, including both hardware and software aspects that contribute to effective training. It aims to ensure a comprehensive understanding of the technical and functional requirements that will make the VR training environment successful.

2. **How can AI be integrated effectively into a VR learning tool to enhance data analysis?**

   This question explores the integration of AI into the VR training system, particularly focusing on how AI can be used to improve data analysis. By leveraging AI methods like visual and voice recognition, the goal is to enrich the training process and provide deeper insights into performance and outcomes.

3. **How can the training effectiveness of the VR learning tool be evaluated and improved?**

   This question is centered around evaluating and enhancing the training efficiency of the VR learning tool. It aims to explore how factors such as accuracy and the duration of training sessions can be measured, with the objective of refining the system to provide more effective and time-efficient training experiences.

   To delve deeper into this evaluation process, the following sub-questions further explore specific methods and aspects of how training outcomes and user interactions in VR environments can be assessed:

   3.1 **How can data from VR learning tool be used to assess the competence of participants in MCI scenarios?**

   This question focuses on the utilization of VR to assess participant competence in MCI training scenarios. By analyzing metrics such as task completion time and AI-driven performance evaluations, this question seeks to establish an objective method for measuring the skill levels and preparedness of participants.

   3.2 **Can a VR environment convince participants to behave in a way that will allow their performance to be assessed?**

   This question investigates the extent to which the VR environment can influence participant behavior to enable accurate performance assessments. It aims to understand if the immersive nature of VR can prompt participants to engage with the simulation realistically enough to allow for valid evaluations of their performance.

3.3 **How well do participants accept VR technology for triage training in MCIs?**

This question examines how well participants accept and engage with VR technology in the context of triage training for MCIs. It seeks to explore whether the immersive and interactive features of VR can foster realistic participation in training exercises, which in turn can lead to reliable assessments of participant skills.

Based on the above-mentioned research questions, the research unfolds in several phases:

- Defining Training Tasks: Conducting an extensive literature review to determine the training tasks, ensuring the VR learning tool aligns with contemporary emergency healthcare training practices.
- Constructing a Data Collection Platform: Developing a platform inclusive of both hardware and software. Hardware components include VR head-mounted displays, controllers, sensors, and cameras, while the software comprises various VR scenarios simulating major incidents.
- Conducting Experiments: Engaging participants with the VR learning tool to collect data such as timespan, audio, and video streams, capturing interactions with the training scenarios.
- Analysing Data: Analyzing the data using both ML and traditional statistical methods to evaluate the tool's effectiveness.

In the literature review chapter, we undertakes an extensive exploration of the integration of VR and AI in emergency healthcare training, focusing on healthcare fields, especially MCIs.

To achieve a comprehensive understanding of the field, three distinct rounds of literature review are conducted, employing rigorous methodologies to ensure the depth and breadth of the research. The first round of the literature review encompasses a broad examination of the existing knowledge and applications of VR and AI technologies. This initial phase is instrumental in establishing a foundational understanding of the technological landscape and its potential applications in various domains, including healthcare. By assessing the broader context, this round of review identifies key areas where VR and AI intersect with emergency healthcare training, particularly highlighting the emerging trends and technological advancements. Subsequently, to refine the scope and pinpoint specific research gaps, a second round of literature review is executed. This phase employs the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, a systematic approach that ensures thoroughness and accuracy in literature analysis. The PRISMA methodology guides the selection and scrutiny of academic works, leading to a more targeted and relevant collection of literature. This methodical approach is particularly beneficial in identifying the most pertinent and recent research findings, which directly inform the objectives and direction of the thesis. Finally, a third round of literature review is conducted to gather the latest research findings in the relevant field, using the same search criteria as in the second round.

In synthesizing these findings, the literature review sets the stage for the research. It highlights the need for enhanced interaction methods in VR training and the integration of AI for objective evaluation and analysis. Building upon the existing studies, the research aims to develop an advanced VR learning tool for emergency healthcare training, evaluated using AI techniques to address the identified gaps. This comprehensive literature review thus provides the necessary context and justification for the research objectives and methodologies adopted in the thesis.

In the methodology chapter, we employs a Design Science Research (DSR) methodology to develop and evaluate the VR learning tool integrated with AI for emergency healthcare training, focusing specifically on MCIs. DSR is a systematic approach that emphasizes the creation and evaluation of innovative artifacts to address real-world problems. In this study, the VR learning tool serves as the artifact, designed to enhance decision-making, triage skills, and communication in MCI scenarios. The methodology is structured into three main phases: Problem Identification, Research Design and Development, and Training Experiment and Evaluation.

The first phase involves Problem Identification, which mentions the limitation of current research and presents the research question identified by the research. These methods are often limited by high costs, infrequency, and the unpredictability of real-life emergencies. As a result, emergency healthcare professionals may lack the necessary hands-on experience to make rapid, accurate decisions during real-world incidents. To overcome these limitations, this research proposes a VR learning tool that integrates AI to simulate realistic MCI scenarios, offering an immersive and scalable training solution. The research is guided by three key questions: What are the essential components required to build an effective VR learning tool? How can AI be effectively integrated to enhance data analysis? How can the tool's training effectiveness be evaluated and improved? These questions form the foundation for developing a VR-based solution that addresses the need for more effective MCI training.

The second phase is Research Design and Development, which consist of two subsections: the development of tools and the implementation of the VR learning tool. The first subsection focuses on configuring the necessary VR and AI development tools. Using VR modeling software and the Unity Engine, detailed and interactive environments that mimic real-world emergency conditions are created. In parallel, AI tools such as Whisper for speech recognition and BERT for semantic analysis are used to develop intelligent evaluation algorithms, ensuring the VR environment is both immersive and responsive. These tools allow the system to monitor participant interactions and provide real-time feedback on decision-making and communication during the training sessions. This phase directly addresses the research questions related to the development of essential VR components and the integration of AI to enhance the data analysis process.

In the VR Learning Tool Implementation subsection, the focus shifts to the hardware and software integration necessary for an immersive VR experience. This stage includes the setup of VR headsets, motion sensors, and other equipment to facilitate user engagement in realistic MCI scenarios. The VR software is designed to track user performance, including decision-making accuracy and task completion times, while AI-driven models analyze the collected data to provide insights into participant performance. These data insights are crucial for evaluating how effectively the VR tool improves decision-making and communication skills, addressing the research question concerning the tool's effectiveness. By implementing these tools, the research ensures that the learning tool not only provides an immersive experience but also generates valuable data for performance assessment.

The third phase involves the VR Training Experiment and Evaluation, where the developed VR learning tool is tested and evaluated in a practical setting. Ethical approvals were obtained, and participants were recruited to engage in VR-based MCI scenarios. During the training sessions, motion tracking, verbal communication, and decision-making data were collected to assess how well participants responded to the simulated scenarios. This phase directly evaluates the tool's ability to simulate real-world conditions and how participants interact with the VR environment, thereby addressing the research questions regarding training effectiveness. The data is analyzed using both quantitative and qualitative methods, with AI models assessing key metrics such as task completion times and the accuracy of triage decisions. Additionally, semantic analysis of the METHANE reports generated during the training allows for the evaluation of communication skills, contributing to a comprehensive understanding of the tool's impact.

The final step in this methodology involves Evaluation as well as Improvement and through iterative feedback. Based on the collected feedback, the VR learning tool keep refinement to improve its user experience and effectiveness.

In summary, this research follows a DSR methodology to develop a VR learning tool that addresses the critical need for more effective MCI training. By systematically identifying the problem, developing a solution, and testing it through rigorous evaluation, the research contributes a scalable and immersive tool that improves decision-making, triage skills, and communication in emergency healthcare settings. The iterative nature of this approach ensures continuous refinement, making the VR tool a valuable asset in modern healthcare training.

In the results chapter, the thesis presents an in-depth analysis of data collected from VR training sessions focusing on pre-hospital triage in emergency scenarios, involving ten participants. This analysis is structured into three comprehensive sections: Sensor Data Results, Survey Results, and Speech Recognition and Semantic Similarity Evaluation, each offering distinct insights into the effectiveness and areas for improvement of the VR learning tool. In the **Sensor Data Results** section, the results of participants' performance across two simulated scenarios—a car crash and an earthquake—are presented. The detailed analysis encompasses metrics such as the time taken for triage decisions, the accuracy of triage categorization, and the sequence of actions taken by participants. The findings reveal a notable variation in participants' approaches, with some prioritizing speed over accuracy and others adopting a more methodical but slower approach. This diversity in performance not only reflects the variability encountered in real-world emergency responses but also underscores the varying strategies and skills participants may utilize in such high-pressure situations. The **Survey Results** section gathers insights from two surveys conducted before and after the VR training sessions. The pre-training survey provides background information on the participants, who have varied levels of MCI experience and previous exposure to

VR technology. The post-training survey, consisting of 16 questions, offers feedback on the VR training experience. Participants report high levels of immersion and engagement, suggesting the VR learning tool's effectiveness in simulating real-life emergency scenarios. However, reports of physical discomfort, such as nausea and dizziness, are noted, indicating a need for ergonomic improvements in VR technology. Overall, the feedback is positive regarding the potential of VR in enhancing emergency medical education, with a majority expressing optimism about its integration into paramedic training. The **Speech Recognition and Semantic Similarity Evaluation** section utilizes advanced AI models to process and evaluate the audio data collected during the training. The analysis using the Bidirectional Encoder Representations from Transformers (BERT) model, a state-of-the-art natural language processing framework, reveals improvements in the consistency and accuracy of METHANE reports over time. METHANE is an acronym used in emergency management, representing Major incident declared, Exact location, Type of incident, Hazards, Access, Number of casualties, and Emergency services required. The findings indicate a progressive enhancement in structured communication among participants. Additionally, the evaluation conducted using the Generative Pre-trained Transformer (GPT) model, another advanced AI tool, assesses the coverage of METHANE components. This evaluation shows a noticeable advancement in participants' ability to provide comprehensive reports, moving from initial inadequacies to more detailed and accurate descriptions. These findings underscore the critical importance of clear and structured communication in emergency situations and highlight the potential of AI tools in refining emergency response training methodologies.

The discussion chapter addresses the research questions proposed in the previous chapter, drawing on findings from the results and feedback gathered during the VR training experiments. It explores the essential components of the VR learning tool and the role of AI in enhancing data analysis, while also evaluating training effectiveness, participant competence, and the tool's overall impact. The chapter highlights how the VR environment encouraged participants to engage in decision-making and communication, though it notes areas needing refinement, such as advancing AI models to process more complex scenarios. Limitations, including the restricted participant recruitment scope, small sample size, and the limited realism of casualty models and interaction methods, are discussed in detail. Additionally, the evaluation on audio data indicating a need for more comprehensive data collection in future studies. The chapter outlines several potential pathways for future research to enhance VR-based triage training tools, including developing more sophisticated vital sign models, improving decision-making mechanisms, expanding the range of scenarios, and refining casualty animations and interactive objects. Through these improvements, future research can optimize the tool's technological and educational effectiveness, better preparing healthcare professionals for the complexities of emergency medical care.

This thesis makes contributions to the field of emergency healthcare training by integrating advanced VR and AI technologies to enhance triage training for MCIs. The research introduces an innovative VR learning tool, enabling emergency professionals and paramedic students to practice triage and decision-making in a controlled, risk-free setting, addressing the limitations of traditional training methods such as high costs and logistical challenges. AI is integrated to objectively evaluate participant performance through data analysis, including timespan, speech, and decision-making processes, which aims to achieve the subjectivity in VR training evaluations. Additionally, the speech recognition and analysis tool enhances communication skills under stress, a critical component in MCI scenarios. Through experimental trials, the study demonstrates the VR learning tool's effectiveness in improving triage accuracy and decision-making speed. The thesis fills a gap in emergency healthcare training by focusing on pre-hospital MCI triage, proposing a robust framework for developing and evaluating VR-based MCI training tools, offering a scalable, practical solution for improving emergency preparedness. Overall, this research extends the application of VR and AI in emergency training, providing valuable contributions to both academic knowledge and practical emergency response applications.

The findings and contributions of this research have been presented through multiple publications. These include a conference paper presented at the 2nd Information Systems for Crisis Response and Management Asia Pacific Conference (ISCRAM 2022) [158], which discusses the integration of AI and VR in MCI training. Another paper, published in the Australasian Computer Science Week, further elaborates on the tool's design and its implications for MCI training [124].

1.1 Summary

The introduction chapter of this thesis sets the stage for an in-depth investigation into the use of VR and AI technologies in emergency healthcare training, with a special focus on MCIs. The research is motivated by the growing pressures on healthcare systems globally, driven by demographic changes and amplified by challenges like the COVID-19 pandemic, which have underscored the urgent need for innovative solutions to meet these increasing demands. The chapter emphasizes the need for innovative training solutions to prepare emergency personnel for MCIs, proposing a VR learning tool integrated with AI as a cost-effective and practical alternative to traditional training methods.

This innovative tool aims to simulate real-life MCI scenarios, providing an immersive environment for emergency responders to practice casualty triage, assessment, and critical decision-making. The integration of AI and Machine Learning (ML) techniques plays a crucial role in evaluating participants' performance, leveraging data from various sources to offer detailed insights into their training outcomes.

To structure the research, the Chapter 1 Introduction presents key research questions concerning the development and effectiveness of the VR learning tool, the integration of AI in VR-based healthcare training, and methods for evaluating and improving training efficiency.

These questions guide the research through several phases, including defining training tasks, constructing a data collection platform, conducting experiments, and analyzing the collected data.

The Chapter 2 Literature Review follows, providing a comprehensive exploration of VR and AI applications in emergency healthcare training, particularly for MCIs. Through three rounds of literature review, the chapter identifies key trends, technological advancements, and gaps in current research, setting a solid foundation for the project. The review highlights the potential of VR and AI to enhance training outcomes but also notes existing limitations in interaction methods and evaluation techniques.

The Chapter 3 Methodology details the systematic approach taken to develop and implement the VR learning tool, describing the development of VR and AI tools, the creation and application of the learning tool, and the execution of training experiments.

In the Chapter 4 Results, an extensive analysis of the data collected from VR training sessions is presented, focusing on sensor data results, survey findings, and speech recognition and semantic similarity evaluations. This analysis uncovers variations in participants' performance, levels of immersion and engagement, and the effectiveness of structured communication in training scenarios. These findings demonstrate the VR learning tool's potential to simulate emergency scenarios realistically and identify areas for improvement in VR interface design and communication training.

The Chapter 5 Discussion addresses the research's limitations and suggests avenues for future studies to enhance the VR-based triage training tool's effectiveness and applicability. By acknowledging these challenges and proposing targeted research directions, the thesis contributes to the ongoing development of more effective, immersive, and realistic VR training environments. This work aims to improve emergency medical personnel's preparedness and response capabilities, ultimately enhancing casualty outcomes in emergency situations.

Finally, the conclusion chapter summarizes the research output.

Chapter 2

Literature Review

In this research, three distinct rounds of literature review were conducted with the objective of refining the research scope, identifying research gaps, and accessing the most recent research findings. The initial round of literature review encompassed a comprehensive examination of the broader landscape within the domains of VR and AI technology, as well as their various applications. Subsequently, the second round of literature review was executed using the PRISMA methodology to systematically identify and scrutinize a more targeted selection of academic works. The last round of literature review covered the latest works in the relevant research fields. To clarify, the three rounds of literature review were undertaken for different purposes. The first round of literature review was completed before the COVID-19 pandemic, which provided a broad scoping review but did not fully follow the PRISMA methodology. The second round therefore reexamined the relavent topics systematically under PRISMA guidelines and extended the database coverage (e.g., Scopus, IEEE, ACM, MedLine, Springer), which had not been included in the first round literature review. Finally, the third round literature review is an update to capture the most recent advances in VR and AI applications in emergency healthcare, ensuring that the thesis reflects the latest developments.

2.1 First Round Literature Review

2.1.1 The VR technology

Eric Howlett [119] developed the Large Expanse, Extra Perspective (LEEP) optical system in 1979, revolutionizing the field of virtual reality. This system was pioneering in creating a stereoscopic image with a wide field of view, essential for a convincing virtual environment. It laid the groundwork for many of the VR technologies seen today, including the headsets used in modern VR devices. The LEEP system's influence extends beyond just the realm of gaming; it has applications in various fields such as training simulations and educational tools.

In the early 1990s, VR became a popular concept of interactive computer-generated experience taking place within a simulated environment. In [106], Milgram and Kishino introduced the concept of the 'Virtuality Continuum.' This continuum spans from the real environment at one end to a fully virtual environment at the other. The authors also discussed the concepts of augmented reality (AR) and augmented virtuality (AV), which blend real and virtual elements in different proportions. Their work has been fundamental in understanding and categorizing the various types of Mixed Reality (MR) experiences and has had a significant impact on how AR and VR technologies are designed and interacted with. Most current popular VR devices use 'head-mounted displays' (HMD) and fit into the 'virtuality continuum.' For example, Oculus Rift, released by Facebook in 2015 [118], represents a significant milestone in virtual reality, offering a fully immersive experience that places users in a completely virtual environment. On the other end, Microsoft's Hololens [68], also released in 2015, represents a major advancement in augmented reality technology, overlaying digital information onto the real world. Both devices have broad applications, from gaming and entertainment to practical uses in business and education, illustrating the diverse potential of VR and AR technologies. The applicable fields of this hardware include general entertainment, gaming, business, manufacturing, quality assurance, and telemedicine [65].

Based on these classifications, Billinghurst and Kato [16] developed two prototype mixed reality interfaces: 'WearCom' and 'Collaborative Web Space.' These interfaces were groundbreaking in demonstrating how MR could be used to enhance interaction in specific areas. Their work has been influential in exploring the practical applications of MR technology, paving the way for future developments in the field, particularly in collaborative and interactive environments.

Boletsis [20] explored VR locomotion through a systematic literature review investigating empirical studies of VR locomotion techniques from 2014–2017. The review identified various locomotion methods in virtual environments, such as real-walking, walking-in-place, and joystick-based movement. This research was crucial in understanding how different locomotion techniques affect user experience in VR, contributing to the development of more intuitive and immersive VR systems. Boletsis' work also proposed a typology for these locomotion methods, helping to categorize and understand their usages.

2.1.2 Previous work of healthcare with VR

In the past 20 years, some researchers have explored the possibility of applying VR technology in healthcare, mainly in the field of healthcare training and mental healthcare treatment. Besides healthcare, other research areas such as safety training and disaster simulation have also integrated VR technology to increase training efficiency.

Applying VR in healthcare training

Most early work applying VR in healthcare training used customized hardware. Due to the limited hardware performance, researchers primarily explored the feasibility of applying VR in this field. With the development of commercial-level VR hardware since 2015 [118, 68], more research has focused on solving specific problems in the healthcare training field.

- Emergency response training

  Gout et al. [59] introduced a VR-based training system for disaster medicine on the Second Life platform. This system's adaptability and versatility in creating scenarios tailored to disaster medicine training needs were underlined as its key strengths, enhancing healthcare providers' readiness and response capabilities in disaster situations. These studies collectively underscore the transformative potential of VR in emergency response and disaster preparedness training, advocating for overcoming current limitations to fully leverage VR's benefits.

- Healthcare skill training

  Mantovani et al.'s study in 2003 laid the groundwork for understanding the innovative applications of VR technology in training healthcare professionals [100]. By emphasizing VR's ability to simulate complex, real-world scenarios in a virtual setting, they highlighted how VR technology could make training more accessible, diverse, and engaging. The study pointed out VR's unique advantages in enhancing learning through increased motivation and active engagement, alongside its flexibility and programmability, which allow for a broad spectrum of training options. However, the authors also noted the potential side effects in immersive VR environments, indicating the necessity for further research. Building on this, Gaba's research [53] underscored the significance of realistic scenarios, including both mannequin-based and computer simulations, in developing healthcare professionals' skills. He detailed the benefits of simulations in providing a secure learning environment and improving patient outcomes, advocating for the integration of simulation technologies into medical education and healthcare systems. Nunes and Costa's 2008 research [117] discussed the emergence of VR as a transformative tool in medical education, surgical training, rehabilitation, and the development of interactive 3D applications for enhancing healthcare delivery. The discussion included the challenges faced in adopting VR technologies, such as the need for interdisciplinary research, the development of specific hardware and software solutions, and the importance of overcoming ethical and practical obstacles. Their research emphasized the potential of VR to revolutionize medical training and patient care by providing immersive, interactive environments that facilitate learning, improve treatment outcomes, and support the rehabilitation process. They also argued that further research and investment in VR technology are crucial for realizing its full potential in healthcare, underscoring the need for collaborative efforts among healthcare professionals, computer scientists, and policymakers to address the grand challenges identified by the Brazilian Computer Society (SBC) and ensure the successful integration of VR into health practices.

On the practical side, Aggarwal et al. [3] successfully developed an evidence-based VR laparoscopic training curriculum tailored for novice laparoscopic surgeons. By structuring a curriculum that involved both experienced and novice surgeons in completing a series of tasks across different levels of difficulty in virtual sessions, the study demonstrated VR's effectiveness in ensuring that junior trainees acquire essential skills before entering the operating room. Larsen and his team's research [88] further validated the impact of VR on the learning curve of laparoscopic training, showing significant improvements in the performance of novice surgeons. This study not only confirmed VR's role in enhancing surgical skills but also addressed the critical issue of skill transfer from virtual environments to real-world surgical practice, highlighting VR's potential to bridge this gap. Additionally, Caroline et al.'s 2018 discussion [50] on the use of VR in communication skills training, particularly relevant to medical ethics, based on Pan et al.'s research [120], shed light on VR's utility in scenarios involving complex interpersonal dynamics, such as managing patient demands for antibiotics. Their findings suggest that VR can be a powerful tool for inducing behavior change and enhancing doctors' self-awareness, thereby improving therapeutic interactions.

- Healthcare education

  There are some practical studies that apply VR in the healthcare education field. Kilmon et al. [83] explored the use of virtual reality as an educational strategy in nursing. The focus was on improving the speed and accuracy of nurses' responses in emergency situations. The team aimed to develop compact and user-friendly VR platforms to make the software more accessible and easier for nurses to use in their training. This approach reflects a growing recognition of the need for innovative and effective training tools in healthcare education, especially for preparing nurses for high-pressure emergency scenarios. The move towards more accessible VR platforms signifies a shift in training methods, potentially leading to broader adoption and enhanced training outcomes in nursing education.

  Izard and Méndez's research [74] showcased the potential of VR in teaching human anatomy, a critical component of medical education. Their VR system enabled users to immerse themselves in a detailed virtual representation of the human bone structure, including the ability to visualize the cranium from both inside and outside. The researchers noted that while various VR applications exist for exploring the human body, many are limited by their reliance on 360-degree camera recordings, which restrict movement and user interaction. Therefore, their system focused on enhancing the interactive capabilities between the user and the software, offering a more engaging and educational experience. This approach to anatomy education using VR represents a significant advancement in medical training, providing a more interactive and immersive learning environment.

Applying VR in healthcare treatment

Parsons and Rizzo's [122] comprehensive review marked a significant milestone in the application of VR technology in mental health treatment. Analyzing 21 studies involving 300 subjects, they provided substantial evidence of virtual reality exposure therapy's (VRET) effectiveness in treating anxiety and specific phobias across various affective domains. This study not only underscored VR's therapeutic potential but also opened new avenues for using technology in mental health treatments, suggesting a technologically advanced approach to managing anxiety and phobias. Building on Parsons and Rizzo's work, Meyerbroker and Emmelkamp in 2010 systematically reviewed VRET's use in treating anxiety disorders [105], finding significant support for its efficacy, particularly in fear of flying and acrophobia. Their research pointed towards the exploration of remote VR treatments via the internet, signaling a step forward in making VR-based treatments more accessible for anxiety disorders. Cho et al. [34] explored the potential of VR tasks combined with physiological monitoring to detect stress levels, employing metrics such as heart rate variability and skin conductance. Their research demonstrated the feasibility of using VR alongside physiological signals to develop devices for accurate stress level detection, opening new pathways in healthcare and wellness monitoring. Rizzo et al. [134] developed VR software for treating PTSD in patients affected by the Iraq war, highlighting VR's benefits in treating PTSD with improvements in cost and efficiency. This study extended VR's application realm to mental health treatments, offering a novel approach to addressing complex psychological conditions like PTSD with technology.

Applying VR in disaster emergency response

The MERITS program [113], developed by the National Institute for Occupational Safety and Health in Pittsburgh in 2000, marked a significant leap forward in emergency response training for the mining sector. As a pioneering VR simulator, it allowed stakeholders from mining companies, labor, and government agencies to engage in simulated emergency scenarios collaboratively. This initiative underscored the importance of multi-agency cooperation and showcased VR's capability to simulate realistic, high-risk environments for effective safety training, representing a crucial advancement in the field. Similarly, Schofield and Dasys's 2009 investigation [137] into VR technology's role in emergency response training within the mining sectors of Australia and Canada highlighted its underutilized potential despite existing limitations. They emphasized VR's capacity to offer more realistic, engaging, and safe training experiences, crucial in high-risk industries like mining. Costa et al. [39] evaluated the essential aspects of mine rescue operations, discussing the critical role of well-trained and equipped personnel in executing miner rescues under hazardous conditions. The study emphasized the responsibility of local and state governments in ensuring the readiness of mine rescue teams and introduced innovative training methodologies. Specifically, their research on the application of heart rate monitors and GPS sport watches in training mine rescuers at the INCDINSEMEX Laboratory illustrated the potential of integrating modern technology to advance mine rescue training practices.

Hsu et al. [70] analyzed the advantages and limitations of VR-based disaster response training across various sectors in the United States. Their findings highlighted VR's substantial potential to enhance disaster preparedness and response training's effectiveness and efficiency, signaling a shift towards more innovative training methodologies compared to traditional approaches.

In a novel approach to medical emergency response training, Stone et al. [142] developed a cost-effective training application utilizing VR and MR technology. This application, blending real-world objects with virtual reality scenes, demonstrated MR's capacity to create immersive and realistic training environments. Their research indicated the feasibility of developing a credible, simulation-based training system using VR and MR technology that is both affordable and adaptable. Such advancements present practical solutions for meeting broad training needs, offering new avenues for enhancing the quality and accessibility of training in medical emergency response."

2.1.3 Machine Learning

In the meantime, machine learning (ML), as a significant technique in AI research, started to play an important role in healthcare. ML was first introduced by Arthur Samuel in 1959 and later discussed by Kohavi and Provost in 1998 [129]. The formal definition of ML was given by Mitchell in 1997 [110]. Christopher Bishop [18] classified machine learning algorithms into four categories: 'Supervised Learning' (SL), 'Semi-supervised Learning,' 'Unsupervised Learning' (UL), and 'Reinforcement Learning' (RL). In recent years, SL, UL, and RL have become the most popular subareas in ML. This review examines these major learning patterns as well as typical algorithms of each pattern.

Supervised Learning (SL)

In supervised learning, the input dataset is divided into training and test datasets, and there is an output variable in the training dataset that needs to be predicted or classified. The decision tree is one of the most typical patterns in SL. The foundational work by Quinlan [130] on decision tree induction and Breiman et al. [22] on the classification and regression trees (CART) methodology laid the groundwork for the development of decision tree algorithms, a cornerstone in the field of machine learning and data analysis. Quinlan's study reviewed the strategies for constructing decision trees, providing valuable insights into their applicability across various domains. Breiman et al.'s seminal work introduced the fundamentals of CART, discussing their construction and application in classification and regression tasks. Loh [91] further expanded on this research, offering a comprehensive examination of CART's progression and transformation over fifty years, emphasizing its enduring importance in statistical modeling, machine learning, and data analysis.

Jensen [76] underscored the challenges of decision-making in the presence of incomplete and unreliable information, presenting Bayesian networks as a robust theoretical framework for managing uncertainty. Building on this, Lowd and Domingos [93] highlighted the effectiveness of Naive Bayes models, especially when trained using the Expectation-Maximization (EM) algorithm, demonstrating their accuracy and efficiency in probabilistic learning and inference tasks. These studies collectively illustrated the applications and potential of Naive Bayes models and Bayesian networks in diverse fields, including medical diagnosis and finance.

Cortes and Vapnik [38] introduced SVMs, which have since become foundational in machine learning for classification and regression tasks. Their research detailed SVMs' theoretical foundations and practical utilization, emphasizing their role in optimizing data point separation. Hofmann et al. [67] provided a comprehensive investigation into kernel methods, essential tools for nonlinear data analysis and tasks related to classification and regression, offering insights into their practical implementation and significance in contemporary machine learning.

Srinivas et al. [141] showcased the application of data mining techniques in healthcare, particularly for predicting heart attacks, highlighting data mining's potential as a tool for early detection and prevention of cardiovascular diseases. Yassine et al. [160] introduced a model utilizing smart home big data for discerning human activity patterns, employing techniques like frequent pattern mining and cluster analysis for effective pattern identification and prediction. Velikova et al. [151] proposed a probabilistic model for predicting pregnancy-related disorders using Bayesian networks, demonstrating the model's novelty and integral role in daily pregnancy care.

Semi-Supervised Learning

Semi-supervised learning represents a crucial paradigm within machine learning that effectively leverages a mix of a small amount of labeled data alongside a larger volume of unlabeled data during the training process. Dey [44] elucidated the significance of semi-supervised learning, especially in contexts where unlabeled data are abundant, but labeling is labor-intensive and challenging. This approach is particularly beneficial in machine learning and data mining scenarios where obtaining labeled data can be a daunting task, offering a viable solution for exploiting the readily available unlabeled data to improve learning accuracy and efficiency.

Building on the foundation of semi-supervised learning, Goodfellow et al. [58] introduced a groundbreaking framework for generative models known as adversarial nets. In this setup, a generative model (G) is trained to emulate the data distribution, while a discriminative model (D) is tasked with differentiating between the actual training data and the samples produced by G. The goal for G is to enhance the likelihood of D making classification errors, effectively forming a mini-max two-player game. This adversarial approach facilitated an optimized solution where G accurately represents the data distribution, and D approaches a classification probability of 1/2 across the board. Notably, when the system is articulated through multi-layer perceptions, it enabled training via back-propagation, circumventing the need for Markov chains or approximate inference networks. The efficacy of this framework was underscored by both qualitative and quantitative assessments of the samples generated.

Further expanding on the concept of GANs, Shaikh [138] presented a comprehensive guide to understanding these powerful generative models. The research explored the intricacies of GANs, shedding light on their applications and the promising avenues they open in artificial intelligence and data analytics. This work aimed to demystify GANs for readers, providing a foundational knowledge base that captures the essence of these models within the broader landscape of deep learning techniques.

Goodfellow's subsequent research [57] further elaborated on GANs, emphasizing their role in approximating complex cost functions through supervised learning. The study acknowledged the novelty of GANs and the ongoing need for research to refine their training processes, particularly in managing high-dimensional, non-convex optimization challenges. Success in this domain not only enhanced GANs' capabilities but also holds promise for broadening their application beyond current image generation and manipulation technologies, indicating potential for significant impact in future applications."

Unsupervised Learning (UL)

Unsupervised learning represents a critical category of machine learning techniques designed to identify patterns within unlabeled data, primarily employed for clustering and feature reduction. Dey [44] elucidated unsupervised learning's capacity to autonomously discover underlying structures in data, a capability that is especially beneficial for handling the vast and complex datasets typically encountered in healthcare and biomedicine. In this context, Yoo et al. [161] conducted a thorough survey on the application of data mining, a subset of unsupervised learning, within the healthcare and biomedicine sectors. Their research underscored data mining's ability to extract meaningful insights from extensive datasets, covering its principles and its important role in healthcare for tasks such as fraud detection and disease diagnosis. Despite its potential, they also addressed the challenges related to the clinical adoption of these technologies by health professionals.

[45] leveraged partition-based clustering algorithms, specifically K-Means and K-Medoids, to analyze lung cancer data, demonstrating the practical applications of these methods in identifying patterns within complex health-related datasets. Complementing clustering approaches, Principle Component Analysis (PCA) is highlighted as a powerful technique for dimensionality reduction. Dash and Subudhi [41] and Das [40] described PCA as a method for identifying the most significant linear combinations of variables that maximize variance,

thereby simplifying the data structure without substantial loss of information. Smith [140] further elaborated on the advantages of PCA, noting its capacity for data compression with minimal information loss, a feature that significantly enhances data analysis efficiency. Harrington [63] provided practical insights into utilizing PCA for reducing data dimensions, illustrating its applicability across various domains.

Expanding on the applications of PCA, Harrou et al. [64] proposed an advanced PCA-based anomaly detection method tailored for monitoring emergency department (ED) demands. By applying the multivariate cumulative sum (MCUSUM) control chart to the uncorrelated residuals obtained through PCA, their approach demonstrated superior performance over traditional PCA-based methods, especially in identifying minor anomalies within pediatric ED data.

This narrative synthesizes the core aspects and contributions of unsupervised learning in the realms of healthcare, biomedicine, and beyond. By detailing the applications of clustering algorithms and dimensionality reduction techniques, such as PCA, it highlights the transformative potential of unsupervised learning in extracting valuable insights from unlabeled datasets, enhancing anomaly detection, and streamlining data analysis processes.

Reinforcement Learning (RL)

Reinforcement Learning represents a important approach in machine learning, where the aim is to maximize long-term rewards through actions taken in specific situations without initial knowledge of the best actions. This technique, approximating traditional optimal control methods, allows learners to adapt their actions based on the outcomes and the evolving situation, thereby affecting future scenarios and potential rewards [146].

Building upon the foundational principles of RL, Kao et al. [80] introduced a novel context-aware hierarchical reinforcement learning method specifically designed to enhance the functionality and user interaction of online symptom checkers, such as those provided by WebMD and Mayo Clinic. Their approach sought to improve diagnostic accuracy and reduce the number of queries necessary, marking a significant advancement over traditional symptom checking systems by leveraging RL's capability to adapt and refine strategies over time.

In a unique application of RL within VR, Rovira et al. [136] explored its potential to guide participants through virtual environments without prior instructions. The study focused on enabling participants to navigate spaces while avoiding virtual hazards, with the RL agent optimizing their movements towards achieving specific goals. This research illustrated how RL can be effectively applied in VR settings to enhance user interaction and goal orientation, demonstrating the versatility of RL in creating intuitive and responsive virtual experiences.

Mnih et al. [111] further pushed the boundaries of RL through the development of a deep Q-network artificial agent. This agent, trained using end-to-end reinforcement learning, demonstrated an exceptional ability to master a wide array of complex tasks by processing high-dimensional sensory inputs. Achieving performance levels comparable to professional human game testers across a challenging suite of 49 Atari 2600 games, this work underscored RL's profound impact in bridging the divide between sensory data processing and strategic action planning. The success of this agent highlighted the transformative potential of RL in enhancing decision-making processes across diverse and complex environments.

2.1.4 Applying AI in healthcare

Jiang et al. [77] provided a comprehensive review of the current state and future prospects of AI applications in healthcare, particularly focusing on how ML techniques are utilized to process structured data, including images, electrophysiological (EP) data, and genetic information. Their findings underscored the significant impact of AI in enhancing healthcare outcomes through the analysis of complex data sets. Similarly, Dua et al. [47] and Kavakiotis et al. [81] conducted explorations of the applications of machine learning and data mining within diabetes research, highlighting the predominance of supervised learning methods and the notable success of support vector machines (SVMs) in extracting insights from clinical datasets for diabetes research. Miotto et al. [109] explored the burgeoning field of deep learning within healthcare, suggesting its potential to transform big biomedical data into actionable health insights, despite existing challenges in interpretability for domain experts and citizen scientists.

On the practical side, Nguyen et al. [114] introduced an automated system leveraging supervised learning techniques for classifying imaging examination results into reportable and non-reportable cancer cases, demonstrating the utility of AI in improving diagnostic accuracy. Manogaran et al. [99] addressed the challenge of detecting DNA copy number changes—a key factor in cancer diagnosis—by proposing a Bayesian Hidden Markov Model (HMM) with Gaussian Mixture (GM) Clustering. This approach was shown to be more effective in analyzing large DNA sequences and identifying genetic variations related to cancer and other diseases, compared to traditional methods. Lastly, Ismaeel et al. [73] developed an Extreme Learning Machine (ELM) algorithm to model factors common among heart disease patients, offering a cost-effective early warning system with approximately 80% accuracy in diagnosing potential heart conditions based on real data from the Cleveland Clinic Foundation.

2.1.5 Automatic Speech Recognition (ASR)

The following papers present inspirations for applying ASR in healthcare training. Ghulam Muhammad [69] proposed a novel cloud-based framework for speech enabling healthcare, which allows patients or any healthy person seeking medical assistance via speech commands. Nicholas et al. [153] proposed an ASR-based training module for healthcare professionals, which employed automatic speech recognition technology, pronunciation assessment, and video clips of a simulated medical history interview with a minority language patient. The module was tested in five Quebec nursing training institutions, and it was easy to operate and addressed anticipated language learning needs.

There are four main open-source implementations of ASR since the 1990s: CMU Sphinx, Julius, Kaldi, and Whisper. CMU Sphinx and Julius are based on the Hidden Markov Model (HMM), Kaldi is based on neural networks (NN), and Whisper is based on transformers. Our review is focused on the latter two. Daniel Povey et al. [128] designed a free, open-source toolkit named Kaldi for speech recognition research. Kaldi provides a speech recognition system based on finite-state transducers (using the freely available OpenFst), together with detailed documentation and scripts for building complete recognition systems, and it has become an established framework used to develop state-of-the-art speech recognizers. Mirco Ravanelli et al. [133] aimed to bridge the gap between popular toolkits, trying to inherit the efficiency of Kaldi and the flexibility of PyTorch, and finally designed an assembled toolkit named PyTorch-Kaldi. They claimed that the PyTorch-Kaldi also supports combinations of neural architectures, features, and labels, allowing users to employ complex ASR pipelines. The experiments confirmed that PyTorch-Kaldi can achieve state-of-the-art results in some popular speech recognition tasks and datasets.

OpenAI Whisper

Whisper, developed and released by OpenAI in September 2022 [131], represented a significant advancement in the field of speech recognition as an open-source machine learning model. It is proficient in transcribing English as well as a variety of other languages and possesses the capability to translate multiple non-English languages into English. The model's exceptional performance was largely attributed to its utilization of a diverse and extensive dataset, which enhanced its ability to accurately handle accents, background noises, and specialized jargon. Whisper employed a weakly-supervised learning strategy within an encoder-decoder transformer architecture, offering several distinct advantages over traditional speech recognition systems:

1. High Accuracy Rate:

   Whisper was trained on 680,000 hours of multilingual audio data, enabling it to achieve remarkable accuracy across a broad spectrum of dialects and acoustic conditions. This training approach ensured high performance even in zero-shot scenarios, where the model had not been explicitly fine-tuned, making it invaluable for real-world applications that encounter a wide range of linguistic inputs.

2. Multilingual and Multitask Capability:

   In contrast to conventional models that often require separate specialized systems for each language and task, Whisper integrated support for multiple languages and various speech processing tasks, such as language identification and speech-to-text translation, within a single cohesive model framework. This capability significantly simplified deployment and increased utility across diverse global scenarios.

3. Enhanced Contextual Understanding:

   Whisper excelled in maintaining contextual coherence over extended speech sequences, a substantial improvement over traditional models, which typically process speech in shorter, disjointed segments. This advanced contextual understanding was particularly beneficial for processing long-form content such as academic lectures, legal proceedings, and lengthy meetings, ensuring that transcriptions are not only accurate but also contextually consistent throughout.

These features of Whisper indicate its potential as a robust, versatile, and precise tool in the field of speech recognition, offering substantial improvements in the technological capabilities of machine learning models in natural language processing.

Chen et al. [31] addressed the critical need to enhance speech recognition capabilities for Hakka, a minority language with limited resources. By incorporating OpenAI's Whisper model with readily available online Hakka speech resources, the study not only advanced the development of a specialized ASR model for Hakka but also demonstrated its effectiveness. The project explored how this technology could serve as a tool for cultural preservation and the promotion of Hakka culture, highlighting its significance in the broader context of language sustainability. The synthesized Hakka ASR model showcased a detailed workflow for training and deployment, underscoring its potential applications in digital education and smart living technologies. This approach not only fostered linguistic diversity but also enhanced access to technology-driven educational resources for Hakka speakers, positioning the Whisper-based ASR model as an important innovation in the field of language technology.

Calbert and Nathan [60] examined the performance of Whisper across a variety of native and non-native English accents. The findings showed that Whisper performed better with American English accents compared to British and Australian accents, while demonstrating comparable accuracy with Canadian English. Native English accents consistently yielded higher accuracy compared to non-native accents. The study further investigated how speaker characteristics such as sex, native language typology, and second language (L2) proficiency influenced the WER, uncovering significant associations. Additionally, Whisper was found to be more accurate in transcribing read speech than conversational speech and showed variations in accuracy based on speaker gender. These results underlined the need for further refinement of ASR systems to handle diverse speech patterns and accent variations effectively, pointing towards potential improvements in training methodologies to enhance inclusivity and accessibility in communication technologies.

Wang et al. [156] explored the in-context learning capabilities of OpenAI's Whisper ASR models. It introduced a new method known as speech-based in-context learning (SICL) for adapting during test time, which effectively lowered the word error rates (WERs) with just a few labeled speech samples, avoiding the need for gradient descent. Experiments adapting to language levels with Chinese dialects demonstrated that SICL, when applied to isolated word ASR, consistently achieved significant WER reductions with any size of Whisper models across two dialects, averaging 32.3% improvement. The efficiency of SICL was further enhanced by a k-nearest-neighbours-based technique for selecting in-context examples, which raised the average WER reduction to 36.4%. These results were confirmed through tasks involving speaker adaptation or continuous speech recognition, both showing notable WER reductions. Additionally, comprehensive quantitative analyses were provided, highlighting SICL's ability to handle phonological differences and dialect-specific lexical peculiarities.

2.1.6 Natural Language Processing (NLP)

Generations of NLP

NLP has evolved significantly over the decades, progressing through three distinct generations, each marked by different methodologies and technological advancements. These generations encompass Symbolic-based NLP, Statistical-based NLP, and NN-based NLP. Each generation contributed uniquely to the development of the field, addressing the complexities and challenges inherent in processing human language.

1. Symbolic-based NLP (First Generation):

   The Symbolic-based NLP relied on handcrafted rules and linguistic knowledge. Techniques in this era were driven by the logic and rules formulated by linguists, which were then encoded into computer systems. Symbolic NLP was dominant in the early days of artificial intelligence, focusing on logic, grammar, and expert systems. It emphasized the use of grammar rules for parsing and the manipulation of symbols, hence the name. The systems developed during this period were quite rigid, often failing to handle the variability and ambiguity inherent in human languages effectively. Foundational works such as [35, 157, 4] collectively underscored the significant role of structured rules in the development of early NLP. Chomsky's transformational grammar [35] introduced a theoretical framework that revolutionized linguistic structure analysis, greatly influencing early NLP systems. Building on this, Winograd's exploration [157] of rule-based parsing and Allen's [4] detailed examination of symbolic NLP demonstrated practical applications of these theories. Together, these texts highlighted a important place in NLP characterized by a focus on predefined linguistic rules to interpret and generate human language, establishing a systematic approach that was integral to both linguistic research and computational applications.

2. Statistical-based NLP (Second Generation):

   Emerging in the late 1980s and gaining prominence in the 1990s, statistical NLP introduced mathematical models to handle uncertainties and partial truths in language processing. This approach leveraged large amounts of data and statistical methods to learn from this data, focusing on probabilities rather than hard-coded rules. Machine learning models, particularly those based on Bayesian methods and hidden Markov models, became central to NLP tasks such as speech recognition, machine translation, and text summarization. The flexibility of statistical methods allowed for more nuanced understanding and generation of language compared to the rigid rule-based methods of the previous generation. Seminal works such as [23, 98] significantly shaped the landscape of statistical NLP. Brown et al.'s introduction [23] of n-grams as a fundamental concept for language modeling illustrated the practical application of statistical methods to predict word sequences effectively by grouping words into classes based on usage. Complementarily, Manning and Schütze [98] expanded on this foundation by providing an exhaustive exploration of the underlying algorithms and theories that drive these statistical approaches, establishing a scholarly resource that deepened the understanding of statistical NLP's capabilities and applications. Together, these works not only highlighted innovative techniques but also marked a critical evolution in the field, from rudimentary methods to sophisticated statistical models that advanced the comprehension and processing of language.

3. Neural Network-based NLP (Third Generation):

   The latest generation, flourishing with the advent of deep learning in the 2010s, used neural networks to model language tasks. These models learned representations and patterns directly from the data without explicitly programmed rules. The introduction of architectures like recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and Transformers significantly advanced NLP tasks, leading to state-of-the-art performance across various benchmarks. This generation leveraged massive datasets and powerful computational resources to train complex models that could capture subtle nuances of language, perform context-aware understanding, and generate human-like text. Landmark contributions by [66, 11, 149] propelled significant advancements in neural network methodologies and their efficacy in natural language processing. Hinton et al. [66] developed a fast learning algorithm that was important in the effective training of deep neural networks, laying the groundwork for subsequent innovations in deep learning. Bengio et al. [11] expanded on this by detailing the mechanisms of representation learning, which became critical in how neural networks understand and process complex data structures. Complementing these advancements, Vaswani et al. [149] introduced the Transformer model, which utilized the novel attention mechanism to dramatically improve the handling of sequence data by focusing on relevant parts of the input data. Collectively, these works not only enhanced our understanding of deep learning capabilities but also set new standards for designing NLP systems that are both powerful and efficient.

Bidirectional Encoder Representations from Transformers (BERT)

Within the landscape of the third generation of NLP, which leverages neural network-based methodologies, the introduction of BERT by Jacob Devlin et al. [43] marked a pivotal advancement. As a subcategory of this generation, BERT epitomized the utilization of advanced neural architectures, specifically the Transformer model, to significantly enhance language understanding. This model innovatively pre-trained deep bidirectional representations from unlabeled text by conditioning concurrently on both left and right contexts, a departure from traditional sequential text processing approaches. By integrating the entire sequence of words at once, BERT was uniquely equipped to capture a more nuanced understanding of context, which is critical for complex language processing tasks. This approach set a new benchmark in the field, enabling state-of-the-art performances across diverse NLP tasks such as question answering and language inference, with only minimal architectural adjustments required for task-specific adaptations. BERT's development not only represented a major stride in NLP capabilities but also solidified the critical role of Transformer-based models in driving forward the boundaries of what machines can understand and achieve with human language.

Following BERT's introduction, several studies explored its applications and modifications. Bahman Tahayori et al. [145] utilized BERT in predicting patient disposition from emergency triage notes with high accuracy, demonstrating ML and NLP's combined power in healthcare. Yen-Pin Chen et al. [32] proposed a BERT-based model for diagnoses-extractive summarization in hospital systems, showing that character-level tokens could reduce model size without significantly impacting performance.

Parallel to BERT, the development of Generative Pre-trained Transformer models, particularly GPT-2 by Radford et al. [132] and GPT-3 by Brown et al. [24], showcased the potential of unsupervised multitask learning. These models, trained on extensive internet text, excelled in generating coherent and contextually relevant text across a variety of language tasks, highlighting the advancements in NLP and the effectiveness of unsupervised learning.

The foundational architecture for these advancements, the Transformer model introduced by Vaswani et al. [149], revolutionized NLP with its attention mechanism-based approach. This architecture was crucial in improving machine translation, text generation, and understanding, setting the stage for the development of models like BERT and GPT.

Generative Pre-trained Transformer (GPT)

Within the evolving domain of the third generation of NLP, which capitalizes on neural network-based methodologies, the advent of Generative Pre-trained Transformer models, particularly GPT-3.5 [24] and GPT-4 [2], by OpenAI, represented a monumental progression. As key exemplars of this generation, these models demonstrated the sophisticated use of the Transformer architecture to dramatically enhance generative text capabilities. Unlike prior models that relied solely on directional context, GPT-3.5 and GPT-4 leveraged a broader, more nuanced contextual understanding, allowing for more sophisticated and coherent text generation across various domains.

GPT-3.5 and GPT-4 were characterized by their massive scale and the extensive pre-training on diverse internet text, which equipped them with a profound capacity to generate human-like text and solve complex language tasks with little to no task-specific data. This ability was a significant leap from earlier iterations, enabling applications in content creation, conversation generation, summarization, and more complex reasoning tasks that previous models could not handle as effectively. The size of these models, with GPT-4 featuring an even larger number of parameters than GPT-3.5, allowed them to perform at unprecedented levels of sophistication, setting new benchmarks in the field.

The development of GPT-3.5 and GPT-4 not only underscored the incredible strides in NLP capabilities but also cemented the critical role of advanced Transformer-based models in pushing the boundaries of what artificial intelligence can achieve with language. These models' ability to understand and generate text across a spectrum of styles and topics revolutionized the potential applications of NLP technologies, making them pivotal in the ongoing evolution of machine learning and artificial intelligence in understanding and interacting with human language.

As the adoption of foundation models like GPT-3 grows, discussions around their transformative impact, ethical considerations, and potential risks have also emerged. Bommasani et al. [21] and McGuffie and Newhouse [104] explored the capabilities and challenges posed by these models, from their application in various domains to the risks of facilitating radicalization and spreading misinformation. These discussions emphasized the importance of responsible AI development and the need for robust mitigation strategies. Furthermore, the GLUE benchmarking platform introduced by Wang et al. [154] and ethical considerations raised by Bender et al. [10] highlighted the ongoing efforts to evaluate the performance of NLP models and address the societal and environmental implications of deploying large-scale language models. These contributions underscored the need for sustainable and ethical AI practices.

Jarou et al. [75] evaluated the capabilities of AI models like ChatGPT in emergency medicine by using questions from the American College of Emergency Physicians (ACEP) PEERprep In-Training Self-Test and Study Guide. Clinical image-based questions were excluded, as ChatGPT couldn't analyze them. The study found that human test takers had a mean correct response rate of 83.7%, compared to 61.4% for GPT-3.5 and 82.1% for GPT-4. GPT-4's performance was notably closer to human levels, indicating significant improvements over its predecessor. However, the study acknowledged limitations, including a small question set, suggesting the need for further research with more extensive data for comprehensive evaluation. This study was the first to test large language models on emergency medicine-specific board preparation questions, revealing that newer models could perform comparably to human practitioners in this field.

Wang et al. [155] assessed ChatGPT's effectiveness in primary screening for mild cognitive impairment. It involved 174 participants from the DementiaBank screening, with 70% in the training set and 30% in the test set. Variables considered included vocabulary, syntax, grammar, and semantics, based on published studies and statistical analyses. The GPT-4 model demonstrated high sensitivity, specificity, and area under the curve (AUC) in both sets. The prompt design consisted of five parts: character, scoring, indicator, output settings, and explanatory information. The findings suggested that ChatGPT is effective in mild cognitive impairment screening, with performance potentially enhanced by prompt standardization by professional clinicians, though it's not a substitute for clinical diagnosis.

Gan et al. [54] compared the performance of ChatGPT, Google Bard, and medical students in executing START triage in MCI. A cross-sectional analysis using a validated questionnaire with 15 MCI scenarios assessed triage accuracy in four categories. Google Bard showed higher accuracy (60%) compared to ChatGPT (26.67%), with the difference being statistically significant (p = 0.002). Medical students had previously achieved 64.3% accuracy, with no significant difference noted between Google Bard and medical students (p = 0.211). The qualitative content analysis revealed Google Bard's superior performance over ChatGPT in all assessed categories.

Nori et al. [116] conducted a comprehensive evaluation of GPT-4, a general-purpose LLM, in the medical domain, highlighting its remarkable capabilities. GPT-4 notably exceeded the passing score on the USMLE by over 20 points, outperforming both GPT-3.5 and specialized medical models like Med-PaLM. Its improved calibration indicated a more reliable prediction of answer correctness. GPT-4 also excelled in explaining medical reasoning, personalizing explanations for students, and creating counterfactual medical scenarios. The study suggested potential applications in medical education, assessment, and practice, emphasizing the importance of addressing challenges in accuracy and safety. Mao et al. [101] provided a comprehensive review of studies evaluating ChatGPT and

GPT-4. The survey focused on the language and reasoning abilities, scientific knowledge, and ethical considerations of these models. It addressed the strong curiosity and speculation surrounding ChatGPT's potential impact on social and economic systems. The paper also critically examined existing evaluation methods for large language models and offered recommendations for future research. This survey aimed to consolidate collective assessment findings and guide further investigations into large language models.

2.1.7 The MCI standard

The MCI standard typically refers to any large number of casualties produced in a relatively short period of time, usually as the result of a single incident such as a military aircraft accident, hurricane, flood, earthquake, or armed attack that exceeds local logistic support capabilities. Different countries or organizations have different guidelines to control mortality in an MCI, and triage is one of the most important parts of such guidelines. There are four major MCI triage systems that have been used by Australia, Canada, the UK, and the United States, respectively

1. St John New Zealand's triage System

   The St John New Zealand's triage category [5] system was a critical tool utilized in emergency medical settings to efficiently prioritize patient care based on the severity of their medical needs. This system divided patients into four levels: Category 1 (Immediate/Red) for life-threatening conditions requiring immediate intervention; Category 2 (Very Urgent/Orange) for serious conditions that could potentially escalate but were not immediately life-threatening; Category 3 (Urgent/Yellow) for conditions that needed prompt treatment but did not pose an immediate critical threat; and Category 4 (Standard/Green) for non-urgent conditions that could afford to wait without causing harm to the patient. This triage method ensured that medical resources were used effectively, allowing healthcare providers to deliver timely and appropriate care based on the urgency of each case.

2. Australia Triage System (ATS)

   ATS had 5 categories [121], and each category indicated the status of a single patient. The category was decided by a combination of major and minor features. Major features included direct vital signs such as Respiratory Rate (RR), Blood Pressure (BP), Skin color, etc. Minor features included urgent statuses such as Airway risk and Behavioral states like violent or aggressive behavior.

3. Canadian Triage and Acuity Scale (CTAS)

   Similar to ATS, CTAS also used 5 categories to indicate the current status of a single patient [25]. CTAS used a combination of First Order Modifiers and Second Order Modifiers to decide the category of each patient. First Order Modifiers included Respiratory Distress, Hemodynamic Stability, Level of Consciousness, etc. Second Order Modifiers included Blood Glucose Level, Dehydration Severity, and Blood Pressure.

4. Manchester Triage System (MTS)

   MTS was a 4-category, flow-chart-based triage system. According to the MTS guideline, several scenarios such as chest pain and traumatic brain injury (TBI) were used to indicate the basic situation of the patient. Once the scenario was confirmed, there would be a few flow-charts to measure the severity of the condition and then lead to a specific triage category eventually [8].

5. Simple Triage and Rapid Treatment (START)

Similar to the UK, the United States also used a 4-category triage system, but the scenario-based workflow did not apply in its system. The US's triage system also used vital signs as a major feature while applying triage [52]."

2.2 Second Round Literature Review

As we found that the first round of the literature review was not enough to narrow down the research details, we conducted a second round of the literature review to extend our knowledge base. In the second round, we used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to assess the proposed papers. The flow diagram of PRISMA guidelines consisted of three phases, which were Identification, Screening, and Inclusion, respectively. The databases we used to search the literature included Scopus, Directory of Open Access Journals, IEEE Xplore Digital Library, ACM Library, MedLine, and Springer Link. In this research, the included papers had to meet the following criteria: 1) The topic should focus on or be relevant to the emergency healthcare field, especially in the sub-field of triage; 2) they should use either VR or AI technology; 3) they must come from a peer-reviewed conference or journal; 4) they must be written in English.

We identified 24 papers among 508 papers (113 in the VR field and 408 in the AI field). Nine of them were highly relevant to VR and healthcare, and 15 were highly relevant to AI and emergency healthcare. The publication dates of the identified papers were between 2010 and 2021. Figures 2.1 and 2.2 show the process of paper selection in the VR and AI fields, respectively.

2.2.1 Search Criteria

In the Identification phase, we used several keyword combinations to acquire the literature. The main keyword combinations included "Virtual Reality, Triage", "Virtual Reality,

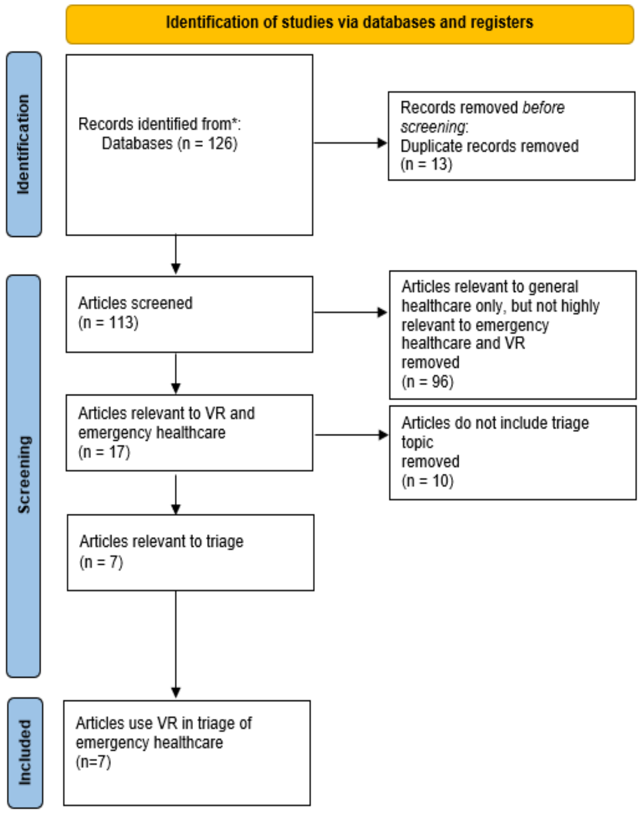

Fig. 2.1 VR Paper selection flow according to PRISMA guidelines

Emergency Healthcare","Artificial Intelligence, Triage", and "Artificial Intelligence, Emergency Healthcare". All articles were retrieved from six databases before 11 November 2021. The articles were reviewed by authors in the field. A total of 126 VR and 418 AI papers were included. More specifically, by searching the keyword combination "Virtual Reality, Triage", 116 articles were included. By searching the keyword combination "Virtual Reality, Emergency Healthcare", 10 articles were included. By searching the keyword combination "Artificial Intelligence, Triage", 403 articles were included. By searching the keyword combination "Artificial Intelligence, Emergency Healthcare", 15 articles were included. The detailed distribution of article types in each database is shown in Figure 2.3 and 2.4.

According to Figure 2.3, the majority of VR research papers were sourced from the ACM and Scopus databases, with 58 and 40 papers respectively. Springer Link contributed 18 papers, while other databases accounted for 10 papers. In contrast, AI research papers

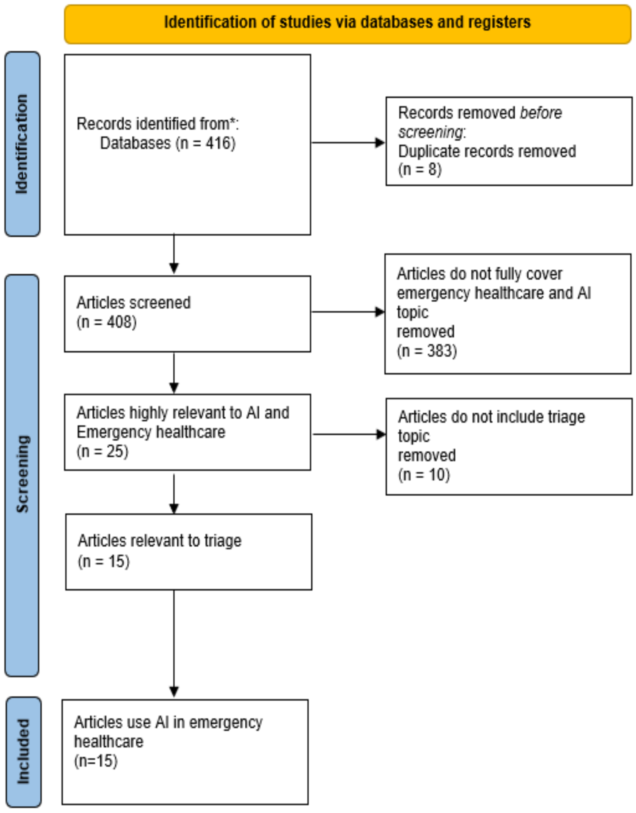

Fig. 2.2 AI Paper selection flow according to PRISMA guidelines

primarily came from Scopus and Springer Link, with contributions of 246 and 159 papers respectively. Contributions from other databases were minimal, totaling only 13 papers.

During the PRISMA Screening phase for VR papers, 13 duplicates were removed. An additional 96 papers were excluded for only addressing general healthcare topics without specific relevance to emergency healthcare and VR. Ten papers were eliminated for not addressing the triage topic, ultimately leaving 7 VR papers that specifically relate to healthcare.

Regarding AI papers, as detailed in Figure 2.4, Scopus was again a major source with 245 papers, followed closely by Springer Link with 153 papers. ACM contributed only 2 selected papers. Following the PRISMA protocol, 8 duplicated AI papers were removed. A further 383 papers were excluded for not adequately covering the intersection of AI and

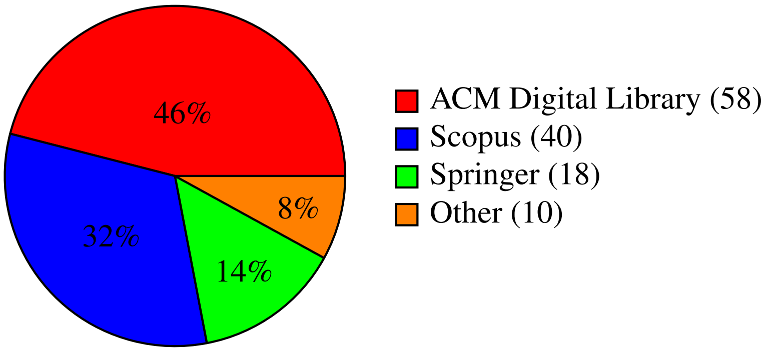

Fig. 2.3 Source database of VR and healthcare papers (Total: 126)

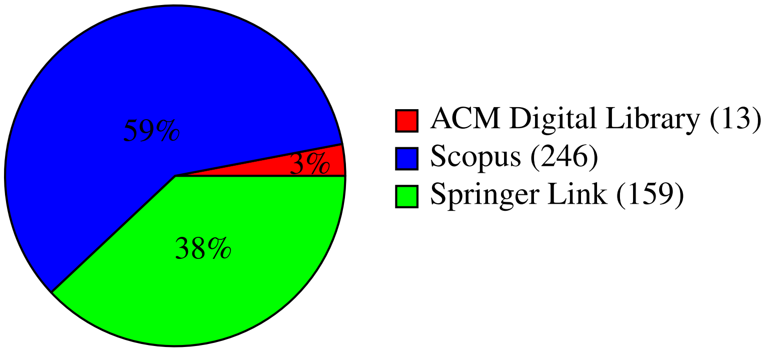

Fig. 2.4 Source database of AI and healthcare papers (Total: 418)

emergency healthcare, and an additional 10 were discarded for not including the triage topic. Consequently, 15 AI papers relevant to emergency healthcare were retained for further study.

2.2.2 VR Literature Overview

As we discovered the significant advancements in emergency healthcare by reviewing those selected papers, a notable trend was the increasing reliance on VR technologies. These innovations included 1) Disaster Triage Training and MCI Management; 2) VR and AR Technology for Emergency Preparedness and Training; and 3) Specialized VR Systems for Training and Education.

In the realm of disaster triage and mass casualty management, VR technologies have proven instrumental. Andreatta et al. [6] evaluated VR against standard patient drills for emergency medicine residents, highlighting its potential as an on-demand training tool. Cone et al. [37] showed that VR systems could effectively simulate real-world triage scenarios, enhancing paramedic students' accuracy in victim assessment. Similarly, Ingrassia et al. [72] found no significant differences in triage accuracy between VR simulations and live exercises, underscoring VR's utility in medical training. These studies collectively suggested that VR could augment traditional training methods, offering flexible and immersive learning experiences that are critical in emergency preparedness [51, 13, 107, 15].

Furthermore, VR and AR technologies extended their influence into broader aspects of emergency preparedness and training. Yu, X and Ganz, A [162] developed MiRTE, a mixed reality game for MCI management, utilizing immersive environments to enhance interactivity and integration with external systems. Caballero and Niguidula [27] explored VR's potential in disaster risk management, creating simulations that improve disaster awareness and capacity building. Koutitas et al. [86] demonstrated how VR and AR could prepare EMS personnel for large-scale emergencies, facilitating remote training that could be important during actual disaster responses.

On the educational front, specialized VR systems were tailored to meet the specific needs of medical training. Tayama et al. [147] introduced a VR triage training system that adjusted difficulty based on the learner's proficiency, enhancing skill acquisition through customized scenarios and feedback. Lourdeaux et al. [92] focused on non-technical skills, crucial for leadership in high-stress medical situations, using VR to simulate complex interpersonal interactions and decision-making processes. Similarly, Lowe et al. [94] investigated the feasibility of a 360 VR platform for pediatric MCI decision-making, highlighting its acceptability and potential as a training tool in disaster preparedness scenarios. These applications illustrated VR's capacity to not only simulate technical skills but also to foster the soft skills necessary for effective medical leadership and crisis management.

2.2.3 VR literature statistical result

To categorize the selected seven VR and healthcare articles, we used Hardware (e.g., Oculus Rift), Experiment Scale, and Scenario (e.g., MCI training) as the main categories. A summary is given in Table 2.1

Reference	**Hardware**	**Scale**	**Scenario**
Andreatta et al. [6]	CAVE (Customized VR environment)	Medium (14)	MCI triage training
Berndt et al. [13]	Oculus Rift and HTC VIVE	Small (10)	MCI triage training
Koutitas et al. [86]	Microsoft Hololens	Large	Equipment training
Caballero et al. [27]	Oculus Rift	Large	Disaster simulation
Mills et al. [108]	Oculus Rift and HTC VIVE	Small (10)	MCI triage training
Lowe et al. [94]	Oculus Rift and Google Daydream VR	Large (206)	MCI triage training
Bilet et al. [15]	HTC VIVE	Large (41)	MCI triage training

Table 2.1 Summary of VR Hardware and Scenarios in Healthcare Research

Hardware Utilization in VR Healthcare Studies

Based on the information detailed in Table 2.1, the research scrutinizing virtual reality's application within healthcare contexts revealed a varied deployment of head-mounted displays (HMDs), with a marked preference for Oculus Rift and HTC Vive. These devices were favored due to their immersive capabilities and robust performance features, which are critical for realistic simulation environments in medical training. Specifically, the Oculus Rift was utilized independently in one study and collaboratively in three others [13, 27, 108, 94], highlighted for its high-resolution display and integrated audio system, which notably enhanced the user's sensory experience. Its repeated appearance in four studies underlined its popularity and reliability as a tool for immersive healthcare education and training.

Similarly, the HTC Vive was featured in three studies [13, 108, 15], chosen for its precision in spatial tracking and its support for extensive, room-scale virtual experiences. The consistent selection of the HTC Vive in these studies underscored its suitability for environments where accurate motion tracking and expansive interactive spaces were necessary, demonstrating its widespread acceptance and utility in the field.

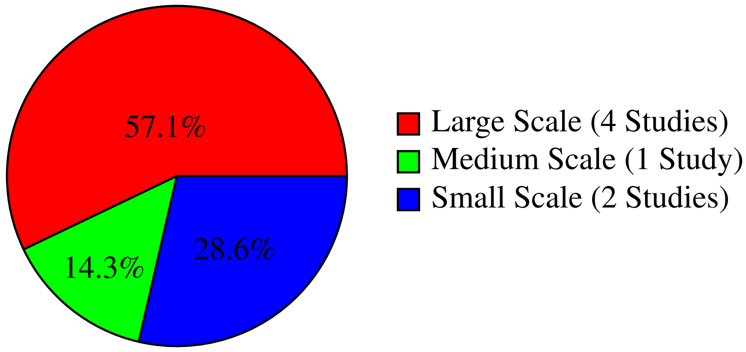

Fig. 2.5 Distribution of Experimental Scales in VR Studies

Additionally, the Microsoft Hololens [86] and Google Daydream VR [94] were utilized in particular scenarios, emphasizing their roles in enhancing mixed reality interactions and providing accessible mobile VR experiences, respectively. These instances highlighted the diversification of VR technologies in adapting to different educational and training needs within healthcare settings. Another study referenced the use of a customized VR environment, further illustrating the innovative and tailored approaches being explored in VR applications to meet specific research and training objectives.

The frequent incorporation of the Oculus Rift and HTC Vive across multiple studies indicated their significant role in advancing VR methodologies in healthcare education. This trend not only reflected the technological preferences of current research but also suggested a broader adoption and trust in these systems to deliver effective and immersive training solutions.

Experiments scale and Scenario analysis

The experimental scale of the included VR papers was sorted into three groups according to the number of participants (n): small scale ($n \leq 10$), medium scale ($10 < n < 20$), and large scale ($n \geq 20$). Out of these, five studies were classified as large scale, one as small scale, and one paper did not detail its experimental setup. A summary is given in Figure 2.5.

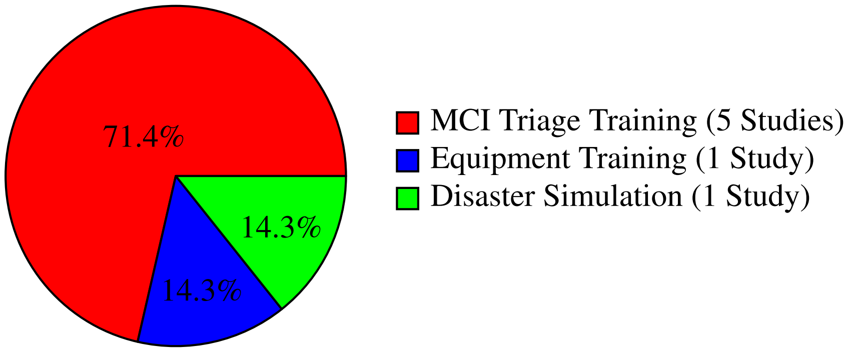

Fig. 2.6 Distribution of VR scenarios in healthcare studies

In terms of scenarios, out of the seven VR papers centered on healthcare, five [108, 13, 7, 15, 94] were related to MCI triage training. These papers each offered scenarios such as vehicular accidents or firearm incidents to train triage skills. The other two studies [86, 27] discussed ambulance bus training and a disaster simulation system involving various natural disasters, respectively. A summary is given in Figure 2.6.

Advantage and disadvantage of VR in healthcare

Several advantages brought by VR are highlighted in the included papers. These advantages include:

- Low Implementation Cost

  Many studies have pointed out that one of the most significant benefits of VR technology is its low implementation cost. For example, in [108], the authors conducted a detailed cost analysis comparing the expenses associated with VR-based MCI training scenarios to those of traditional live simulations. Their findings revealed that the costs associated with running VR scenarios were less than 10% of the costs required for comparable live simulations over the course of a teaching year. This cost-effectiveness makes VR an attractive option for frequent training sessions and for institutions with limited budgets.

- High Immersion

  The ability to create highly immersive environments is another key advantage of VR technology, as discussed in papers such as [7] and [86]. This high level of immersion can engage participants in a more compelling learning experience, simulating real-life scenarios and environments that are either too dangerous or impractical to recreate in the real world. For instance, medical students can perform surgical procedures or respond to emergency situations without the risks associated with actual operations.

- Similar Efficiency and Satisfaction

  The effectiveness of VR-based training in replicating real-world training experiences is well documented. Several studies, including those using assessment tools such as the NASA-TLX framework ([7]) and Presence Scales ([13]), have evaluated the workload and satisfaction levels of participants in VR training environments. The results indicated that VR training not only matched the efficiency of real-world training but also enhanced participant satisfaction. Participants reported high levels of engagement and satisfaction, attributing this to the interactive and immersive nature of VR training.

These points underscore the potential of VR as a transformative tool in education and training, offering cost-effective, engaging, and efficient alternatives to traditional methods. The included papers also mentioned several disadvantages encountered during experiments with VR technology, which include:

- Low Fidelity

  A frequently noted limitation in the included papers is the low fidelity of VR training scenarios. Current VR technology often fails to replicate the full spectrum of sensory feedback required for a fully immersive experience. For example, as mentioned in [107], VR was unable to replicate human interaction and emotional immersion as effectively as live simulations, although it allowed for better focus on triage skills without the emotional distractions. Such limitations may not only affect the learner's ability to perform in virtual scenarios but also impact the transferability of these skills to real-world applications. [12] also mentioned the technical issues such as connection loss and the simulation halting unexpectedly, which could disrupt the training process.

- Dizziness and Discomfort
  The occurrence of dizziness or vertigo, also known as "virtual reality sickness," is another significant concern highlighted in the literature. This condition is a form of motion sickness that results from a conflict between the visual motion cues one perceives and the physical motion cues one experiences. This discord can lead to severe disorientation and nausea, detracting from the user's ability to engage effectively with the VR environment. The severity of these symptoms can vary widely among individuals, complicating the standardization of VR training modules. Prolonged exposure to VR under such conditions can also lead to longer-term health issues, such as persistent postural-perceptual dizziness, which could deter users from frequent use of VR for training. [107] also discussed the discomfort brought by VR technology.

These challenges underscore the need for ongoing research and technological improvements in VR systems to enhance the realism and user comfort in educational and training settings. Addressing these issues is crucial for wider adoption and acceptance of VR as a valuable training tool.

2.2.4 AI Literature Overview

After reviewing the selected papers, we categorized them into several groups based on their research topics, which are 1) AI and ML in Emergency Healthcare and Triage Systems; 2) Immersive Training Technologies; 3) Decision Support and Patient Flow Management; 4) AI in Emergency Disaster Response and Training; and 5) Patient Preference and Healthcare in Smart Cities.

Advancements in artificial intelligence and machine learning have revolutionized emergency healthcare by enhancing the efficiency and accuracy of triage processes. These technologies provide crucial support in decision-making and patient management, particularly in high-stress environments such as emergency departments (EDs). Yan et al. [159] proposed applying ML methods to improve the manual triage process in Australian Emergency Departments (EDs), leveraging patient history and clinical data for enhanced triage accuracy. Yun et al. [163] developed an ML algorithm for predicting critical care outcomes in adult patients using initial ED triage information, showcasing the model's superior discriminative performance. Kerr et al. [82] proposed a fuzzy triage system using visual and tactile sensing for healthcare robotics, demonstrating potential applications in emergency situations and home assistance for the elderly and disabled. Kang et al. [79] developed an AI algorithm to predict the need for critical care in emergency medical services (EMS), demonstrating its accuracy over traditional triage tools.

Immersive technologies are gaining traction as effective tools for training and simulation in emergency healthcare scenarios. By creating realistic, controlled environments, these technologies help improve the preparedness of first responders and healthcare professionals. Sherstyuk et al. [139] explored hand-assisted viewing in virtual environments, improving performance in visual tasks and altering user behavior, suggesting implications for training in emergency scenarios. Surer et al. [143] developed a scenario-based game generation framework to convert structured emergency training scenarios into serious games for Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNe) domain training, highlighting the potential of gamification in emergency preparedness. Mossel et al. [112] introduced VROnSite, a platform for immersive training of first responder squad leaders, emphasizing the importance of navigation technologies in simulating stress and exhaustion for decision-making training. Bjørn et al. [19] proposed using immersive Cooperative Work Environments in Virtual Reality for cooperative patient resuscitation scenarios, revealing functional errors not apparent in other architectural representations and emphasizing the importance of design dimensions in immersive environments.

The integration of decision support systems and advanced patient flow management strategies in emergency departments can significantly improve service delivery and patient outcomes. These systems utilize AI to enhance operational efficiency, reduce waiting times, and optimize resource allocation. Chonde et al. [36] employed discrete event simulation to compare patient flow models, providing insights into optimizing patient management in EDs. Billis et al. [17] outlined IntelTriage, a smart triage system for EDs that automatically prioritizes patients and continuously monitors vital signs, addressing ED overcrowding and enhancing care quality. Dehghani Soufi et al. [42] proposed a DSS for patient triage using a hybrid approach, achieving high accuracy and significantly reducing mis-triage, enhancing emergency nursing triage accuracy. Georgopoulos and Stylios [56] explored advancements through the integration of AI and machine learning technologies, aimed at enhancing triage, patient care, and disaster response efficiency. Cabrera et al. [29, 28] conducted studies focusing on improving Healthcare EDs through the use of Agent-Based Modeling (ABM) and DSS, optimizing ED performance and addressing issues like overcrowding and the quality of care.

Artificial intelligence frameworks have become increasingly notable in emergency disaster response and training, offering dynamic solutions that adapt to complex and evolving scenarios. These technologies facilitate rapid decision-making and enhance training through realistic simulations. Nistor et al. [115] reviewed AI frameworks for disaster response, highlighting the potential of AI in dynamic environments. Donevant et al. [46] validated an informatics tool for triage in mass casualty incidents involving irritant gas syndrome agents, providing insights for future research. Marchiori et al. [102] introduced a certified triage system using ML and NLP, operational at a major European tele-medicine provider, for personalized care recommendations.

As urban areas evolve into smart cities, AI-based systems are becoming essential in managing healthcare delivery and aligning it with patient preferences. These systems are tailored to enhance the accessibility and efficiency of medical services in densely populated areas. Kong [85] analyzed AI-based online triage models for hospitals within sustainable smart cities, exploring patient preferences for medical treatment. Zijie et al. [165] discussed the application progress of smart glasses for triage during MCI, emphasizing the technology's promise in enhancing triage efficiency and patient survival rates.

2.2.5 AI literature statistical result

To categorize the AI and healthcare articles, we used application scenario (e.g., in-hospital or pre-hospital), objective, and technique as the main categories.

Scenario analysis

Among the 15 papers on AI and healthcare, 11 targeted research on the in-hospital phase, aiming to enhance triage efficiency using various AI technologies. Two papers explored the pre-hospital phase, and two others focused on theoretical frameworks, enriching the discussion around the application of AI in emergency healthcare environments. A summary is given in Figure 2.7.

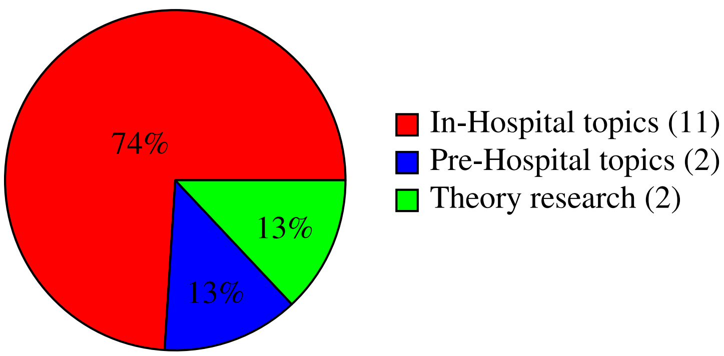

Fig. 2.7 AI and healthcare scenarios (Total: 15)

Objectives of AI and healthcare

The AI in healthcare primarily focused on three key objectives: 1) auto-triage decision making, 2) Emergency Department (ED) resource optimization, and 3) ED patient prediction. A summary is given in Figure 2.8. The majority of studies [84, 42, 56, 55, 102, 144, 167, 166, 82, 85] concentrated on utilizing AI for auto-triage or providing supplementary support for triage decision making.

Following this, several studies [17, 28, 30] aimed to optimize resource allocation in the ED and minimize waiting times. For instance, Cabrera et al. [28] focused on optimizing staff configurations to reduce patient stay lengths in the ED.

Finally, the remaining studies [126, 79] emphasized predictive applications of AI within the ED context. Pereira et al. [126] explored predictive models to forecast triage waiting times, enhancing patient flow efficiency. Additionally, Kang et al. [79] developed an AI algorithm to predict critical care needs in prehospital settings, enabling quicker and more precise decision-making in emergency medical responses.

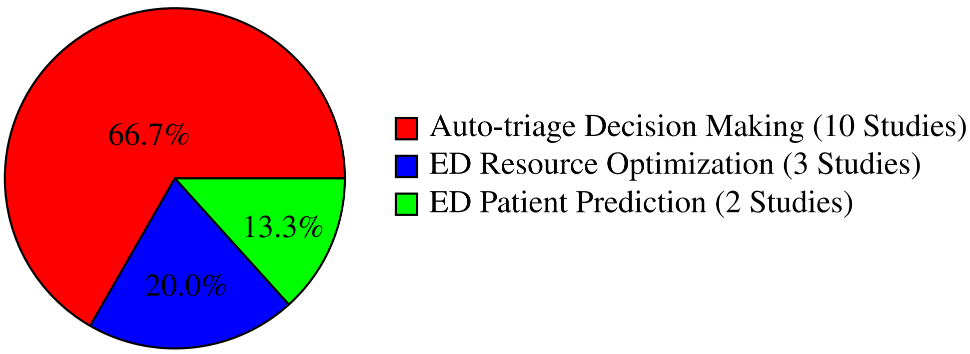

Fig. 2.8 Objectives of AI Research in Healthcare (Total: 15 Studies)

Techniques of AI and healthcare

Different techniques were used in the included papers. Most of them employed traditional machine learning algorithms. For instance, fuzzy logic emerged as a powerful tool in several studies aimed at enhancing decision-making frameworks within medical settings, particularly in emergency departments where imprecision and clinical uncertainty are rampant. In [82], a robotic system employed a fuzzy triage system to assess vital signs and triage patients accurately, utilizing fuzzy logic to interpret sensory data for nuanced decision-making in emergency scenarios. Similarly, [42] developed a decision support system that integrated fuzzy logic with rule-based reasoning, enhancing the system's capability to accurately categorize patient urgency by processing vital sign data. Furthermore, [56, 55] used Fuzzy Cognitive Maps (FCMs) to support triage decisions for elderly patients, applying fuzzy logic to model complex patient data and improve the triage process by considering various health factors and their interconnections.

The comparative analysis of these fuzzy logic applications highlighted several key aspects. All three studies harnessed fuzzy logic's capacity to handle uncertainty and partial truths, allowing for more nuanced assessments than those afforded by binary or discrete systems. [42] uniquely combined fuzzy logic with rule-based reasoning to create a structured yet adaptable decision framework. In contrast, [56, 55] integrated fuzzy logic within cognitive maps to enhance the predictive capabilities based on complex interdependencies. Additionally, while [82] focused on immediate physical assessments using a robotic system, studies [42] and [56, 55] deployed fuzzy logic more broadly within decision support systems to enhance overall triage and patient management processes. This diversity in application underscored the adaptability of fuzzy logic to various aspects of emergency medical care, improving outcomes by facilitating precise and flexible decision-making.

Bills et al. [17] used Logistic Regression (LR) and general machine learning approaches to assist in triage decision-making. Cabrera et al. [28] used Monte Carlo and Pipeline schemes to optimize ED staff configuration. New techniques, such as neural networks (NN), were also employed in many studies, including [30, 125, 144, 79]. We noted that a few studies compared the performance of traditional machine learning algorithms with NNs (e.g.,

[30]), and the results showed that NNs generally achieved the best performance among other algorithms.

2.3 Most recent Literature Review

We have conducted a new round of literature review in September 2024 to collect new development during 2022 and 2024. We used the same search criteria used in the previous round. After completing the screening process, we selected a total of 10 relevant papers (5 focused on VR and 5 on AI), that provided fresh insights into their application in emergency healthcare.

The integration of these new papers into the review ensures that the most recent advancements are reflected in the study. This updated literature not only reinforces existing knowledge but also offers new perspectives on how both VR and AI technologies can enhance triage accuracy, decision-making, and training effectiveness in emergency medicine. The incorporation of these papers helps to maintain the relevance and rigor of the research, ensuring that it is grounded in the latest developments in the field.

2.3.1 VR Literature Overview

VR and Triage

Harada et al. [62] examined the effectiveness of using VR for training paramedic students in the START method. Seventy students were divided into two groups, one receiving VR-based training and the other traditional live lectures. The results showed that the VR group performed significantly better in practical tests, particularly in scenarios that required more detailed observations and interactions, while both groups demonstrated similar levels of knowledge retention in written exams. The study concluded that combining VR with traditional teaching methods, such as live lectures and simulations, offers an ideal approach for enhancing practical skills in emergency medical training, providing a more immersive and engaging learning experience.

Behmadi et al. [9] conducted a study comparing VR-based medical education with traditional lecture-based methods for teaching START triage lessons to emergency medical students. The study involved 44 students who were divided into two groups: one group received lecture-based education, while the other group was taught using virtual simulation. Although the VR group's performance scores were slightly higher, the difference was not statistically significant. However, students expressed higher satisfaction with the VR-based learning experience. The study concluded that while VR can enhance student engagement and satisfaction, it is not necessarily superior to traditional methods in improving learning outcomes. Further research with larger sample sizes is needed to confirm VR's effectiveness in medical education

Vogt et al. [152] explored the use of VR for training paramedics in handling mass-casualty incidents, which are challenging and rare but require a rapid and efficient response. The study followed a user-centered design approach, engaging 32 paramedics and trainers in the development process. Additionally, a business model for the training was evaluated by ten ambulance service coordinators. Results showed that paramedics were highly motivated and rated the training positively in terms of user experience. Trainers viewed the VR training as a valuable complement to existing practices rather than a replacement. The study concluded that co-developing VR training with paramedics and trainers offers an effective method for practicing triage in complex scenarios, although further refinements are needed. The research also provided a framework for designing similar VR training environments for medical professionals.

VR and Emergency Healthcare

Abbas et al. [1] conducted a systematic review to assess the role of VR in simulation-based emergency skills training. The review included 34 studies published between 1998 and 2022, covering a wide range of emergency care skills across different clinical settings. While the studies demonstrated the educational benefits of VR, such as knowledge retention, skill performance, and high levels of user satisfaction, there was no clear evidence of its impact on patient outcomes. Additionally, although VR was considered a valuable and immersive training tool, the review found insufficient evidence regarding its cost-effectiveness. The authors concluded that while VR offers significant potential in emergency skills training, further research is needed to assess its long-term impact on clinical outcomes and cost benefits.

Mahling et al. [96] conducted a large-cohort study to evaluate medical students' perceptions of VR in emergency medicine training. The study included 129 fourth-year medical students at the Medical Faculty in Tübingen, Germany, who voluntarily participated in a VR-based training session. Results showed that most students believed VR was useful for quickly conveying complex information, complementing mannequin-based courses, and even being used for exams. However, female students were less positive in their feedback compared to male students. While 69% of participants felt immersed in the VR experience, only 3% felt confident with the medical content presented. Additionally, factors like gender, age, and prior experience with VR or emergency medicine did not influence the students' test scores. The study highlighted the overall positive attitude toward VR-based training, but it noted lower confidence levels and potential gender differences that should be addressed when incorporating VR into the medical curriculum.

2.3.2 AI Literature Overview

AI and Triage

Levine et al. [89] and Pasli et al. [123] both explored the diagnostic and triage capabilities of AI models in healthcare settings. Levine et al. compared GPT-3's diagnostic and triage accuracy to that of physicians and lay individuals using clinical vignettes. GPT-3 was able to correctly diagnose 88% of cases, surpassing lay individuals (54%) but trailing slightly behind physicians (96%). Its triage accuracy (71%), however, was similar to that of lay individuals (74%) and considerably lower than physicians (91%). Despite its near-physician-level diagnostic performance, Levine et al. emphasized that GPT-3's triage capabilities needed improvement before clinical deployment. Pasli et al. took this investigation further by evaluating GPT-4, specifically in an emergency department (ED) setting. The study involved 758 patients and compared GPT-4's triage decisions, trained on local ED rules, with those of ED physicians. The results showed near-perfect agreement between GPT-4 and physicians, with a Cohen's Kappa score of 0.899, suggesting that GPT-4 could be highly effective in supporting triage decisions. However, the authors stressed that GPT-4 should be used as a complementary tool rather than a replacement for human triage teams. Both studies highlight the potential of GPT models in healthcare, with GPT-4 showing more promise in triage accuracy, though further research is needed to validate its broader implementation.

Joudar et al. [78] conducted a systematic review focused on the use of AI in triage and priority-based diagnosis for autism spectrum disorder (ASD) and gene contributions. The review analyzed 363 articles published between 2017 and 2022, ultimately narrowing down to 18 relevant studies. These were categorized into three groups: triage-based diagnosis methods, prioritization of risky genes, and e-triage using telehealth. The review highlighted the potential of AI and machine learning in improving the accuracy and efficiency of ASD diagnosis by prioritizing patients based on severity levels and identifying significant genetic factors. However, it also identified challenges, such as the complexity of integrating AI technologies with ASD diagnosis and the need for further research to address gaps in knowledge. The study concluded by proposing a multi-criteria decision-making approach for triaging ASD patients according to their severity and emphasized the need for more focused research on AI's role in ASD diagnosis.

AI and Emergency Healthcare

Chenais et al. [33] conducted a narrative review exploring the applications of ML in emergency medicine, highlighting advancements and challenges in the field. The review focuses on ML's role in triage, disease-specific risk prediction, medical imaging, and ED operations. It also addresses how ML could enhance causal inference in clinical research. Despite the promise of ML in improving patient care and operational efficiency, the authors discuss several barriers to real-world implementation, including dataset biases, proprietary systems, and regulatory hurdles. The review concludes by calling for the development of quality controls and ethical frameworks to ensure safe and effective integration of AI and ML in emergency medical practice.

Vearrier et al. [150] explored the integration of AI in emergency medicine, focusing on its benefits, risks, and recommendations for its use. AI, or "augmented intelligence" as preferred by the American Medical Association, offers significant potential to enhance physician skills without replacing them. The paper emphasizes that AI technologies, such as those used in radiography and diagnosis, can improve patient care by providing cognitive assistance to emergency physicians (EPs). However, the authors stress that AI should be carefully vetted for legal and ethical standards, with an emphasis on patient confidentiality and professional liability. EPs are encouraged to partner with AI, using it as a supportive tool while maintaining the clinical context and humanistic goals of patient care. Ultimately, AI can help streamline workflow and enhance decision-making in emergency settings, but it must be implemented with caution and adequate safeguards.

2.4 Summary

This chapter deliberately adopted three separate rounds of literature review to balance breadth, methodological rigor, and recency. The pre-COVID scoping review, the systematic PRISMA review with broader database coverage, and the 2024 update each added distinct value to refining the research scope. The three rounds of literature review identified notable limitations and gaps within the fields of VR, AI, and emergency healthcare. Despite some exploratory efforts in applying VR to emergency healthcare, these studies often fell short in certain areas:

1. Evaluation Methodology Predominantly Based on Interviews:

   Many studies focusing on VR in healthcare primarily used interviews as their evaluation method. These interviews were often structured around various assessment schemas, like the NASA Task Load Index (TLX), with an emphasis on measuring user experiences. This approach may have compromised the objectivity of the research outcomes. Moreover, while these studies explored VR applications in healthcare, they frequently lacked comprehensive incorporation of AI techniques in their evaluation. Our proposed solution is a hybrid approach, combining traditional interviews with AI-based evaluation methods. This not only captures user experiences but also objectively assesses the effectiveness of the training system. Leveraging AI in evaluation provides a more nuanced understanding of both the user experience and the actual efficacy of the VR training.

2. Limitations in Interaction Methods:

   Current research also revealed constraints in interaction methods used within VR environments. Predominantly, interactions were facilitated through simple methods like gazing or pointing, which limited the scope of data collection. To overcome this, our study proposed an enhanced interaction approach that incorporates VR controllers. This modification allows for more complex and realistic interactions within the VR environment. Additionally, this approach broadens the scope of data collection, extending it to include not only traditional metrics but also audio data. This comprehensive data collection facilitates a deeper analysis and understanding of user interactions and responses within the VR environment.

In summary, these identified limitations highlight the need for more robust, objective, and interactive methodologies in the intersection of VR, AI, and emergency healthcare research. The suggested enhancements aim to address these gaps, offering a more comprehensive and effective approach to studying and applying VR in emergency healthcare training and evaluation.

Chapter 3

Methodology

3.1 Introduction

This research adopts a Design Science Methodology to develop and evaluate a VR learning tool integrated with AI for emergency healthcare training, specifically focusing on MCIs. Design science is a systematic approach that emphasizes the creation and evaluation of innovative artifacts aimed at solving real-world problems. In this study, the VR learning tool serves as the artifact, designed to address the need for more effective and scalable MCI training solutions.

By using the design science method, this research follows a structured process involving problem identification, VR learning tool development, evaluation, and iterative refinement. The aim is to create a solution that not only enhances the decision-making and triage capabilities of healthcare professionals but also allows for a detailed analysis of training outcomes through AI-driven performance metrics. This approach ensures that the tool is fully tested, improved through multiple iterations, and assessed for its effectiveness in preparing healthcare workers for real-world emergency scenarios.

3.2 Research Process and Design Science Phases

This research followed the DSR methodology, which structures inquiry into problem identification, artefact design and development, evaluation, and reflection. To improve clarity and cohesion, this section provides an integrated overview of the research process and explicitly aligns the phases undertaken in this thesis with the DSR framework.

Phase 1 – Problem Identification and Literature Exploration

The research began with iterative rounds of literature review (see Chapter 2). The first round, conducted before the COVID-19 pandemic, was a broad scoping review that did not fully adopt systematic standards. The second round followed the PRISMA methodology and expanded the search across additional databases. The third round was a recent update (2023–2024), ensuring the work reflects the latest advances in AI and VR for emergency healthcare training. Together, these reviews clarified the knowledge gaps in current triage training and informed the research questions.

Phase 2 – Conceptual Framework and Study Design

Based on the identified gaps, a conceptual framework integrating VR and AI for pre-hospital triage training was developed. This phase included defining objectives, scope, and design requirements, supported by both literature and expert input.

Phase 3 – Artefact Development (VR Training Tool and AI Evaluation Module)

The VR learning tool was developed through an *iterative, design-cycle approach*, consistent with DSR. Initial prototypes were designed, tested, and refined based on expert feedback. Parallel to this, AI methods (e.g., BERT- and GPT-based models) were integrated to provide automated evaluation of METHANE reports. This development process reflects the iterative nature of DSR, where artefacts evolve through cycles of construction and refinement.

Phase 4 – Experiment Design and Implementation

Following development, experimental procedures were established to evaluate the artefact.

This included defining participant criteria, recruitment, and the setup of controlled scenarios. The experiment design was guided by both methodological rigour and practical feasibility.

Phase 5 – Evaluation and Data Analysis

The artefact was evaluated through structured experiments involving participants performing triage in VR scenarios. Quantitative and qualitative data were collected, including performance scores, METHANE reports, and feedback. Analysis combined AI-based assessments with statistical and thematic methods. This aligns with DSR's emphasis on artefact evaluation.

Phase 6 – Reflection and Contribution

The final phase synthesised findings, assessed the artefact's effectiveness, and reflected on contributions to theory and practice. Limitations were acknowledged, and future research directions were proposed.

By structuring the research in these six phases, this thesis ensures that the process is transparent and coherent, directly addressing concerns about scattered discussions across Chapters 3 and 4. Each subsequent section of this chapter and Chapter 4 corresponds to the phases outlined above.

3.3 Problem Identification

The research process begins with Phase 1: Problem Identification. This phase highlights the fundamental challenges in current MCI training and introduces the research questions that guide the subsequent phases of the study. By clearly articulating the problem space, Phase 1 provides the foundation for the design, development, and evaluation activities that follow. Traditional MCI training methods are not only costly but also limited in their frequency due to the unpredictable nature of such events. As a result, emergency healthcare professionals often lack sufficient hands-on experience to make rapid and accurate decisions when confronted with real emergencies.

In the introduction chapter, we have identified the research questions guiding this study, which are:

- Research Question 1: What are the essential components required to build an effective VR learning tool?
- Research Question 2: How can AI be integrated effectively into a VR learning tool to enhance data analysis?
- Research Question 3: How can the training effectiveness of the VR learning tool be evaluated and improved?
  - Research Question 3.1: How can data from VR learning tool be used to assess the competence of participants in MCI scenarios?
  - Research Question 3.2:Can a VR environment convince participants to behave in a way that will allow their performance to be assessed
  - Research Question 3.3: How well do participants accept VR technology for triage training in MCIs?

By addressing these questions, our research aims to overcome the limitations of traditional training methods. The integration of VR and AI technologies offers an innovative solution to simulate real-world MCI scenarios, providing immersive and dynamic training environments. This study's focus on improving decision-making capabilities through VR learning tools directly responds to the increasing need for more effective and accessible MCI training options.

3.4 Research Design and Development

This section corresponds to Phase 2 (Conceptual Framework and Study Design) and Phase 3 (Artefact Development) of the research process. In this stage, the problem definition, research

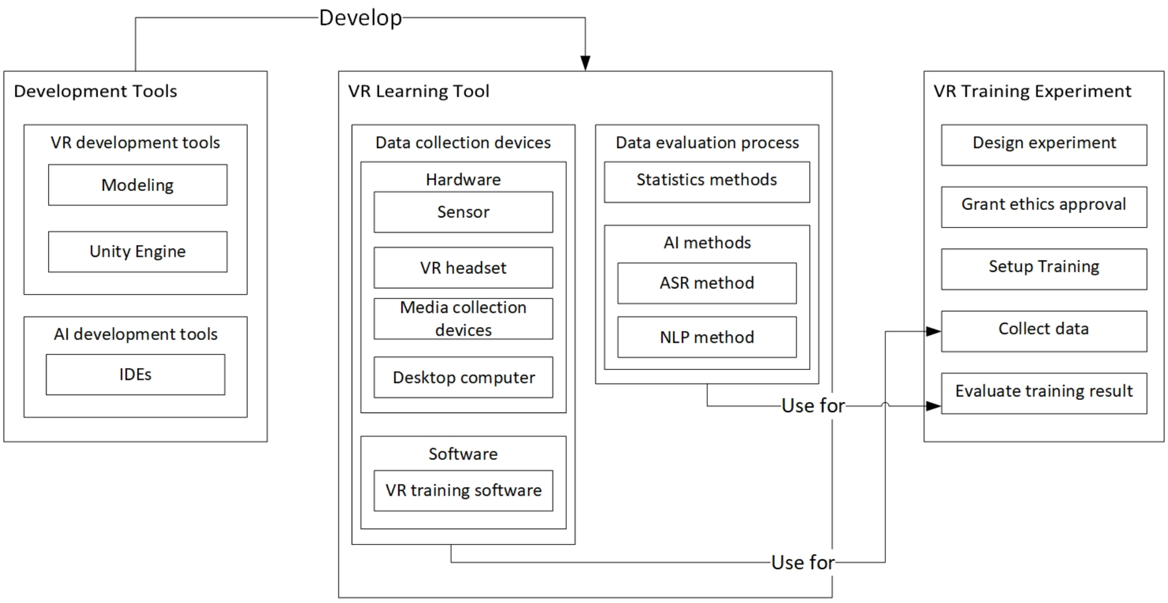

Fig. 3.1 Research Framework

questions, and system requirements are translated into the iterative design and construction of the VR training tool and the AI evaluation module. The comprehensive research framework outlined in Figure 3.1 demonstrates a systematic approach to developing and executing a VR learning tool and training experiment, which we followed to ensure efficacy and educational value. There were three major components in our research framework: the Development Tools, the VR Learning Tool, and the VR Training Experiment.

Our process began with the use of VR development tools specifically for modeling and building interactions. This step aims to address **Research Question 1** by identifying the essential components required to create an effective VR learning tool. We developed the virtual environments and objects by Using the Unity Engine which participants would interact with. AI development tools were also employed to construct evaluation methods, crucial for the educational aspect of the simulation.

Moving to the VR Learning Tool phase, we incorporated a suite of hardware and software to deliver the VR experience. This phase also directly addresses **Research Question 1** by focusing on creating an intuitive and immersive experience. Data collection devices, such as sensors and headsets, tracked and recorded user interactions, while desktop computers served as the backbone for data processing and storage. The VR training software was designed to enhance user engagement and interaction with the virtual environment.

The Data Evaluation process employed statistical methods and AI-driven tools to analyze the data collected. This step aligns with **Research Question 2**, which examines how AI can be integrated into VR-based healthcare training to enhance data analysis. AI models such as Whisper for speech recognition and BERT/GPT for semantic analysis allowed us to capture detailed participant interactions and evaluate performance, addressing **Research Question 3.1** by assessing how data such as task completion time and AI-based performance evaluations can be used to gauge competence in MCI scenarios.

For the VR Training Experiment, we designed an experiment to measure specific learning outcomes, directly addressing **Research Question 3** and its sub-questions. The setup allowed us to explore whether the VR environment could simulate real-world scenarios and engage participants in a way that allowed for accurate performance assessments (**Research Question 3.2**). Through continuous data collection during the sessions, including motion tracking and decision-making inputs, we gathered a comprehensive dataset on participant engagement and interaction. This data helped us evaluate **Research Question 3.3** by assessing user acceptance and engagement with the VR learning tool in the context of MCI training.

Finally, after each session, we conducted an evaluation based on predefined metrics and learning outcomes. This iterative process allowed us to refine the tool continuously and enhance its training efficiency, contributing to a deeper understanding of how to improve VR training and its applicability in real-world healthcare settings.

In the following sections, we describe the details of each major component in depth:

3.4.1 Component 1: Development Tools

We had two types of development tools. One was for VR, and the other was for AI. VR development tools consisted of modeling tools and the Unity Engine. We used modeling tools to build the 3D objects and assembled those objects in Unity Engine to build the VR scenario. AI development tools included various development environments. We implemented the AI algorithms based on those environments. In the following sections, we present details of these development tools.

VR development tools

- Modelling

  VR development tools consisted of modeling tools and the Unity Engine. As VR modeling involved a full process and utilized various tools, we demonstrated our modeling process in Figure 3.2.

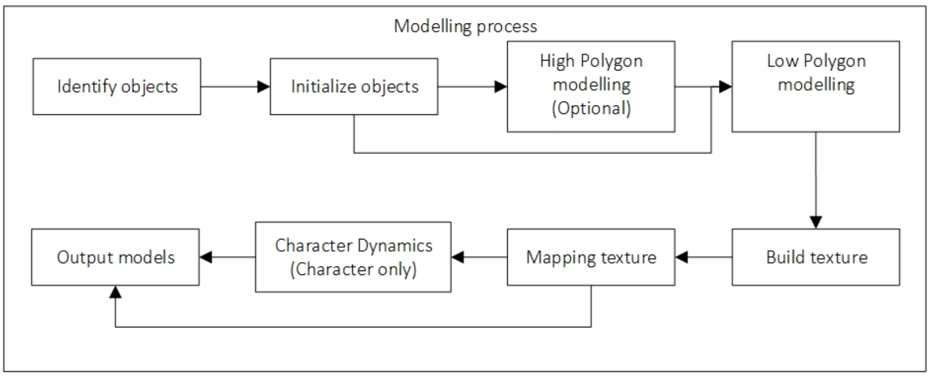

Fig. 3.2 Modeling workflow

  Figure 3.2 indicated a structured modeling process we used for creating various objects within a virtual environment. The initial step involved identifying the necessary objects for our project, which included items like vehicles, buildings, and humanoid characters. These objects were categorized as either static or dynamic, depending on their role within the simulation. For each object, we established a target model and sought out reference images from diverse sources to accurately represent its appearance.

Once we identified our objects, we proceeded to initialize the models using Blender and Google SketchUp. These tools allowed us to draft and build static objects, like cars and streetlights, which populated our virtual scene.

Beyond static models, dynamic objects, specifically humanoid characters, underwent a process called polygon optimization. Our project distinguished between two types of humanoid character modeling. The first type was standard, similar to other 3D character models. The second type required special attention to detail, such as articulating joint connections or mesh modifications to represent medical conditions like trauma or bone fractures. For standard models, we began with low polygon modeling. For the more complex models, we initially crafted high polygon models to incorporate specific details like scars, then reduced these models to a low polygon format. For these steps, we utilized Adobe Fusion for low polygon modeling, ZBrush for high polygon modeling, and Autodesk Maya for both high and low polygon modeling.

As we developed the polygon models, we simultaneously created textures. This was particularly intricate for virtual casualties, where we needed to tailor the textures to accurately represent medical conditions. Adobe Photoshop was our tool of choice for this task. Following texture creation, we used Maya or Unity to merge the textures with the polygon models.

To animate our characters, we introduced dynamics, also known as ’bones’ or ’joints,’ into the humanoid figures. Each character was equipped with around 20 joint groups to emulate human movement accurately. For example, the ’Shoulder’ joint group, which included two joints, connected to the ’Upper Arm’ and ’Chest’ joint groups, allowing us to simulate the movement and rotation of a human shoulder. The construction of these dynamics was performed in Maya and the Unity Engine.

Finally, after meticulous construction and texturing, we exported the models as *.FBX files. These files were then imported into the Unity Engine, ready to be integrated into the virtual environment we were crafting. The process captured in the image ensured that each model, whether a static background object or a dynamic character with complex animations, was created with precision and attention to detail, resulting in a rich and engaging virtual experience.

- Unity Engine

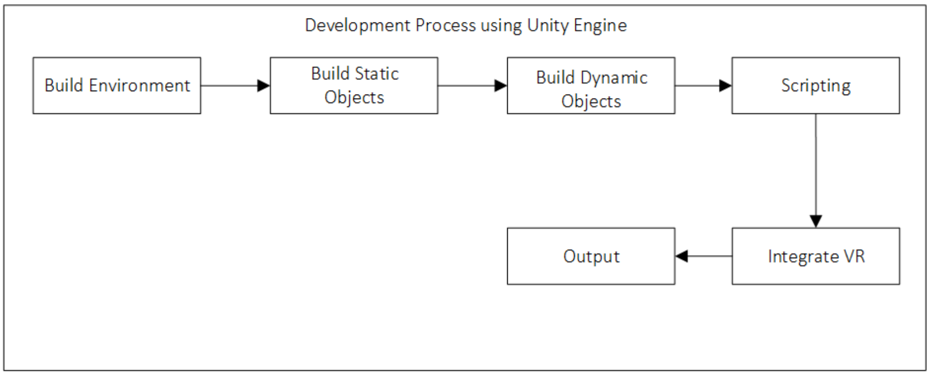

Fig. 3.3 Unity workflow

Figure 3.3 showed the development process using Unity Engine. Our development process began with the construction of the environment. This foundational step was pivotal as it established the setting for the simulation. We merged a variety of assets from the Unity asset marketplace with our own custom designs to create a detailed urban landscape. This landscape set the stage for all the interactive elements that followed.

Once the environment was established, we proceeded to place static objects within it. Items such as vehicles, rocks, and roads were strategically positioned to portray a scene of accident or disaster. These static objects, while non-interactive, played a crucial role in adding context and depth to the scenario, enhancing the visual fidelity and providing a realistic backdrop for the simulation.

The third phase of our workflow was the development of dynamic objects, which was significantly more intricate than placing static objects. Dynamic objects were essential for interactivity in the simulation. These often included humanoid characters, so-called virtual casualties, that the participant could interact with. We carefully planned their properties, such as their placement, interaction methods, and behavior patterns, to ensure that they behaved in believable and appropriate ways within the context of the training scenario.

To fine-tune the performance of these dynamic objects, we utilized scripting and blend animations. Scripting was an essential component, where we used tools such as Microsoft Visual Studio to craft the code that governed the interactions and behavior of the dynamic objects. For example, a dynamic object designed to represent a paramedic might be programmed to open the airway to an injured person when the participant performed the appropriate action.

Upon completing these steps, we developed a fully functional digital simulation. Yet, the process extended beyond this point. The resulting simulation was then integrated with VR technology to further enhance its immersive qualities. By integrating VR, we allowed participants to interact with the environment in a more intuitive way, significantly narrowing the gap between a virtual exercise and real-life practice.

- AI development tools

  The AI development tools were employed to craft machine learning (ML) algorithms, including those focused on speech. Our primary Integrated Development Environment (IDE) of choice was Visual Studio Code, running on the Windows Operating System (OS), for the development and testing of AI algorithms. Additionally, we employed Pycharm and Ubuntu as alternative IDE and OS options, respectively, for further debugging purposes. Python served as the predominant programming language for most ML algorithms, and its widespread adoption was attributed to several key advantages:

1. Python enjoyed robust support from a vast ecosystem of third-party add-on toolkits and libraries.
2. Its highly readable programming style facilitated code comprehension and collaboration.
3. Python's high portability ensured seamless adaptation to diverse operating systems, while its extensibility enhanced development flexibility.
4. Python was both free to use and boasted an abundance of open-source projects available to end-users.

Given these considerations, we chose Python as our primary programming language for AI development. Visual Studio Code was selected as the primary IDE due to its strong support for Python programming, which enhanced development efficiency. In addition to Visual Studio Code, we leveraged Pycharm to compile and test certain C++ ML projects.

3.4.2 Component 2: VR learning tool

VR has been shown to be useful in healthcare training by previous research (Fertleman et al., 2018 [50]; Gunn et al., 2018 [61]; Kyaw et al., 2019 [87]; Bhagavathula et al., 2018 [14]; Butt et al., 2018 [26]). On the one hand, VR can provide an immersive training experience and create scenes in a cost-effective way. On the other hand, VR sensors can record various data and provide a more precise evaluation method for the training results.

Dealing with MCI triage training is a key skill in paramedics training for emergency responses. Various triage systems (Zachariasse et al., 2019 [164]) have been developed for the prioritization of casualties in the emergency department. High-cost methods such as medical mannequins are used to train paramedic students. However, some signs, such as the color of the skin, are difficult to customize on those mannequins. In this project, we aimed to design a VR learning platform to simulate MCI scenes and train triage skills for paramedics, which also addressed Research Question 2. The VR learning platform consists of a Data Collection Platform and a set of Data Evaluation Processes. Details of these two major components are given below:

Data collection platform

The data collection platform consisted of hardware and software. The hardware included four sub-components: motion sensors, a VR headset (HMD), media data collection devices, and a desktop computer.

- VR device

    1. Oculus Rift Virtual Reality Headset

        The Oculus Rift comprised a pair of virtual-reality goggles compatible with desktops or laptops. It featured dual screens, each presenting a separate image for the left and right eyes. Over these screens, lenses were positioned to concentrate and reshape the visuals, forming a stereoscopic 3D display. Embedded sensors within the goggles tracked the user's head movements, dynamically adjusting the visuals. This created an immersive 3D environment where users could freely explore. Figure 3.4 shows a visual representation of the Oculus Rift headset.

        

        Fig. 3.4 Oculus Rift Headset

    2. Oculus Sensors

        The Constellation system's design offered significant flexibility by utilizing a single camera, enabling users to experiment with various sensor configurations while enhancing tracking capabilities. This design allowed for seamless integration with Oculus Touch controllers. In the context of this research, the use of three sensors was crucial for supporting immersive room-scale VR training sessions. Figure 3.5 illustrates the Oculus Sensor.

Fig. 3.5 Oculus Sensor

3. Oculus Touch Controller

   Oculus Touch stood as a groundbreaking VR input system launched alongside the Oculus Rift for consumers. Comprising two intricately monitored controllers, it delivered an unparalleled feeling of "hand presence" to users, granting developers a wide array of natural interaction capabilities. Figure 3.6 shows the Oculus controller.

Fig. 3.6 Oculus Rift Controller

- Media data collection devices

  We used several cameras and smartphones to collect media data from the participants. These media data collection devices were hosted at different positions around the training ground. Specifically, one wide-angle camera was set at the front of the participants at a height of 2000mm. Side cameras were set at the left and right sides of the user, at chest-level height (1600mm). An additional voice recording device was placed in front of the participant.

- Desktop

  During the software development phase, a high-performance desktop computer was installed to meet the specifications of both Oculus VR systems and the development platform. This same computer was also utilized for the training sessions. Table 3.1 shows the main specifications of the desktop.

Part	Model
CPU	Intel Core i5-10500 @ 3.10 GHz (6 Cores)
RAM	32GB DDR4
Storage Device	256GB SSD and 320GB HDD
Graphics Card	NVIDIA GeForce RTX 2070 Super
Operating System	Windows 10
Monitor	22-inch Wide LCD Monitor

Table 3.1 Desktop Configuration

Data evaluation process

This subsection links Phases 3 and 4, outlining the evaluation methods that informed the design of the training experiment. A workflow for performance evaluation was formulated to gauge the effectiveness of the VR learning tool, as illustrated in Figure 3.7. The workflow encompassed two stages: the first stage involved turning raw data into processed data, and the second stage involved transforming the processed data into performance evaluation results.

Initially, raw data were retrieved from four distinct sources: Sensor data, Survey data, Audio data, and Researcher observations. The Sensor data included Controller data, which captured actions triggered by the controller, such as measuring vital signs or categorizing casualties. Additionally, Scenario data were collected at specific points during the training

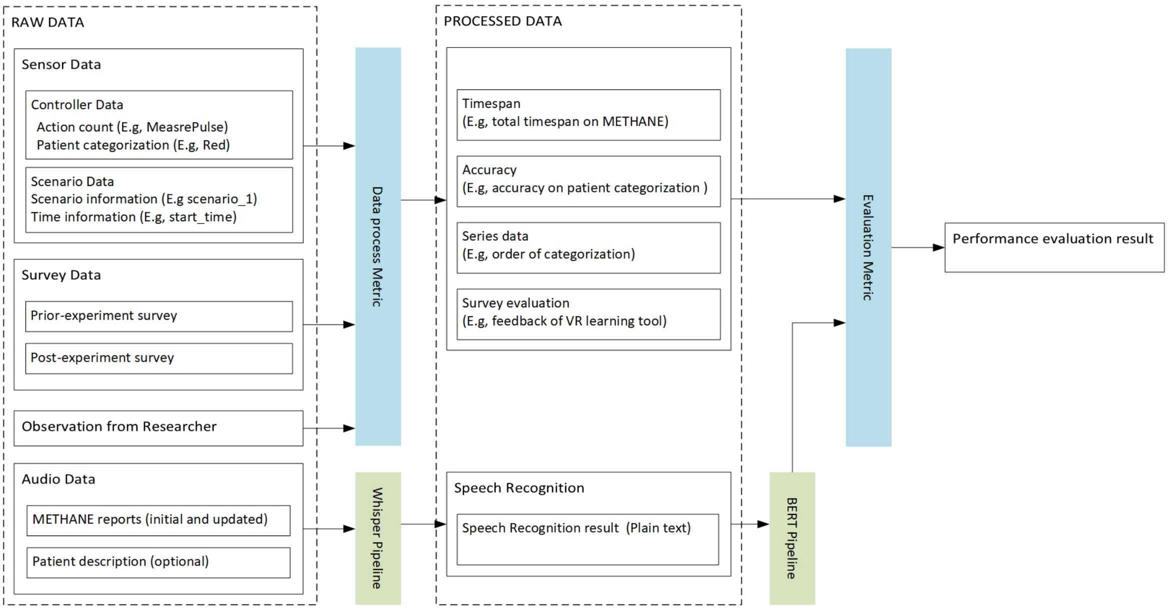

Fig. 3.7 Data Process Workflow

session, encompassing scenario information like scenario index and timestamp. The processing and integration of these diverse data sources, especially through AI tools like Whisper and BERT models, align with **Research Question 2** by demonstrating how AI enhances the analysis within the VR learning tool.

Two surveys were used to collect the background and feedback data from the participants. The initial survey gathered participants' age, educational backgrounds, prior VR exposure, and familiarity with MCI. The post-experiment survey was designed to gather feedback on the current learning tool's effectiveness, coupled with suggestions for its potential enhancement in future applications. This evaluation process is central to **Research Question 3**, as it allows us to assess the training efficiency of the VR learning tool by analyzing participants' performance, engagement, and feedback to continuously improve the system.

Noteworthy participant-specific insights emerged from researcher observations during the training sessions. For instance, certain participants exhibited a proactive inclination to engage in interactive conversations with the virtual casualties, underscoring their intent for immersive experiences. Conversely, others opted to communicate solely the essential vital signs as and when required. This observation highlighted the potential incorporation of individual personality traits into future teaching methodologies, offering an enriched and tailored learning approach. By recognizing these distinctions, educational strategies could be refined to align with various learning preferences, contributing to a more personalized and effective learning environment that catered to a diverse range of participant behaviors and tendencies.

The final dataset obtained from the experiment was audio data. In each training session, participants were instructed to provide an initial METHANE report at the commencement of each scenario, followed by an updated METHANE report once all casualties had been categorized. As previously mentioned, during the triage process, certain participants consistently reported vital signs, while others aimed to minimize oral communication. The extraction of METHANE reports using AI tools like Whisper (ASR) and the BERT model for semantic analysis provides a clear example of how AI can be used to enhance the performance of evaluation, addressing **Research Question 2**.

The raw data underwent analysis using conventional statistical techniques to process Sensor data, Survey data, and observations. The resultant processed data incorporated various aspects, including Timespan perspective (e.g., total timespan of METHANE report), Accuracy of performance (e.g., precision in casualty categorization), Series data (e.g., sequence of categorization), and survey evaluation. Additionally, the audio data were subjected to examination through the utilization of the auto speech recognition model "Whisper." This advanced model facilitated the transformation of audio content into textual time-series data, providing a structured and interpretable format for further analysis and insights. The analysis of these data not only enhances the understanding of participant performance (relevant to **Research Question 2**) but also provides insights into how the training tool can be continuously improved, thus addressing **Research Question 3** by evaluating the tool's efficiency and impact.

Following the extraction of METHANE reports from the audio data, we utilized the BERT language model to assess semantic congruence with predetermined scenario descriptions. This process aimed to elucidate the extent of precision achieved in participants' reports, thereby offering insights into the authenticity of their responses. Employing the gathered statistical data, we constructed an evaluation metric. It was our conviction that through the application of this evaluation metric, we could quantitatively appraise and assess the participants' performance, directly addressing **Research Question 3**, which focuses on improving training efficiency through data-driven assessments.

3.4.3 Component 3: Training experiment

This subsection corresponds to Phase 4 (Experiment Design and Implementation), describing participant recruitment, experimental setup, and procedural steps.

Setup training

- Recruitment plan

  Our recruitment strategy was designed with precision to identify and engage individuals who were not only capable of performing triage in urgent medical situations but were also keenly interested in contributing to the research project and adept at handling the physical demands of a Virtual Reality (VR) training environment. Targeting individuals aged 20 to 65 who were active paramedic students, fluent in English, and immersed in emergency healthcare practices, we aimed to recruit 10 to 15 participants. This was graphically represented in the recruitment workflow, as shown in Figure 3.8.

  To assemble a representative and diverse cohort, we selected individuals based on a balanced combination of their background experience in emergency healthcare and their proficiency with VR technology. This mix was crucial to effectively gauge the VR training's impact across a spectrum of knowledge and skills. The study, however,

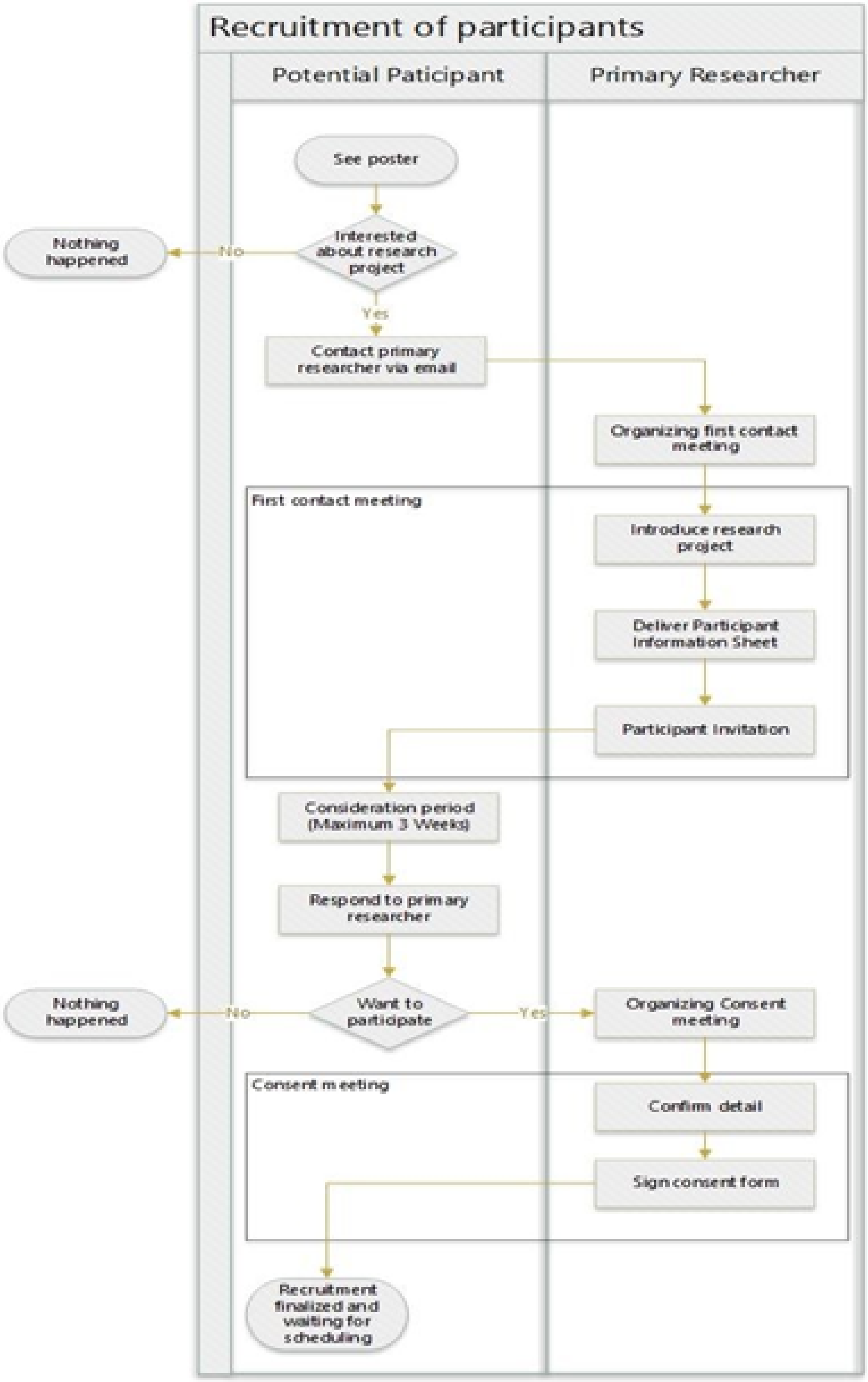

Fig. 3.8 Recruitment Process

instituted exclusion criteria—such as visual impairments, photosensitive epilepsy, or a lack of triage knowledge—to mitigate risk and preserve the study's scientific integrity.

Our recruitment campaign was conspicuous across AUT's campuses, with strategically placed posters designed to inform and attract potential participants. These posters directed interested candidates to contact the primary researcher, Peng Xia. Xia then coordinated personal meetings to provide further details about the study using an extensive Participant Information Sheet.

Upon the initial engagement, candidates were afforded up to three weeks to deliberate on their participation, after which they were expected to communicate their decision to the researcher. Absence of a response was interpreted as a decline. Should the first recruitment round have failed to attract sufficient participants, a second 14-day recruitment phase would have been initiated. The total planned duration for recruitment spanned 64 days, split between the two phases. However, the unforeseen advent of the Covid-19 pandemic and subsequent lockdowns in Auckland necessitated a recalibration of the recruitment timeline. After two years of challenges, including lost connections, the recruitment process recommenced in 2023. This renewed effort resulted in successfully enlisting 10 eligible participants within a constrained 28-day window, taking into account the academic calendar's holiday periods. Table 3.2 shows the schedule arrangement for recruitment in in 2023.

Task Name	**Duration**	**Expected Start**	**Expected End**
First Round Participant Recruitment	Total 28 Days	02/06/2023	30/06/2023
Recruitment	21 Days	02/06/2023	23/06/2023
Contact Meeting	7 Days	23/06/2023	30/06/2023
First Round Recruitment Complete	0 Days		
Second Round Participant Recruitment	Total 14 Days	27/10/2023	10/11/2023
Recruitment	7 Days	27/10/2023	03/11/2023
Contact Meeting	7 Days	03/11/2023	10/11/2023
Second Round Recruitment Complete	1 Day	10/11/2023	

Table 3.2 Second Recruitment in 2023

- Setup training environment

The training sessions took place in a dedicated classroom within AUT South Campus, occupying a minimally required room size of approximately 45-50 square meters, as dictated by the optimal configuration for virtual reality setups. Participants in this study, wore Oculus Rift headsets and held Oculus controllers, engaging in emergency healthcare procedures guided by custom virtual reality software developed by our research team. The Oculus Rift headset and Oculus controller are labeled as Numbers 4 and 5 in Figure 3.9.

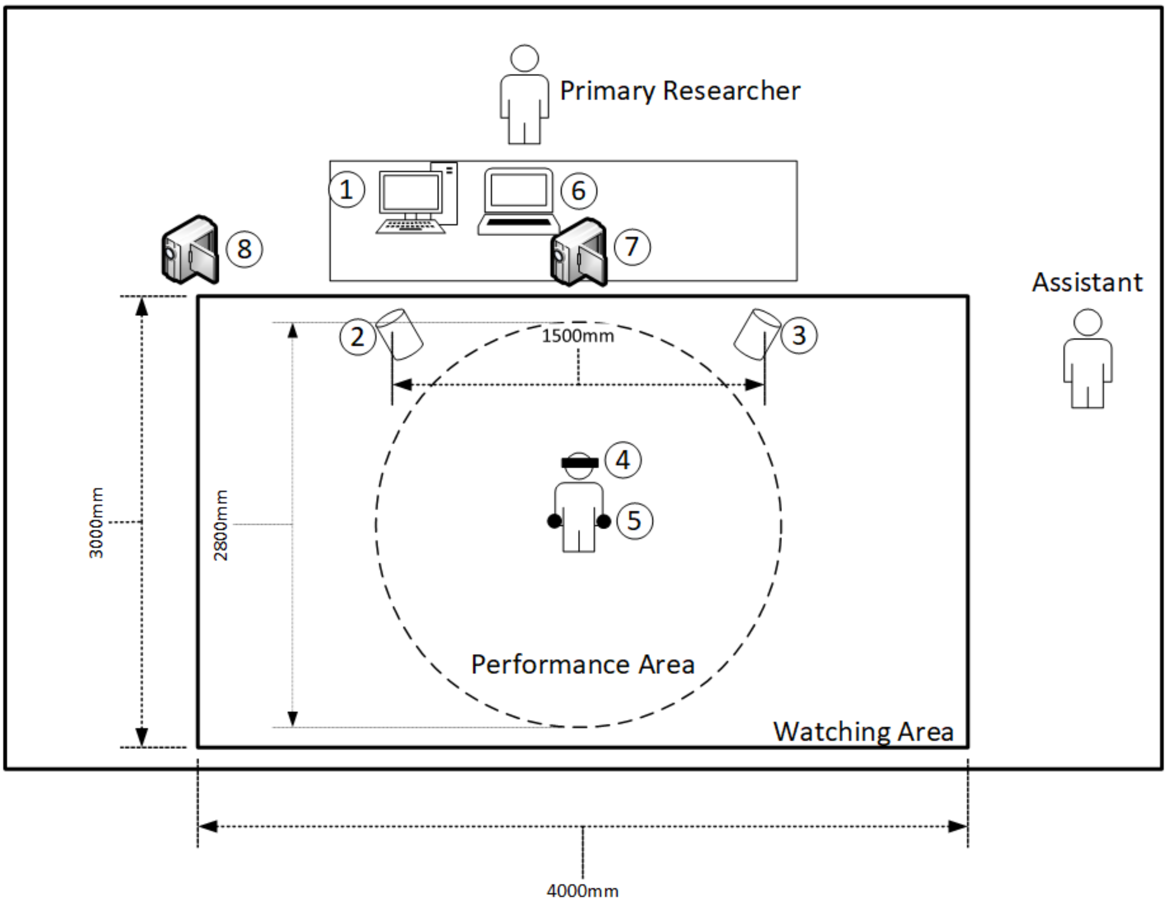

Fig. 3.9 Training Environment Setup

A desk accommodated the desktop systems running the VR learning tool and was positioned in a corner of the room, marked as Number 1 in Figure 3.9. Another laptop

(Number 6) was utilized by the primary researcher to record observations during the training sessions.

Participants performed procedures within a circumscribed Performance Area spanning roughly 2800 mm. This area included a larger Watching Area with no physical obstructions. If a participant's movements exited the Performance Area or if tracking by the recording sensors became inconsistent, a research staff member provided vocal reminders and could enter the Watching or Performance Area to assist in repositioning.

Three Oculus motion capture sensors, represented as Numbers 2 and 3 in Figure 3.9, were situated at the front-left and front-right positions. These sensors tracked the constellations of IR LEDs from the headset and controllers, translating participants' movements and actions into the virtual environment. The minimum distance between the two front-side sensors was 1500mm.

Additionally, two extra digital cameras (Numbers 7 and 8) were placed at the front-central and front-right positions. These cameras recorded audio and video data, serving as backup devices to fine-tune and correct the performance timeline. The front-central camera was set at approximately 2200mm in height to capture the entire Watching Area, while the front-right camera's height ranged from 1600mm to 1700mm.

Each training session spanned a duration of 25-30 minutes, encompassing three distinct scenarios: a tutorial room, a car crash scenario, and an earthquake simulation. Following the completion of each scenario, participants were afforded a break, during which the principal researcher engaged in a brief discussion to gather insights about their experiences and any discomfort. This interaction guided the decision to continue or conclude the training session. The culmination of these efforts yielded the experiment's results.

Experiment procedure

The experimental procedure unfolded in a structured process to ensure consistency and to safeguard the validity of the results. Once the recruitment advertisements were released, interested candidates reached out directly to the primary researcher. Each candidate was then screened against the predetermined inclusion and exclusion criteria, and those who met the inclusion criteria were provided with experiment documents. Before participating, they were required to review and sign the informed consent form along with the experiment documents, thereby formalizing their participation.

For scheduling purposes, the experiment were divided into two days. On each day, participants were allocated individual time slots to undertake the training experiment. During these sessions, participants entered the experimental environment one at a time. To avoid contamination of results and reduce the risk of learning effects, participants who were waiting for their turn were not allowed to observe the sessions of others. This arrangement ensured that each participant engaged with the VR training scenarios independently, free from external influence, and under comparable experimental conditions.

Data collection

- Data collection items

  In this research, we aimed to gather diverse datasets, encompassing sensor data, survey data, observation data, and media data, including audio waves. These data were collected at various stages of the training sessions. To elucidate further, sensor data were acquired while participants utilized the VR headset to engage in triage scenarios. Survey data were collected both before and after the training sessions, serving distinct research objectives. Concurrently, observation and media data were gathered while participants performed VR triage tasks. Table 3.3 introduces the collecting status terminologies, and the following tables indicate the detailed information regarding the data elements.

Table 3.3 Terms of Collecting Status

Collecting Status	Description
ER	Data collected at the end of the recruitment
T	Data collected during the training session
ET	Data collected at the end of the training session

- Sensor dataset

The raw sensor dataset comprised three distinct subsets: pulse-action data, airway and breathing action data, and categorization data. These subsets not only independently depicted specific actions, such as pulse measurement or airway status assessment, but also could be combined to reconstruct the performance of individual participants over time. Table 3.4 illustrates the data elements within the pulse-action subset.

Table 3.4 Pulse-Action Subset

Data Item	Collecting Status	Description	Example
Scenario_id	T	Id S_i of current scenario	*1*
Start_time	T	Starting time T_s of S_i	*11:06:56 a.m.*
Controller_time	T	Current time T_c	*11:08:29 a.m.*
Casualty_id	T	Casualty id P_i of S_i	*Casualty_2*
Pulse_rate	T	Pulse rate of Pi at time T_c	*30*
Action	T	What action is done at T_c	*PULSE_CHECK*

With the data provided in this subset, one could glean specific details pertaining to pulse measurement actions. For instance, at 11:13:30 am (T_c), Participant 1 measured the pulse rate of casualty 2 (P_i) in Scenario 1 (S_i), resulting in a pulse rate of 30. Table 3.5 shows the data elements of the airway and breathing-action subset.

Within the dataset subset presented, one could extract precise information regarding airway and breathing measurement actions. For instance, at 11:09:41 am (T_c), Participant 1 assessed the airway status of casualty 2 (P_i) in Scenario 1 (S_i), resulting in an

Table 3.5 Airway and Breathing-Action Subset

Data Item	Collecting Status	Description	Example
Scenario_id	T	Id S_i of current scenario	*1*
Start_time	T	Starting time T_s of S_i	*11:06:56 a.m.*
Controller_time	T	Current time T_c	*11:08:29 a.m.*
Casualty_id	T	Casualty id P_i of S_i	*Casualty_2*
Airway_status	T	Airway status of P_i at time T_c	*1*
Resp_rate	T	Respiration rate of P_i at time T_c	*6*
Action	T	What action is done at T_c	*PULSE_CHECK*

airway status of 1, indicating normal airway conditions. Concurrently, the casualty's respiration rate was recorded as 6. However, it was essential to note that this respiration data was not visible to the participants, as they were focused on assessing the airway status at that moment. Table 3.6 shows the data elements of the categorization subset.

Table 3.6 Categorization Subset

Data Item	Collecting Status	Description	Example
Scenario_id	T	Id S_i of current scenario	*1*
Start_time	T	Starting time T_s of S_i	*11:06:56 a.m.*
Controller_time	T	Current time T_c	*11:08:29 a.m.*
Casualty_id	T	Casualty id P_i of S_i	*Casualty_2*
Assigned_cate	T	Category assigned to P_i at time T_c	*1*
True_cate	T	Ground truth category of P_i at time T_c	*2*
Result	T	If assigned_cate is equal to True_cate. *FALSE*	

Contained within the subset of the dataset, one could meticulously extract information concerning triage categorization actions. For example, at 11:11:10 am (T_c), Participant 1 assigned a red category (equivalent to 1 in the VR learning tool) to casualty 2, signifying an immediate urgent level. However, it was crucial to note that, at this juncture, casualty 2's actual category was 2, denoting a normal urgent level or Orange. Consequently, the outcome of the current assessment was deemed False.

3.4.4 VR Learning Tool Development

Prior to initiating the development of the VR learning tool, we had outlined the primary requirements at a high level as follows:

1. Functionality for MCI Training:

   Our preparatory work, which had included a thorough review of relevant literature, guided us to utilize development tools for constructing several VR scenarios representative of MCIs, such as a car accident. These scenarios were designed to feature a variety of virtual casualties exhibiting diverse medical conditions. participants would engage with these casualties, assessing their conditions, and making critical triage decisions. Additionally, participants were tasked with the efficient allocation of resources. Throughout the training, our software was designed to collect detailed, quantifiable data (like the time spent on each task, sensor data, etc.), which would be integrated with qualitative data (such as audio recordings) for a comprehensive assessment of the participant's performance.

2. Flexibility and Scalability:

   The VR learning tool we had envisioned was adaptable, allowing for the simulation of virtual casualties with a range of conditions and vital signs. Moreover, we aimed for the software to be scalable, facilitating future enhancements with minimal additional investment. To this end, we had taken specific measures: firstly, we had evaluated add-on tools to ensure they could be seamlessly integrated into the training software and function autonomously; secondly, we had encapsulated core functionalities (like data input classes) to remain distinct from the main operational logic; thirdly, we had been refining the software architecture to adhere to principles of high cohesion and low coupling, with the foresight that this would minimize the impact on the system when new features were added or existing functionalities were modified.

3. Cost-Effective Deployment:

   It was imperative that the training software we proposed could be deployed efficiently on standard desktop systems and be readily accessible to users without incurring substantial costs.

Through these stipulated requirements, we intended to create a VR learning tool that was not only effective in teaching MCI skills but was also robust, versatile, and economically viable for widespread use.

Prototype development

The development of the VR Learning Tool commenced with a structured and sequential approach. In this phase, we focused on translating the requirements into several prototypes.

This had involved establishing the core functionalities, such as building VR MCI scenarios, enabling user interaction, and integrating built-in data collection methods. Given the complexity and novelty of the VR Learning Tool, it was crucial to have a clear and well-thought-out blueprint, which the Waterfall model had provided.

While prototypes were built, we also developed a few modules to verify the feasibility that might satisfy the major requirements. Some modules were integrated into the later developed VR learning tool, and the rest of them were archived for future use. Following this, we present two of the archived modules that could be used in future development and one prototype scenario that was used to verify the user interaction.

- Vital sign generator

  In our first prototype, we had developed a vital sign generator module comprising various scripts to manage the essential characteristics of the virtual casualty. These scripts were capable of creating a multitude of virtual casualties, each with unique vital signs, which in turn influenced the casualties' visual appearance. Upon examining

Component	Features
Respiratory	Respiratory Rates (RR)
Respiratory	Respiratory Status (E.g. Airway Blocked)
Blood Pressure	Blood Pressure (BP)
Glasgow Coma Score	Glasgow Coma Score (GCS)
Seizure status	Seizure status
Circulation	Skin Color
Circulation	Heart Rate (HR)
Circulation	Blood Loss
Pain	Chest
Pain	Other type pain

Table 3.7 Triage Feature Schema

the Advanced Triage System (ATS), we derived a schema for triage features. A basic representation of this triage feature schema was shown in Table 3.7.

We had observed that certain elements of the triage schema were measurable. For example, the ATS guidelines specified that a Heart Rate (HR) of 80 or less was a critical cutoff for classifying a casualty as ATS-1 or ATS-2 category. To accommodate this, we engineered a script to generate random data for each measurable characteristic. Figure 3.10 displays an example of the output from this script, with red arrows highlighting the measurable features.

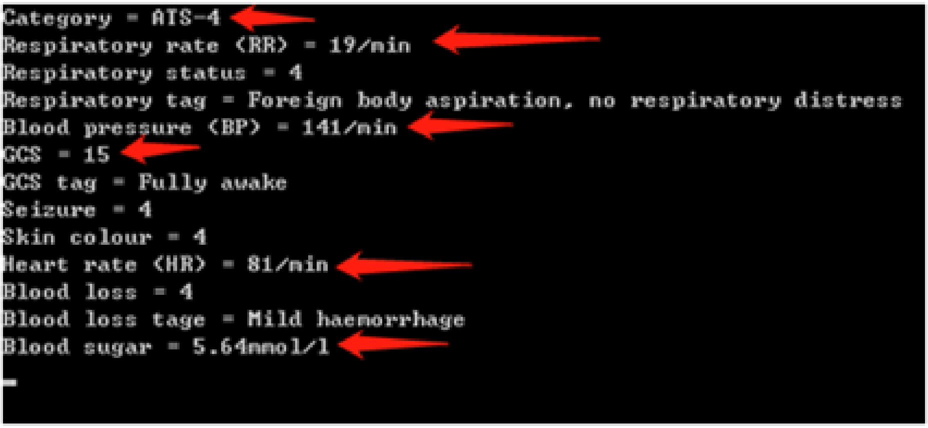

Fig. 3.10 Vital sign generator (console view)

imultaneously, we developed the scripts for features that define precise measurement, such as Seizure and Respiratory status. We had based our approach on three suppositions: 1) That such features, while not quantifiable, manifest visually on a casualty's body; 2) It was unlikely for all features to be present in a single casualty at the same time; and 3) More severe symptoms could suggest a higher ATS level, implying that casualties with higher ATS levels were likely experiencing more serious conditions.

With these considerations in mind, we designed several scripts incorporating a binary-tree structure and a weighting system. The binary-tree dictated the severity of the casualty's condition, while the weighting system influenced the visual representation of each non-measurable feature. In conjunction with the scripts for quantifiable traits, this framework enabled the vital signs to visibly affect the virtual casualty's appearance. We then applied some animations to dynamic objects that responded to signals from the script, allowing for diverse reactions to different interactive events.

In this prototype, we had integrated three weighted features: Seizure, Respiration, and Pain. Each of these conditions was managed by scripts and was associated with specific animations Figure 3.11 illustrates the weighted features on the left and the binary-tree model on the right. The blue progress bars represent the weight percentage for each feature, ranging from 0 to 1. By blending these animations, we were able to generate virtual casualties with a variety of visual expressions. Figure 3.12 displays a selection of these casualties: a casualty with a high Pain weight is marked by a red arrow, one with a high Respiration weight by a yellow arrow, and one with a high Seizure weight by a green arrow.

After adding scripts and animations to the objects, we integrated them with the Oculus VR toolkit and adapted them to fit the VR device. During the integration step, we first adapted the scene view from screen-based to VR, and then fitted the control methods to the VR device. Finally, we output the entire scene as a VR application.

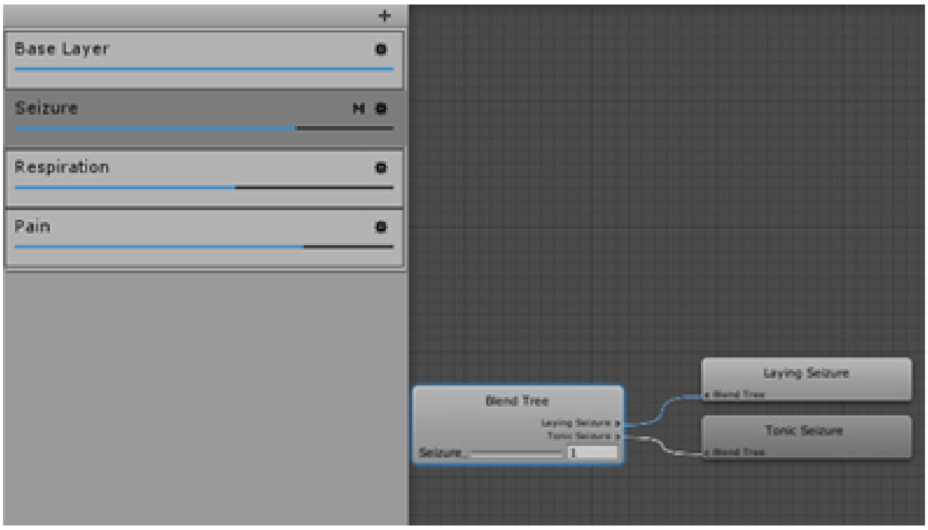

Fig. 3.11 Weighted parameter and Binary tree view

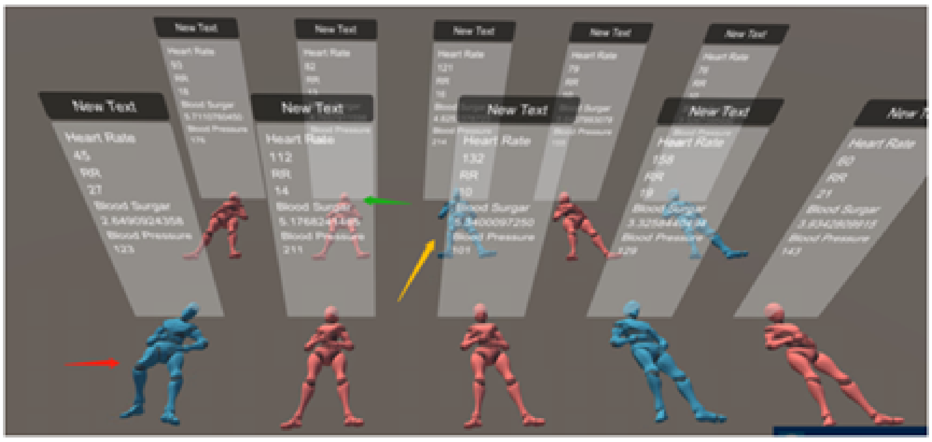

Fig. 3.12 Visual expression of sample casualties

- State transition based dynamic vital system

  In an earlier iteration, we designed a dynamic narrative module for each casualty. In this design, the casualty's different statuses were described by a state-transition graph, and this state-transition graph conducted the casualty's animations in the training software. For instance, in this graph, the casualty's animation would start from the state 'initialization' (category = 3) and directly transition into state 'S5' (category = 3). In this state, the casualty would obtain a leg injury animation. After a reasonable random time, the state would have the chance to transition to either state 'S6' or 'S7'. 'S6' represented the casualty's condition getting better and it might transition into 'S2' (category = 2) in a short time, while 'S7' represented the casualty's condition worsening and it might increase the triage category eventually. At each state, the casualty would obtain a new animation that could represent the casualty's current condition. Figure 3.13 shows how the state-transition graph was mapped in the Unity animation engine. We are currently trying to optimize this design and fit it into a future version.

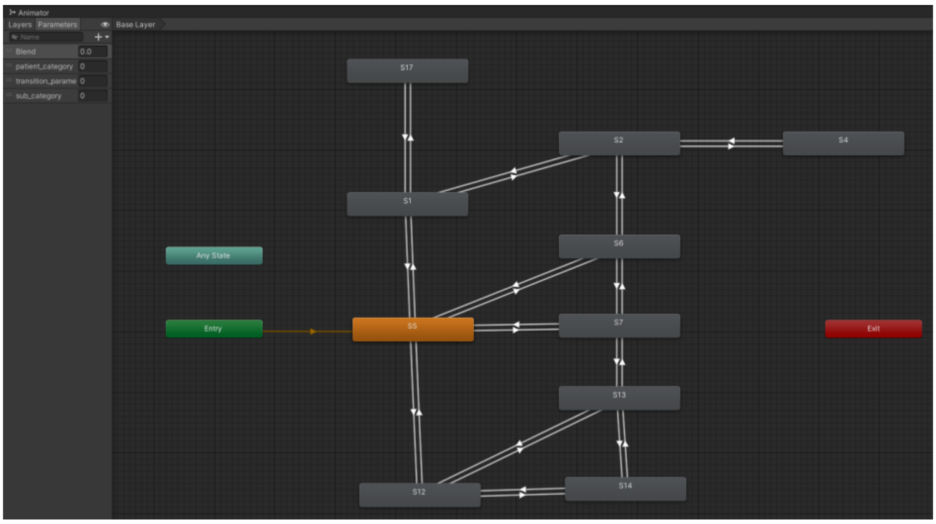

Fig. 3.13 State-transition graph of dynamic narrative mechanism in Unity

In the last prototype before the actual scenarios were developed, we built two distinct scenarios: a tutorial for beginners and a car accident simulation.

The tutorial scenario was aimed at acclimating new participants to the software. This scenario, which involves just a single casualty, was intended to quickly introduce users to the system's functionality, including the use of indicators and wristbands. As shown in Figure 3.14, it presents a single casualty setup, while Figure 3.15 displays the tutorial's message panel, which guides the participant through the learning process with instructive prompts. This single-casualty scenario ensures that new users can efficiently learn the software before advancing to more complex simulations like the car crash scenario.

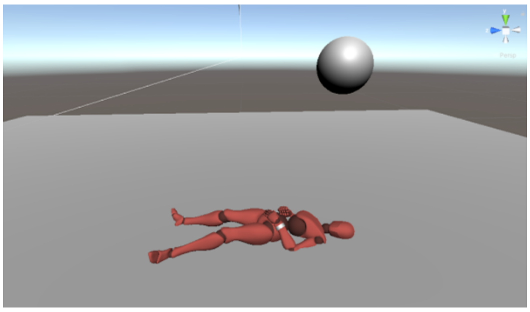

Fig. 3.14 Tutorial scenario, single casualty

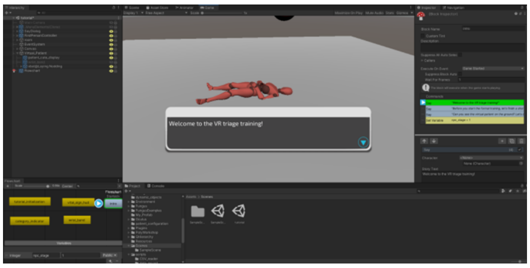

Fig. 3.15 Tutorial scenario, message panel

Figures 3.16 and 3.17 demonstrate the virtual setting of our prototype for a car accident scenario. This digitally created environment features a traffic collision with four onlookers and four simulated casualties. Each casualty was initially programmed with a set of conditions that align with the ATS categories: one with a life-threatening hemorrhage, one with a severe traumatic brain injury (TBI), and two with various trauma-related injuries. These conditions correspond to the 'immediate' and 'urgent/delay' triage categories. Moreover, these virtual casualties were modeled after actual cases from the AUT Massive Incident (MI) day training (Casualty 1, Casualty 3, Casualty 4, and Casualty 5).

In our training module designed for triage in MCIs, we developed a sophisticated system for categorizing virtual casualties, a vital feature highlighted in Figure 3.18. This system

Fig. 3.16 Prototype car accident scenario (Front view)

Fig. 3.17 Prototype car accident scenario (Back view)

incorporates color-coded indicators positioned above each casualty, aligning with the Australasian Triage Scale (ATS) standards. These indicators are color-coded as follows: Red signifies 'Immediate', Yellow/Orange represents 'Urgent', Light Yellow indicates 'Delay', and Green denotes 'Non-Urgent'. These indicators serve a dual purpose. In the tutorial/trial mode, they are visible to assist participants in their learning process. However, for a more challenging and realistic experience, they become invisible during the actual training/testing phase.

Fig. 3.18 Indicators for MCI training demonstration

Fig. 3.19 Wristband for MCI training demonstration

In the context of MCIs, color-coded wristbands are a crucial element in triage and casualty management. These wristbands, allocated based on the severity of injuries, facilitate the swift and effective sorting of casualties—a critical process in scenarios where medical resources are often stretched thin. Commonly, colors such as red, yellow, green, and black categorize casualties into immediate, delayed, minor, and deceased groups, respectively.

Figure 3.20 and 3.21 show the triage slap and wristband used by St. John New Zealand. This categorization enables first responders and medical staff to rapidly identify and give priority to casualties requiring urgent care. These wristbands not only provide a visual summary of each casualty's condition but also aid in efficiently directing casualties to suitable medical facilities and keeping track of injury types and numbers, which is essential for effective resource distribution and ongoing management of the incident.

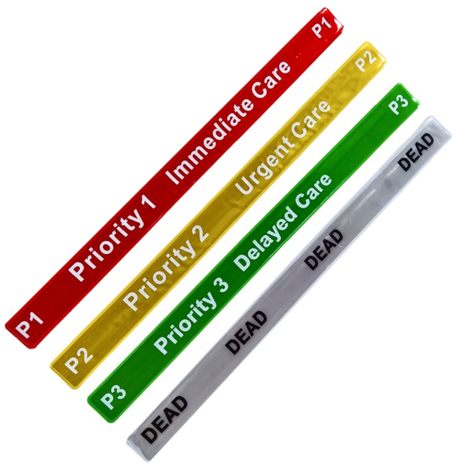

Fig. 3.20 Triage Slap used by St.John New Zealand

Fig. 3.21 Wristband used by St.John New Zealand

Furthermore, our simulation includes an interactive triage tool that mirrors the wristband system employed in actual triage situations. This tool allows participants to assign triage categories to casualties, with each category being represented by a distinct color on the wristband, as depicted in Figure 3.19. The software then evaluates the participants' triage decisions for accuracy by comparing them against the actual conditions of the casualties, as indicated by the category indicators. This integration adds an invaluable dimension to our training program, equipping participants with the skills needed to manage MCIs effectively.

Feature Development and User Feedback (Agile):

With a solid prototype in place, we adopted an Agile approach for further development. This phase was characterized by rapid iterations, user-centered design, and continual improvement. We expanded the VR scenarios, added more sophisticated interaction capabilities, and enhanced data collection features.

User feedback was integral to this phase. We conducted user testing sessions, gathering insights on the usability, engagement, and educational effectiveness of the VR Learning Tool. This feedback was then used to inform subsequent development sprints. For example, if users found a particular VR scenario confusing or less engaging, we would refine it in the next sprint. This iterative process ensured that the VR Learning Tool was not only technically sound but also met the practical needs and expectations of its users.

Completed VR learning tool

Following extensive phases of prototype development and close collaboration with medical professionals specializing in MCI, we successfully amassed a comprehensive set of requirements and harnessed the requisite techniques to bring the VR learning tool to its fruition. A milestone in this journey was the establishment of a software framework that endows us with the agility to effortlessly tailor casualty vital signs across a myriad of scenarios. This framework plays a role in expediting the creation of new scenarios and configuring casualty parameters, thereby significantly mitigating the time and effort traditionally associated with these tasks.

Additionally, we have enriched the VR learning tool with novel functionalities, all inspired by the invaluable insights provided by esteemed experts in the realm of emergency healthcare. Notably, we have integrated the oropharyngeal airway (OPA), a critical medical device that finds indispensable application in emergency medicine, anesthesia, and critical care settings. The core purpose of this device is to establish and diligently sustain a clear and unobstructed airway within the casualty's oropharynx region, ensuring optimal respiratory function and the provision of adequate oxygenation. Within the intricately designed scenarios, we have incorporated instances where casualties experience airway obstructions, necessitating the active intervention of the participant. By employing the OPA, the participant is tasked with skillfully opening the casualty's airway, and our dynamic vital sign mechanism responds in real-time to these actions, effecting pertinent alterations in the casualty's vital signs. This immersive approach enhances the overall realism and educational value of the training experience. To provide a visual representation of this, please refer to Figures 3.22 and 3.23, which vividly depict the OPA and its application within the immersive VR scenario, thereby elevating the effectiveness and engagement of our VR learning tool to unprecedented levels.

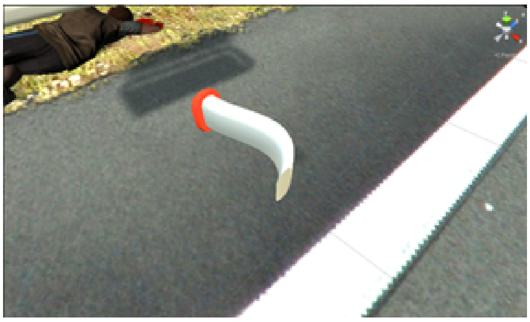

Fig. 3.22 Oropharyngeal airway tool

Fig. 3.23 Use oropharyngeal airway tool

Moreover, we improved the performance of the Unity triggers positioned on the casualty's body. Unity triggers play a central role in Unity game development because they are intricately linked with collider components and serve as advanced event sensors that swiftly respond to particular in-game situations and interactions. By configuring specific colliders to function in "trigger" mode, game developers can choreograph events that activate when GameObjects enter, exit, or remain within their designated boundaries. These versatile triggers have a wide range of applications across various gameplay mechanics, including enabling interactions with in-game objects, streamlining item collection processes, monitoring progress towards achieving quest objectives, and facilitating seamless transitions between different game levels. Figure 3.24 illustrates the pulse check trigger, while Figure 3.25 represents the trigger for evaluating the airway and breathing status. When participants employ their virtual hand to interact with these triggers, the corresponding vital signs are displayed on their HMDs. Additionally, Figure 3.26 shows the vital sign display following a participant's interaction with the airway and breathing trigger.

Fig. 3.24 Pulse trigger

Fig. 3.25 Airway and Breathing trigger

Fig. 3.26 Vital Sign shows on after triggering

Suggested by emergency healthcare experts, we incorporated a bleeding simulation into the VR learning tool. When a scenario begins, a timer is activated. In certain situations, casualties may present with severe bleeding issues. If a tourniquet is not applied to address this bleeding promptly, their pulse rate will undergo changes, and the visual representation of bleeding will gradually accumulate around their body. Figure 3.27 and 3.28 show the bleeding texture is changed over the time.

Subsequently, to enhance the realism of our VR learning tool, we incorporated location features like road signs into the simulation. This addition is vital, particularly in the context of MCIs, where the specifics of location and egress routes hold immense significance.

Fig. 3.27 Major bleeding (scenarios starts)

Fig. 3.28 Major bleeding (Over times)

Efficient access to the incident site is essential for prompt response by emergency personnel. Simultaneously, ensuring clear and unobstructed egress paths is crucial for the safe evacuation of casualties and the effective execution of triage operations. In these scenarios, road signs serve a critical function. They not only expedite the emergency services' navigation to the site but also aid in controlling traffic to facilitate the rapid movement of rescue vehicles. Furthermore, these signs provide indispensable directional guidance for both the emergency responders and evacuees. Therefore, in the context of MCIs, the thoughtful integration of road signs, along with meticulously planned location and egress strategies, becomes instrumental in efficiently managing the incident and significantly reducing the potential for further casualties. Figure 3.29 shows an example of the road sign, which indicates the incident happens on the southbound of Auckland Motorway, and before the exit of Papakura.

In our scenario, we have implemented a METHANE banner within the VR learning tool, serving a similar purpose as the traditional METHANE card used in emergency response, especially in MCIs. Figure 3.31 shows the detail of the METHANE banner. This is exemplified by St. John New Zealand's METHANE card (Figure 3.30), as depicted in a reference figure. The METHANE banner in our VR learning tool is presented in two distinct forms, tailored to the requirements of the training exercise. Participants are tasked with compiling a METHANE report both at the start and end of the scenario. The final display of the METHANE banner is uniquely designed to include additional details, specifically the number of categorized casualties, thus providing a comprehensive overview of the incident's

Fig. 3.29 Example of road sign

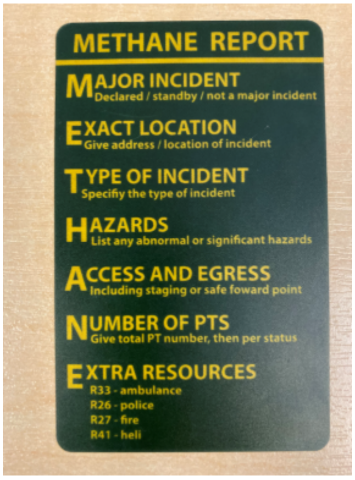

Fig. 3.30 METHANE card used by St.John New Zealand

status at its conclusion. This enhanced METHANE banner, as shown in the referenced figure, offers a detailed summary at the scenario's end, aiding in the effective communication and management of the incident.

Upon developing the essential software framework, we proceeded to tailor the virtual casualties for each specific scenario, drawing upon the expertise of emergency healthcare professionals to map out the triage process for each individual case. Figure 3.32 illustrates the triage procedure for a particular casualty. This process began by evaluating whether the

Fig. 3.31 METHANE banner in VR learning tool

casualty could move; mobile casualties were guided to a safe zone and assigned a status of 3, indicating they were non-critical and could mobilize without urgent aid. The depicted triage process features an image of a casualty who had suffered a leg fracture, moderate bleeding, and was presumably experiencing pain. The casualty, unable to move, was given a status of 0, which tragically marked them as deceased. Descriptively, the casualty was a virtual representation of an approximately 30-year-old male. Adjacent to the figure, a box specified the casualty's expected condition: immobility, an open airway requiring no intervention, a respiration rate of 35 breaths per minute, and a critically elevated pulse rate of 140. The triage goal was to reclassify the casualty to a status of 2, indicative of an urgent condition that, while serious, was not immediately life-threatening.

In cases where the casualty was immobile, the protocol advised a verification of a clear airway. If the casualty was found breathing, the focus shifted to assessing circulation by checking the pulse rate, which was noted as a critical 140. Subsequent steps involved determining the presence and severity of bleeding. A casualty without active bleeding remained at a non-urgent status of 3. Conversely, if bleeding was present, the process evaluated whether a tourniquet had been successfully applied to halt the bleeding, potentially

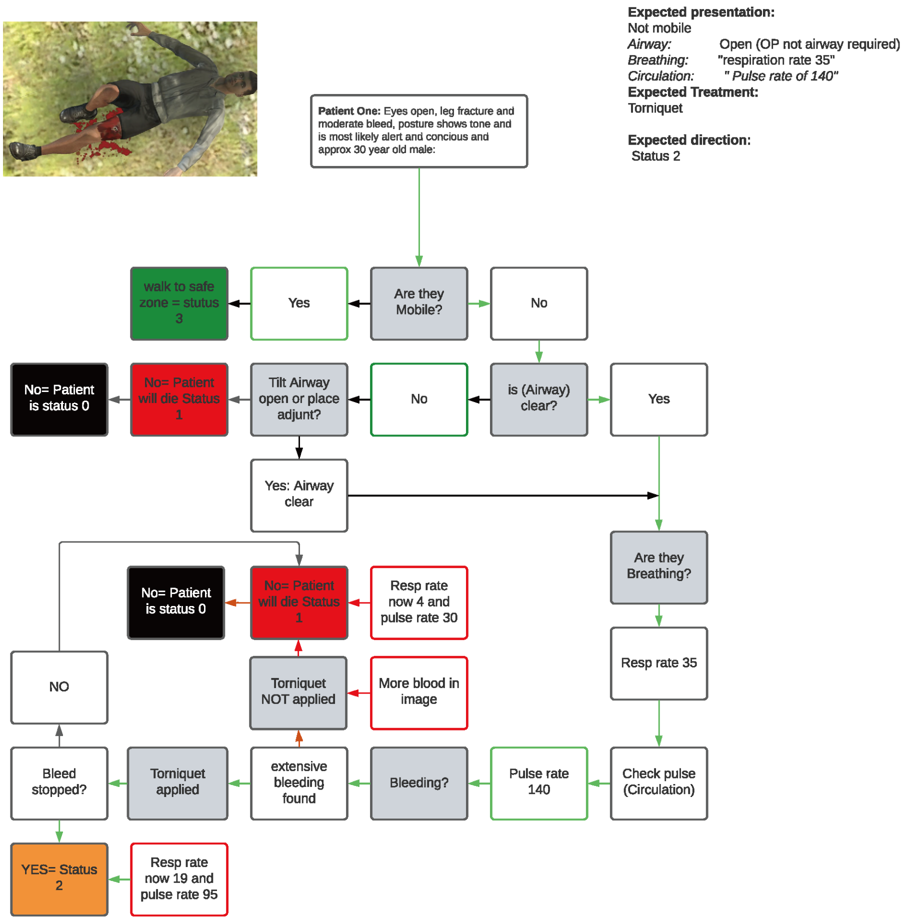

Fig. 3.32 Example triage process of a casualty

upgrading the casualty's status to 2, thereby flagging them as urgent and in need of prompt care. Should there have been extensive bleeding without a tourniquet in place, the casualty's status was elevated to 1, indicating a dire, life-threatening condition that demanded immediate intervention.

Eventually, we built 9 casualties in scenario 1 (Car crash), and 6 casualties in scenario 2 (Earthquake). Figure 3.33 and 3.34 show the overall view of each scenario.

Fig. 3.33 Overall view of car crash scenario

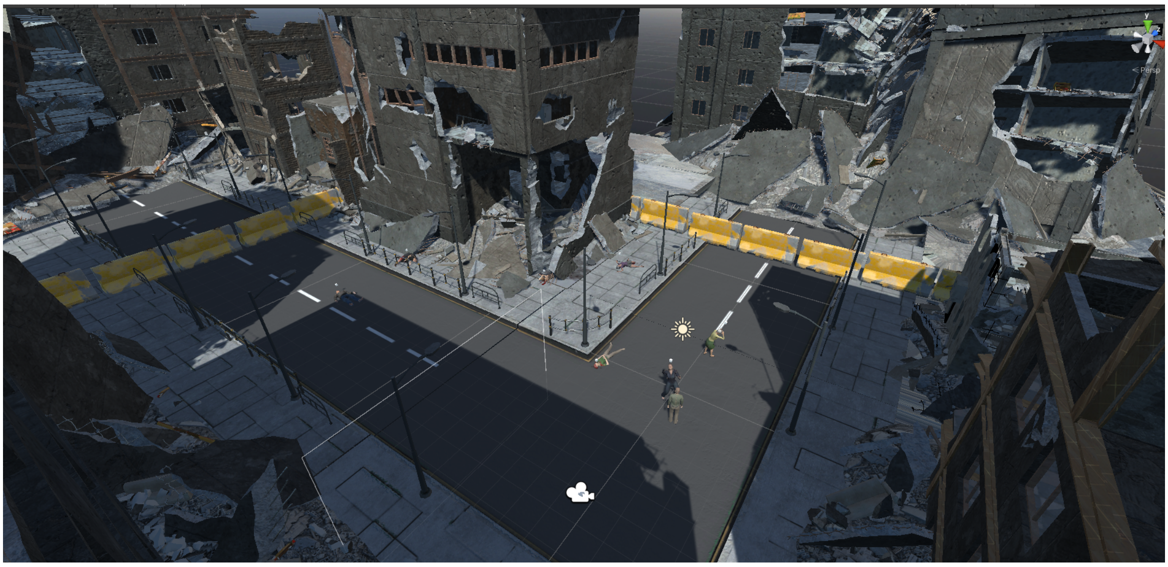

Fig. 3.34 Overview of earthquake scenario

3.4.5 Development of VR and AI Tools

Development of VR and AI Tools The initial stage of developing the VR and AI tools had been approached with the meticulousness and linear structure of the Waterfall model. We had begun with an exhaustive requirements analysis, which involved different departments (e.g., AUT paramedicine department) and institutions (e.g., St John New Zealand), thoroughly understanding the needs and expectations from our VR and AI components. The requirements collection had occurred after the gap analysis of the literature review.

This phase had been crucial in identifying the type of 3D objects needed, the level of interactivity required in the VR scenarios, and the complexity of the AI algorithms to be developed. We had documented specifications for modeling tools, AI development environments, and hardware requirements.

Following the requirements gathering, we had embarked on a detailed design phase. This had involved creating high-level designs for the VR environments, outlining the architecture for the AI algorithms, and setting up a framework for integrating these components. The design phase had been pivotal in ensuring that all subsequent development efforts were aligned with our initial vision. This linear approach, characteristic of the Waterfall model, had provided a clear roadmap and had helped in establishing a solid foundation for the complex development that lay ahead.

Transitioning to the Agile methodology, we had embraced its iterative and collaborative nature for the development phase. This shift had been critical as it had allowed for greater flexibility and responsiveness to emerging challenges and new insights. We had organized our work into sprints, with each sprint focusing on a specific set of features or improvements. For instance, one sprint might have focused on developing a set of 3D models for the VR environment, while another might have concentrated on implementing a specific AI algorithm.

During each sprint, continuous testing and integration had been key components. This approach had ensured that issues were identified and resolved promptly, preventing any significant deviations from our objectives. The iterative sprints also had facilitated regular feedback loops with stakeholders, including end-users and technical experts. Their insights had been invaluable in refining the tools, making adjustments based on usability feedback, and ensuring the final product met the needs of its users. This phase had exemplified the adaptive and user-centered approach of Agile, allowing for a dynamic and responsive development process.

3.5 Data Evaluation Methods

The evaluation of the VR learning tool was a critical aspect of this study, designed to assess its effectiveness in enhancing emergency healthcare training for MCIs. The evaluation process was methodical, incorporating both qualitative feedback from participants and quantitative data derived from ML models applied to sensor and audio data. This comprehensive evaluation aimed to measure the VR learning tool's impact on key areas such as decision-making, triage accuracy, task efficiency, and communication clarity.

3.5.1 User Testing and Survey Feedback

A significant part of the evaluation process involved gathering direct feedback from the participants. To achieve this, a survey instrument was developed, drawing on established frameworks like the NASA Task Load Index (NASA-TLX), which is widely used in human performance assessments. Additionally, the design of the survey considered previous studies in both VR and MCI training, ensuring that the questions were relevant to the specific context of this research. The initial survey draft contained 40 questions, carefully crafted to cover multiple dimensions of the user experience.

After consultations with emergency healthcare professionals, the survey was refined to 16 questions. These questions were designed to capture detailed insights into the specific elements of the VR training experience that were most relevant to our study. The survey was taken after each training session, allowing participants to reflect on their experience in real-time. The qualitative feedback gathered through this survey provided crucial information regarding the tool's strengths and areas needing improvement, particularly regarding user comfort and the intuitiveness of the control system.

3.5.2 Sensor and Audio Data Analysis

To complement the subjective feedback from the participants, objective data was collected using a variety of sensors embedded in the VR system. These sensors tracked key metrics such as time and correctness.

The sensor data was analyzed using ML algorithms. The Whisper toolkit, known for its robust audio processing capabilities, was employed to process the participants' verbal communications during the training sessions. The Whisper toolkit was customized to handle complex environments with various types of background noise, ensuring that speech recognition remained accurate even during dynamic MCI scenarios. Through several rounds of experimentation, a hybrid configuration of 'small' and 'medium' pre-trained models was selected, optimizing the trade-off between computational efficiency and recognition accuracy.

In addition to task-related metrics, audio data provided valuable insights into how effectively participants reported the circumstance of the virtual casualties. For instance, during the training simulations, participants were required to make verbal triage reports following the METHANE protocol (Major incident, Exact location, Type of incident, Hazards, Access, Number of casualties, and Emergency services required). The ability to report the casualty's condition clearly and effectively under pressure is a critical component of real-world MCI scenarios, and the VR learning tool's ability to simulate and measure this communication was an important part of the evaluation.

3.5.3 Semantic Analysis

The study also utilized advanced semantic analysis techniques to evaluate the quality of METHANE report during the VR training. Communication is a vital skill in emergency healthcare, and structured communication protocols like METHANE are critical for ensuring that paramedics or first-responders can provide accurate and concise information in high-pressure environments. To analyze this, the GPT model was used in the VR learning tool to process and assess participants' spoken words during the simulations.

The GPT-based semantic analysis focused on identifying the completeness and accuracy of participants' METHANE reports. For example, when participants provided METHANE reports in the training sessions, the system evaluated whether they correctly addressed all required elements, such as the number of casualties or the nature of the incident. The goal was to determine not only the participants' adherence to the METHANE protocol but also their ability to convey critical information clearly and effectively.

The semantic analysis was fine-tuned to ensure reliability and consistency. Different configurations of the GPT model were tested to determine the optimized approach for evaluating participant communications in the context of the training scenarios. This allowed the VR learning tool to assess the communication performance, highlighting any areas where improvements were needed.

3.6 Iterative Improvement

After completing the first major version of the VR learning tool, several trials were conducted to evaluate its performance. Based on participant feedback and data analysis, multiple iterations were implemented to address key areas, including reducing physical discomfort, refining user interactions, and optimizing the AI-driven data analysis. These refinements were crucial in enhancing the overall user experience and ensuring that the tool was as effective and immersive as possible.

3.6.1 Addressing Physical Discomfort

One of the primary concerns identified during the testing phase was the physical discomfort experienced by participants, especially nausea and dizziness, which are common in VR environments. These symptoms were attributed to factors such as rapid movements and long-time use of the VR system. To resolve these issues, we made the following changes:

To address this, the following changes were made:

1. Adjustment of movement controls

   Initially, we experimented with different movement control methods in the VR environment, including teleportation, running, and walking. Through early trials with researchers and emergency healthcare professionals, we found that participants experienced nausea and disorientation when using faster movement speeds, particularly during running or extended walking sequences. Based on this feedback, we fine-tuned the movement controls by adjusting the parameters such as gravity, moving speed and accelerations for both walking and running modes in each scenario. Based on this feedback, we fine-tuned the movement controls by adjusting parameters such as gravity, moving speed, acceleration, and velocity for both walking and running modes in each scenario. More specifically, we made adjustments to the tracking of position and rotation, as well as to the speed, acceleration, and velocity of movement. For instance, we found that the default settings for speed, acceleration, and velocity were too fast and caused participants to feel dizzy or uncomfortable. As a result, we slowed down these parameters to create a smoother, more comfortable movement experience. This recalibration allowed participants to navigate the virtual environment more comfortably, minimizing discomfort and motion sickness, particularly during longer training sessions.

2. Refined User interaction

   Another aspect contributing to participant discomfort was the complexity of interacting with virtual casualties. Initially, participants found it challenging to precisely interact with specific parts of the virtual casualties, which led to confusion and unnecessary movements. To improve the user experience, interaction indicators were introduced. These visual cues highlighted which parts of the casualty could be interacted with, simplifying the process and reducing the need for guesswork. Figure 3.25, 3.24, 3.27 and 3.28 show the interaction indicators. By clearly marking interactive areas, we were able to streamline the user interactions, thereby reducing unnecessary physical movement and potential discomfort. This change not only made the tool more intuitive but also enhanced the efficiency of completing tasks, as participants could focus on critical actions without distraction.

3.6.2 Improvements to AI and Data Analysis

The second key area of iteration focused on enhancing the AI-driven data analysis system, which played a crucial role in evaluating participant performance, particularly in communication and decision-making.

1. Optimized Whisper Toolkit for Audio Processing

   The Whisper toolkit, which was used to process the audio data from participants during training, initially struggled with varying noise levels and complex sound environments typical of MCI scenarios. This affected the accuracy of the ASR model in capturing verbal communications. To address this issue, we experimented with different configurations of the Whisper toolkit. After testing various setups, a hybrid configuration of 'small' and 'medium' network models was selected, which provided a balance between computational efficiency and speech recognition accuracy. This optimization improved the tool's ability to accurately capture and analyze verbal communications, particularly during high-pressure scenarios where background noise was present. The improved audio processing allowed us to better assess participants' METHANE reports and other verbal interactions, providing more stable feedback on communication performance.

2. Enhanced Semantic Analysis with GPT:

   The initial feedback indicated that the system occasionally struggled with the consistency of the reports, particularly when participants deviated from standard formats or used less structured communication. To improve this, we fine-tuned the GPT model's prompts and parameters to ensure greater stability and accuracy in evaluating the semantic content of the reports. This adjustment made the system more robust in analyzing the clarity, completeness, and relevance of the information provided by participants during the training.

By adopting this iterative approach, we ensured that the final version of the VR learning tool was not only functional and effective but also user-friendly and comfortable for extended use. The ongoing process of evaluation, refinement, and re-evaluation resulted in a training tool that was able to meet the needs of emergency healthcare professionals, providing an immersive and realistic training experience that improved decision-making and communication skills in MCI scenarios.

3.7 Summary

This chapter has diligently delineated the comprehensive research framework that underpinned the development and execution of a VR learning tool and training experiment. The systematic approach adopted in this study reflects the meticulous planning and strategic execution necessary for ensuring the efficacy and educational value of the VR learning tool.

Initially, the chapter elaborated on the Development Tools, an integral component of the research framework. This section provided a detailed exposition of the VR development tools, highlighting the pivotal role of the Unity Engine in creating interactive virtual environments. Additionally, the employment of AI development tools was underscored, emphasizing their significance in constructing humanoid character interactions and advanced evaluation methodologies. This integration of VR and AI tools was central to the development of an immersive learning environment, pivotal to the success of this research.

Subsequently, the VR Learning Tool phase was explicated. This segment of the chapter revealed the intricate assembly of hardware and software designed to deliver an engaging VR experience. Particularly notable was the Data Evaluation Process, which employed sophisticated statistical methods and AI techniques to process and analyze collected data, transforming it into actionable educational insights.

Further, the chapter meticulously outlined the VR Training Experiment phase, detailing the comprehensive preparation and execution of the training sessions. Key aspects of this phase included ensuring ethical compliance and orienting participants effectively to the VR environment. The breadth of data collection during these sessions was emphasized, illustrating the commitment to capturing a comprehensive view of participant interactions within the VR setting.

A defining characteristic of the research framework was its iterative nature, as highlighted in the chapter. This iterative approach was essential for the ongoing refinement of the VR learning tool, driven by data analysis and focused on delivering measurable educational value. The ability to adapt and evolve the tool based on iterative feedback underscored the dynamic nature of this research.

In summation, this chapter has not only meticulously set forth the foundational aspects of the research but also covered Phases 2 to 4 of the research process, from conceptual framework design through artefact development to experiment implementation. The depth and systematic nature of the approach adopted in this study are indicative of a strong commitment to advancing the field of VR in educational tools. These phases established the foundation for the evaluation presented in chapter 4, where Phase 5 (Evaluation and Data Analysis) is addressed. This research holds the promise of revolutionizing training methodologies, particularly in emergency healthcare, through the introduction of immersive and impactful learning experiences, thereby making a significant contribution to the domain of VR-enhanced education.

Chapter 4

Results

This chapter corresponds to Phase 5 (Evaluation and Data Analysis) of the research process, presenting quantitative and qualitative findings from the VR training tool and AI evaluation module. In this study, data were collected from a group of 10 participants. Unfortunately, 2 out of the 10 participants were only able to partially complete the scenario due to discomfort and dizziness in the VR environment during training. However, the remaining 8 participants successfully completed the entire training. This section presents the outcomes of the data evaluation process.

4.1 Sensor Data Result

As mentioned above, we built scripts into the VR learning tool to collect key actions from the controllers and consolidate them as sensor data. We evaluated the performance of each participant from the perspectives of timespan, correctness of triage categorization, and triage orders.

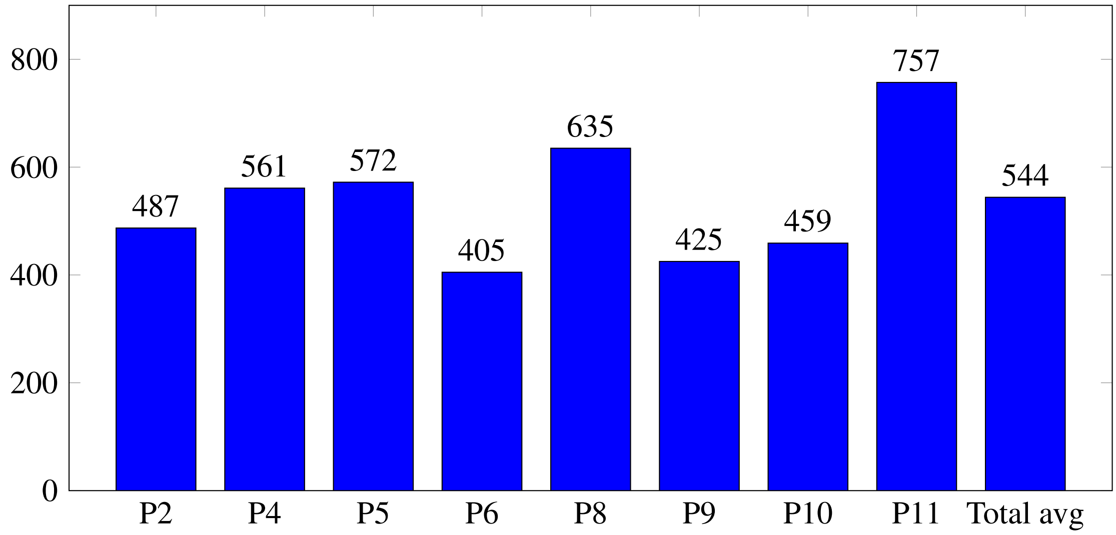

Fig. 4.1 Total Timespan (Sceanrio 1, second)

4.1.1 Scenario 1 - Car crash accident

Figure 4.1 shows the total timespan of the car crash scenario for each participant as well as the overall average. The longest timespan was 757 seconds (participant 11), and the shortest timespan was 405 seconds (participant 6). The total average was 544 seconds (approximately 9 minutes), which satisfied the general triage timespan requirement.

Figures 4.3 through 4.7 provide a detailed breakdown of the individual Casualty timespans in pairs.

Participant 2 devoted a total of 487 seconds to the triage exercise, demonstrating a solid start by allocating a moderate amount of time to the initial METHANE report and maintaining relatively uniform times for each casualty. A notable increase in time spent on Casualty 2 and the concluding METHANE report suggested a detailed approach to these critical components. This participant showed variability in their assessment durations, indicating inconsistency in the triage process.

Participant 4 spent a total of 561 seconds, with less time on the initial METHANE report, possibly indicating a greater familiarity with the procedure. This participant dedicated more time to Casualties 1 and 2, which could signify a more in-depth assessment of these cases. A considerable portion of time was spent on the final METHANE report, possibly reflecting a thorough wrap-up. The participant exhibited moderate timespans, with the initial phase taking 30 seconds and the updated phase notablely longer at 113 seconds, indicating a more thorough or potentially more complex assessment process in the updated phase.

Participant 5 took a total of 572 seconds, the highest time on the initial METHANE report among the group, potentially indicating thoroughness or a need for acclimation. Time spent on Casualty 2 and the final METHANE report was balanced, showing a consistent and focused approach. This participant had the highest timespan during the initial METHANE phase (101 seconds) and a much lower timespan in the updated phase (67 seconds). This reduction could indicate improved efficiency or changes in triage procedures.

Participant 6 completed the triage in the shortest total time of 405 seconds, which may suggest efficient time management. However, this rapid completion could also hint at possible oversights in detail due to the uniformly lower times across casualties. The participant showed consistency with moderate timespans, with the initial phase at 69 seconds and the updated phase slightly higher at 73 seconds, indicating a stable triage process.

Participant 8 used 635 seconds, the second-longest time, with a significant investment in Casualties 1 and 2, and no time recorded for Casualty 3, possibly indicating an error or omission. This participant's time investment in the final METHANE report and Casualty 8 indicates a focus on more complex aspects of the scenario. The participant exhibited variability with an initial phase timespan of 82 seconds and an updated phase significantly lower at 52 seconds, suggesting improvements in the assessment process.

Participant 9, with a total time of 425 seconds, had a relatively shorter duration compared to others. This participant's consistent approach across most casualties and a more substantial focus on the initial METHANE report suggested a thorough initial assessment, which was maintained through to the final METHANE report. The participant had consistent timespans with 83 seconds in the initial phase and 67 seconds in the updated phase, indicating steady performance with slight improvements.

Participant 10's time allocation was even across casualties, totaling 459 seconds. A spike in time for Casualty 4 could imply a challenging case, and their balanced time on the final METHANE report suggests an even-handed summarization. The participant showed lower timespans in the initial phase (41 seconds) compared to the updated phase (40 seconds), suggesting stable performance with minimal changes.

Participant 11, with a total of 757 seconds, had the longest duration, particularly concerning given no time was spent on the initial METHANE report, suggesting a missed critical step. The extended times for Casualties 1 and 2 and the final report suggest a detailed yet inefficient process. The participant had a dramatic increase in the updated phase (156 seconds) compared to 0 seconds in the initial phase, indicating a notable change in the assessment process or complexity.

It was observed that Participants 4, 5, and 8 spent more time overall, with Participant 8 dedicating no time to Casualty 3. In contrast, Participants 6 and 9 were more efficient. Participant 11's approach was the most time-consuming, particularly in the final METHANE report, where they spent the most time compared to others. This cross-comparison indicates that while some participants prioritized thoroughness, others balanced efficiency with detail, and some may need to improve their time management or understanding of the triage process.

Figure 4.2 illustrates a comparison between the initial METHANE phase and the updated METHANE phase. The initial METHANE timespan varied significantly across different participants, with a peak at Participant 5 (101 seconds) and a notable drop to 0 seconds for Participant 11. The updated METHANE phase exhibited a more consistent pattern, with the highest timespan observed for Participant 11 (156 seconds). This comparison indicates a more structured and perhaps more time-intensive approach in the updated METHANE phase.

Figure 4.3 presents the individual timespans for Casualty 1 and Casualty 2. Casualty 1's timespan ranged from 40 seconds (Participant 9) to 148 seconds (Participant 11), with an average of 82.25 seconds. Casualty 2, however, had slightly higher times, ranging from 33 seconds (Participant 9) to 175 seconds (Participant 8), averaging 88.75 seconds. These results suggest that Casualty 2's triage assessments were consistently more time-intensive than those of Casualty 1.

Figure 4.4 shows the timespans for Casualty 3 and Casualty 4. Casualty 3 had relatively shorter timespans, with a minimum of 0 seconds (Participant 8) and a maximum of 89 seconds (Participant 11), averaging 31.5 seconds. Casualty 4, on the other hand, showed higher variability, with times ranging from 14 seconds (Participant 2) to 74 seconds (Participant 10), averaging 42.5 seconds. This variation highlights differences in the complexity or the approach taken for these casualties.

Figure 4.5 details the timespans for Casualty 5 and Casualty 6. Casualty 5's timespans were more evenly distributed, ranging from 18 seconds (Participant 9) to 51 seconds (Participant 4), with an average of 32.62 seconds. Casualty 6 had slightly shorter timespans, ranging from 15 seconds (Participant 9) to 35 seconds (Participant 11), averaging 28.15 seconds. This data indicates a relatively efficient triage process for both casualties.

Figure 4.6 illustrates the timespans for Casualty 7 and Casualty 8. Casualty 7's timespans varied between 15 seconds (Participant 11) and 32 seconds (Participant 9), averaging 24.25 seconds. Casualty 8 exhibited higher variability, with times ranging from 14 seconds (Participant 2) to 64 seconds (Participant 8), averaging 34.62 seconds. This suggests that while Casualty 7's assessments were quicker, Casualty 8 required more detailed evaluations.

Figure 4.7 focuses on Casualty 9's timespans. Casualty 9's times were relatively consistent, with the lowest at 17 seconds (Participant 6) and the highest at 77 seconds (Participant 2), averaging 39.25 seconds. This consistency implies a stable triage process for Casualty 9.

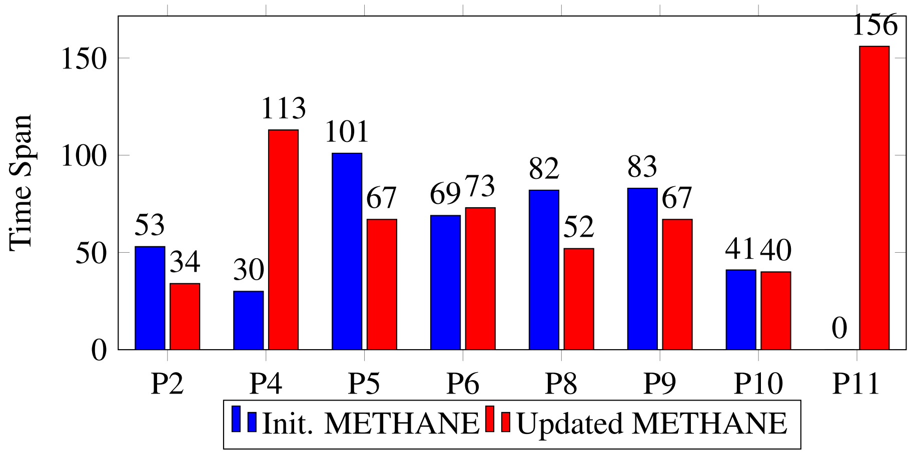

Fig. 4.2 Comparison of Init. METHANE and Updated METHANE Time Spans

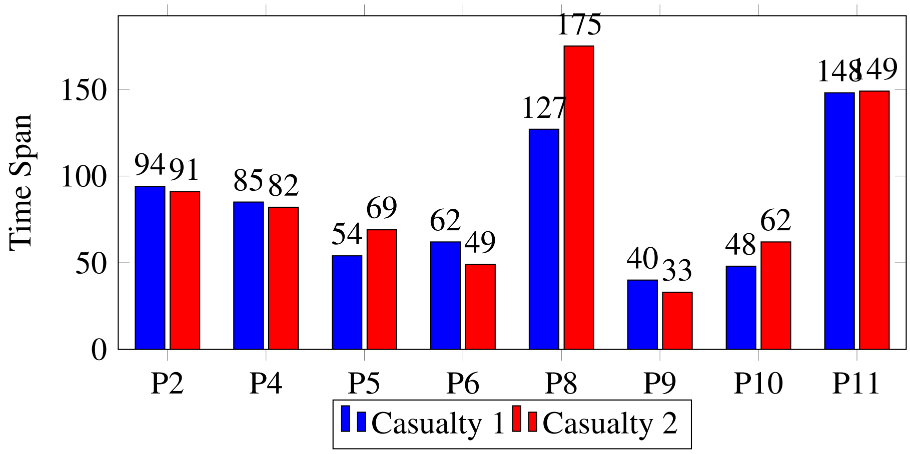

Fig. 4.3 Individual Casualty Timespan - Casualty 1 and Casualty 2

These detailed visualizations offer a comprehensive view of the individual timespans for each casualty, emphasizing the variability in triage efficiency and the distinct characteristics of each assessment. This analysis can provide valuable insights into improving triage protocols and understanding the factors contributing to the time required for casualty evaluations.

The extended time taken for triage assessments of Casualties 1 and 2, as shown in Figure 4.8, can be explained by several key factors. Primarily, these casualties were the first and second individuals encountered by participants in their VR sessions. This initial engagement meant participants were still familiarizing themselves with the virtual reality setting, including tasks like using the controller for airway assessment, which understandably

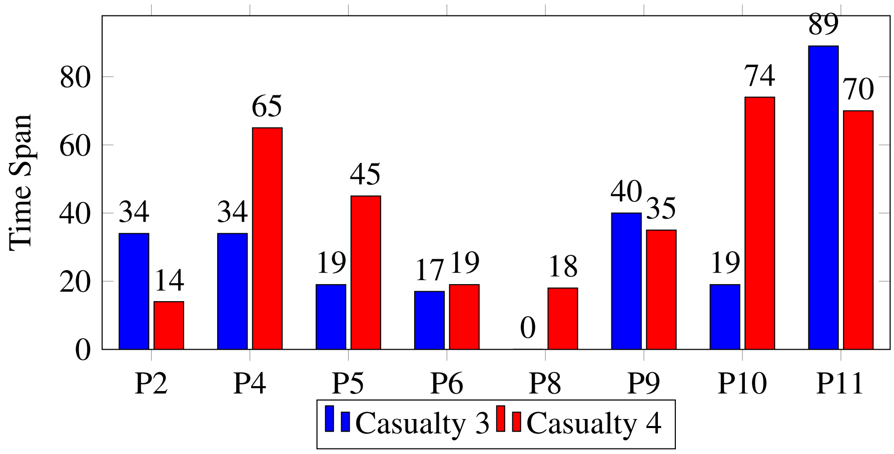

Fig. 4.4 Individual Casualty Timespan - Casualty 3 and Casualty 4

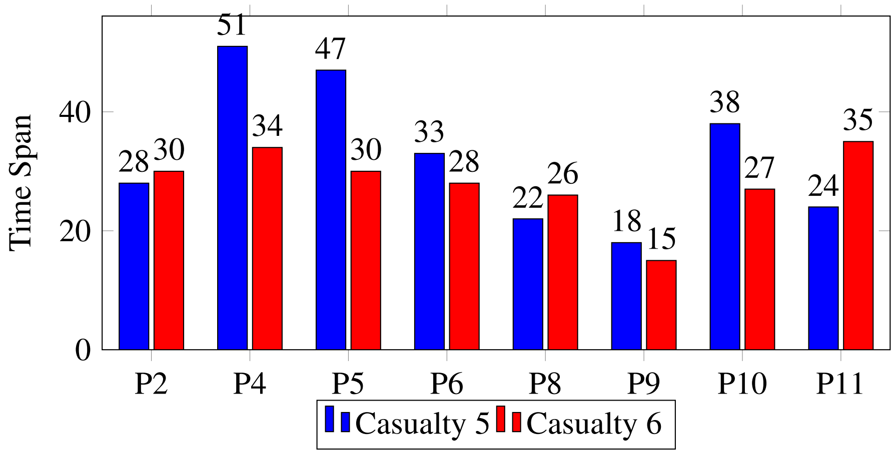

Fig. 4.5 Individual Casualty Timespan - Casualty 5 and Casualty 6

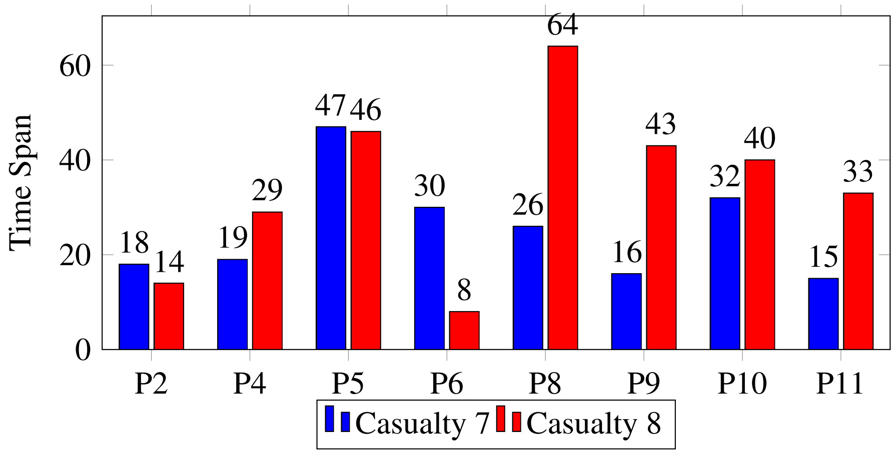

Fig. 4.6 Individual Casualty Timespan - Casualty 7 and Casualty 8

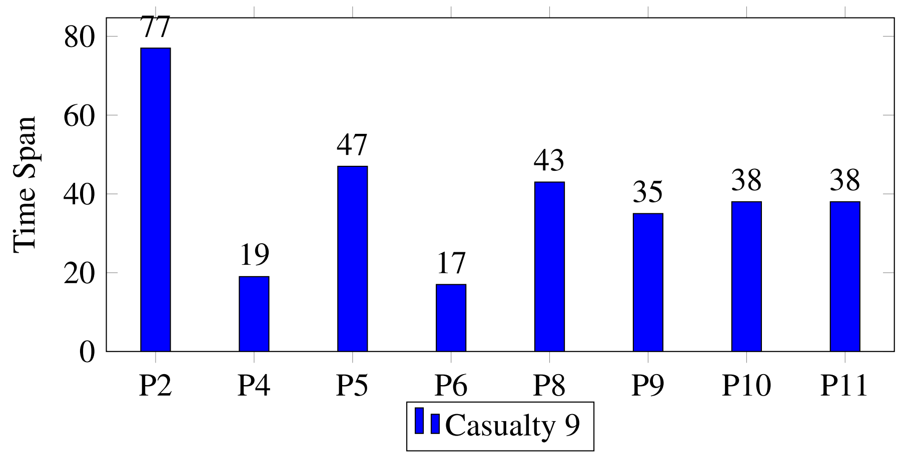

Fig. 4.7 Individual Casualty Timespan - Casualty 9

took more time. Moreover, the specific actions required for these casualties further prolonged their assessments. Notably, Casualty 1's situation required the application of a tourniquet, and Casualty 2's case involved additional procedural steps. These unique requirements contributed significantly to the longer duration observed in their assessments.

Additionally, the study evaluated the precision of the triage process. As depicted in Figure 4.9, the graph illustrates the overall accuracy of triage decisions in the first scenario. The majority of participants successfully identified at least six casualties correctly during the triage. Nonetheless, there were two participants who correctly triaged only three casualties. Further insights were gained from analyzing the responses in the post-experiment questionnaire. Notably, these two participants reported experiencing symptoms such as nausea and dizziness while participating in the experiment. This physical discomfort likely impacted their performance, contributing to their lower success rate in triage accuracy.

In light of these findings, it is essential to consider the impact of the virtual reality environment on participants' physical well-being and its subsequent effect on their decision-making capabilities. The correlation between the physical side effects experienced by the participants and their triage performance underscores the need for more ergonomic and user-friendly VR interfaces, especially in training simulations that demand high concentration and

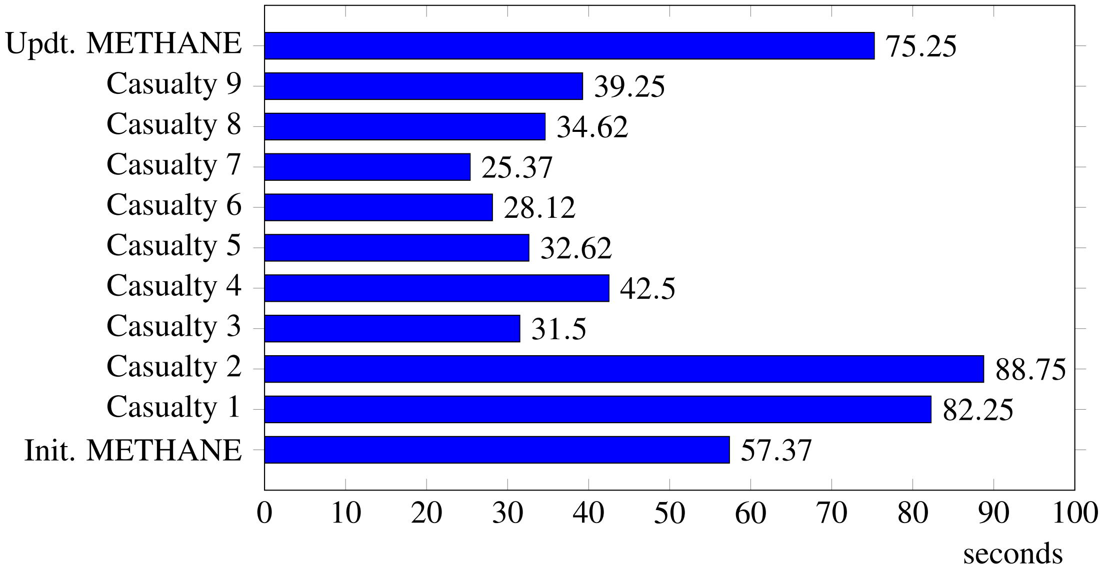

Fig. 4.8 Average Timespan for Individual Casualties and METHANE report(Scenario 1)

accuracy. Future research could focus on minimizing these adverse effects to ensure more consistent performance across all participants.

Summary of Scenario 1 Sensor data evaluation

Analyzing the sensor data from the VR training sessions provided valuable insights into the participants' performance, correlating their actions with the training's effectiveness. The key findings were as follows:

- Timespan Analysis:

  The total timespan for completing the triage varied significantly among participants, with the longest being 757 seconds and the shortest 405 seconds, against an average of 544 seconds. This variation suggested differing levels of efficiency and comfort with the VR environment among participants.

- Triage Assessment Time:

  The data indicated that Casualty 2's triage assessment took the longest time on average, followed closely by Casualty 1's assessment. This could be attributed to these casual-

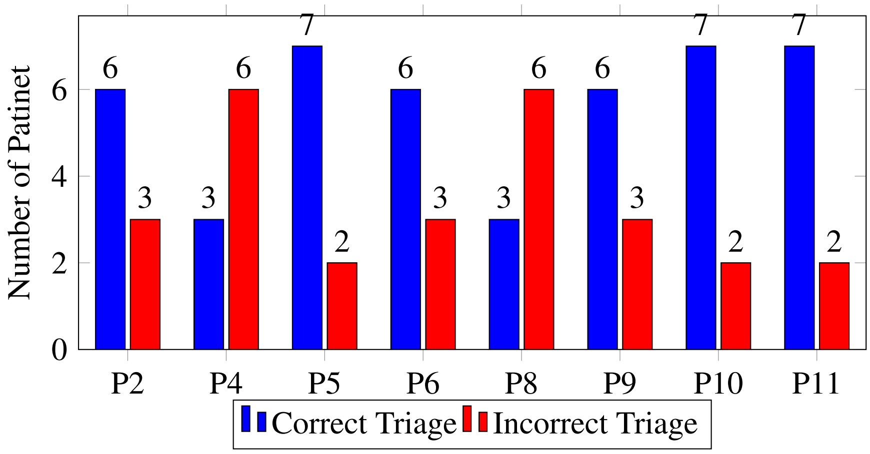

Fig. 4.9 Correctness of total triage (Scenario 1)

ties being the first ones encountered in the VR sessions, which required participants to familiarize themselves with the VR controls and environment. The specific requirements for these casualties, such as applying a tourniquet for Casualty 1 and additional steps for Casualty 2, also added to the time taken.

- Triage Accuracy:

  The overall accuracy of triage decisions varied, with most participants correctly identifying at least six casualties. However, a couple of participants had a notable lower accuracy rate. These participants reported experiencing physical discomfort like nausea and dizziness, which likely impacted their performance.

- Correlation with Physical Well-being:

  The correlation between participants' physical side effects and their performance in triage pointed to the importance of considering the ergonomic design of VR interfaces. The physical discomfort experienced by some participants suggested a need for VR environments that are more user-friendly and minimize adverse effects to ensure consistent performance across all users.

- Impact on Decision-making:

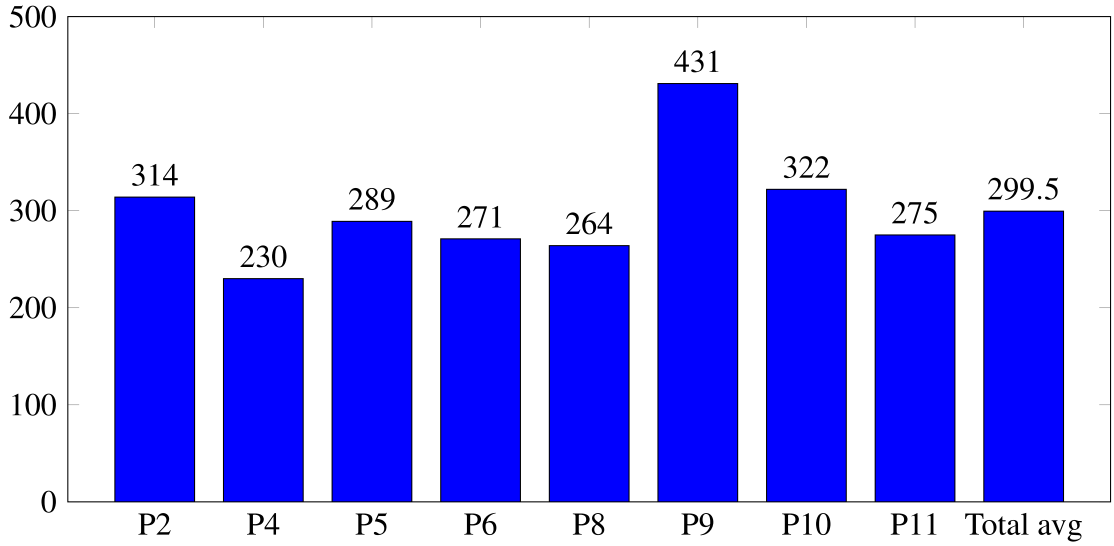

Fig. 4.10 Total Timespan (Sceanrio 2, second)

The physical well-being of participants in a VR setting appeared to directly affect their decision-making capabilities. This was particularly crucial in training simulations that required high concentration and accuracy, underscoring the need for VR systems that are both immersive and comfortable.

4.1.2 Scenario 2 - Earthquake

Figure 4.10 presented the total timespan of the earthquake scenario for each participant, as well as the overall average. The longest timespan was 431 seconds (participant 9), while the shortest was 230 seconds (participant 4). The total average was 300 seconds (approximately 5 minutes), which satisfied the general triage timespan requirement.

Figures 4.11 through 4.14 provide a detailed breakdown of the individual casualty timespans in pairs.

Participant 2 devoted a total of 314 seconds overall, starting with a thorough 71 seconds on the initial METHANE report and displaying a balanced approach with consistent times across casualties. The most time, 43 seconds, was allocated to Casualty 4, indicating a targeted focus on potentially more complex cases. The session was rounded off with a relatively high 81 seconds on the updated METHANE report, suggesting a comprehensive and reflective conclusion, emphasizing the participant's attention to both initial and final assessments.

Participant 4 completed the exercise in 230 seconds, demonstrating efficiency throughout the process. They began with 45 seconds on the initial METHANE report, a moderate amount hinting at familiarity with the protocol. The times were fairly uniform across casualties, with a slight emphasis on Casualty 4 (42 seconds), and concluded with 34 seconds on the updated METHANE report, maintaining a swift yet balanced approach that ensured all areas received adequate attention.

Participant 5 took 289 seconds, indicating a thorough and methodical approach. They spent 51 seconds on the initial METHANE report and dedicated the most considerable time, 60 seconds, to Casualty 4, likely reflecting a detailed and meticulous assessment. The final METHANE report also received notable attention with 65 seconds, showcasing a consistent strategy from start to finish, highlighting the participant's methodical nature and focus on detailed evaluations.

Participant 6 had a total time of 271 seconds, characterized by a uniform distribution of times across the exercise, highlighting a balanced method. They spent 47 seconds on the initial METHANE report and ended with a substantial 71 seconds on the updated METHANE report, indicating a strong focus on both the beginning and end phases. This approach suggests thorough initial assessments and reflective final evaluations, ensuring comprehensive coverage throughout the session.

Participant 8 spent 264 seconds, starting with a notable 62 seconds on the initial METHANE report, the highest among the participants, indicating a careful and detailed start. The time was evenly distributed across casualties, except for an outlier of 22 seconds on Casualty 3, which might indicate a more straightforward case. The session concluded with 55 seconds on the updated METHANE report, reflecting a focused and detailed end, indicative of the participant's consistent and methodical approach.

Participant 9 had the longest total time of 431 seconds, with a meticulous 94 seconds dedicated to the initial METHANE report, indicating an extremely detailed and thorough evaluation. An exceptional 103 seconds were spent on Casualty 1, suggesting a particularly complex or critical case. The participant's approach was otherwise consistent, with the session concluding with 87 seconds on the updated METHANE report, underscoring the participant's detailed and comprehensive evaluation throughout.

Participant 10 used 322 seconds, beginning with 57 seconds on the initial METHANE report and showing pronounced focus on Casualty 1 (60 seconds) and Casualty 4 (54 seconds). This focus might indicate the need for detailed examinations of these complex cases. The session ended with a balanced 42 seconds on the updated METHANE report, reflecting an equitable summarization and the participant's ability to manage time effectively across various assessment stages.

Participant 11 spent 275 seconds, the shortest among the participants, starting with only 31 seconds on the initial METHANE report, potentially indicating high confidence or a missed critical step. A notable 105 seconds were spent on Casualty 4, suggesting an intensive assessment of a particularly challenging case. The exercise concluded with a thorough 52 seconds on the updated METHANE report, indicating a detailed final review, highlighting the participant's focus on addressing complex cases and ensuring comprehensive evaluations.

Figure 4.11 illustrated a comparison between the initial METHANE phase and the updated METHANE phase. The initial METHANE timespan varied significantly across different participants, peaking at Participant 8 (94 seconds) and dropping notably to 31 seconds at Participant 11. The updated METHANE phase exhibited a more consistent pattern, with the highest timespan observed at Participant 11 (52 seconds). This comparison indicated a more structured and perhaps more time-intensive approach in the updated METHANE phase.

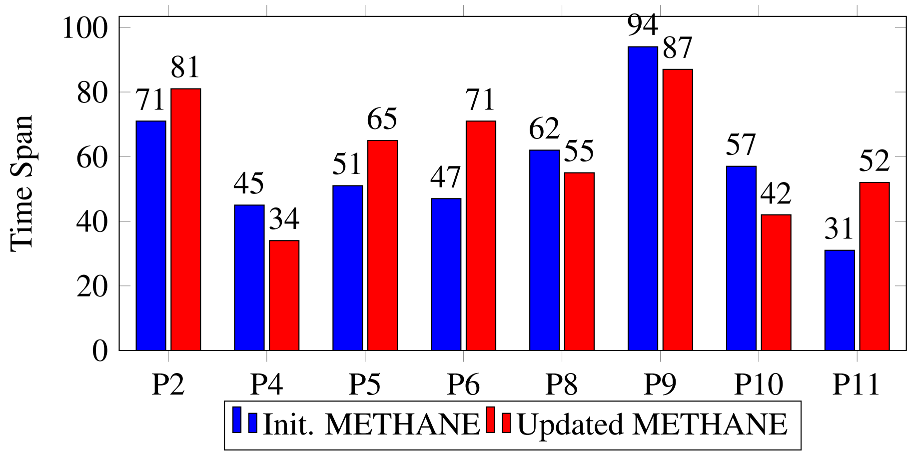

Fig. 4.11 Comparison of Initial METHANE and Updated METHANE Time Spans

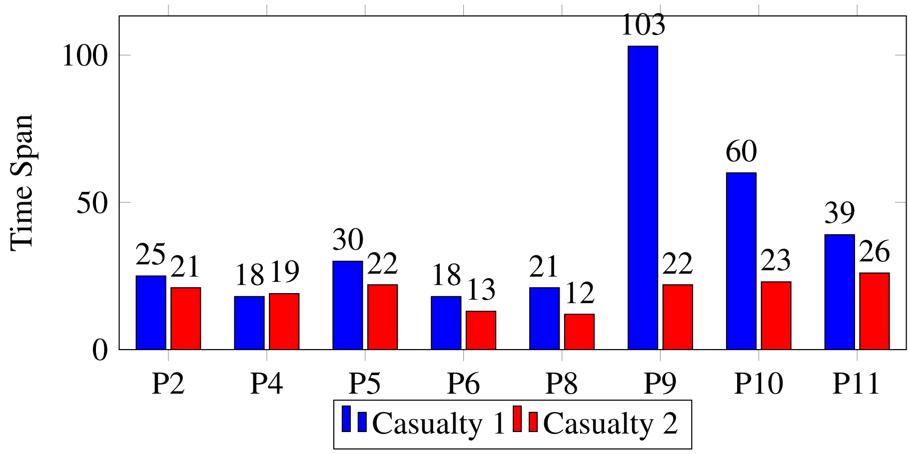

Fig. 4.12 Individual Casualty Timespan - Casualty 1 and Casualty 2

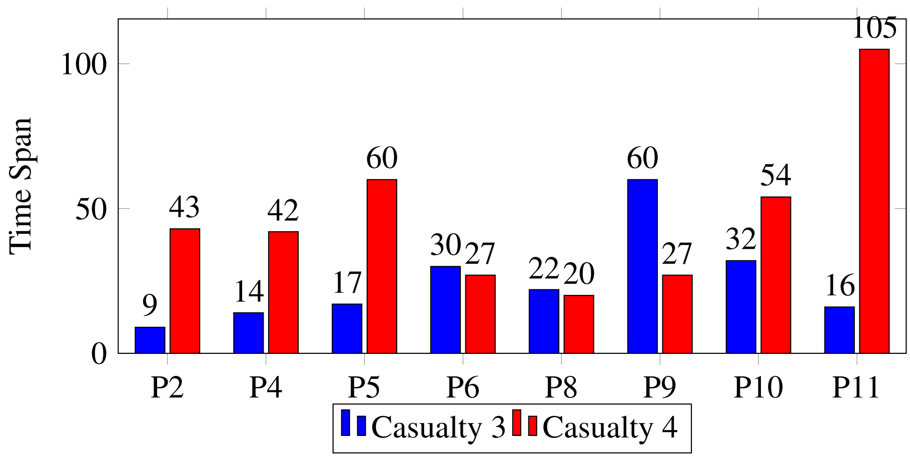

Fig. 4.13 Individual Casualty Timespan - Casualty 3 and Casualty 4

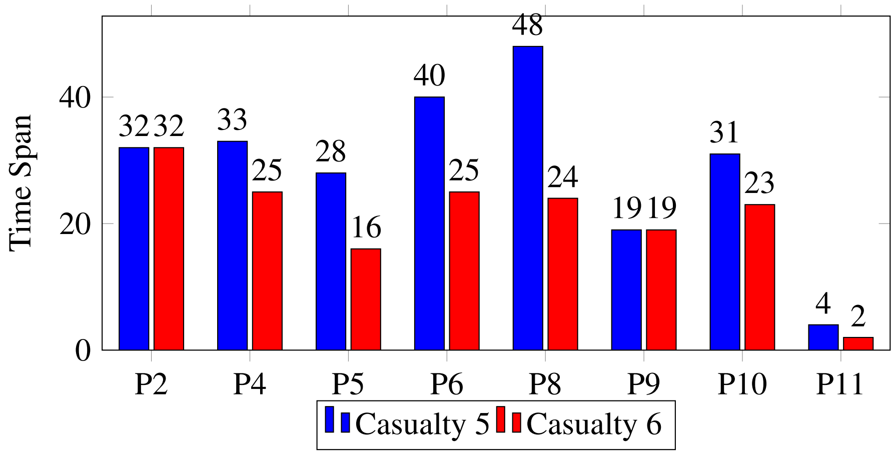

Fig. 4.14 Individual Casualty Timespan - Casualty 5 and Casualty 6

Figure 4.11 presented the individual timespans for Casualty 1 and Casualty 2. Casualty 1's timespan ranged from 18 seconds (Participant 4) to 103 seconds (Participant 9), with an average of 39 seconds. Casualty 2 had slightly higher times, ranging from 12 seconds (Participant 8) to 26 seconds (Participant 11), averaging 20.25 seconds. These results suggested that Casualty 2's triage assessments were consistently less time-intensive than those of Casualty 1.

Figure 4.13 showed the timespans for Casualty 3 and Casualty 4. Casualty 3 had relatively shorter timespans, with a minimum of 9 seconds (Participant 2) and a maximum of 60 seconds (Participant 9), averaging 24.75 seconds. Casualty 4, on the other hand, showed higher variability, with times ranging from 20 seconds (Participant 5) to 105 seconds (Participant 11), averaging 47.5 seconds. This variation highlighted differences in the complexity or the approach taken for these casualties.

Figure 4.14 detailed the timespans for Casualty 5 and Casualty 6. Casualty 5's timespans were more evenly distributed, ranging from 4 seconds (Participant 11) to 48 seconds (Participant 8), with an average of 28.25 seconds. Casualty 6 had slightly shorter timespans, ranging from 2 seconds (Participant 11) to 33 seconds (Participant 4), averaging 21.5 seconds. This data indicated a relatively efficient triage process for both casualties.

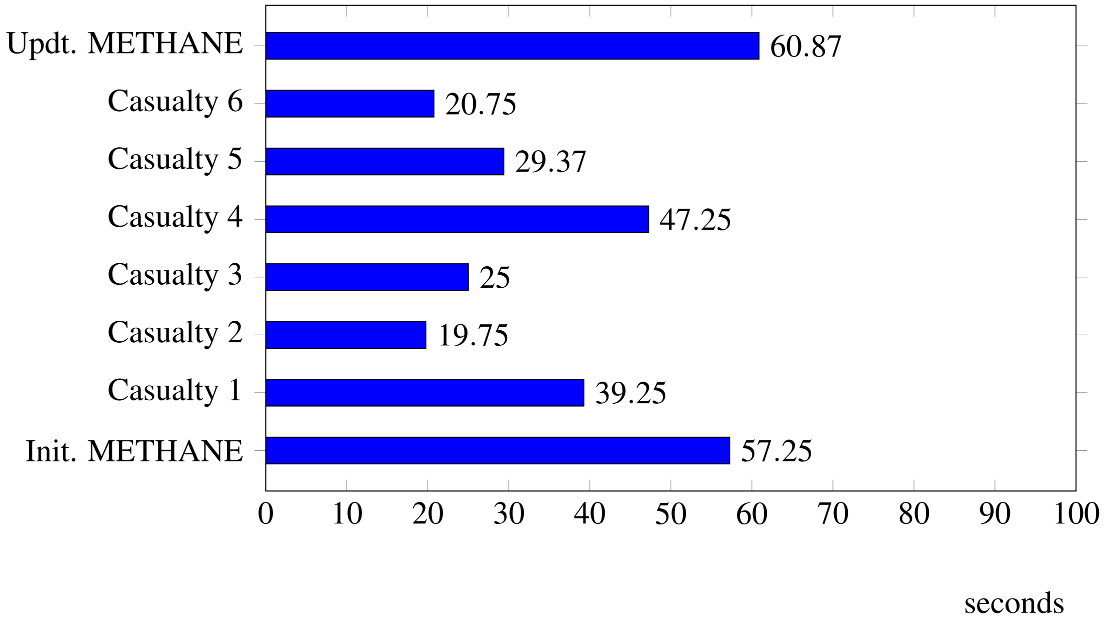

Fig. 4.15 Average Timespan for Individual Casualties and METHANE report(Scenario 2)

These detailed visualizations offered a comprehensive view of the individual timespans for each casualty, emphasizing the variability in triage efficiency and the distinct characteristics of each assessment. This analysis could provide valuable insights into improving triage protocols and understanding the factors contributing to the time required for casualty evaluations.

Figure 4.15 illustrated the average timespan allocated to each casualty and the METHANE phases during Scenario 2. The initial METHANE phase saw a moderate duration of 57.25 seconds, indicating a balanced approach to the initial situational assessment. Casualty 1 required an average of 39.25 seconds, reflecting a notable but efficient investment of time, while Casualty 2, with the shortest average timespan of 19.75 seconds, suggested a straightforward assessment. Casualty 3 had a moderate average timespan of 25 seconds, indicating a balanced approach to managing its complexity. Casualty 4 stood out with the longest average duration of 47.25 seconds, highlighting its higher complexity and the need for a thorough assessment. Casualty 5 and Casualty 6 had average timespans of 29.37 seconds and 20.75 seconds, respectively, indicating relatively efficient evaluations. The updated

METHANE phase, with the longest average timespan of 60.87 seconds, underscored the importance of the final situational assessment and debriefing, reflecting a comprehensive and thorough wrap-up by the participants. Overall, the figure revealed variability in time spent on different casualties and phases, showcasing the diverse complexity of cases and the prioritization strategies employed by participants. The detailed evaluation of Casualty 4 and the comprehensive approach to the updated METHANE phase emphasized the critical aspects of these components within the triage exercise, providing insights into the efficiency and thoroughness of the participants' assessment processes. This analysis could inform improvements in triage protocols and training programs.

The evaluation of triage precision in Scenario 2, as detailed in Figure 4.16, presented a distinct distribution of triage accuracy among the participants. In this scenario, the majority of participants (Participant 2, Participant 4, Participant 5, Participant 9, and Participant 10) correctly triaged three casualties. However, the performance did not significantly exceed this number, with Participant 6 and Participant 11 only correctly triaging two casualties each, a decrease compared to the first scenario. Notably, no participant was able to correctly triage more than three casualties.

The consistency of the triage correctness, with half of the participants incorrectly assessing the same number of casualties as they correctly assessed, suggested potential gaps in the training or the complexity of the task (e.g., the removal of the METHANE reminder) that could have contributed to the challenges faced by the participants.

Summary of Scenario 2 Sensor data evaluation

- Timespan Analysis:

  Scenario 2 displayed a more consistent timespan for triage completion among participants compared to Scenario 1, with the longest at 431 seconds and the shortest at

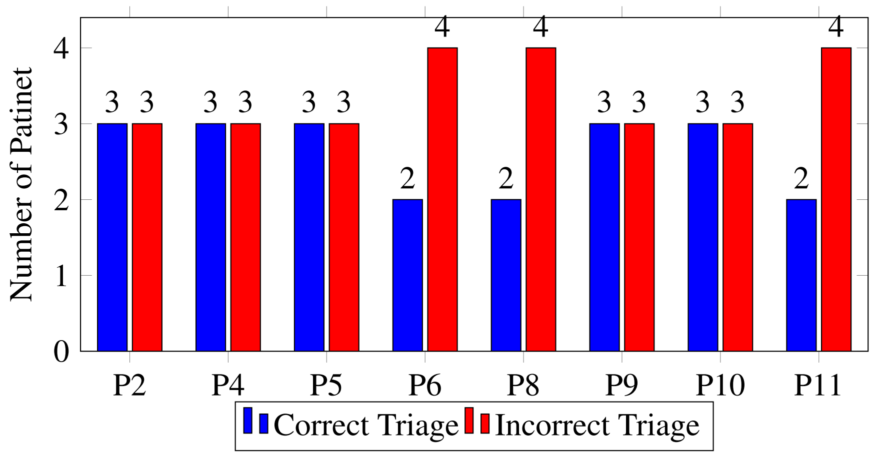

Fig. 4.16 Correctness of total triage (Scenario 2)

230 seconds. The average of 300 seconds met the triage time requirements, indicating improved efficiency and familiarity with the VR environment.

- Triage Assessment Time:

  The average time spent on each casualty varied, with Casualty 4 receiving the most attention at 47.25 seconds, likely due to the complexity of their condition. In contrast, Casualty 2 and Casualty 6 had shorter assessment times, suggesting a streamlined process or potential areas where thoroughness might have been compromised.

- Triage Accuracy:

  Triage accuracy was generally lower in Scenario 2, with most participants correctly identifying only three casualties. This result may reflect the increased complexity of the scenario or the absence of decision aids like the METHANE reminder, highlighting the potential need for additional training or support.

- Impact on Decision-making:

  The variation in performance emphasized the significance of the VR interface and scenario design in participant decision-making. The findings suggest that the clarity of the triage process and the user-friendliness of the VR technology are crucial for achieving accurate triage outcomes.

Participant	SC1 Timespan (s)	SC1 Rank	SC2 Timespan (s)	SC2 Rank
P2	378	4th	314	6th
P4	494	6th	230	1st
P5	432	5th	289	5th
P6	350	3rd	271	3rd
P8	502	7th	264	2nd
P9	331	1st	431	8th
P10	349	2nd	322	7th
P11	671	8th	275	4th

Table 4.1 Timespan and Rankings of Participants in Scenario 1 and Scenario 2

The performance rankings of participants across Scenario 1 and Scenario 2 revealed notable patterns of consistency and variability. Participants P5 and P6 demonstrated high consistency in their performance, maintaining their rankings at 5th and 3rd positions, respectively, across both scenarios. This consistency indicates a stable and efficient approach to the triage tasks in both settings.

In contrast, participants P2, P4, P8, P9, P10, and P11 exhibited notable variability in their performance. Participant P4 showed a remarkable improvement, moving from 6th place in Scenario 1 to 1st place in Scenario 2, suggesting a substantial enhancement in efficiency or adaptability to the different scenario. Similarly, Participant P8 improved from 7th place in Scenario 1 to 2nd place in Scenario 2, indicating better performance under varying conditions.

Conversely, Participant P9, who ranked 1st in Scenario 1, dropped to 8th place in Scenario 2, signaling a notable decline in performance. Participants P10 and P11 also experienced considerable shifts in their rankings, with P10 dropping from 2nd to 7th place and P11 improving from 8th to 4th place.

These findings highlight the varying degrees of adaptability and consistency among the participants. The improvements observed in some participants suggest that certain individuals may benefit from specific training or increased familiarity with the task, while the decline in others indicates potential challenges in adapting to different scenarios. This analysis underscores the importance of flexibility and continuous improvement in triage performance and can inform targeted training programs to enhance overall efficiency and effectiveness in emergency response scenarios.

In conclusion, the sensor data from the VR training sessions revealed important correlations between the time taken for triage assessments, the accuracy of these assessments, and the participants' physical comfort in the VR environment. These findings suggest that while VR technology is effective for training purposes, attention to user comfort and interface design is critical for enhancing the overall quality and effectiveness of the training.

4.2 Survey Result

As previously indicated, we conduct two surveys, one before and one after each training session. A total of 10 responses for each survey across all participants.

Initial Survey

The initial survey comprises 4 questions designed to collect background information from participants, including details about their age group, education background, MCI experience, and VR exposure. Most questions are single-choice format, with only one question being of the multiple-choice type.

Among these 10 participants, 7 fall within the age group of 18 to 24, while the remaining 3 belong to the age group of 25 to 34. The distribution of ages across all participants is illustrated in Figure 4.17. Regarding participants' educational back-grounds, 9 individuals

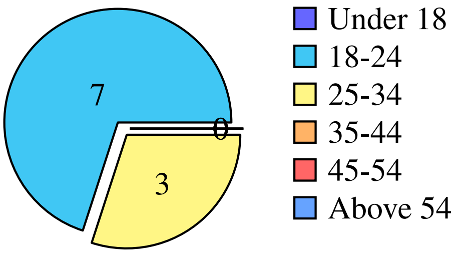

Fig. 4.17 Age Distribution

Fig. 4.18 Educational background distribution

possess at least a bachelor's or Undergraduate diploma, whereas the remaining 1 participant has completed a certificate course or program. Figure 4.18 illustrates the distribution of educational backgrounds.

Regarding the MCI knowledge (multiple choice) question, 5 out of 10 participants reported deriving theoretical knowledge from lectures and textbooks, while 6 out of 10 participants engaged in training through paper simulations. Additionally, 8 out of 10 participants gained experience through live simulations like AUT MI Day. Interestingly, none of the participants had encountered VR-based MCI simulations. The distribution of knowledge sources is illustrated in Figure 4.19.

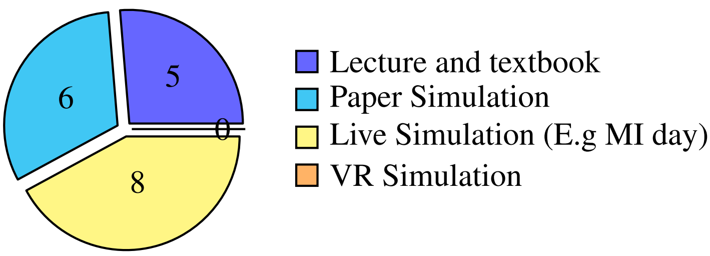

Fig. 4.19 Experience of MCI training

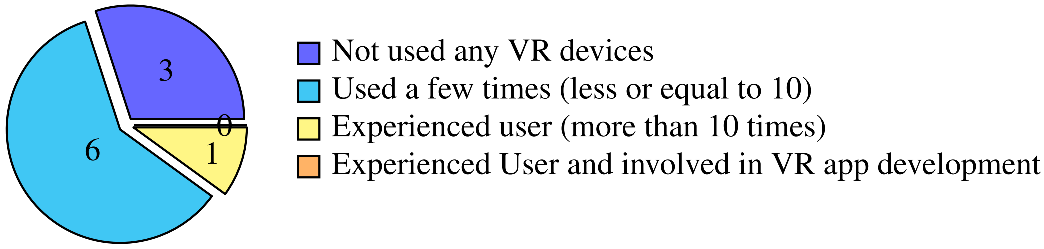

Fig. 4.20 Experience of VR Technology

The final question in the initial survey pertained to participants' familiarity with VR devices. Notably, 6 out of 10 participants indicated that they had used VR devices 10 times or less, highlighting limited exposure. One participant reported being well-versed with more than 10 instances of VR usage, while another participant had never interacted with a VR device. This range of familiarity with VR equipment underscores the diversity of experience levels among the participants. Figure 4.20 presents the results of VR device familiarity.

Post-experiment survey

The post-experiment survey consisted of a comprehensive set of 16 questions, thoughtfully organized to serve distinct data collection purposes, encompassing immersion, presence, engagement, emotion, usability, skill, experience consequences, technology adoption, and other relevant domains. The majority of these questions were designed to elicit participant feedback regarding their specific experiences during the experimental session. Additionally, a few questions sought participants' valuable recommendations for potential future enhancements. The survey primarily employed a Likert scale to facilitate participant responses, offering a spectrum from "strongly agree" to "strongly disagree." Furthermore, some questions were structured in an open-ended format to accommodate diverse feedback.

Question 1 asked about the immersion of the VR learning tool. Five participants strongly agreed that VR is more immersive than paper-based simulations in MCI triage, while the remaining four participants agreed with this sentiment. The results indicated that most participants felt immersed in the VR learning tool. Table 4.2 shows the result of question 1.

Table 4.2 Feedback on Immersion

	Question (Immersion)
Strong Agree	5
Agree	4
Neutral	0
Disagree	0
Strong Disagree	0

Questions 2 and 3 inquired about the perception of the VR training tool's presence. In response to question 2, three participants expressed strong agreement with feeling a sense of naturalness during their interaction with the VR environment. An additional four participants also agreed, while the remaining three participants were neutral. In response to question 3, three participants strongly agreed that they could examine objects closely, six participants agreed, and the remaining one participant was neutral. The results suggest that the majority of participants felt a natural interaction in the VR environment. Table 4.3 shows the result of question 2 and 3.

Table 4.3 Feedback on Presence

	Question 2 (Presence)	Question 3 (Presence)
Strong Agree	3	3
Agree	4	6
Neutral	3	0
Disagree	0	1
Strong Disagree	0	0

Questions 4, 5, and 6 were structured to gauge aspects related to engagement, emotion, and usability. Regarding question 4 (engagement), six participants strongly agreed with their sense of engagement during the VR experience, while an additional four participants also agreed. In response to question 5 (emotion), two participants reported feeling a sense of tension within the VR environment, while five participants expressed neutrality in their emotional state, and the remaining three did not report feeling tense. For question 6 (usability), four participants agreed that the VR device was user-friendly, with one participant strongly agreeing. Meanwhile, five participants maintained a neutral stance on the usability of the device. The results indicated that all participants felt engaged with the VR learning tool, the majority did not feel tense while completing tasks in the VR environment, and all found the VR learning tool easy to use. Table 4.4 shows the result of question 4, 5 and 6.

Table 4.4 Feedback on Engagement, Emotion, and Usability

	Question 4 (Engagement)	Question 5 (Emotion)	Question 6 (Usability)
Strong Agree	6	2	1
Agree	4	5	4
Neutral	0	3	5
Disagree	0	0	0
Strong Disagree	0	0	0

Questions 7 and 8 aimed to collect feedback on skill satisfaction. In response to question 7, four participants reported strong confidence in selecting objects in the VR environment, with an additional four participants agreeing. Two participants remained neutral. In response to question 8, four participants strongly agreed that they were confident in moving around in the VR environment using the controller, while four participants agreed, and one remained neutral. One participant did not feel confident moving around. The results suggest that most participants were confident while performing actions in the VR environment. Table 4.5 shows the result of question 7 and 8.

Table 4.5 Feedback Skill

	Question 7 (Skill)	Question 8 (Skill)
Strong Agree	4	4
Agree	4	4
Neutral	2	1
Disagree	0	1
Strong Disagree	0	0

Questions 9 and 10 specifically addressed potential side effects from using the VR device. In response to question 9, two participants strongly reported experiencing nausea during their interaction with the VR environment, one participant partially acknowledged this sensation, and the remaining six participants either did not report feeling any nausea or remained neutral. In response to question 10, the results indicated that six participants did not experience dizziness when using the VR device, while two participants reported feeling dizzy, and two participants remained neutral. These outcomes suggest that the current state of VR technology still has limitations in addressing issues of nausea and dizziness. However, it is noteworthy that most participants were able to successfully complete the training sessions despite these challenges. Table 4.6 shows the result of question 9 and 10.

Table 4.6 Physical Well-being

	Question 9 (Nausea)	Question 10 (Dizziness)
Strong Agree	2	0
Agree	1	2
Neutral	1	2
Disagree	3	3
Strong Disagree	3	3

Questions 11, 12, and 13 focused on technology adoption for the VR learning tool. Impressively, all participants strongly agreed with the statements for all three questions: "The VR system can be an integral tool for MCI training," "I would like to see more VR-based MCI scenarios (e.g., different incidents or disasters)," and "I would like to see more VR for paramedic training (e.g., secondary survey, AED training)." The results indicate that all participants were confident in the future potential of VR technology. Table 4.7 shows the result of question 11, 12 and 13.

Question 14 was an open-ended question that sought to identify the positive aspects of the current experience. Participants in this training expressed highly positive feedback, focusing on several key aspects of their experience. A notable highlight was the realism and immersion of the VR environment. For example, one participant mentioned, "You could

Table 4.7 Feedback on Technology adoption

	Question 11	Question 12	Question 13
Strong Agree	10	10	10
Agree	0	0	0
Neutral	0	0	0
Disagree	0	0	0
Strong Disagree	0	0	0

move around and see the whole scene... it feels more realistic than a classroom setting." This ability to simulate real-life scenarios, particularly in contrast to traditional classroom settings, was highly valued.

The risk-free nature of the training was another notable point of appreciation. As one participant put it, the training session provided a "risk-free and easily accessible way to simulate what an MCI would be like," allowing for safe practice without real-world consequences. The educational value of the training was also commended, with participants noting it as a "good introduction to the MCI environment," offering a "safe learning environment for any student paramedic to practice in."

The enhanced training environment was seen as a major improvement over previous MCI training days. One participant expressed, "I liked the environment. I find in MCI training days we've had they've just been outside and there haven't been any things to immerse me in the experience." The immersive nature of the VR setup, therefore, made the training feel more engaging and realistic.

Participants also appreciated the welcoming atmosphere created by the research team and the opportunity to explore the simulation. As one stated, "I enjoyed the warm welcome I received from the research team. And I enjoyed looking around and getting a feel for the simulation." Furthermore, the training was recognized for its practical skills development, particularly in learning procedures like the METHANE report. A participant noted, "Doing more simulations like this allows us to practice large jobs that we don't go to very often."

Question 15 sought feedback on the negative aspects of the training experience. In response to this open question, participants highlighted several areas for improvement. A recurring issue mentioned by multiple participants was the sensation of nausea or dizziness, often attributed to the movement speed within the VR environment. For example, one participant noted, "I did start to get quite nauseous doing it due to the movement speed," reflecting a common concern about the physical comfort level while using the VR headset.

Another point of critique was the limitation in the simulation's scope and realism. A participant mentioned, "Just that there were aspects not able to be assessed that would be vital to triage, such as casualty ability to mobilize, but I understand this is in development/next step." This highlighted the need for more comprehensive features in the simulation to enhance the training's effectiveness and realism.

The unnatural movement speed and style, causing discomfort, were also highlighted as detracting from the experience. One response mentioned, "Movement speed and style were unnatural, resulting in nausea. Unrealistic object interaction and physics resulted in delayed treatment and triage." This feedback pointed to the need for improved interaction and physics within the simulation to create a more realistic and immersive training environment.

Motion sickness was a standalone issue for some participants, with comments like, "Just the motion sickness," indicating that this was a notable barrier to fully engaging with the VR training. Additionally, some felt the scenarios were not adequately intense or realistic, with one participant suggesting a need for "more blood and screaming from the casualties" to better simulate real emergency situations.

The learning curve associated with using VR technology was another challenge. Participants who were new to VR mentioned struggling with controls initially, which they felt hindered their performance. One participant shared, "Because it was my first time using a VR headset, I struggled with the controls; however, eventually I got the hang of it."

Question 16 asked about potential improvements to the training. In response to this question, participants offered a variety of constructive ideas. A common theme was the need for more control and customization of movement within the VR space. One participant suggested the ability to "slow the movement" and proposed improvements to the wristband interaction in the simulation, noting challenges in grasping and retaining them.

Another practical suggestion was the ability to physically manipulate casualties within the simulation, such as "being able to turn over casualties and position them to assist with opening airways." This feature would add a layer of realism and practical skill application to the training.

Enhancing the realism of movement and object physics was mentioned as a key area that could reduce nausea and increase the time students can train effectively. A participant emphasized that "more realistic movement and object physics will definitely reduce nausea." Additionally, interactive elements such as moving, talking casualties with live injuries were suggested to make the triage and treatment processes more dynamic and realistic.

Reducing the walking speed to alleviate motion sickness was a straightforward yet crucial suggestion from several participants. One simply stated, "Slower walking speed for nausea."

While some were content with the current setup, stating, "Nothing from me. I look forward to what's to come," others focused on the logistics of using VR in training. The idea of wireless VR headsets and larger spaces to move around safely was brought up to prevent accidental injuries from bumping into physical objects like tables and walls.

The incorporation of sound was also a popular suggestion. Participants felt that adding background noise would make the simulation more intense and realistic, better mimicking real-world emergency scenarios. As one participant put it, "Maybe use sound? Without background noise I was able to concentrate on the task, but in a real-world situation, I imagine it would be very noisy."

4.2.1 Summary of Survey evaluation

The analysis of the survey results uncovered distinct correlations among the participants' backgrounds, their experiences with the VR training, and their proposed enhancements. The key findings are summarized in the following points:

- Background Correlation with VR Experience:

  1. Age and Education: The majority of participants were aged 18-24 and held at least an undergraduate diploma. This demographic, typically more familiar with digital technology, reported high engagement and immersion in the VR environment. Their educational background likely contributed to their positive perception of the educational value of the training.
  2. MCI Knowledge Sources: Most participants had experience with live simulations and theoretical knowledge from lectures, but none had prior VR simulation experience. This lack of VR exposure might explain why many found the VR environment novel and immersive, yet also challenging in terms of motion sickness and control navigation.

- Experience During Training Correlation with Feedback:

  1. Immersion and Realism: The strong agreement on the VR learning tool's immersive quality correlated with the positive feedback on the VR environment's realism. Participants appreciated the contrast with traditional training methods, finding VR more engaging and realistic.
  2. Presence and Engagement: The feeling of naturalness in interaction and high engagement levels corresponded with the positive feedback on the VR environment's ability to effectively simulate real-life scenarios.

3. Skill Confidence and Learning Curve: While most participants felt confident in performing actions in the VR environment, there was also a noted learning curve, especially for those new to VR. This was reflected in the suggestions for improved controls and user-friendliness.

- Negative Feedback and Improvement Suggestions Correlation:

  1. Motion Sickness and Movement Speed: Many participants reported nausea or dizziness, which directly correlated with suggestions to reduce movement speed and improve motion physics in the VR environment.
  2. Realism and Scenario Intensity: The desire for more intense and realistic scenarios, including the suggestion for more blood and screaming, was reflected in the feedback where some found the current scenarios lacking in realism.

The analysis of the survey results from the VR training sessions suggested a strong influence of the participants' educational and age backgrounds on their engagement and receptivity to the training. A notable aspect was the lack of prior VR experience among most participants, which, while adding to the novelty and engagement, also posed challenges in navigation and comfort. This was evidenced by issues such as motion sickness, prompting suggestions for slower movement and more realistic physics in the VR setup.

Despite these challenges, the overall response to the VR environment's immersive and educational qualities was overwhelmingly positive, indicating a successful implementation of VR technology in training. However, the correlation between the negative experiences and the proposed improvements pointed to clear areas needing enhancement, particularly in terms of physical comfort and realism.

In conclusion, the surveys reflected that while the VR training was highly valued for its immersion and educational potential, there are significant opportunities for improvement. Enhancing aspects like physical comfort, realism, and user-friendliness could notably elevate the overall effectiveness of the training. Addressing these issues is key to maximizing the benefits of VR in educational settings, especially in scenarios as complex and demanding as the ones simulated in these sessions.

4.3 Speech recognition and semantic similarity evaluation result

We used an AI-based model to facilitate the processing of audio recordings derived from the training sessions. The initial phase involved converting audio waveforms into text-based datasets. In this crucial step, the algorithm incorporated OpenAI Whisper as a pivotal component. Table **??** illustrates example outcomes of the speech recognition algorithm. Importantly, the temporal alignment with the controller was maintained, offering the capability to establish a clear linkage between participants' spoken interactions and the actions executed within the VR learning tool.

TimeSeries	Speech content
11:09:42 AM	go to this guy. Hello sir. Let's see. OPA. Breath 25. Cool. You're breathing good.
11:09:59 AM	Let's see. I can see this. Moaning and groaning. Your leg. Yeah bro. Your leg is sore.
11:10:06 AM	And there's one two. Cool. Where's your arm? Do. Go ahead and just get out. Index finger.

Result of BERT-based model

In our study, we utilized the BERT model to evaluate the semantic similarity between predefined and participant-provided METHANE reports in a VR training scenario. The METHANE reports were essential at both the start and end of each scenario, referred to as the initial and updated reports, respectively. We analyzed these using BERT to assess similarity, with the results presented in Figures 4.21 and 4.22.

In the analysis of the initial METHANE reports (Figure 4.21), the BERT-based semantic similarity scores, which indicate how closely the reports aligned with the ground truth meaning of the sentences (with higher scores representing greater alignment), revealed the following trends:

- Random One-Off BERT-base: The Random One-Off BERT-base scores had a mean value of 0.39, with a standard deviation of 0.21. These scores ranged from a minimum of 0.00 to a maximum of 0.66. This wide range and higher standard deviation indicated considerable variation in one-off comparisons, reflecting the unpredictability and inconsistency inherent in such measures.

- 50 loop BERT-base Average: For the 50-loop BERT-base Average scores, the mean was observed to be 0.43, accompanied by a standard deviation of 0.18. The scores varied from 0.00 at their lowest to 0.53 at their highest. The relatively lower standard deviation compared to the one-off scores suggested more consistency in the 50-loop average, although there remained notable variation.

- 50 loop BERT-base Max: The 50-loop BERT-base Max scores showed a mean of 0.70, with a standard deviation of 0.28, and ranged from 0.00 to 0.85. The high maximum value suggested that some participant reports were highly similar to the standard report at their best, but the wide range indicated inconsistency across different loops.

- 50 loop BERT-base Min: Lastly, the 50-loop BERT-base Min scores had a mean of 0.16 and a standard deviation of 0.10, with scores ranging between 0.00 and 0.31. The low minimum scores indicated instances of very low similarity in certain loops, highlighting the variability in how participant reports aligned with the standard across different iterations.

In the updated METHANE report (Figure 4.22),the BERT-based semantic similarity scores demonstrated the following characteristics:

- Random One-Off BERT-base: The Random One-Off BERT-base scores exhibited a higher level of consistency, with a mean value of 0.58 and a relatively low standard deviation of 0.07. The range of these scores was from 0.48 to 0.68. This narrower range and lower standard deviation, compared to the initial reports, suggested a notable improvement in the stability and reliability of one-off random comparisons in the updated reports.

- 50 loop BERT-base Average: For the 50-loop BERT-base Average scores, the mean was 0.51, with a very low standard deviation of 0.03, indicating a high level of consistency. The scores ranged from a minimum of 0.47 to a maximum of 0.55. This tight range further emphasized the consistency and reliability of the 50-loop average scores in capturing semantic similarity in the updated reports.

- 50 loop BERT-base Max: The 50-loop BERT-base Max scores showed a mean of 0.83, with a standard deviation of only 0.04, and scores varied between 0.78 and 0.90. These high scores, along with the narrow range, indicated that the maximum similarity scores in the 50-loop comparisons were consistently high, reflecting a strong alignment with the standard report in the best cases.

- 50 loop BERT-base Min: Lastly, the 50-loop BERT-base Min scores had a mean of 0.20 and a standard deviation of 0.03, with scores ranging from 0.16 to 0.25. The low minimum scores, although higher than those in the initial reports, still pointed to instances of lower similarity in certain loops, but the overall increase and lower standard deviation suggested an improvement in alignment with the standard report across different iterations.

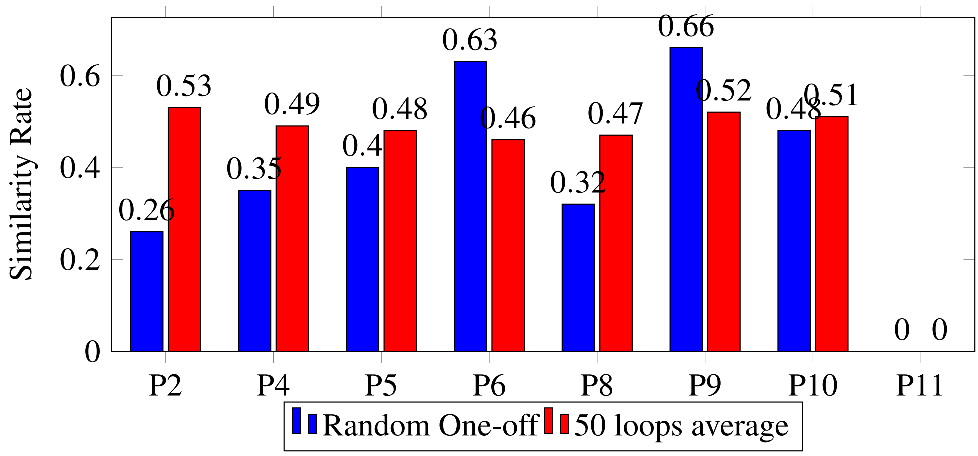

Fig. 4.21 Initial METHANE report result (Scenario 1), BERT-based model

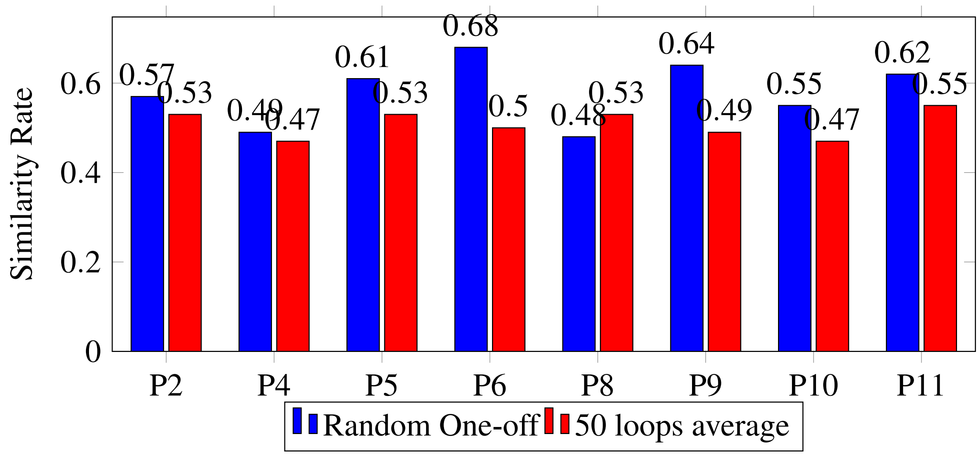

Fig. 4.22 Updated METHANE report result (Scenario 1), BERT-based model

In addition, we conducted a thorough analysis of the highest and lowest scores recorded in the 50-loop evaluations, as outlined in Table [reference table]. This examination offered a more nuanced understanding of the spectrum of semantic similarity encountered by the participants.

Table 4.8 50 Loops max and min value, BERT-based model

	50 loops max	50 loops min
Participant 2 initial	0.82	0.05
Participant 2 updated	0.87	0.21
Participant 4 initial	0.74	0.31
Participant 4 updated	0.82	0.2
Participant 5 initial	0.79	0.1
Participant 5 updated	0.9	0.16
Participant 6 initial	0.85	0.21
Participant 6 updated	0.78	0.16
Participant 8 initial	0.81	0.19
Participant 8 updated	0.85	0.22
Participant 9 initial	0.79	0.21
Participant 9 updated	0.81	0.25
Participant 10 initial	0.76	0.22
Participant 10 updated	0.80	0.19
Participant 11 initial	0	0
Participant 11 updated	0.84	0.22

Reflecting upon these results, it is evident that the analysis using BERT for semantic similarity did not uncover substantial variances in the semantic elements of the METHANE reports, both in their initial and updated forms, after tokenization. This indicates that the application of a BERT-based model for evaluating semantic distinctions in detailed, paragraph-length comparative analyses in such scenarios may not yield optimal effectiveness.

Result from GPT-based model

We also employed a GPT-based model to analyze the content coverage of participants' speech reports. First, the predefined METHANE report was divided into individual components.

Table 4.10 shows the dispatched components of pre-defined initial and updated METHANE reports.

The GPT-based model was then utilized to evaluate whether participants' reports sufficiently covered the content described in these dispatched components. The level of coverage was categorized into three levels: 'Full' (fully covered), 'Part' (partially covered, meaning the main content was covered, but some details were missed), and 'None' (not covered). Table 4.9 shows the result for each participant.

Table 4.9 Individual METHANE component result

	M	E	T	H	A	N	E
Participant 2 initial	Part	Full	Full	Full	Part	Full	Part
Participant 4 initial	None	None	None	None	None	Part	None
Participant 5 initial	Part	Full	Full	Full	Part	Full	Full
Participant 6 initial	Full	Full	Full	Full	Full	Full	Full
Participant 8 initial	None	Part	Part	Part	Part	Full	Full
Participant 9 initial	Part	Part	Full	Part	Part	Part	Full
Participant 10 initial	Full	Full	Full	Full	None	Part	Full
Participant 11 initial	None	None	None	None	None	None	None
Participant 2 updated	Full	Full	Full	Full	None	Full	Full
Participant 4 updated	Full	Full	Full	Full	Full	Full	Full
Participant 5 updated	Full	Full	Full	Full	Full	Full	Full
Participant 6 updated	Full	Full	Full	Full	Full	Full	Full
Participant 8 updated	None	Full	Part	Part	Part	Part	Full
Participant 9 updated	Full	Full	Full	Full	Full	Full	Part
Participant 10 updated	Full	Full	Full	Full	Full	Full	Full
Participant 11 updated	Full	Part	Part	Full	None	Full	Full

In the table, it is evident that some participants did not cover all components in the initial report; however, most of them did cover all components in the updated report. The following sections will discuss the performance of each individual participant.

Participant 2

Participant 2's initial METHANE report addressed the key components of the standard METHANE report, but not all were presented with sufficient clarity. The 'Major Incident'

Table 4.10 METHANE example for Scenario 1

	Initial METHANE report	**Updated METHANE report**
M	Ambulance to control, I would like to declare a major incident.	Ambulance to control, I would like to confirm major incident.
E	The location of the incident is on state highway 1 between junction '458' and '457' heading northbound.	The location of the incident remains on state highway 1 between junction '458' and '457' heading southbound.
T	The incident is a multi-vehicle Accident.	The incident is a multi-vehicle Accident involving '3' number of vehicles.
H	there is glass on the road.	there is glass on the road.
A	Access is southbound junction 457 and egress at junction 458.	Access is northbound junction 457 and egress junction 458.
N	There is an estimated nine patients.	There is an estimated nine casualties. We have the following casualties' status codes: '2' status zero casualties '2' status Ones casualties '3' status Twos casualties '2' status Three casualties.
E	There are currently no other emergency services on scene. Please send fire to contain hazards and Police to control traffic and eight further ambulances.	Police and fire have not yet arrived. We require an addition '2' number of ambulances and please send Westpac HEMs (helicopter).

component was acknowledged but without the immediacy and exactness expected in the standard report. The phrase 'Major incident declared' was included, yet it missed the sense of urgency and specific action implied in the predefined declaration. In the 'Access and Egress' component, while there was some description of access at junction 457, the report omitted a precise definition of the egress at junction 458. Furthermore, the 'Emergency Services' component was incompletely articulated; Participant 2 mentioned "multiple cops" and "at least four ambulances," which did not effectively communicate the original directive for fire services and the need for additional police support for traffic and scene management.

In the updated report, Participant 2 retained the core content from their initial submission but omitted the 'Access and Egress' component in their follow-up. The report implied continuity from the earlier information provided, with consistent 'Location' and 'Type of Incident' components that matched the initial details. The 'Hazards' component also reflected the initial report, with no change in the described dangers. A notable enhancement in the updated report was the 'Number of Casualties' component, which not only aligned with the original estimate but also provided a more detailed account of casualty conditions. Similarly, the 'Emergency Services' component in the updated report mirrored the initial request by specifying the need for a transport ambulance, additional ambulances, and helicopters, thus fulfilling the initial report's call for more extensive resources.

Participant 4

Participant 4's initial METHANE report lacked a direct declaration of a major incident, a critical initial step in the METHANE protocol. The report fell short in specifying the location on State Highway 1 between junctions '458' and '457', a detail that was clearly articulated in the standard METHANE report. Additionally, the nature of the incident was not distinctly characterized as a multi-vehicle accident (MBA), as it was in the original report. There was also an absence of specific mentions of hazards like glass and oil on the road, which were prominent in the original report. Information about access and egress points was omitted, and the number of casualties was noted informally, lacking the structured detail of the original estimate. Furthermore, the report did not request specific emergency services such as fire and police, which were essential for managing the incident.

In the updated report, Participant 4 included a "Standby for declaration of major incident," clearly signaling the initiation of the METHANE report. The update provided precise location details, indicating the exit before junction 458 towards Papakura on the southbound side. The type of incident was correctly identified as an MCI or MVA (Motor Vehicle Accident) involving three vehicles, aligning with the initial report's description of a multi-vehicle accident. The updated hazards section now acknowledged the presence of debris from cars and potential smoke, correlating with the hazards of broken glass and oil from the initial report. The updated section on access and egress suggested that the north access (Papakura exit) might be preferable, considering the motorway debris. The report confirmed the count of nine casualties and added a detailed breakdown of casualty statuses, enhancing the clarity and detail of the casualty information. Lastly, the updated report specifically requested additional resources, including ambulances, ILS, CCPs, ICPs, fire services, and police, in line with the initial report's directives for managing the incident.

Participant 5

For Participant 5's initial METHANE report, the mention of a car accident initiated the statement, but it lacked the clarity and explicitness found in the standard METHANE report's major incident declaration. Location details were provided, with reference to the southbound motorway and exit 458, yet it fell short of specifying the exact positioning between junctions '458' and '457' as the original report did. The nature of the incident was correctly identified as a multi-vehicle crash, in line with the original report. Hazards were noted, including glass and fire, but these lacked the comprehensive detail of the original report, which specified glass and oil on the road. The report gave some access details from exit 457, aligning with the original access point, but clarity on egress was missing. The number of casualties was tallied at nine, consistent with the original report. A request for nine ambulances was made, along with the requisite for fire and police services, mirroring the original report's emergency services directive.

In the updated segment, while there was no initial declaration of a major incident, the context made it clear that this was a continuation and update of the previous report. The updated portion provided more precise location details, describing "the southern motorway, just north of Papakura," which was consistent with the original report. The reiteration of "MVA hazards" maintained the consistency of the incident type with the original report.

Additional traffic hazards were mentioned, which aligned with the initial report's mention of glass and fire hazards. Access information was confirmed as "northern motorway heading south," matching the original description. The number of casualties was affirmed as nine, with an updated breakdown of casualty statuses now included. Specific emergency resources were requested, including extra ambulances, fire services, and police, in accordance with the initial request for emergency services.

Participant 6

In the initial METHANE report from Participant 6, a major incident was promptly declared, aligning with the standard METHANE report's protocol. The location was approximated as before the Papakura exit 458 on State Highway 1, which matched the original report's detailed location. The type of incident was correctly identified as an MVA, akin to the original report. The hazards section of the statement included potential dangers such as broken glass, fuel, chemicals, and people running around, which were consistent with the original report's mention of glass, oil, and hazards. The report also provided a plausible access route through the Takanini exit and egress via the Papakura exit, in accordance with the original report's access and egress points. In terms of casualties, the statement provided an accurate count of nine casualties, mirroring the original report. Additionally, the request for extra resources, including additional ambulances and a Casualty Collection Point (CCP), was in line with the original report's request for supplementary services.

The updated portion of Participant 6's report continued to assert the ongoing major incident, preserving the gravity and urgency of the situation. An updated location was described as "just before the Papakura exit heading south on the motorway," which was consistent with the initial report's location. The type of incident remained described as a motor vehicle accident, ensuring consistency with the original report. Updated hazards were also mentioned, including moving vehicles, broken glass, car parts, and potential fuel and chemicals, which aligned with the original report's mention of similar hazards. The access route was confirmed as the Takanini on-ramp heading south, and egress was identified as the Papakura exit on the motorway heading south, which maintained alignment with the original report's guidance. The number of casualties was reaffirmed as nine, with an added breakdown of casualty statuses provided. Finally, the updated report requested additional resources, including ambulances and the major incident team, aligning with the initial report's calls for additional emergency support.

Participant 8

Participant 8's initial report began without a definitive declaration of a major incident, an element deemed critical for emergency communications. The statement referred to the Papakura Highway heading south but omitted precise location details such as exit numbers, which were specified in the standard report. It described the situation as a motor vehicle accident involving three cars but failed to elaborate on the incident type beyond this broad characterization. Hazards were referenced with the mention of traffic, but the report did not elaborate on other potential hazards like fire, chemicals, or injuries. Access and egress were mentioned in relation to activity on the motorway, yet the report did not offer clear access and egress points as the standard report did. The count of casualties was given as approximately six or seven, but without the detailed breakdown of casualty statuses provided in the initial report. The emergency services section mentioned the need for "another ambulance" and "at least two helis," but the specific emergency services required, their roles, and the rationale for their necessity were not clarified. While the statement referred to additional ambulances and helicopters, it did not explain the reasoning behind these requests as was detailed in the initial report.

The updated statement from Participant 8 used "Major incident" to signal an update but did not specify whether it was an initial or subsequent report and lacked a clear declaration of a major incident. The location was mentioned as Papakura, exit 458, but did not reaffirm its position on the southbound motorway or provide additional location details as before. The incident was described as a "big car crash," offering some insight into the situation but still lacking detail. Hazards were again mentioned vaguely, with a reference to "a whole bunch of cars involved" and traffic, without detailing other potential hazards or the full scope of the situation. Access and egress were discussed in broad terms, noting that traffic removal would lead to good access and egress, but without specifying exit numbers or directions as in the initial report. The number of casualties was stated as nine, with a breakdown by severity (black, red, orange, green), which, while more detailed than the initial statement, still did not align fully with the original report's categorization of casualty statuses. Lastly, the statement mentioned the need for additional ambulances and resources such as an R50, an engine, and police and fire services but lacked a structured rationale for these resource requests.

Participant 9

Participant 9's initial statement commenced with a declaration of a major incident, a positive start in line with emergency reporting protocols. However, the report quickly became muddled, and the speaker appeared to struggle with counting and maintaining focus on the casualties involved. The location was referenced as the Papakura Highway heading towards Hamilton, offering some locational context but without the precise detail included in the standard report. The type of incident was categorized as an MVA car accident, giving a basic understanding of the event but omitting finer details. Potential hazards were identified as traffic and the cars involved in the crash, yet a thorough assessment of the hazards was lacking. Access and egress were mentioned, yet the report did not provide detailed conditions or specific access points. The speaker's attempt to tally the casualties resulted in confusion and repetition, casting doubt on the accuracy of the casualty count. Additionally, the report mentioned that the speaker's ambulance was the only one present but failed to offer clear information on the need for further resources or emergency services.

The updated statement from Participant 9 maintained the initial declaration of a major incident, which was beneficial for the clarity of the report. It furnished details about the location, type of incident, hazards, and the number of casualties involved. The location, described as "Papakura Highway heading south to Hamilton," remained consistent with the initial report. The incident was clearly defined as a car crash involving three vehicles, offering more clarity on the type of incident. Traffic and the presence of other cars were recognized as potential hazards, in alignment with the initial report's findings. The report touched on access and egress, suggesting good passage despite traffic, which implied an understanding of the access points. Regarding casualties, the statement categorized casualties into different severity levels (black, red, orange, green), providing a structured overview of casualty conditions. The report also acknowledged the requirement for additional resources but lacked specificity on the exact number needed, an area that could be enhanced for greater clarity.

Participant 10

Participant 10's initial statement commendably started with a clear declaration of a major incident. The location was aptly identified as "State Highway One, towards South Hamilton and Papakura," providing the necessary locational context. The type of incident was detailed as a "multiple casualty incident of a road vehicle accident," which conveyed a notable amount of information. Hazards were relevantly described with mentions of "broken glass" and that "traffic has been stopped." However, the statement lacked explicit details on access and egress points, which are critical for emergency response coordination. The report did share the number of casualties involved, listed as four.

In the updated statement, the clarity persisted with a continued declaration of a major incident. The location was reiterated as "State Highway 1 leading to South Papakura," ensuring that situational awareness was maintained. The incident was now described as a "road network accident," which remained informative although the specific nature of the accident could be elaborated upon. Hazards were again identified, this time noting "people on the road" and "unregulated traffic," which were pertinent issues that needed addressing in emergency situations. The statement also improved by mentioning both access and egress points on the road. The number of casualties had increased to nine in the updated report, which provided a more severe context for the incident. Emergency services were addressed with a request for additional resources, including fire, police, and four extra ambulance crews, demonstrating an awareness of the needs on the scene.

Participant 11

Participant 11 did not report an initial METHANE at the beginning of the scenario. In the updated statement, Participant 11 started well with a clear declaration of a major incident. However, the location information provided was vague, mentioning only "on the road" without specific details. The type of incident was referred to as an "MVA," which lacked clarity and may not be a recognized term in emergency response contexts. The hazards were identified as glass on the road and the potential for incoming cars, which were relevant concerns in the incident scenario. The access and egress conditions were described in broad terms, likened to a typical motorway situation, which offered some context but lacked the specificity typically required in an emergency response. The number of casualties was given as nine, with a categorization into different statuses such as black, red, orange, and green, providing a structured overview of casualty conditions. Finally, the call for additional resources included police, firefighters, and backup personnel, aligning with standard emergency response needs, but the specifics of these resource requests could be further detailed to match the METHANE report standards.

Summary of GPT-based model evaluation

The key findings from the GPT-based model analysis include:

- Coverage Improvement: Initial reports from some participants lacked full coverage across METHANE components, but there were notable improvements in the updates. For example, Participant 2 and Participant 4 initially missed several components but provided full coverage in their updated reports.
- Variability in Detail: In the initial METHANE report evaluation, Participant 2 addressed most required elements, yet their descriptions of 'Access and Egress' and 'Emergency Services' lacked specificity. Conversely, Participant 4's initial submission was deficient across various METHANE categories, but they significantly improved in their subsequent report, which comprehensively detailed each component. Participant 5 consistently provided thorough coverage on most METHANE aspects in both their initial and subsequent reports, although they initially offered only partial details regarding 'Access and Egress.'
- Consistency with METHANE Format: Participant 6 demonstrated a robust understanding and application of the METHANE reporting format, consistently delivering complete coverage across all elements in both their initial and updated reports. This consistency indicated a strong alignment with the structured communication protocols of METHANE. On the other hand, Participant 8 faced challenges in thoroughly covering the METHANE components in their initial submission. However, there was a notable improvement in their updated report, with an increase in the number of fully covered components, although some areas still only achieved partial coverage.

- Quality of Communication: The reports submitted by Participant 9 initially exhibited a lack of structure and precision, but there was a noticeable improvement in their subsequent update. Despite this progress, certain elements of the report were still not thoroughly addressed. In contrast, Participant 10 demonstrated a commendable understanding of the METHANE format in their updated report, which closely adhered to the predefined standards, showcasing a comprehensive grasp of the METHANE reporting criteria.

- Incomplete Reports: Participant 11 did not provide an initial report, which is a significant omission in emergency reporting. The updated report, although covering more components, still lacked clarity in 'Location' and 'Type of Incident.'

In summary, the GPT-based model analysis highlighted the critical need for detailed and structured reporting in emergency scenarios. The participants showed varying levels of progress in their updates, but achieving the clarity and completeness exemplified by the predefined METHANE standard remains crucial for effective communication in emergencies.

Summary of speech recognition and semantic similarity evaluation

Our study employed a dual-model approach, utilizing both BERT-based and GPT-based models to evaluate METHANE reports in a VR-based MCI training scenario. This comprehensive analysis provided a multifaceted view of the participants' abilities in emergency reporting.

Using the BERT-based model, we assessed the semantic similarity between the standard METHANE reports and those created by participants. This analysis revealed a notable range in the consistency and accuracy of the participants' reports. In the initial METHANE reports, one-off comparisons showed notable variability, indicating the inherent inconsistency of single-instance measurements. However, the 50-loop average scores painted a more stable picture, suggesting that repeated measures could more accurately reflect the semantic alignment with the METHANE format. Interestingly, the updated METHANE reports displayed an enhanced level of consistency and accuracy. The narrowed range and reduced standard deviation in both one-off and 50-loop measures indicated better alignment of participants' reports with the METHANE format over time.

In contrast, the GPT-based model focused on the content coverage within the METHANE reports. The initial reports from some participants showed gaps in covering all METHANE components, but their updated reports demonstrated notable improvements. The analysis highlighted variability in the level of detail and adherence to the METHANE format. While certain participants initially missed crucial elements or provided only partial details, their subsequent reports showed marked advancements in coverage and depth. The quality of communication among participants varied; some initially struggled with the structure and precision of their reports but showed noticeable improvement in their updates. However, the ultimate objective remained to reach the clarity and completeness exemplified by the predefined METHANE standard.

Overall, the results from both the BERT-based and GPT-based models underscored the critical importance of structured and detailed reporting in emergency scenarios. The BERT model provided insights into the semantic similarity of the reports, whereas the GPT model assessed the extent of content coverage, offering complementary perspectives on the participants' reporting performance. The participants exhibited diverse degrees of progress in enhancing the quality of their reports, with some achieving notable alignment with the METHANE standards in their updated submissions. This study not only highlighted the effectiveness of advanced natural language processing tools in training environments but also demonstrated their potential in evaluating and improving essential communication skills in critical scenarios like emergency response.

4.4 Summary

This chapter provided a detailed examination of the outcomes from a VR training session involving 10 participants, focusing on pre-hospital triage in emergency scenarios. The comprehensive analysis was divided into three key areas: Sensor Data Results, Survey Results, and Speech Recognition and Semantic Similarity Evaluation.

The Sensor Data Results section analyzed the participants' performance in two simulated scenarios: a car crash and an earthquake. This analysis covered the time taken for triage, the accuracy of triage categorization, and the sequence of triage actions. The results showed notable variation in the participants' performance, with some displaying speed but less accuracy, and others being more methodical but slower. This diversity in approach reflected the real-world variability in emergency response, emphasizing the different skills and strategies that participants might employ.

In the Survey Results section, insights were gathered from two surveys—an initial survey for background information and a post-experiment survey for feedback on the VR training. The initial survey indicated that the participant group was predominantly aged between 18-24, with varying levels of MCI experience and VR exposure. The post-experiment survey, through its 16 questions, revealed strong immersion and engagement with the VR learning tool, suggesting its effectiveness in mimicking real-life emergencies. However, some participants experienced physical discomfort, such as nausea and dizziness, highlighting the need for ergonomic improvements in VR design. The majority expressed optimism about the integration of VR in paramedic training, reflecting the technology's potential to enhance emergency medical education.

The final section on Speech Recognition and Semantic Similarity Evaluation utilized AI models to process and evaluate audio data from the training. The BERT model assessed semantic similarities in the METHANE reports, revealing improved consistency and accuracy over time. The GPT model evaluated the coverage of METHANE components, indicating progression in the participants' reports from initial inadequacies to more comprehensive coverage in updates. These findings stressed the importance of structured and clear communication in emergency situations and demonstrated the role of AI tools in refining training methods.

In conclusion, this chapter provided a nuanced understanding of the efficacy of VR training, highlighting its strengths in simulating emergency scenarios and identifying areas for improvement, particularly in VR interface design and communication training. The diverse participant responses and the AI model analysis collectively offered valuable insights for future advancements in VR-based emergency response training.

Chapter 5

Discussion

This chapter addresses Phase 6 (Reflection and Contribution) of the research process, synthesising the findings, identifying contributions to theory and practice, and outlining limitations and future work.

5.1 Addressing Research Questions

In this section, we revisit the key research questions outlined in Chapter 1 and 3 and provide detailed answers based on the findings from the results and discussion sections. Each question is addressed in light of the empirical data collected during the VR training experiments, as well as the insights gained from participant feedback and AI-driven performance evaluations. We begin by examining the fundamental components of the VR learning tool and then proceed to explore how AI integration, training efficiency, and participant performance were evaluated. The sub-questions related to Research Question 3 are also discussed in detail to provide a comprehensive understanding of the training's effectiveness.

5.1.1 Research Question 1

RQ1: What are the essential components required to build an effective VR learning tool?

The essential components of the VR learning tool, as outlined in the Methodology chapter, include hardware (Oculus Rift and sensors), software (Unity Engine, Whisper for ASR), and AI-driven tools for data analysis. These components were crucial for creating an immersive and interactive environment that simulated emergency scenarios, such as the car crash and earthquake scenarios used in this study (see Section 3.4). The results showed that the system provided realistic training, with participants responding naturally to the VR environment (see Section 4.1). Similar outcomes have been reported in previous VR-based emergency training studies, which found that immersive simulation environments improved realism and participant engagement [7, 37, 72]. However, the results also highlighted limitations in the realism of casualty models and interaction methods, which require further development to fully simulate real-world emergency scenarios (see Section 5.2.5).

5.1.2 Research Question 2

RQ2: How can AI be integrated effectively into a VR learning tool to enhance data analysis?

The AI integration, particularly using Whisper for speech recognition and BERT/GPT models for semantic analysis, allowed for detailed data capture and evaluation of participant performance. These tools enhanced the ability to assess communication quality and provided evidence of improvements in consistency during METHANE reporting. These outcomes are consistent with prior research showing Whisper's robustness across languages and accents [31, 155] and BERT's effectiveness in analyzing medical communication [145]. In addition, the application of GPT-based models [24] for semantic evaluation represents a novel approach that extends the literature into the specific context of structured METHANE reporting in real-time VR training. Further refinement of these tools is still required to ensure full accuracy and reliability across diverse emergency scenarios.

5.1.3 Research Question 3

RQ3: How can the training effectiveness of the VR learning tool be evaluated and improved?

Training efficiency was evaluated through task completion times, survey feedback, and AI-driven performance assessments. The Results chapter indicated that the average time to complete triage tasks met general requirements, though there was significant variability among participants (see Section 4.1). Similar variability in decision-making speed and accuracy has been reported in other VR triage training studies [37, 72, 96, 1, 152], suggesting that such differences are a common feature of immersive simulations. The survey results reflected that participants found the tool engaging but some experienced discomfort, which may have affected their efficiency (see Section 4.2). Improving user comfort and refining the VR controls, as suggested in participant feedback, could enhance training efficiency (see Section 5.3.3). Additionally, expanding the range of scenarios could provide more comprehensive assessments of participants' performance under different conditions (see Section 5.3.4).

1. Research Question 3.1

   RQ3.1: How can data from VR learning tool be used to assess the competence of participants in MCI scenarios?

   The sensor data, including task completion times and AI-based evaluations of speech reports, provided objective metrics to assess participant competence (see Section 4.1 and 5.2.6). The Results chapter highlighted that task completion times varied widely, indicating differences in decision-making speed and accuracy. Similar challenges in assessing competence have been noted in prior studies, which emphasized the need for objective and standardized evaluation methods in simulation-based training [78, 33]. AI models such as BERT and GPT successfully evaluated the quality of METHANE reports, showing that participants improved their reporting over time (see Section 5.2.6). This suggests that AI tools are valuable for providing detailed, quantitative assessments of participant performance. However, the study also identified the need for more sophisticated models to analyze complex behaviors and interactions (see Section 5.2.4 and 5.3.1).

2. Research Question 3.2

   RQ3.2: Can a VR environment convince participants to behave in a way that will allow their performance to be assessed?

   The immersive nature of the VR environment did prompt participants to engage realistically with the scenarios, as indicated by their actions during triage and METHANE report generation (see Section 4.1 and 5.2.6). The system's ability to capture task completion times and AI evaluations of speech reports demonstrated that the VR environment effectively facilitated performance assessments (see Section 4.1). Similar observations have been reported in earlier studies, which showed that VR environments can elicit realistic behaviors and decision-making patterns from participants [14, 135, 127]. However, the variability in participant behavior, particularly due to discomfort or unfamiliarity with VR technology, also reflects challenges highlighted in prior work on user experience and ergonomics in VR healthcare training [97]. This suggests that while the environment is generally convincing, enhancements in realism and user experience are needed to ensure more consistent behavior across participants (see Section 5.2.3 and 5.3.3).

3. Research Question 3.3

RQ3.3: How well do participants accept VR technology for triage training in MCIs?

The post-experiment survey results indicated high levels of acceptance of VR technology for triage training, with participants expressing enthusiasm for its immersive and interactive nature (see Section 4.2). Similar findings have been reported in earlier studies, which highlighted positive user experiences and acceptance of VR in healthcare education [97]. However, some participants reported physical discomfort, such as nausea and dizziness, which impacted their overall experience (see Section 5.3.1 and 5.2.5). This feedback reflects challenges noted in prior work regarding ergonomics and user adaptation in VR environments, and suggests that while there is strong potential for VR to be integrated into training, further improvements in ergonomics and user interface design are necessary to increase participant comfort and ensure broader acceptance (see Section 5.3.3).

5.2 Study Limitations

Like any research, this study faced several limitations that influenced the generalizability and depth of the findings. These limitations primarily relate to the scope of participant recruitment, sample size, the realism of the VR models, and the methods used for data analysis. Identifying these challenges is essential for understanding the constraints of the current research and guiding future improvements. The following subsections provide a detailed analysis of these limitations and their impact on the study.

5.2.1 Scope of Participant Recruitment

In this study, the focus on pre-hospital triage training led to a restricted recruitment process, guided by specific inclusion and exclusion criteria. This limitation reduced the diversity and range of participants, potentially affecting the broader applicability of the research findings. A notable concern was the impact this had on assessing participant competence using VR data, which is directly related to **Research Question 3.1**: How can data collected from VR systems, such as task completion time and AI-based performance evaluation, be utilized to assess the competence of participants in MCI scenarios? The limited diversity in the participant pool meant that conclusions regarding the use of data collected from VR systems may not be fully generalizable. Future research should aim to adopt a more inclusive recruitment approach, thereby encompassing a broader range of experiences and backgrounds. This would enhance the robustness and applicability of the findings, particularly in evaluating how effectively VR systems measure participant performance across different demographic groups.

5.2.2 Sample Size and Range of Training Sessions

Our study was approved for 10 to 15 participants, but ultimately, only 10 were recruited. We conducted 18 training sessions, some covering different scenarios such as car crashes and earthquakes. However, the relatively small sample size and limited range of training sessions introduced constraints on fully exploring the influence of the VR environment on participant behavior. This is particularly relevant to **Research Question 3.2**: Can a VR environment convince participants to behave in a way that will allow their performance to be assessed? While the study demonstrated that immersive VR environments encouraged behavior reflective of real-world responses, the accuracy of performance assessments was impacted by the limited complexity of scenarios and interactions. Increasing both the participant pool and the range of training scenarios in future studies would allow for a more comprehensive exploration of how participant behavior in VR translates to real-life decision-making and triage effectiveness.

Additionally, the small sample size restricted the ability to draw notable conclusions on **Research Question 3.3**: What is the level of participant acceptance of VR technology for triage training in the context of MCIs? Participants with varying degrees of familiarity with VR responded differently to the technology, but a broader and more diverse participant group would be needed to more accurately assess acceptance across a wider spectrum of users. Future research should expand the participant pool to better understand how diverse experiences with technology influence participant comfort and engagement with VR in emergency medical training.

5.2.3 Vital Sign Model Formulation

The development of a vital sign model capable of accurately reflecting various medical conditions posed a significant challenge in this research. The initial models were either too complex or overly simplistic, which affected their ability to represent a casualty's actual triage category. This limitation was particularly notable in the context of addressing **Research Question 3.2** on the influence of VR environments on participant behavior and performance assessment. With limited realism in the vital signs, participants' decision-making processes were likely not as authentic as they would be in more complex and dynamic simulations. While the study successfully developed a more focused model based on key vital signs, future research should refine the model to ensure it dynamically reflects real-life conditions, allowing for more accurate participant behavior analysis and performance evaluations in VR-based emergency scenarios.

5.2.4 Quantification of the Triage Process

Accurately defining and quantifying the entire triage process within the VR learning tool presented another challenge, particularly given the diverse triage systems used globally. While the simplified scenarios were effective for more straightforward emergency situations, they did not sufficiently capture the complexity needed to fully evaluate the competence of participants in performing triage. The data collected from these simplified scenarios, while useful, did not provide enough variation to assess how participants might perform in more complicated triage situations. Moving forward, research should aim to develop a more comprehensive and adaptable model for the VR environment, capable of simulating a broader range of medical emergencies. This would allow for a more thorough evaluation of participant competence based on VR data and a deeper understanding of how different triage methods affect decision-making and performance.

5.2.5 Realism in Casualty Modeling and Interactions

Achieving a high level of realism in casualty models and their interactions within the Unity environment was a significant challenge in this study. The current models, though functional, lacked the complexity required to fully capture the intricacies of various medical conditions. This limitation impacted the study's ability to assess how accurately VR environments influence participant behavior for performance assessments. Realistic interactions are crucial for ensuring that participants respond as they would in real-life medical emergencies. In the absence of high-fidelity models, the behaviors exhibited by participants may not fully represent what would occur in an actual MCI scenario. Future research should focus on improving the detail and accuracy of casualty models to create a more immersive and educationally valuable VR experience. This will provide more reliable data for evaluating participant behavior and performance in high-pressure, realistic medical scenarios.

5.2.6 Analysis Methods and Data Utilization

The research primarily relied on audio data for analysis, as resource constraints limited the ability to comprehensively analyze video data, particularly in relation to specific triage-related actions. This presented a challenge in accurately evaluating how effectively VR data, such as task completion times and participant actions, can be used to assess participant competence in emergency medical scenarios. The absence of a specialized triage pose recognition model further limited the depth of performance evaluation. Future research should integrate advanced analysis tools, such as pose recognition models tailored for triage scenarios, to capture a more detailed and accurate picture of both audio and video data. By expanding the analytical tools available, future studies will be better equipped to fully assess competence based on VR data and offer richer insights into how VR can be utilized as an effective training and assessment tool in emergency medical education.

5.3 Recommendations to future research

Based on the limitations identified in this study, several areas of improvement have emerged that could enhance the effectiveness and applicability of VR-based emergency training tools. By addressing these gaps, future research can push the boundaries of what is possible in VR training for emergency medical education. The following subsections outline key recommendations for advancing vital sign modeling, decision-making mechanisms, interaction in VR environments, and the expansion of training scenarios.

5.3.1 Enhancement of Vital Sign Modeling in Emergency Simulation Training

Developing a dynamic vital sign model is a crucial direction for future research in emergency simulation training. This advanced model should integrate cutting-edge simulation technologies and machine learning algorithms to accurately represent a broad spectrum of real-life emergency conditions, thereby enhancing the realism and educational value of training simulations.

To create a comprehensive and realistic model, it is essential to cover a diverse range of MCI conditions, each tailored to specific emergency scenarios. For instance, in a car crash scenario, the model would simulate conditions ranging from simple traumas like lacerations or concussions to more severe cases such as Traumatic Brain Injury (TBI), bone fractures, or cardiac arrest. Similarly, in high-altitude emergency scenarios, the model should include unique conditions like hypothermia, altitude sickness, or frostbite, enhancing its applicability and relevance to these specific environments.

Beyond these scenarios, the model should also encompass conditions common in other emergency contexts. This includes simulating chemical burns, inhalation injuries, or electrical shocks in industrial accidents; crush injuries, drowning-related complications, or trauma from debris in natural disasters; and urban emergency situations like asthma attacks, or diabetic emergencies. This broad range of conditions ensures that the training reflects the diversity and complexity of real-world medical emergencies.

A critical feature of this refined model is its adaptability to scenario-specific vital signs. In cases of trauma, the model would simulate vital signs indicative of shock, blood loss, or pain levels. For environmental emergencies, it would adjust vital signs to reflect conditions like dehydration or heatstroke, relevant to high-temperature scenarios.

The proposed vital sign model's ability to learn and adapt in real-time is a key characteristic that sets it apart. Through the integration of machine learning algorithms that analyze extensive medical data, the model can dynamically simulate changes in vital signs, reacting realistically to the treatments and interventions performed by participants. This enhances the immersion of the training experience and provides immediate feedback on the effectiveness of participant actions.

The foundational element of this dynamic model is the concept of a random vital sign generator, initially explored in our research. By enhancing this generator with a set of rules and AI-inferred outcomes, it can evolve into a sophisticated tool capable of producing highly accurate and lifelike vital signs. This includes incorporating various condition-specific biases and an element of randomness, ensuring each simulation presents unique and realistic challenges.

In summary, enhancing the vital sign model in emergency simulation training offers a more authentic and educational experience. By covering a wide range of emergency conditions, adapting to specific scenarios, and incorporating real-time learning and feedback mechanisms, this model stands to improve the preparedness and response capabilities of healthcare professionals in emergency medical situations.

5.3.2 Dynamic decision making mechanism

The advancement of a dynamic vital sign model provides the way for improvements in the decision-making mechanisms within VR-based emergency training tools. This progression not only enhances the model's accuracy but also its practical application in training scenarios.

Upon establishing the dynamic vital sign model, the next logical step involves refining the decision-making process through machine learning (ML). Future studies could leverage this by feeding a comprehensive set of vital sign data, along with corresponding triage categorization levels, into ML algorithms for training. By tagging specific scenarios and medical conditions with detailed vital sign data and triage categorizations, the decision-making mechanism of the VR learning tool could be substantially fortified.

For example, in a scenario where a casualty suffers a heart attack and is categorized at a 'red' triage level, the dynamic nature of the vital sign model would play a crucial role. If the casualty receives no first aid, their vital signs would alter, following a pre-trained pattern from the vital sign model, potentially leading to a simulated fatality if no intervention occurs. Conversely, if paramedics perform actions on the casualty, the outcome would depend on the nature of the action and the results generated by the random vital sign model. Depending on these variables, the casualty's condition could worsen, remain the same, or improve, leading to adjustments in their triage categorization (e.g., from 'red' to 'dead' or 'orange').

This dynamic decision-making mechanism represents a important role in training effectiveness. It realistically simulates the consequences of actions (or inactions) by healthcare professionals, providing a more nuanced and impactful learning experience. participants can witness firsthand how their interventions can alter casualty outcomes, teaching them to respond more effectively to real-world emergencies. This approach not only improves the realism of training scenarios but also enhances the participants' decision-making skills, better preparing them for the complexities and pressures of actual medical emergencies.

5.3.3 Enhancing Interaction in the VR Learning Environment

The current VR learning tool offers fundamental interaction between the participant and the virtual environment. To further enhance the learning experience, the following feature extensions are recommended:

- Casualty Audio and Background Audio:

  Participant feedback has highlighted the need for improved audio elements in the VR learning tool. The current casualty audio cues and background sounds are limited, providing only basic information. Enhancing these audio aspects is crucial for increasing the tool's effectiveness and immersiveness. More detailed and realistic audio cues can offer crucial additional information in medical scenarios. For example, incorporating distinct breathing sounds for a casualty with asthma, even when their general vital signs are normal, can significantly improve the tool's fidelity. This enhancement not only increases the realism of simulations but also adds an educational dimension, requiring participants to use both visual and auditory cues for diagnosis and decision-making. Integrating a diverse range of sounds that accurately depict various medical conditions, such as the labored breathing of an asthma casualty or the indications of pain or distress, will enrich the VR experience, better simulating the complexities of real medical emergencies.

- Realistic casualty Animations:

  Currently, the casualty animations in the VR learning tool are limited due to development constraints. Future iterations should focus on expanding and refining these animations to enhance realism. More diverse animations that simulate a range of physical responses and conditions typical of real-life medical scenarios, such as distress, pain responses, or specific symptoms, would greatly improve the training experience. Including subtle movements like eye blinking, facial expressions, and natural limb movements would make the virtual casualtys appear more lifelike, creating a more engaging environment for participants. Animations that react to the participant's actions, such as changes in expression or posture following treatment, would provide immediate visual feedback on intervention effectiveness.

- Interactive Triggers on Casualties:

  The VR learning tool currently includes a basic set of triggers for simulating casualty vital signs and conditions. However, there is potential for further enhancement, particularly in scenarios involving special conditions or detailed measurements. Introducing more nuanced triggers could deepen participants' understanding of the casualty's condition. Smaller triggers, specific to major body parts like the lungs or liver, could indicate injuries and their severity. Specialized triggers, tailored to each casualty's unique conditions, such as TBI-specific triggers on the head, would allow participants to focus on particular conditions post-initial vital sign assessment. These enhancements aim to closely mirror the complexity of real-world medical scenarios, facilitating natural and informed decision-making in training.

- Interactable Environment Objects:

The VR environment presently features static objects. Introducing interactable objects could significantly add depth and context to the training experience. Allowing participants to move or manipulate items commonly found in emergency scenarios, such as medical equipment or casualty belongings, would make the simulation more engaging and realistic. These interactive elements could range from using contents of a first aid kit to moving obstacles in an accident scene. Interacting with environmental objects can play an essential role in the decision-making process during training, such as uncovering vital clues about a casualty's condition or the nature of an emergency through environmental examination. This enhancement aims to create a more dynamic training experience, equipping healthcare professionals more effectively for real-life emergency situations.

In summary, these proposed improvements aim to provide a more authentic and comprehensive training experience. By advancing audio elements, introducing interactive casualty triggers, developing realistic casualty animations, and adding interactable environmental objects, the VR learning tool can more effectively prepare healthcare professionals for the challenges of emergency medical care.

5.3.4 Expansion of Scenarios in VR Learning

The current VR learning tool, which is limited to car crash and earthquake scenarios, possesses a system architecture that offers considerable flexibility for developers. This flexibility enables rapid development and configuration of diverse environmental scenarios and corresponding casualty vital signs. Moving forward, there is a potential to expand the range of scenarios available within the tool. Future developments could include a variety of natural disaster scenarios, enriching the training experience and providing a broader context for learners.

The introduction of additional scenarios, such as floods, hurricanes, or industrial accidents, would not only diversify the learning environment but also challenge participants with a range of different emergency situations. This expansion would enhance the tool's applicability to real-world conditions and prepare healthcare professionals for a wider array of emergencies. By broadening the scope of scenarios, the VR learning tool can offer more comprehensive training, equipping medical responders with the skills and experience necessary to handle diverse and complex emergency situations effectively.

5.4 Summary

In this chapter, we began by revisiting and addressing the key research questions outlined in Chapters 1 and 3. The answers were drawn from the empirical data collected during the VR training experiments and the insights gained from participant feedback and AI-driven performance evaluations. Research Question 1 was addressed by identifying the essential components of the VR learning tool. Research Question 2 focused on the integration of AI to enhance data analysis, and Research Question 3—along with its sub-questions (3.1, 3.2, and 3.3)—was explored by evaluating the training efficiency, how VR environments influence participant behavior, and the overall acceptance of VR technology for triage training.

After addressing these research questions, we discussed the limitations and potential advancements of our study in pre-hospital triage training using VR technology. Several critical aspects were examined, including participant recruitment, the development and implementation of a vital sign model, the quantification of the triage process, the realism of casualty modeling and interactions, and the methods used for data analysis. These limitations influenced the generalizability and depth of the findings and provided insights for future improvements.

A notable limitation highlighted was the constrained participant recruitment process, influenced by specific inclusion and exclusion criteria. This limitation raised concerns regarding the representativeness and generalizability of the study's findings. Furthermore, the study's restricted sample size and the limited scope of training scenarios likely resulted in an incomplete assessment of the tool's overall effectiveness, particularly in gauging how participants' familiarity with VR technology influenced their response accuracy.

We also encountered challenges in developing a vital sign model that accurately reflects the complexity of real medical conditions and in quantifying the triage process amidst the diversity of global triage systems. Moreover, the current models used in the Unity environment did not entirely capture the complexities of specific medical conditions, impacting the realism of casualty modeling and interactions. The predominant reliance on audio data for analysis, constrained by limited video data evaluation due to resource and technology limitations, also posed a significant challenge in achieving a comprehensive understanding of participant performance and the effectiveness of their interactions in the simulated scenarios.

Looking forward, we propose several avenues for future research to enhance the VR learning tool. A crucial development would be the creation of a dynamic vital sign model, integrating advanced simulation technologies and machine learning algorithms. This model should be designed to more accurately represent a wide spectrum of real-life emergency conditions, thus enriching the training's realism and educational value. The model should encompass various MCI conditions, each tailored to specific emergency scenarios, and should be adaptable to scenario-specific vital signs. Improvements to the VR learning tool are also recommended, such as enhancing audio elements, incorporating interactive casualty triggers, developing more realistic casualty animations, and adding interactable environmental objects. These enhancements aim to provide a more authentic and comprehensive training experience. Moreover, the progression of the dynamic vital sign model could significantly refine the decision-making mechanisms within VR-based emergency training tools. Expanding the range of training scenarios to include diverse emergency situations like natural disasters would offer a broader and more varied training experience.

In conclusion, this chapter provided a comprehensive discussion of the study's findings, addressed the research questions, and outlined key limitations. We also offered several recommendations for future research that could further refine the VR learning tool and contribute to better outcomes in emergency medical education.

Chapter 6

Conclusion

This thesis has advanced the understanding of integrating VR and AI technologies into emergency healthcare training, particularly in the context of pre-hospital triage for MCIs. Through a comprehensive and methodologically robust study, key questions regarding the effectiveness of VR-based training and the potential of AI for assessing learning outcomes were addressed. The findings contribute to the growing body of knowledge on VR's role in healthcare education, offering both theoretical insights and practical implications for future training programs.

The first research question concerned how data collected from VR learning tool, such as task completion time and AI-generated performance evaluations, can be used to assess the competence of participants in MCI scenarios. The study demonstrated that VR learning tool, supported by AI, provide objective metrics that offer a more precise and detailed assessment of participant performance. By combining quantitative data with traditional qualitative feedback from interviews, this research revealed that task completion times and AI-driven analyses contribute significantly to a comprehensive understanding of participant competence. This dual approach addresses limitations in conventional training assessments by enhancing objectivity and allowing for a more nuanced evaluation of how participants perform under simulated MCI conditions.

In response to the second research question, which explored how VR environments influence participant behavior to enable accurate performance assessments, the findings highlighted the immersive nature of VR as a critical factor in shaping participant behavior. The use of advanced VR controllers facilitated more natural and realistic interactions within the training environment, resulting in participant behaviors that more closely mirrored those required in real-world scenarios. This increased immersion enabled a more accurate evaluation of performance, as participants engaged with the VR environment in ways that simulated real-world decision-making and actions during an MCI, thus allowing for a more reliable assessment of their readiness.

The third research question addressed the level of participant acceptance of VR technology for triage training in MCI contexts. Feedback from participants indicated that participants generally responded positively to the immersive experience provided by VR. However, the research also underscored the necessity of proper user orientation and support, as initial unfamiliarity with VR technology posed challenges for some participants. Once adequately guided, participants were able to engage effectively with the VR system, suggesting that with appropriate onboarding measures, VR can be widely accepted and successfully integrated into MCI training programs.

Beyond addressing these research questions, the study also identified several broader implications and challenges that are critical to advancing the field. One key contribution is the development of a hybrid evaluation methodology that combines qualitative interviews with quantitative, AI-driven metrics. This approach offers a more holistic understanding of the effectiveness of VR learning tool, as it bridges the gap between subjective user feedback and objective performance data. However, the continued refinement of AI technologies will be essential for enhancing the accuracy and sophistication of these assessments, particularly in capturing the complexity of decision-making processes in healthcare training.

The study also identified limitations related to current VR interaction methods, particularly the restricted depth of user engagement when relying on basic interaction mechanisms, such as gazing or pointing. To overcome this, advanced VR controllers were introduced, enabling a broader range of interactions and a more immersive training experience. While these enhancements improved the quality and realism of the training, they also raised challenges regarding the scalability and accessibility of such advanced technologies in wider healthcare education settings. Future research will need to focus on developing more intuitive and accessible VR learning tool that maintain the level of interaction required for accurate performance assessments while ensuring ease of use across diverse participant populations.

The issue of participant acceptance further underscores the importance of effective user support and orientation. Although the immersive nature of VR was generally well-received, participants' initial unfamiliarity with the technology highlighted the need for more robust training on the use of VR learning tool. This finding suggests that for VR to be successfully integrated into healthcare training, it must be accompanied by comprehensive onboarding processes to ensure that users can fully engage with the technology. Moreover, future research should explore the long-term retention of skills acquired through VR training, as this study primarily focused on immediate outcomes. It remains crucial to determine whether skills learned in VR environments are retained over time and whether they compare favorably to those acquired through traditional training methods.

The study also faced challenges related to participant recruitment, the development of realistic vital sign models, and the quantification of the triage process. These challenges reflect broader issues inherent in simulating real-world healthcare conditions within a VR framework. Although AI algorithms were used to assist in quantifying aspects of the triage process, further refinement is necessary to more accurately capture the nuances of medical decision-making. Future iterations of the VR learning tool should aim to integrate more sophisticated AI models capable of simulating a wider range of scenarios and healthcare responses.

Looking forward, the integration of VR and AI technologies presents a promising future for emergency healthcare training. However, this study highlights the need for continued development and refinement of these technologies to address the limitations encountered. Future research should prioritize the creation of more advanced AI systems that can simulate complex medical scenarios and provide more reliable performance assessments. Additionally, efforts should be made to develop VR learning tool that are both scalable and accessible, ensuring that these technologies can be seamlessly incorporated into existing training frameworks without creating undue burdens on users or institutions.

Another critical area for future research is the investigation of long-term skill retention following VR training. While this study has demonstrated the immediate benefits of VR for MCI training, it remains unclear how well the skills acquired are retained over time. A better understanding of the long-term impact of VR training on healthcare professionals' preparedness and competency will be crucial for assessing the true value of VR in comparison to traditional training methods.

In conclusion, this thesis has made significant contributions to the understanding of how VR and AI can enhance emergency healthcare training, particularly in the context of MCI scenarios. By addressing key research questions and identifying areas for improvement, this study lays the groundwork for future advancements in the field. The insights gained offer a promising outlook for improving training methodologies, with the potential to enhance the preparedness and capabilities of healthcare professionals in critical situations. As technology continues to evolve, the integration of VR and AI holds the potential to revolutionize healthcare training, paving the way for more immersive, effective, and accessible learning experiences in the years to come.

References

[1] Abbas, J. R., Chu, M. M., Jeyarajah, C., Isba, R., Payton, A., McGrath, B., Tolley, N., and Bruce, I. (2023). Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. *Resuscitation Plus*, 16:100484.

[2] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

[3] Aggarwal, R., Grantcharov, T. P., Eriksen, J. R., Blirup, D., Kristiansen, V. B., Funch-Jensen, P., and Darzi, A. (2006). An evidence-based virtual reality training program for novice laparoscopic surgeons. *Annals of surgery*, 244(2):310–314.

[4] Allen, J. (1995). *Natural language understanding*. Benjamin-Cummings Publishing Co., Inc.

[5] Ambulance, W. F. (2019). Clinical procedures and guidelines; comprehensive edition 2019-2022.

[6] Andreatta, P., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., Stern, S., and Frankel, J. (2010a). Virtual reality triage training provides a viable solution for disaster-preparedness. *Academic Emergency Medicine*, 17(8):870–876. cited By 95.

[7] Andreatta, P. B., Maslowski, E., Petty, S., Shim, W., Marsh, M., Hall, T., Stern, S., and Frankel, J. (2010b). Virtual reality triage training provides a viable solution for disaster-preparedness. *Academic emergency medicine*, 17(8):870–876.

[8] Azeredo, T. R. M., Guedes, H. M., de Almeida, R. A. R., Chianca, T. C. M., and Martins, J. C. A. (2015). Efficacy of the manchester triage system: a systematic review. *International emergency nursing*, 23(2):47–52.

[9] Behmadi, S., Asadi, F., Okhovati, M., and Sarabi, R. E. (2022). Virtual reality-based medical education versus lecture-based method in teaching start triage lessons in emergency medical students: Virtual reality in medical education. *Journal of Advances in Medical Education & Professionalism*, 10(1):48.

[10] Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of FAccT '21*.

[11] Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

[12] Berndt, H., Wessel, D., Mentler, T., and Herczeg, M. (2018a). Human-centered design of a virtual reality training simulation for mass casualty incidents. cited By 4.

[13] Berndt, H., Wessel, D., Mentler, T., and Herczeg, M. (2018b). Human-centered design of a virtual reality training simulation for mass casualty incidents. cited By 4.

[14] Bhagavathula, R., Williams, B., Owens, J., and Gibbons, R. (2018). The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 2056–2060. SAGE Publications Sage CA: Los Angeles, CA.

[15] Bilek, N., Feldhofer, A., and Moser, T. (2021). Virtual reality based mass disaster triage training for emergency medical services. pages 506–507. cited By 0.

[16] Billinghurst, M. and Kato, H. (1999). Collaborative mixed reality. pages 261–284.

[17] Billis, A., Papadimitriou, G., Douligeris, C., Papavramidis, T., Krizea, M., Bamidis, P., Logaras, E., Zouka, M., Karanasiou, N., Fourlis, A., Nicopolitidis, P., Lagakis, P., Gialelis, J., and Kallergis, D. (2019). Functional and non-functional requirements of a smart triage system for emergency departments: The case of inteltriage project. cited By 2.

[18] Bishop, C. (2006). Pattern recognition and machine learning. 16(4):049901.

[19] Bjørn, P., Wulff, M., Petræus, M. S., and Møller, N. H. (2021). Immersive cooperative work environments (CWE): Designing human-building interaction in virtual reality. *Computer Supported Cooperative Work (CSCW)*, 30(3):351–391.

[20] Boletsis, C. (2017). The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction*, 1(4):24.

[21] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

[22] Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. wadsworth int. *Group*, 37(15):237–251.

[23] Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., and Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–480.

[24] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

[25] Bullard, M. J., Chan, T., Brayman, C., Warren, D., Musgrave, E., Unger, B., Group, C. N. W., et al. (2014). Revisions to the canadian emergency department triage and acuity scale (ctas) guidelines. *Canadian Journal of Emergency Medicine*, 16(6):485–489.

[26] Butt, A. L., Kardong-Edgren, S., and Ellertson, A. (2018). Using game-based virtual reality with haptics for skill acquisition. *Clinical Simulation in Nursing*, 16:25–32.

[27] Caballero, A. R. and Niguidula, J. D. (2018). Disaster risk management and emergency preparedness: A case-driven training simulation using immersive virtual reality. In *Proceedings of the 4th International Conference on Human-Computer Interaction and User Experience in Indonesia, CHIuXiD '18*, CHIuXiD '18, page 31–37, New York, NY, USA. Association for Computing Machinery.

[28] Cabrera, E., Luque, E., Taboada, M., Epelde, F., and Iglesias, M. (2012). Optimization of emergency departments by agent-based modeling and simulation. pages 423–430. cited By 5.

[29] Cabrera, E., Taboada, M., Iglesias, M., Epelde, F., and Luque, E. (2011). Optimization of healthcare emergency departments by agent-based simulation. volume 4, pages 1880–1889. cited By 51.

[30] Caicedo-Torres, W., García, G., and Pinzón, H. (2016). A machine learning model for triage in lean pediatric emergency departments. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10022 LNAI:212–221. cited By 2.

[31] Chen, C.-Y., Hsu, Y.-H., and Chang, C. (2023). Accelerating hakka speech recognition research and development using the whisper model. In *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, pages 367–370.

[32] Chen, Y.-P., Chen, Y.-Y., Lin, J.-J., Huang, C.-H., Lai, F., et al. (2020). Modified bidirectional encoder representations from transformers extractive summarization model for hospital information systems based on character-level tokens (alphabert): development and performance evaluation. *JMIR medical informatics*, 8(4):e17787.

[33] Chenais, G., Lagarde, E., and Gil-Jardiné, C. (2023). Artificial intelligence in emergency medicine: viewpoint of current applications and foreseeable opportunities and challenges. *Journal of Medical Internet Research*, 25:e40031.

[34] Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N.-K., and Lee, B. (2017). Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*, 17(10):2435.

[35] Chomsky, N. (1957). Syntactic structures. mouton, the hague (1965). aspects of the theory of syntax.

[36] Chonde, S., Parra, C., and Chang, C.-J. (2013). Minimizing flow-time and time-to-first-treatment in an emergency department through simulation. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, WSC '13, page 2374–2385. IEEE Press.

[37] Cone, D., Serra, J., and Kurland, L. (2011). Comparison of the salt and smart triage systems using a virtual reality simulator with paramedic students. *European Journal of Emergency Medicine*, 18(6):314–321. cited By 45.

[38] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

[39] Costa, C., Stelescu, I., et al. (2014). Increasing the efficiency of the training programs for the mine rescue teams. *Annals of the University of Petroşani, Economics*, 14(Part 1).

[40] Das, R. C. (2017). *Social, Health, and Environmental Infrastructures for Economic Growth*. IGI Global.

[41] Dash, S. and Subudhi, B. (2016). *Handbook of research on computational intelligence applications in bioinformatics*. IGI Global.

[42] Dehghani Soufi, M., Samad-Soltani, T., Shams Vahdati, S., and Rezaei-Hachesu, P. (2018). Decision support system for triage management: A hybrid approach using rule-based reasoning and fuzzy logic. *International Journal of Medical Informatics*, 114:35–44. cited By 20.

[43] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[44] Dey, A. (2016). Machine learning algorithms: A review. *vol*, 7:1174–1179.

[45] Dharmarajan, A. and Velmurugan, T. (2016). Efficiency of k-means and k-medoids clustering algorithms using lung cancer dataset.

[46] Donevant, S., Svendsen, E., Richter, J., Tavakoli, A., Craig, J., Boltin, N., Valafar, H., Dinardi, S., and Culley, J. (2019). Designing and executing a functional exercise to test a novel informatics tool for mass casualty triage. *Journal of the American Medical Informatics Association*, 26(10):1091–1098. cited By 0.

[47] Dua, S., Acharya, U. R., and Dua, P. (2014). *Machine learning in healthcare informatics*, volume 56. Springer.

[48] Fan, X. (2012). Decision as choice of potential intentions. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 02*, WI-IAT '12, page 220–227, USA. IEEE Computer Society.

[49] Ferracani, A., Pezzatini, D., Seidenari, L., and Del Bimbo, A. (2015). Natural and virtual environments for the training of emergency medicine personnel. *Universal Access in the Information Society*, 14(3):351–362. cited By 8.

[50] Fertleman, C., Aubugeau-Williams, P., Sher, C., Lim, A.-N., Lumley, S., Delacroix, S., and Pan, X. (2018). A discussion of virtual reality as a new tool for training healthcare professionals. *Frontiers in public health*, 6:44.

[51] Foronda, C., Shubeck, K., Swoboda, S., Hudson, K., Budhathoki, C., Sullivan, N., and Hu, X. (2016). Impact of virtual simulation to teach concepts of disaster triage. *Clinical Simulation in Nursing*, 12(4):137–144. cited By 27.

[52] Franc, J. M., Kirkland, S. W., Wisnesky, U. D., Campbell, S., and Rowe, B. H. (2022). Metastart: a systematic review and meta-analysis of the diagnostic accuracy of the simple triage and rapid treatment (start) algorithm for disaster triage. *Prehospital and disaster medicine*, 37(1):106–116.

[53] Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and safety in Health care*, 13(suppl 1):i2–i10.

[54] Gan, R. K., Ogbodo, J. C., Wee, Y. Z., Gan, A. Z., and González, P. A. (2024). Performance of google bard and chatgpt in mass casualty incidents triage. *The American Journal of Emergency Medicine*, 75:72–78.

[55] Georgopoulos, V. and Stylios, C. (2012). Introducing fuzzy cognitive maps for developing decision support system for triage at emergency room admissions for the elderly. volume 45, pages 484–489. cited By 7.

[56] Georgopoulos, V. and Stylios, C. (2018). Fuzzy cognitive maps for decision making in triage of non-critical elderly patients. volume 2018-January, pages 225–228. cited By 5.

[57] Goodfellow, I. (2016). Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*.

[58] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

[59] Gout, L., Hart, A., Houze-Cerfon, C.-H., Sarin, R., Ciottone, G. R., and Bounes, V. (2020). Creating a novel disaster medicine virtual reality training environment. *Prehospital and Disaster Medicine*, 35(2):225–228.

[60] Graham, C. and Roll, N. (2024). Evaluating openai's whisper asr: Performance analysis across diverse accents and speaker traits. *JASA Express Letters*, 4(2).

[61] Gunn, T., Jones, L., Bridge, P., Rowntree, P., and Nissen, L. (2018). The use of virtual reality simulation to improve technical skill in the undergraduate medical imaging student. *Interactive Learning Environments*, 26(5):613–620.

[62] Harada, S., Suga, R., Suzuki, K., Kitano, S., Fujimoto, K., Narikawa, K., Nakazawa, M., and Ogawa, S. (2024). Usefulness of self-selected scenarios for simple triage and rapid treatment method using virtual reality. *Journal of Nippon Medical School*, 91(1):99–107.

[63] Harrington, P. (2012). *Machine learning in action*, volume 5. Manning Greenwich.

[64] Harrou, F., Kadri, F., Chaabane, S., Tahon, C., and Sun, Y. (2015). Improved principal component analysis for anomaly detection: Application to an emergency department. *Computers & Industrial Engineering*, 88:63–77.

[65] Hasan, M. S. and Yu, H. (2015). Innovative developments in hci and future trends. In *Automation and Computing (ICAC), 2015 21st International Conference on*, pages 1–6. IEEE.

[66] Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.

[67] Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220.

[68] Hololens, M. (2015). Microsoft hololens. *Available from https://www.microsoft.com/microsoft-hololens/en-us*.

[69] Hossain, M. S. and Muhammad, G. (2014). Cloud-based collaborative media service framework for healthcare. *International Journal of Distributed Sensor Networks*, 10(3):858712.

[70] Hsu, E. B., Li, Y., Bayram, J. D., Levinson, D., Yang, S., and Monahan, C. (2013). State of virtual reality based disaster preparedness and response training. *PLOS Currents Disasters*.

[71] Ilgen, J., M. D., M. C. R., Sherbino, J., MEd, M. D., Cook, D., M. D., M. H. P. E., 3, ., and Zehtabchi, S. (2013). Technology-enhanced simulation in emergency medicine: A systematic review and meta-analysis. *Acad Emerg Med*, 20(ce1, 9418450):117–127.

[72] Ingrassia, P., Ragazzoni, L., Carenzo, L., Colombo, D., Gallardo, A., and Corte, F. (2015). Virtual reality and live simulation: A comparison between two simulation tools for assessing mass casualty triage skills. *European Journal of Emergency Medicine*, 22(2):121–127. cited By 26.

[73] Ismaeel, S., Miri, A., and Chourishi, D. (2015). Using the extreme learning machine (elm) technique for heart disease diagnosis. In *Humanitarian Technology Conference (IHTC2015), 2015 IEEE Canada International*, pages 1–3. IEEE.

[74] Izard, S. G. and Méndez, J. A. J. (2016). Virtual reality medical training system. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM '16, pages 479–485, New York, NY, USA. ACM.

[75] Jarou, Z. J., Dakka, A., McGuire, D., and Bunting, L. (2023). Chatgpt versus human performance on emergency medicine board preparation questions. *Annals of Emergency Medicine*.

[76] Jensen, F. V. (1996). *An introduction to Bayesian networks*, volume 210. UCL press London.

[77] Jiang, D., Huang, R., Calyam, P., Gillis, J., Apperson, O., Chemodanov, D., Demir, F., and Ahmad, S. (2019). Hierarchical cloud-fog platform for communication in disaster incident coordination. pages 1–7. cited By 1.

[78] Joudar, S. S., Albahri, A. S., and Hamid, R. A. (2022). Triage and priority-based healthcare diagnosis using artificial intelligence for autism spectrum disorder and gene contribution: a systematic review. *Computers in Biology and Medicine*, 146:105553.

[79] Kang, D.-Y., Cho, K.-J., Kwon, O., Kwon, J.-M., Jeon, K.-H., Park, H., Lee, Y., Park, J., and Oh, B.-H. (2020). Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 28(1). cited By 14.

[80] Kao, H.-C., Tang, K.-F., and Chang, E. Y. (2018). Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning.

[81] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., and Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, 15:104–116.

[82] Kerr, E., McGinnity, T., Coleman, S., and Shepherd, A. (2021). Human vital sign determination using tactile sensing and fuzzy triage system. *Expert Systems with Applications*, 175. cited By 1.

[83] Kilmon, C. A., Brown, L., Ghosh, S., and Mikitiuk, A. (2010). Immersive virtual reality simulations in nursing education. *Nursing Education Perspectives*, 31(5):314–7. Copyright - Copyright National League for Nursing, Inc. Sep/Oct 2010; Document feature - Photographs; ; Last updated - 2016-07-02.

[84] Klug, M., Barash, Y., Bechler, S., Resheff, Y., Tron, T., Ironi, A., Soffer, S., Zimlichman, E., and Klang, E. (2020). A gradient boosting machine learning model for predicting early mortality in the emergency department triage: Devising a nine-point triage score. *Journal of General Internal Medicine*, 35(1):220–227. cited By 14.

[85] Kong, L. (2021). A study on the ai-based online triage model for hospitals in sustainable smart city. *Future Generation Computer Systems*, 125:59–70. cited By 1.

[86] Koutitas, G., Smith, K. S., Lawrence, G., Metsis, V., Stamper, C., Trahan, M., and Lehr, T. (2019). A virtual and augmented reality platform for the training of first responders of the ambulance bus. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, page 299–302, New York, NY, USA. Association for Computing Machinery.

[87] Kyaw, B. M., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., Divakar, U., Masiello, I., Kononowicz, A. A., Zary, N., et al. (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the digital health education collaboration. *Journal of medical Internet research*, 21(1):e12959.

[88] Larsen, C. R., Soerensen, J. L., Grantcharov, T. P., Dalsgaard, T., Schouenborg, L., Ottosen, C., Schroeder, T. V., and Ottesen, B. S. (2009). Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *Bmj*, 338:b1802.

[89] Levine, D. M., Tuwani, R., Kompa, B., Varma, A., Finlayson, S. G., Mehrotra, A., and Beam, A. (2023). The diagnostic and triage accuracy of the gpt-3 artificial intelligence model. *MedRxiv*.

[90] Lineberry, M., Walwanis, M., and Reni, J. (2013). Comparative research on training simulators in emergency medicine: A methodological review. *Simul. healthc.*, 8(4):253–261.

[91] Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348.

[92] Lourdeaux, D., Afoutni, Z., Ferrer, M.-H., Sabouret, N., Demulier, V., Martin, J.-C., Bolot, L., Boccara, V., and Lelong, R. (2019). Victeams: A virtual environment to train medical team leaders to interact with virtual subordinates. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, IVA '19, page 241–243, New York, NY, USA. Association for Computing Machinery.

[93] Lowd, D. and Domingos, P. (2005). Naive bayes models for probability estimation. In *Proceedings of the 22nd international conference on Machine learning*, pages 529–536. ACM.

[94] Lowe, J., Peng, C., Winstead-Derlega, C., and Curtis, H. (2020). 360 virtual reality pediatric mass casualty incident: A cross sectional observational study of triage and out-of-hospital intervention accuracy at a national conference. *Journal of the American College of Emergency Physicians Open*, 1(5):974–980.

[95] Maclure, J. (2021). AI, explainability and public reason: The argument from the limitations of the human mind. *Minds and Machines*, 31(3):421–438.

[96] Mahling, M., Wunderlich, R., Steiner, D., Gorgati, E., Festl-Wietek, T., and Herrmann-Werner, A. (2023). Virtual reality for emergency medicine training in medical school: prospective, large-cohort implementation study. *Journal of Medical Internet Research*, 25:e43649.

[97] Mäkinen, H., Haavisto, E., Havola, S., and Koivisto, J.-M. (2020). User experiences of virtual reality technologies for healthcare in learning: an integrative review. *Behaviour & Information Technology*, pages 1–17.

[98] Manning, C. and Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

[99] Manogaran, G., Vijayakumar, V., Varatharajan, R., Kumar, P. M., Sundarasekar, R., and Hsu, C.-H. (2018). Machine learning based big data processing framework for cancer diagnosis using hidden markov model and gm clustering. *Wireless personal communications*, 102(3):2099–2116.

[100] Mantovani, F., Castelnuovo, G., Gaggioli, A., and Riva, G. (2003). Virtual reality training for health-care professionals. *CyberPsychology & Behavior*, 6(4):389–395.

[101] Mao, R., Chen, G., Zhang, X., Guerin, F., and Cambria, E. (2023). Gpteval: A survey on assessments of chatgpt and gpt-4. *arXiv preprint arXiv:2308.12488*.

[102] Marchiori, C., Dykeman, D., Girardi, I., Ivankay, A., Thandiackal, K., Zusag, M., Giovannini, A., Karpati, D., and Saenz, H. (2020). Artificial intelligence decision support for medical triage. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2020:793–802. cited By 0.

[103] Marx, J., Hockberger, R., and Walls, R. (2013). *Rosen's Emergency Medicine-Concepts and Clinical Practice E-Book: 2-Volume Set*. Elsevier Health Sciences.

[104] McGuffie, K. and Newhouse, A. (2020). The radicalization risks of gpt-3 and advanced neural language models. *arXiv preprint arXiv:2009.06807*.

[105] Meyerbroker, K. and Emmelkamp, P. M. (2010). Virtual reality exposure therapy in anxiety disorders: a systematic review of process-and-outcome studies. *Depression and anxiety*, 27(10):933–944.

[106] Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329.

[107] Mills, B., Dykstra, P., Hansen, S., Miles, A., Rankin, T., Hopper, L., Brook, L., and Bartlett, D. (2020a). Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. *Prehospital Emergency Care*, 24(4):525–536. cited By 5.

[108] Mills, B., Dykstra, P., Hansen, S., Miles, A., Rankin, T., Hopper, L., Brook, L., and Bartlett, D. (2020b). Virtual reality triage training can provide comparable simulation efficacy for paramedicine students compared to live simulation-based scenarios. *Prehospital Emergency Care*, 24(4):525–536.

[109] Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2017). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*.

[110] Mitchell, T. M. et al. (1997). Machine learning. wcb.

[111] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529.

[112] Mossel, A., Schoenauer, C., Froeschl, M., Peer, A., Goellner, J., and Kaufmann, H. (2020). Immersive training of first responder squad leaders in untethered virtual reality. *Virtual Reality*, 25(3):745–759.

[113] National Institute for Occupational Safety and Health Office for Mine Safety and Health Research Pittsburgh Research Laboratory (2000). Mine emergency response interactive training simulation (merits).

[114] Nguyen, D. H. and Patrick, J. D. (2014). Supervised machine learning and active learning in classification of radiology reports. *Journal of the American Medical Informatics Association*, 21(5):893–901.

[115] Nistor, M., Pham, T., Pickl, S., Gaindric, C., and Cojocaru, S. (2020). A concise review of ai-based solutions for mass casualty management. volume 2805, pages 222–232. cited By 0.

[116] Nori, H., King, N., McKinney, S. M., Carignan, D., and Horvitz, E. (2023). Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*.

[117] Nunes, F. L. S. and Costa, R. M. E. M. (2008). The virtual reality challenges in the health care area: A panoramic view. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, SAC '08, pages 1312–1316, New York, NY, USA. ACM.

[118] Oculus, V. (2015). Oculus rift. *Available from http://www. oculusvr. com/rift*.

[119] of Criminal Justice Education, U. S. L. E. A. A. O. and Training. (1978). Leep - a step ahead. *Washington : Dept. of Justice, Law Enforcement Assistance Administration, Office of Criminal Justice Education and Training*.

[120] Pan, X., Slater, M., Beacco, A., Navarro, X., Bellido Rivas, A. I., Swapp, D., Hale, J., Forbes, P. A. G., Denvir, C., de C. Hamilton, A. F., et al. (2016). The responses of medical general practitioners to unreasonable patient demand for antibiotics-a study of medical ethics using immersive virtual reality. *PloS one*, 11(2):e0146837.

[121] Pardey, T. G. M. (2006). The clinical practice of emergency department triage: Application of the australasian triage scale—an extended literature review: Part i: Evolution of the ats. *Australasian Emergency Nursing Journal*, 9(4):155–162.

[122] Parsons, T. D. and Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of behavior therapy and experimental psychiatry*, 39(3):250–261.

[123] Paslı, S., Şahin, A. S., Beşer, M. F., Topçuoğlu, H., Yadigaroğlu, M., and İmamoğlu, M. (2024). Assessing the precision of artificial intelligence in emergency department triage decisions: Insights from a study with chatgpt. *The American Journal of Emergency Medicine*, 78:170–175.

[124] Peng Xia, J. R. and Parry, D. (2022). Enhancing mass casualty incident training with virtual reality and artificial intelligence. *Australasian Computer Science Week.*

[125] Pereira, S., Portela, F., Santos, M., Machado, J., and Abelha, A. (2016a). Predicting pre-triage waiting time in a maternity emergency room through data mining. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9545:105–117. cited By 2.

[126] Pereira, S., Torres, L., Portela, F., Santos, M., Machado, J., and Abelha, A. (2016b). Predicting triage waiting time in maternity emergency care by means of data mining. *Advances in Intelligent Systems and Computing*, 445:579–588. cited By 2.

[127] Pollard, K. A., Oiknine, A. H., Files, B. T., Sinatra, A. M., Patton, D., Ericson, M., Thomas, J., and Khooshabeh, P. (2020). Level of immersion affects spatial learning in virtual environments: results of a three-condition within-subjects study with long intersession intervals. *Virtual Reality*, 24(4):783–796.

[128] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., et al. (2011). The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*, number CONF. IEEE Signal Processing Society.

[129] Provost, F. and Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, 30(2-3):271–274.

[130] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.

[131] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

[132] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.

[133] Ravanelli, M., Parcollet, T., and Bengio, Y. (2019). The pytorch-kaldi speech recognition toolkit. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6465–6469. IEEE.

[134] Rizzo, A., Pair, J., Graap, K., Manson, B., McNerney, P. J., Wiederhold, B., Wiederhold, M., and Spira, J. (2006). A virtual reality exposure therapy application for iraq war military personnel with post traumatic stress disorder: From training to toy to treatment. *NATO Security through Science Series E Human and Societal Dynamics*, 6:235.

[135] Rose, T., Nam, C. S., and Chen, K. B. (2018). Immersion of virtual reality for rehabilitation-review. *Applied ergonomics*, 69:153–161.

[136] Rovira, A. and Slater, M. (2017). Reinforcement learning as a tool to make people move to a specific location in immersive virtual reality. *International Journal of Human-Computer Studies*, 98:89–94.

[137] Schofield, D. and Dasys, P. A. (2009). The use of virtual simulators for emergency response training in the mining industry. *Journal of Emergency Management*, 7(6):1.

[138] Shaikh, F. (2017). Introductory guide to generative adversarial networks (gans) and their promise. Technical report, Analyticsvidhya, https://www.analyticsvidhya.com/blog/2017/06/introductory-generative-adversarial-networks-gans/.

[139] Sherstyuk, A., Vincent, D., and Wang, K. L. (2010). Making first steps in vr: Monitoring user progress in virtual travel. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, page 19–26, New York, NY, USA. Association for Computing Machinery.

[140] Smith, L. I. (2002). A tutorial on principal components analysis. Technical report.

[141] Srinivas, K., Rani, B. K., and Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)*, 2(02):250–255.

[142] Stone, P., Karki, U., and Ganapathy, S. (2019). Agent-based simulation of artificial-intelligence-assisted, transfer of care. cited By 0.

[143] Surer, E., Erkayaoğlu, M., Öztürk, Z. N., Yücel, F., Bıyık, E. A., Altan, B., Şenderin, B., Oğuz, Z., Gürer, S., and Düzgün, H. Ş. (2020). Developing a scenario-based video game generation framework for computer and virtual reality environments: a comparative usability study. *Journal on Multimodal User Interfaces*, 15(4):393–411.

[144] Tahayori, B., Chini-Foroush, N., and Akhlaghi, H. (2021a). Advanced natural language processing technique to predict patient disposition based on emergency triage notes. *EMA - Emergency Medicine Australasia*, 33(3):480–484. cited By 2.

[145] Tahayori, B., Chini-Foroush, N., and Akhlaghi, H. (2021b). Advanced natural language processing technique to predict patient disposition based on emergency triage notes. *Emergency Medicine Australasia*, 33(3):480–484.

[146] Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, pages 330–337.

[147] Tayama, Y., Kato, R., and Okada, K.-i. (2015). Triage training system: Adjusting the difficulty level according to user proficiency. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, page 139–147, New York, NY, USA. Association for Computing Machinery.

[148] United Nations, D. o. E. and Social Affairs, P. D. (2017). world population prospects.

[149] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[150] Vearrier, L., Derse, A. R., Basford, J. B., Larkin, G. L., and Moskop, J. C. (2022). Artificial intelligence in emergency medicine: benefits, risks, and recommendations. *The Journal of Emergency Medicine*, 62(4):492–499.

[151] Velikova, M., Terwisscha van Scheltinga, J., Lucas, P. J., and Spaanderman, M. (2014). Exploiting causal functional relationships in bayesian network modelling for personalised healthcare.

[152] Vogt, P., Boer, R., de Boer, M., Prins, H., Smit, J., Tuinstra, D., Degens, N., Hettinga, M., and Paans, W. (2023). Designing and evaluating a virtual reality training for paramedics to practice triage in complex situations. In *International conference on human-computer interaction*, pages 503–522. Springer.

[153] Walker, N. R., Trofimovich, P., Cedergren, H., and Gatbonton, E. (2011). Using asr technology in language training for specific purposes: A perspective from quebec, canada. *CAlICO Journal*, 28(3):721–743.

[154] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. (2018). Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.

[155] Wang, C., Liu, S., Li, A., and Liu, J. (2023). Text dialogue analysis based chatgpt for primary screening of mild cognitive impairment. *medRxiv*, pages 2023–06.

[156] Wang, S., Yang, C.-H., Wu, J., and Zhang, C. (2024). Can whisper perform speech-based in-context learning? In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 13421–13425. IEEE.

[157] Winograd, T. (1972). Understanding natural language academic press. *New York*.

[158] Xia, P., Ruan, J., Parry, D., Britnell, S., and Yu, J. (2022). Enhancing triage training for mass casualty incidents with virtual reality and artificial intelligence. In *2nd Information Systems for Crisis Response and Management Asia Pacific Conference (ISCRAM 2022)*.

[159] Yan, S., Peng, J., Grain, H., and Yi, M. (2019). Technology road mapping of two machine learning methods for triaging emergency department patients in australia. In *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, PRAI '19, page 60–67, New York, NY, USA. Association for Computing Machinery.

[160] Yassine, A., Singh, S., and Alamri, A. (2017). Mining human activity patterns from smart home big data for health care applications. *IEEE Access*, 5:13131–13141.

[161] Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.-F., and Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, 36(4):2431–2448.

[162] Yu, X. and Ganz, A. (2011). Mirte: Mixed reality triage and evacuation game for mass casualty information systems design, testing and training. pages 8199–8202. cited By 0.

[163] Yun, H., Choi, J., and Park, J. (2021). Prediction of critical care outcome for adult patients presenting to emergency department using initial triage information: An xgboost algorithm analysis. *JMIR Medical Informatics*, 9(9). cited By 0.

[164] Zachariasse, J. M., van der Hagen, V., Seiger, N., Mackway-Jones, K., van Veen, M., and Moll, H. A. (2019). Performance of triage systems in emergency care: a systematic review and meta-analysis. *BMJ open*, 9(5):e026471.

[165] Zijie, P., Tong, X., Yi, Z., Xiang, C., Han, Z., Lili, W., and Li, C. (2021). Application progress of smart glasses for triage during mass casualty incident. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue*, 33(2):244–248. cited By 0.

[166] Zikos, D., Vandeliwala, I., and Makedon, P. (2014). A probabilistic algorithm with user feedback loop for decision making during the hospital triage process. volume 2014-May. cited By 0.

[167] Zlotnik, A., Alfaro, M., Perez, M., Gallardo-Antolin, A., and Martinez, J. (2016). Building a decision support system for inpatient admission prediction with the manchester triage system and administrative check-in variables. *CIN - Computers Informatics Nursing*, 34(5):224–230. cited By 13.

Appendix A

Ethics Approval

Auckland University of Technology Ethics Committee (AUTEC)

Auckland University of Technology
D-88, Private Bag 92006, Auckland 1142, NZ
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

14 July 2021

Ji Ruan
Faculty of Design and Creative Technologies

Dear Ji

Re Ethics Application: **20/419 Enhancing Healthcare with Virtual Reality and Artificial Intelligence.**

Thank you for providing evidence as requested, which satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC).

Your ethics application has been approved for three years until 12 July 2024.

Standard Conditions of Approval

1. The research is to be undertaken in accordance with the Auckland University of Technology Code of Conduct for Research and as approved by AUTEC in this application.
2. A progress report is due annually on the anniversary of the approval date, using the EA2 form.
3. A final report is due at the expiration of the approval period, or, upon completion of project, using the EA3 form.
4. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form.
5. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
6. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.
7. It is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard and that all the dates on the documents are updated.

AUTEC grants ethical approval only. You are responsible for obtaining management approval for access for your research from any institution or organisation at which your research is being conducted and you need to meet all ethical, legal, public health, and locality obligations or requirements for the jurisdictions in which the research is being undertaken.

Please quote the application number and title on all future correspondence related to this project.

For any enquiries please contact ethics@aut.ac.nz. The forms mentioned above are available online through http://www.aut.ac.nz/research/researchethics

(This is a computer-generated letter for which no signature is required)

The AUTEC Secretariat
Auckland University of Technology Ethics Committee

Cc: peng.xia@aut.ac.nz; Dave Parry; ghowie@aut.ac.nz; Wai (Albert) Yeap